Data Sensitivity Analysis Part 2

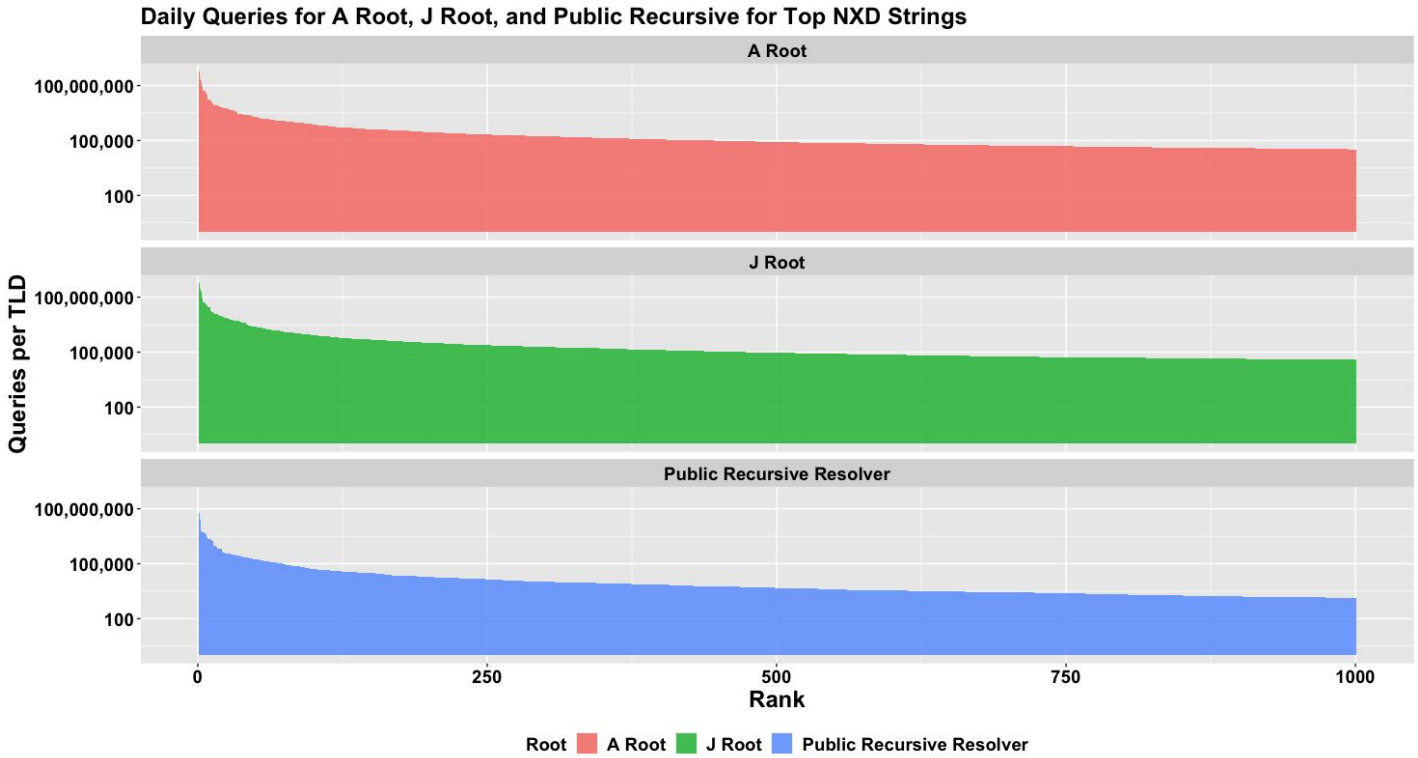
Recursive Resolver Data

Measurements

1. Examine recursive resolver top NXDomain TLDs strings against root data
    a. Compare using a sorting function of total query volume per TLD
    b. Compare using a sorting function of distinct source IPs per TLD
    c. How do these ranking functions compare: a.) between roots and b.) between RRs and roots

2. Examine how recursive resolvers top NXDomain TLDs compare to each other
    a. Rank ordering
    b. Overlap

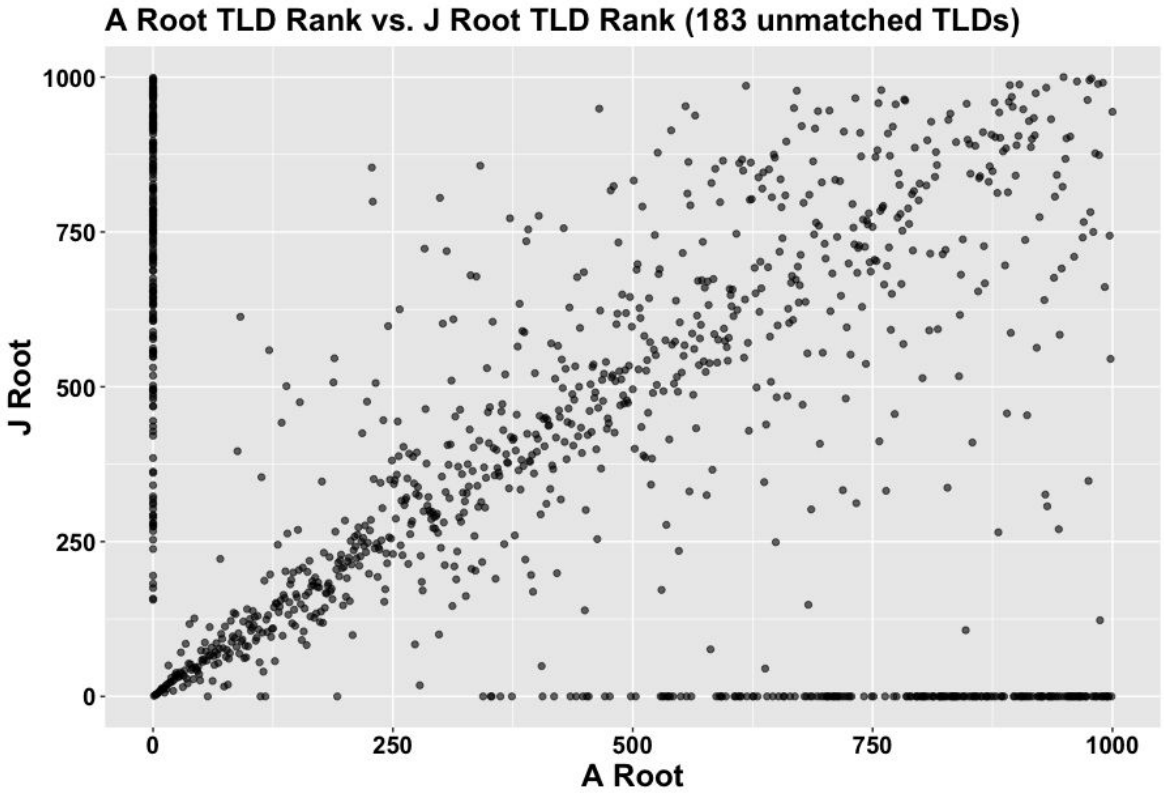
Total Query Volume per TLD and rank

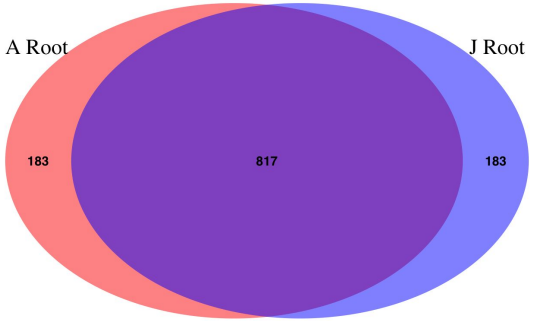


- Power law distribution at all observation points
- Curve flattens significantly after top 50 TLDs based on query volume at all observation points

A and J Root Servers Compared to a Public Recursive Using Total Query Volume per TLD Ranking as a Function

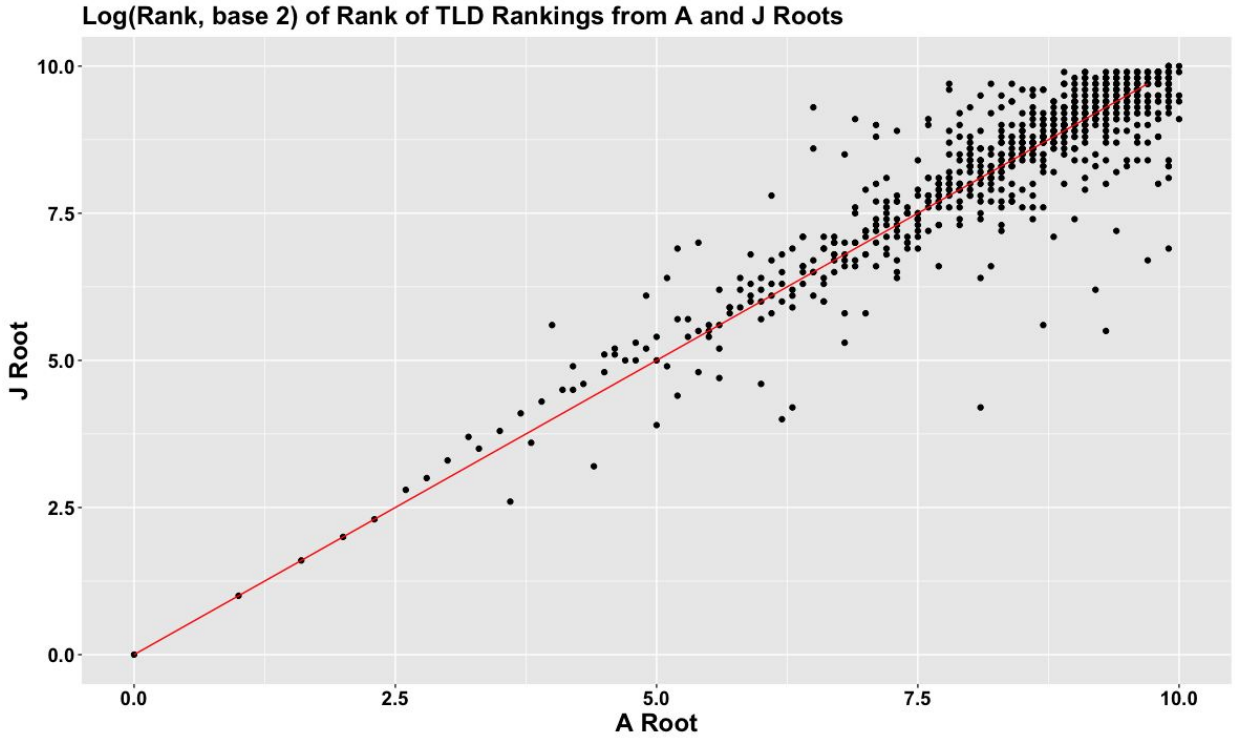
A Root Ranks vs. J Root Ranks





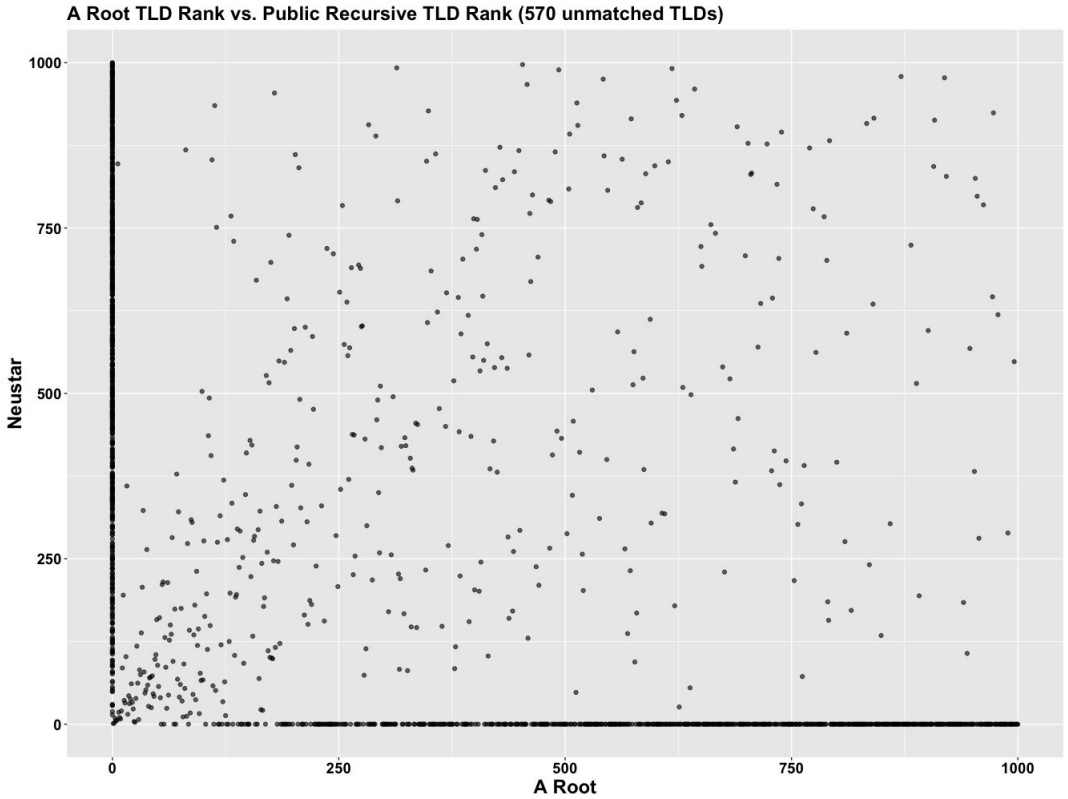
- Decent correlation at lower ranks between A and J. Correlation disperse a similar point in which the power law curve flattened from previous slide.
- Overlap of 183 TLDs between two roots and each root having a unique 183
- Missing TLDs more common as rank approaches 1000

A Root Ranks vs. J Root Ranks (Log Normalized)

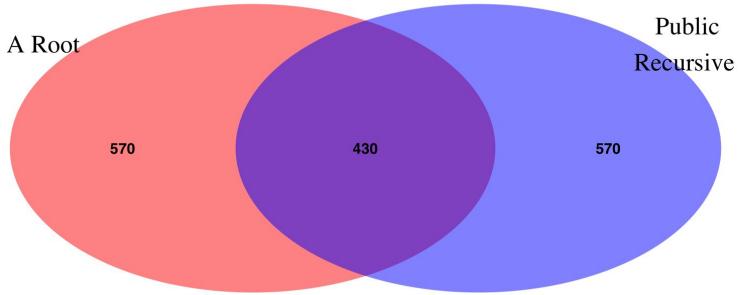


- It really doesn't make sense (or matter) if TLD X is 447 at A and is 572 at J.
- Using a log transformation of the Rank will create more "bins" in which you can better compare rankings with exponentially increasing bin sizes
- Better visualization that there is decent correlation by applying better bins
- Also indicates that measurements trying to do overlap analysis will likely suffer from some noise due to arbitrary cutoff of N (e.g. TLD is rank 997 at A but 1002 at J)

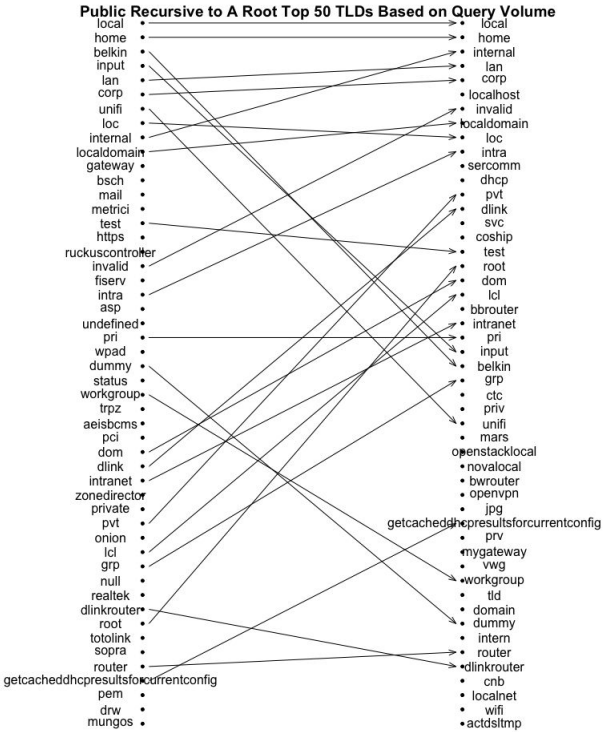
A Root Ranks vs. Recursive Using Query Volume Sorting



- Clearly little to no correlation of top N between a root server and this recursive resolver.
- 570 TLD strings were unmatched in the two top 1K lists
- Some initial evidence things might be very different between these entities in the DNS ecosystem

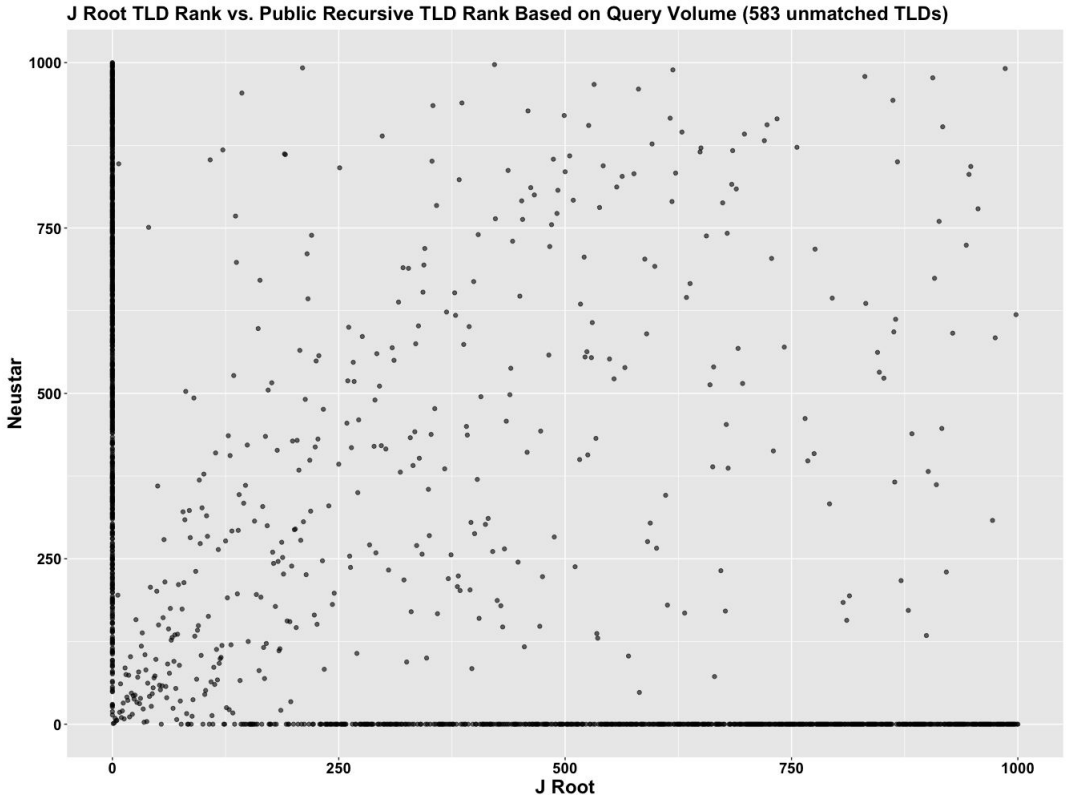


A Root Ranks vs. Public Rec. Using Query Volume Sorting

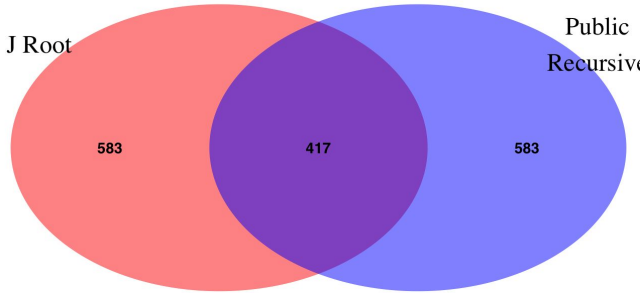


- Mapping of top 50 TLDs seen at public recursive to top 50 TLDs seen at A root
- Some align but a fair amount don’t have any corresponding match

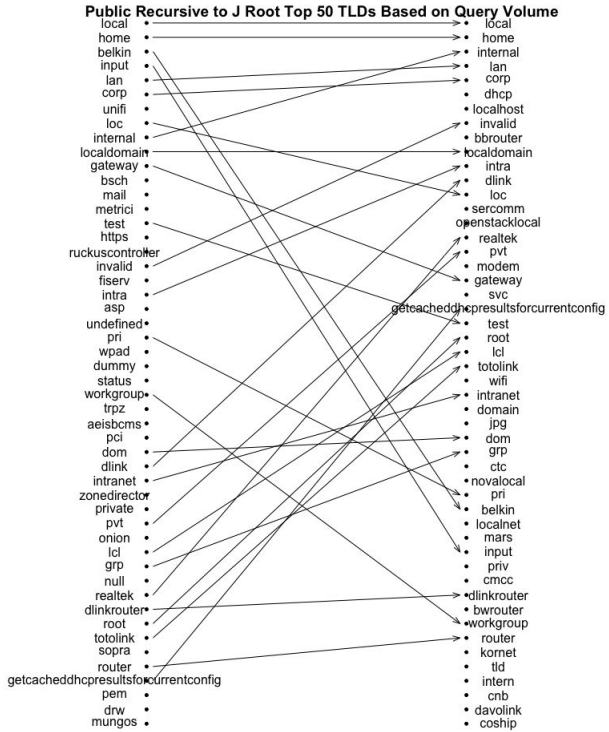
J Root Ranks vs. Public Rec. Using Query Volume Sorting



- Again - clearly little to no correlation of top N between a root server and this recursive resolver.
- 583 TLD strings were unmatched in the two top 1K lists
- more evidence things might be very different between these entities in the DNS ecosystem



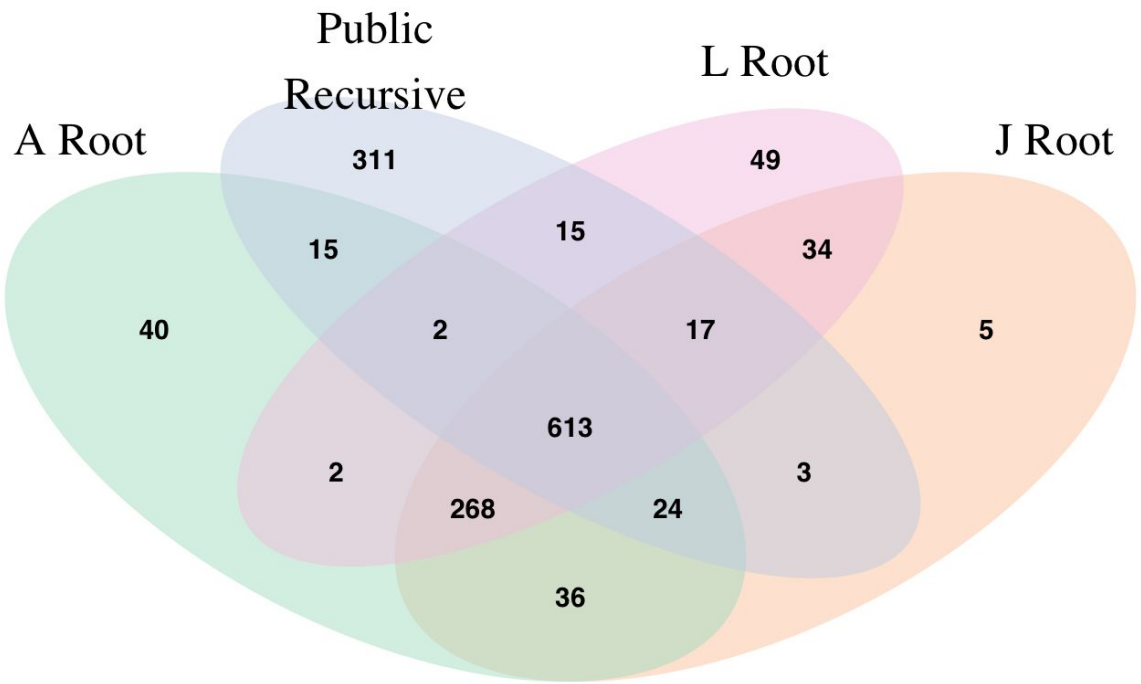
J Root Ranks vs. Public Rec. Using Query Volume Sorting



- Mapping of top 50 TLDs seen at public recursive to top 50 TLDs seen at J root
- Again some align but a fair amount don’t have any corresponding match

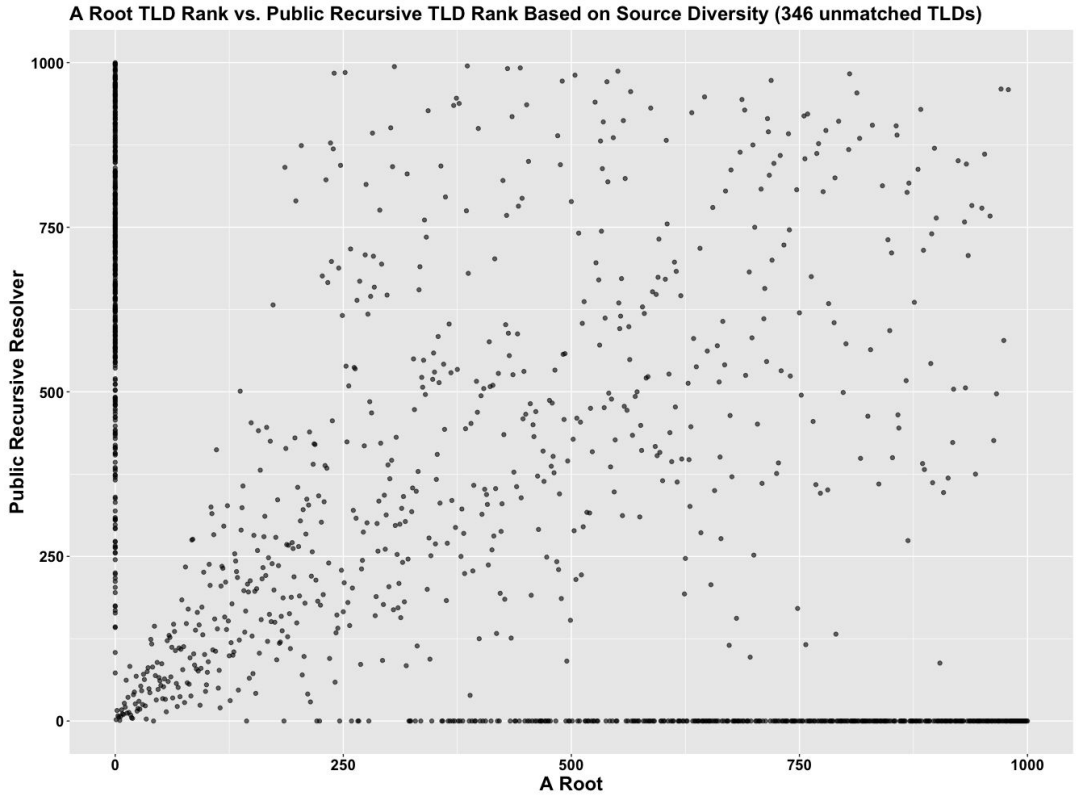
A, J, and L Root Servers Compared To Public Recursive Using Distinct Source IPs per TLD Ranking Function

Venn Diagram of A, J, and L Roots and Public Recursive



- Overlap between root letters is very high (compared to ranking by total query volume).
- Large number of top TLDs observed at public recursive are not observed at any of the root letters
- Venn below shows root overlap a bit easier

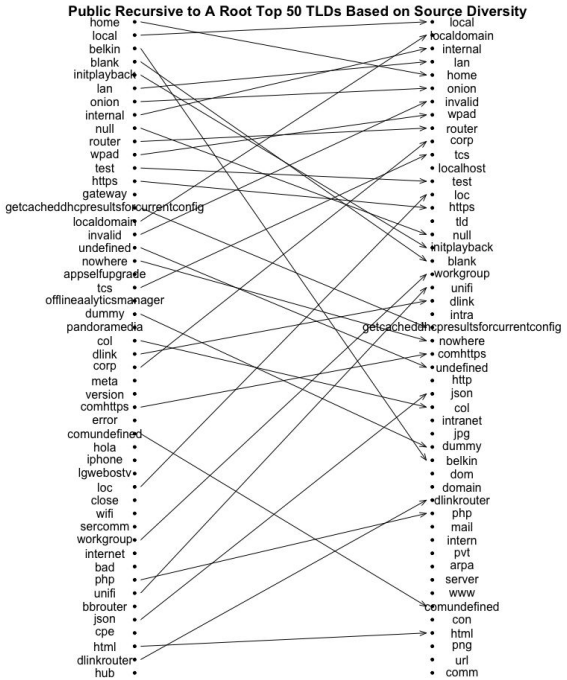
A and Public Recursive Resolver Rank Matches



- Correlation between A root and public recursive top TLDs using the source diversity ranking function is much stronger than total query volume (slide 7)
- Fewer unmatched TLDs (346 vs 570A or 583J).
- Unmatched TLDs are further down in the tail
- Some of the unmatched are because of the change in ranking function
  - The “bwrouter” is ranked 33 at A root when sorted by total query volume yet it doesn’t make the top list for L root by source diversity.
  - High query volume domains with small sources are not going to make this style of list
  - Potential impact to identifying “easier to remediate” domains

```
> a[TLD == "bwrouter", ]
        TLD Queries  IPs Ranks   Root
1: bwrouter 3803023 2569    33 A Root
> sl[TLD == "bwrouter", ]
Empty data.table (0 rows and 2 cols): TLD,Root
>
```

A Root Ranks vs. Public Recursive Using Source Diversity




```
> head(sar[Ranks.x == 0,][order(Ranks.y)], n=50)
                                TLD IPs.x Ranks.x Root.x Queries IPs.y Ranks.y          Root.y
 1:                         mobilis    NA       0   <NA>   24477   508      73 Public Recursive
 2:                        bwrouter    NA       0   <NA>   18233   329     104 Public Recursive
 3:                     formacioncau    NA       0   <NA>   17433   204     142 Public Recursive
 4:                       hingeentry    NA       0   <NA>   56068   202     143 Public Recursive
 5:                     sharkiatoday    NA       0   <NA>     256   162     164 Public Recursive
 6:                           coship    NA       0   <NA>    8457   157     168 Public Recursive
 7:                        classlink    NA       0   <NA>     383   153     174 Public Recursive
 8:                       arundefined    NA       0   <NA>     374   153     175 Public Recursive
 9:                         davolink    NA       0   <NA>   18991   131     195 Public Recursive
10:                          viacode    NA       0   <NA>     174   125     203 Public Recursive
11:                            gcash    NA       0   <NA>     146   120     212 Public Recursive
12:                             hsia    NA       0   <NA>     753   117     217 Public Recursive
13:                          domowy    NA       0   <NA>    6710   110     223 Public Recursive
14:                       newssearch    NA       0   <NA>     138   110     225 Public Recursive
15:                           status    NA       0   <NA>  369296    98     245 Public Recursive
16:                           mungos    NA       0   <NA>  168989    89     255 Public Recursive
17:                       imolivesdk    NA       0   <NA>     175    89     256 Public Recursive
18:                      comhttpshttps    NA       0   <NA>     282    87     263 Public Recursive
19:                        savebuild    NA       0   <NA>     201    86     264 Public Recursive
20:                              sng    NA       0   <NA>    9151    85     266 Public Recursive
21:                              mms    NA       0   <NA>     528    81     272 Public Recursive
22:                              vec    NA       0   <NA>    1004    81     273 Public Recursive
23:                            regus    NA       0   <NA>    6187    70     292 Public Recursive
24:                             alog    NA       0   <NA>   12868    70     293 Public Recursive
25:                       skbroadband    NA       0   <NA>    7759    68     298 Public Recursive
26:                        traductor    NA       0   <NA>     131    66     305 Public Recursive
27:                          anymore    NA       0   <NA>      72    65     306 Public Recursive
28:                         enternet    NA       0   <NA>     564    65     307 Public Recursive
29:                            prdlic    NA       0   <NA>    1020    65     309 Public Recursive
30:                    freedomisnotfree    NA       0   <NA>     337    63     319 Public Recursive
31:                          homelan    NA       0   <NA>   18166    60     332 Public Recursive
32:                          torexit    NA       0   <NA>     324    58     338 Public Recursive
33:                          actuator    NA       0   <NA>     176    58     340 Public Recursive
34:                             netsa    NA       0   <NA>     339    51     366 Public Recursive
35:                yotaaccessinterface    NA       0   <NA>  105394    50     370 Public Recursive
36:                         dedicated    NA       0   <NA>   19996    50     373 Public Recursive
37:                         mtsrouter    NA       0   <NA>    2096    50     374 Public Recursive
38:                            iossss    NA       0   <NA>     385    47     380 Public Recursive
39:                   wirelessinternet    NA       0   <NA>    4708    47     383 Public Recursive
40:                        xalipaynebula    NA       0   <NA>     161    47     385 Public Recursive
41: ckaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa    NA       0   <NA>      74    45     393 Public Recursive
42:                             wind    NA       0   <NA>    4445    44     404 Public Recursive
43:                       iniciarbldaoi    NA       0   <NA>      95    43     406 Public Recursive
44:                        multilaserap    NA       0   <NA>    1743    43     407 Public Recursive
45:                          localhosts    NA       0   <NA>      59    42     413 Public Recursive
46:                             trhttp    NA       0   <NA>      95    42     415 Public Recursive
47:                           tradutor    NA       0   <NA>      97    42     416 Public Recursive
48:                              nude    NA       0   <NA>     135    42     419 Public Recursive
49:                          comconfig    NA       0   <NA>      59    41     429 Public Recursive
50:                                wap    NA       0   <NA>    2408    41     433 Public Recursive
```

- Mapping of top 50 TLDs seen at public recursive to top 50 TLDs seen at A root
- Again some align but a fair amount don’t have any corresponding match
- Top 50 TLDs observed at public recursive but not in top 1K for A root