

# A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists

Quirin Scheitle<sup>1</sup>, Oliver Hohlfeld<sup>2</sup>, Julien Gamba<sup>3,4</sup>, Jonas Jelten<sup>1</sup>,  
Torsten Zimmermann<sup>2</sup>, Stephen D. Strowes<sup>5</sup>, Narseo Vallina-Rodriguez<sup>3,6</sup>

<sup>1</sup>Technical University of Munich (TUM) <sup>2</sup>RWTH Aachen University,  
<sup>3</sup>IMDEA Networks Institute, <sup>4</sup>Universidad Carlos III de Madrid, <sup>5</sup>RIPE NCC, <sup>6</sup>ICSI

## Introduction

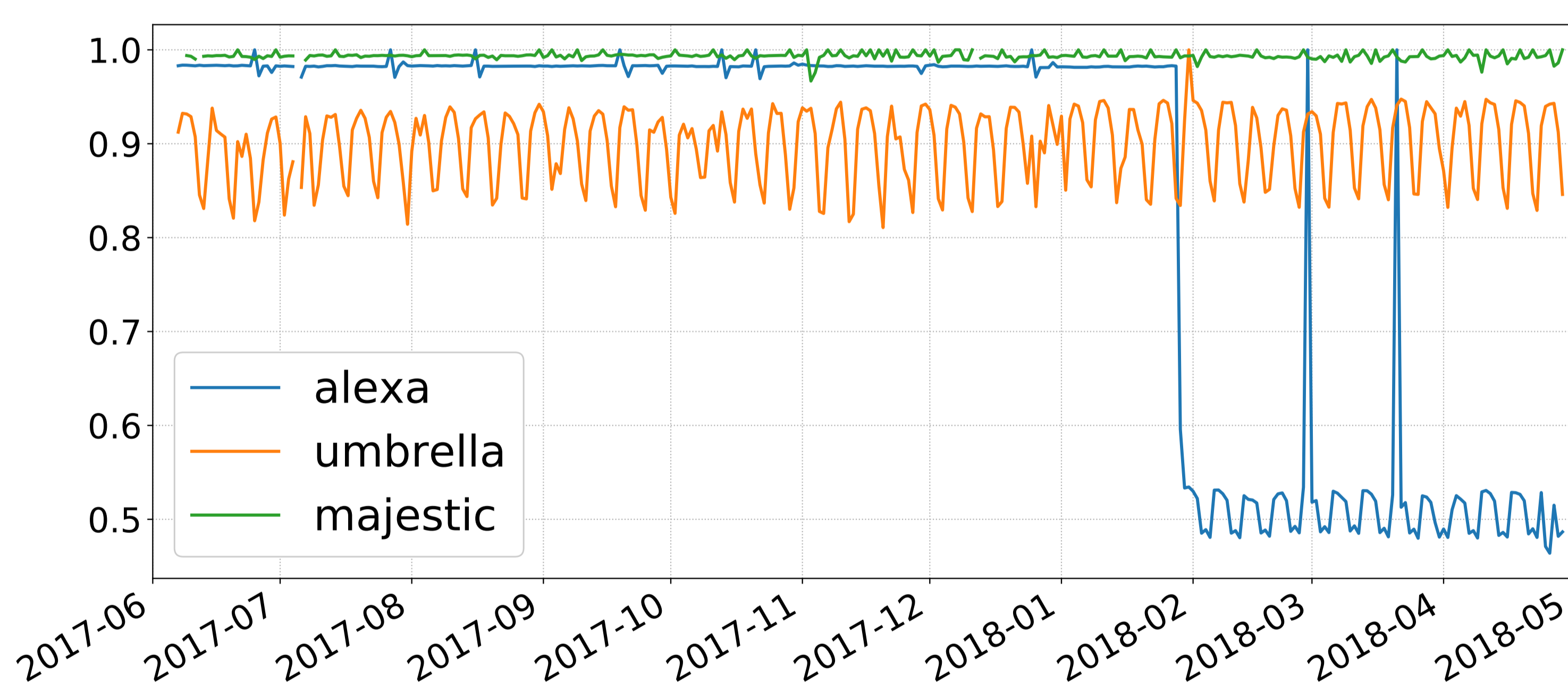
- Internet top lists allegedly give a representative sample of Internet domains in use
- Widely used in literature: 69 measurement or security papers cite them in 2016 and 2017
- Only two of these papers cite the date of list retrieval and date of measurements
- The stability of these lists remains unknown

This raises a number of questions

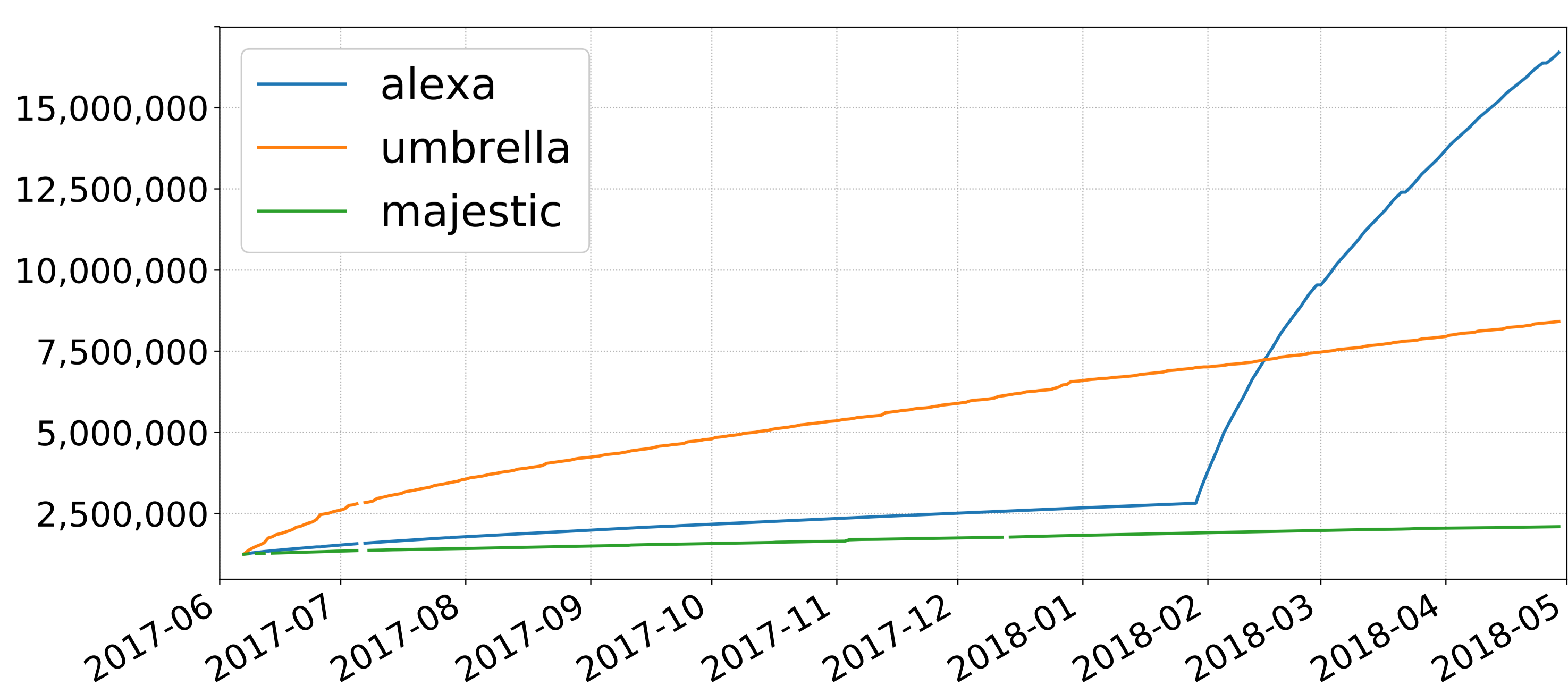
- How stable are these lists?
- How representative of the Internet are they?
- Does using a list has an impact on the results reproducibility?

## Stability

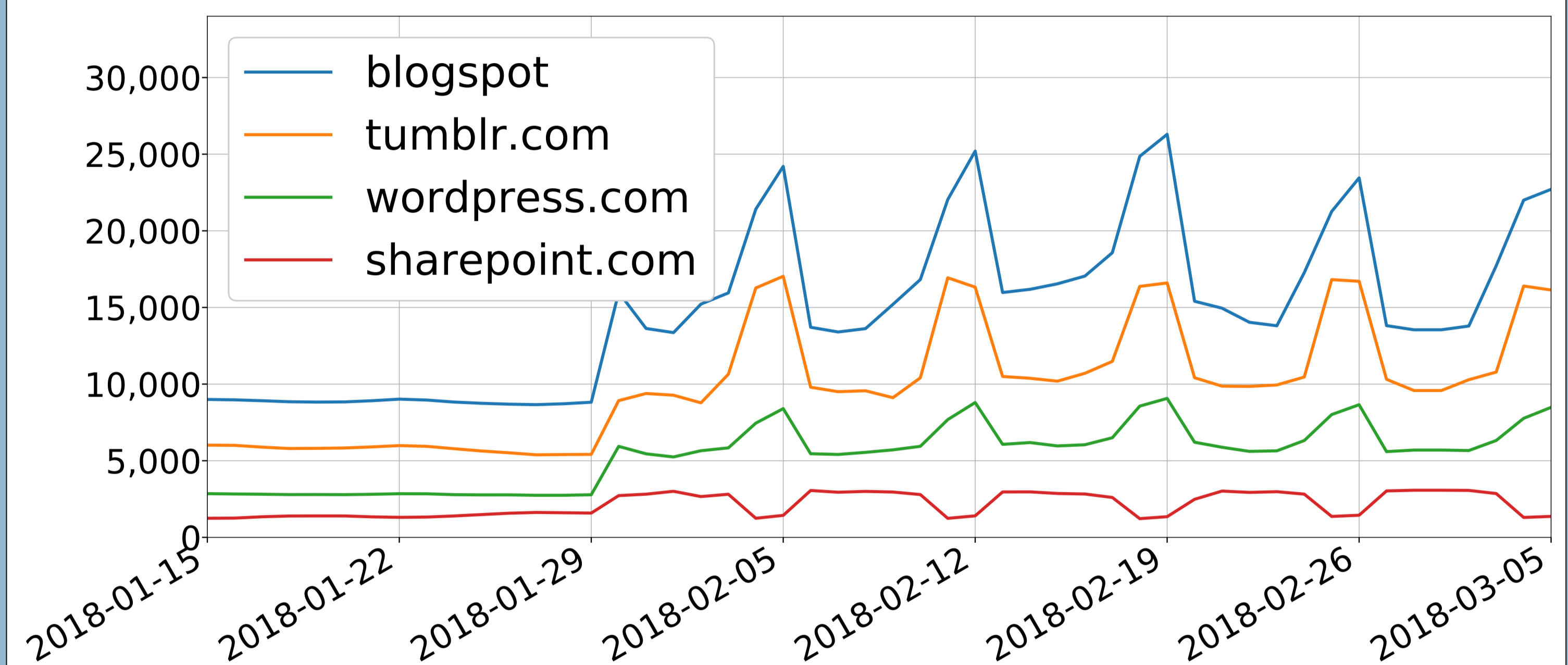
- We find significant changes between two consecutive days in Umbrella and Alexa in 2018
- Majestic stays stable
- Alexa clearly changed its methodology around February 2018



- Daily changes stem in part from new domains appearing in the ranks
- Majestic and Umbrella grow linearly
- Again we see a structural change in the Alexa list in February 2018
- Unstability comes in part from new domains appearing in the top lists



- Ranks change over the week-ends, leisure domains are more visited by users
- The date of the list download can yield a very different ranking



## Reproducibility

Top list, however large, are still a sample of Internet domains, and using a top list can yield biased results, e.g. IPv6 adoption:

- Alexa 1M: ~10% of domains
- Umbrella 1M: ~12%
- .com/.net/.org domains (which represent roughly 45% of the Internet): ~4%

## Conclusions

- Internet top lists construction remains unknown
- Using a list comes with a selection bias
- The choice of a list can mitigate that bias depending on the study

Some recommendations:

- List structure must match study purpose
- Stability should be considered
- List use should be documented
- Biases should be taken into account

## References

Q. Scheitle et. al. *Structure and Stability of Internet Top Lists*, PAM '18 poster session

Alexa. *Top 1M sites*.

<https://www.alexac.com/topsites>

Cisco. *Umbrella Top 1M List*

<https://umbrella.cisco.com/blog/blog/2016/12/14/cisco-umbrella-1-million/>

Majestic, *Majestic Million*

<https://majestic.com/reports/majestic-million/>