

Popularity ranking for domain names

Sebastian Castro - NZRS

DNS-OARC Workshop - Amsterdam 2015

Introduction

- Td-idf: term frequency – inverse document frequency
- Method to determine how important is a word within a document in a collection or corpus (ref from Wikipedia)
- Tf: number of times a word occurs in a document
- Idf: inverse function of the number of documents a word occurs

What if?

- Using authoritative DNS data, replace **words** by domains (query name) and **documents** by the set of queries coming from a given IP address?
- **d** = query name from DNS
- **a** = source address
- **A** = all unique source addresses

Methodology

- For every day, calculate

$$\text{tf}(d, a) = \frac{\sum \text{queries}(d \text{ from } a)}{\sum \text{queries}(a)}$$

$$\text{tfidf}(d, a) = \text{tf}(d, a) \times \text{idf}(d, A)$$

$$\text{idf}(d, A) = \log \frac{|A|}{|a \in A| : a \text{ asked for } d}$$

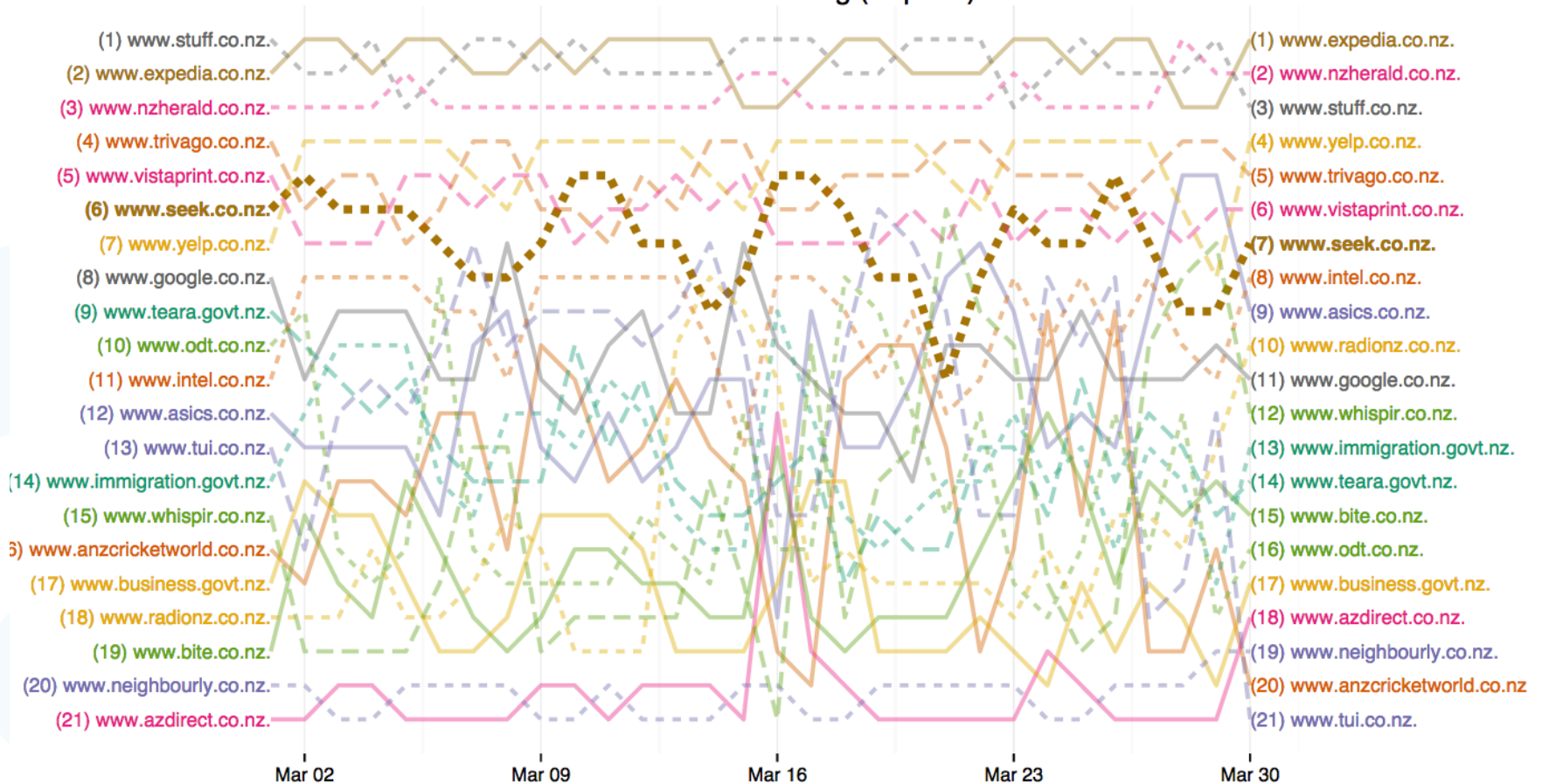
Methodology

$\text{rank}(d, a) = \#d$ sorted desc by $\text{tfidf}(d, a)$

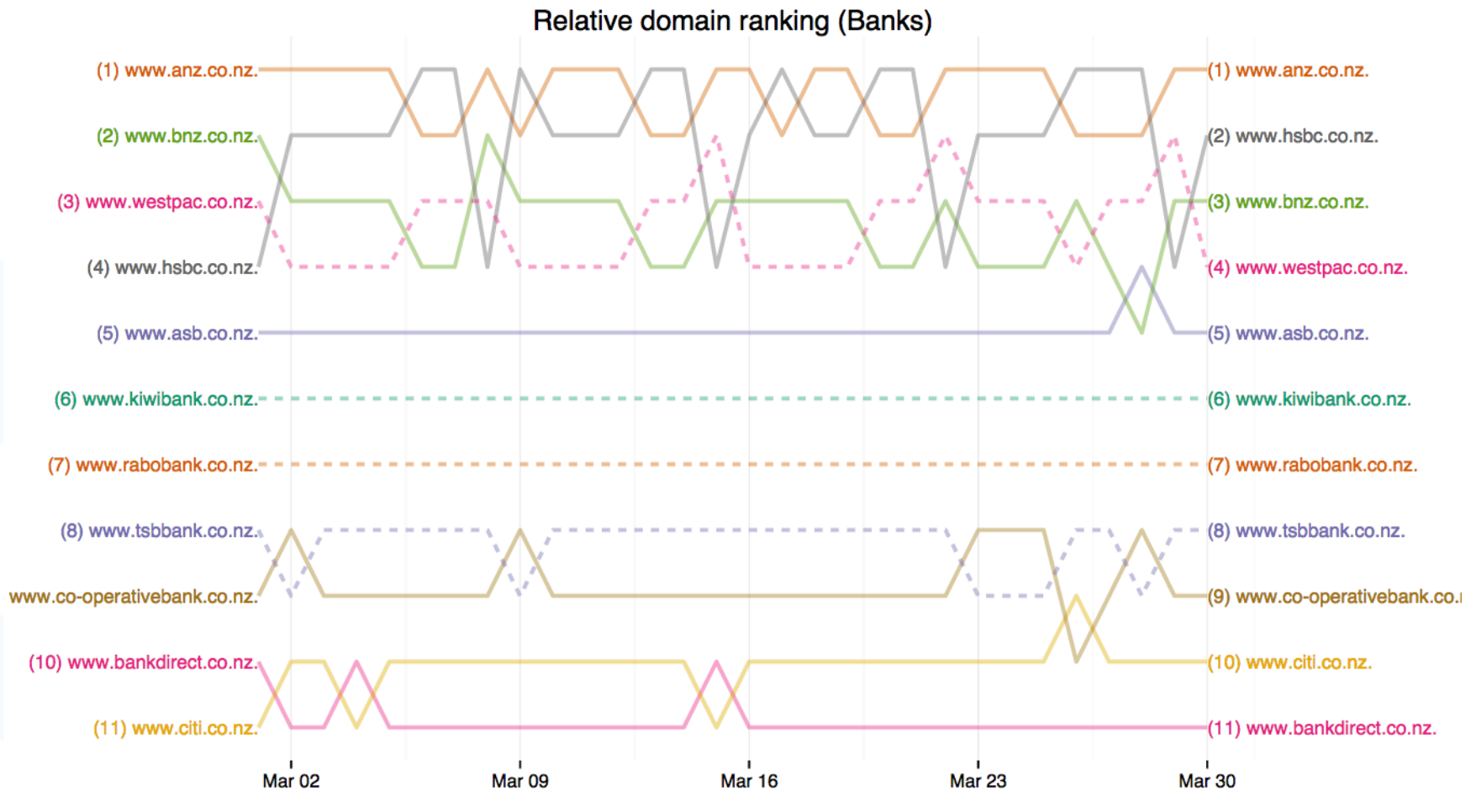
$\text{rank}(d, A) = \#d$ sorted desc by
 $\sum \text{rank}(d, a) \text{weight}(\text{rank}(d, a))$

Pretty pictures

Relative domain ranking (Popular)



More pretty pictures



Interactive visualization demo

- Bump chart, evolution of ranking across time
- Inspired from <http://datatodisplay.com/demonstrations/co2/>
- Bank ranking <http://domain-rank.nzrs.net.nz/bank.html>
- Popular ranking <http://domain-rank.nzrs.net.nz/popular.html>

ToDo List

- Evaluate how TTL affects the results
- Explore other query name and type combinations
- Compare this to some “ground truth”
- Possibly identify addresses by “role” and give different weight to well known sources
- For example, using Verisign’s idea about source address clustering
- Make code available for analysis and visualization

Thanks for your attention

Contact: sebastian@nzrs.net.nz

www.nzrs.net.nz