Evaluating the Long-term Effects of Parameters on the Characteristics of the Tranco Top Sites Ranking

Victor Le Pochat, Tom Van Goethem, Wouter Joosen

CSET 2019, 12 August 2019
“We perform a comprehensive analysis on Alexa’s Top 1 Million websites”

“We collected the benign pages from the Alexa top 20K websites”

“The list of websites we chose for our evaluation comes from the Alexa Top Sites service, the source widely used in prior research on Tor”
Impact of rankings is not well-known

- **Unannounced changes** to methods
- Little **agreement** on most popular domains
- Potentially very **volatile**
- Easily **manipulated**
- **Unknown effects** in composition

Rankings can have a **large impact** on research

[LeP19, Sch18, Rwe19]
We proposed Tranco as a research-oriented ranking

- **Transparent** methods
- **Reproducible** rankings
- **Improved** properties

Daily updated default ranking + custom rankings

https://tranco-list.eu/

[Le Pochat et al. Tranco: a research-oriented top sites ranking hardened against manipulation. NDSS 2019]
We now **evaluate** Tranco's properties and parameters.

- Comparison with existing rankings
- Researcher assumptions
- Stability
- Anomalies
We evaluate Tranco's properties and parameters.

Comparison with existing rankings

Researchers' assumptions

Stability

Anomalies
Tranco has some similarity with each component.
Tranco contains domains popular in **Chrome**
We evaluate Tranco's properties and parameters

Comparison with existing rankings

Stability

Researcher assumptions

Anomalies
Responsive domains guarantee a sufficient sample
Some **malicious** domains are present, but can be filtered out using Google Safe Browsing

<table>
<thead>
<tr>
<th></th>
<th>10K</th>
<th>100K</th>
<th>1M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malware</td>
<td>1</td>
<td>24</td>
<td>187</td>
</tr>
<tr>
<td>Social engineering</td>
<td>1</td>
<td>21</td>
<td>1,486</td>
</tr>
<tr>
<td>Unwanted software</td>
<td>2</td>
<td>34</td>
<td>189</td>
</tr>
<tr>
<td>Potentially harmful application</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total (unique domains)</strong></td>
<td><strong>1,851</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We evaluate Tranco's properties and parameters

- Comparison with existing rankings
- Researcher assumptions
- Stability
- Anomalies
Tranco is very **stable** compared to its components
Aggregating over 30 days leads to balanced stability
Smaller subsets see **higher stability** over one year.
We evaluate Tranco's properties and parameters

Comparison with existing rankings

Researcher assumptions

Stability

Anomalies
Component rankings experience anomalies

![Graph showing list length over time for different rankings systems](image)
Tranco is somewhat affected, but **impact is reduced**
We evaluate Tranco's properties and parameters

Comparison with existing rankings

Stability

Researcher assumptions

Anomalies
We evaluate Tranco's properties and parameters

- Similar to component and external lists
- Researcher assumptions
- Stability
- Anomalies
We evaluate Tranco's properties and parameters

Similar to component and external lists

Mostly responsive and benign

Stability

Anomalies
We evaluate Tranco's properties and parameters

**Similar** to component and external lists

**Mostly responsive and benign**

**Aggregation improves stability**

**Anomalies**
We evaluate Tranco's properties and parameters

- Similar to component and external lists
- Mostly responsive and benign
- Aggregation improves stability
- Impact of anomalies is reduced
We make researchers aware of Tranco's properties

- **30-day aggregation** yields good stability trade-off
- **Apply filters** where appropriate
- Use **full list** to retain at least 1M domains
- Properties improve slightly for **smaller subsets**
- Properly **reference** the specific list used

Default parameters → **representative** set of domains
Download the Tranco ranking:

https://tranco-list.eu/
Thank you!

Victor.LePochat@cs.kuleuven.be
@VictorLePochat
References


