

The Morphology of the Web is Changing!



Martin Lopatka

Data Scientist/Applied Statistician



DATA NATIVES

moz://a

Disclaimer

Martin Lopatka provides this contribution to the Data Natives 2018 conference in a personal capacity. The views expressed are his own and do not necessarily represent the views of Mozilla Corporation or the Mozilla Foundation.

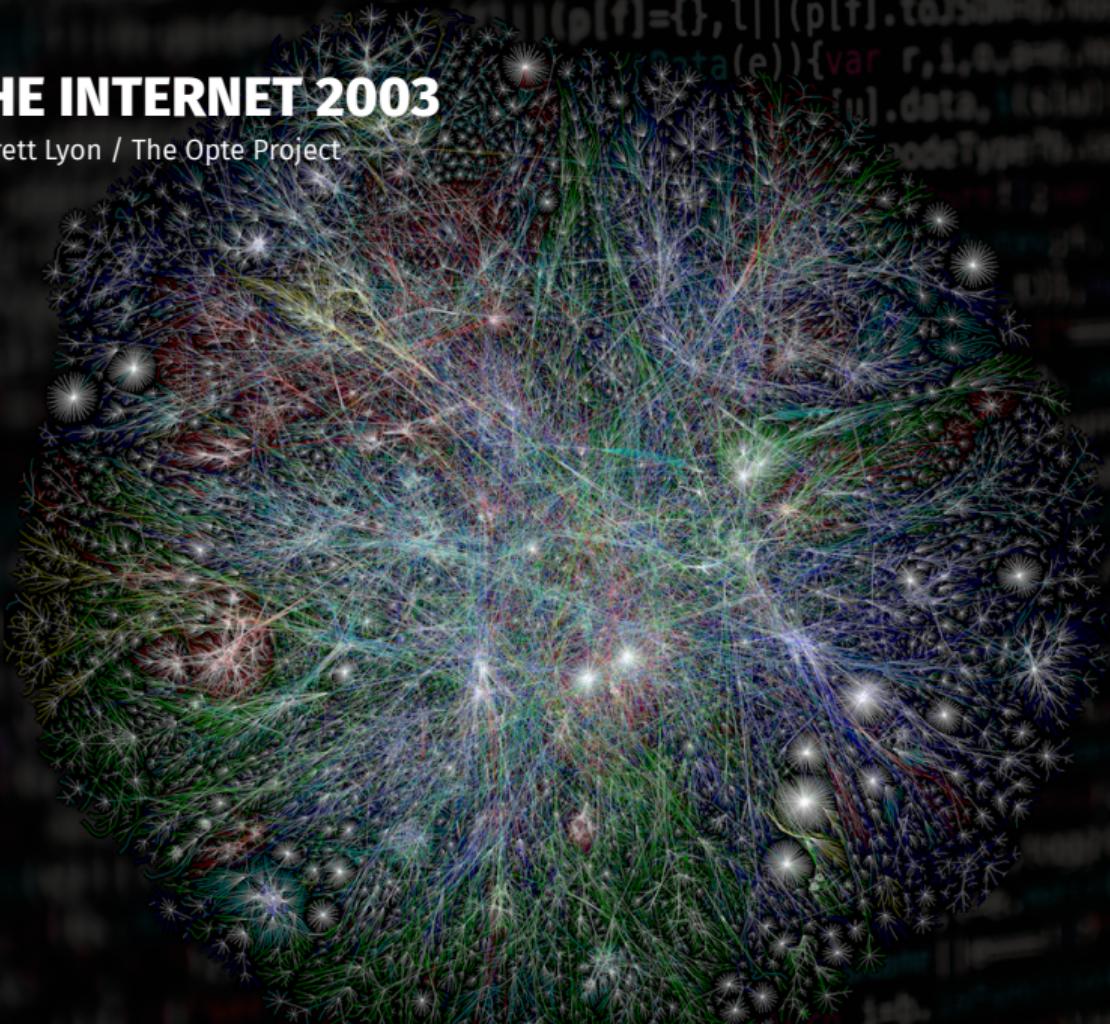


DATA NATIVES

moz://a

THE INTERNET 2003

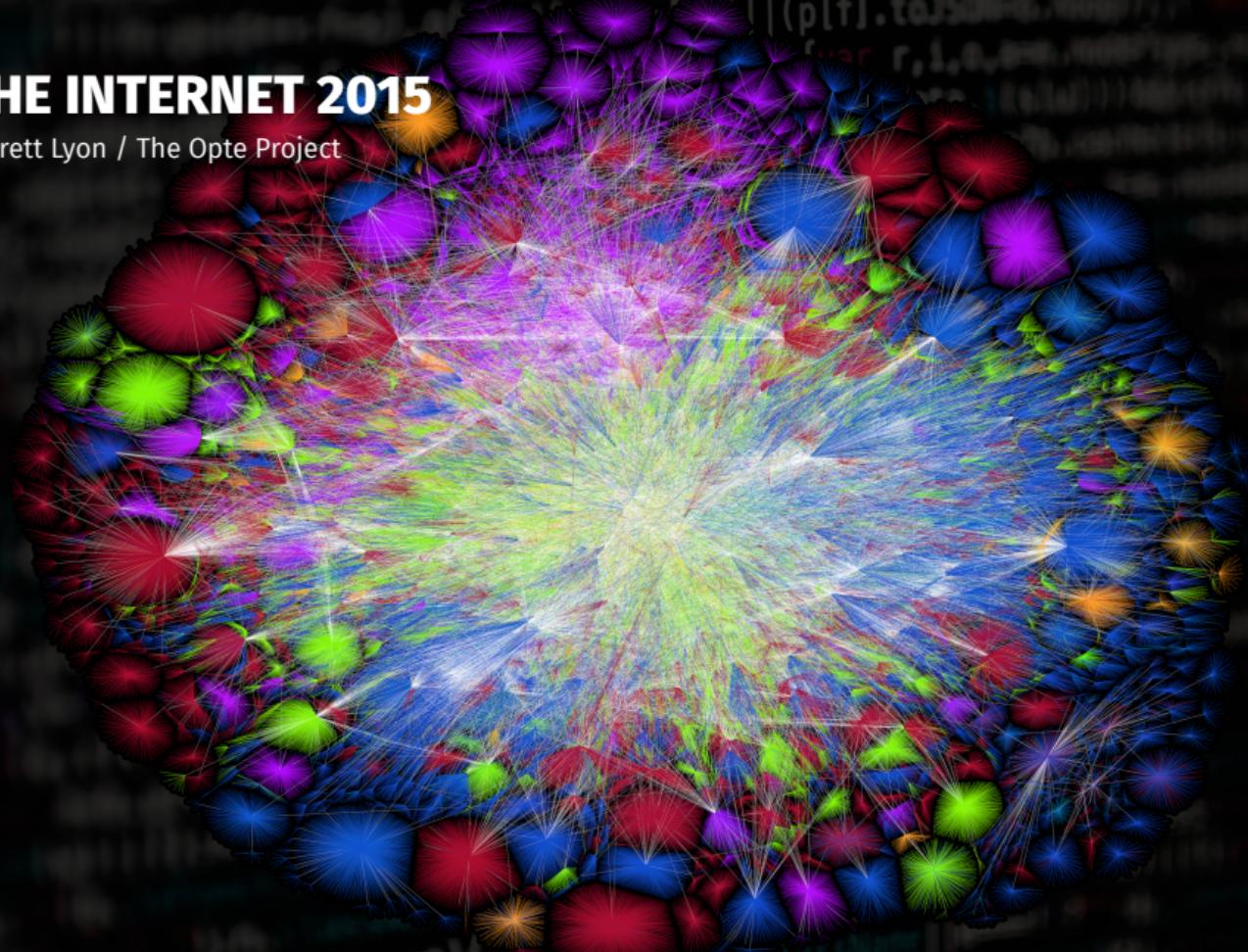
Barrett Lyon / The Opte Project



moz://a

THE INTERNET 2015

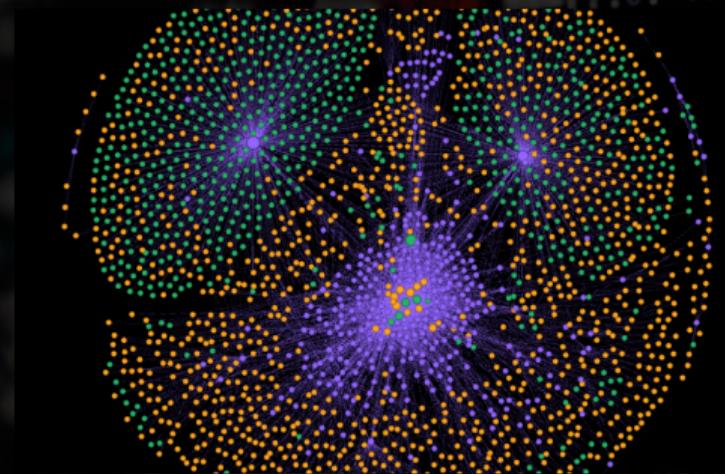
Barrett Lyon / The Opte Project



moz://a

The Centrality of Advertisers

Are trackers the new backbone of the Web?



DATA NATIVES

moz://a

How did we get to an ecosystem of silos?

- ▶ 40% of total Web browsing page views can be attributed to only 65 top level domains (TLD)
- ▶ Five (TLD+1) sites (Google, Facebook, Amazon, Yahoo, and Reddit) make up 22% of all traffic¹
- ▶ Of the 22,310,889 domains processed, 52.63% (11,742,112) were found to serve advertisements²

¹<https://medium.com/firefox-context-graph/are-trackers-the-new-backbone-of-the-web-fb800435da15>

²<https://commeica.com/2018/06/27/web-ad-prevalence/>

Social platforms were designed to facilitate; they became attention brokers, they are designed to help companies reach users as these sort of middlemen where they know a ton about you and that's the service they are providing to the advertisers.

Renee DiResta "The Internet's Original Sin"

27-Oct-2018 12:45



DATA NATIVES

moz://a

The highest trafficked pages on the Web³ are search engines, Social media platforms, and Commerce platforms

- 1. Google.com
- 2. Youtube.com
- 3. Facebook.com
- 4. Baidu.com
- 5. Wikipedia.org *
- 6. Qq.com
- 7. Taobao.com
- 8. Yahoo.com
- 9. Tmall.com
- 10. Amazon.com

³<https://www.alexa.com/topsites>; accessed 20-Nov-2018 11:17

The new ecology of outbound links

- ▶ Retargeting
- ▶ advertisement
- ▶ realtime bidding
- ▶ social linking



DATA NATIVES

moz://a

The problem is that the web is no longer built upon the simple premise of a collection of small static HTML and image files served up with a simple tag structure and readily parsed with a few lines of code. Today's web is richly dynamic, multimedia and increasingly broken into walled gardens and device-specific parallel webs.

Kalev Leetaru "Are Web Archives Failing The Modern Web: Video, Social Media, Dynamic Pages and The Mobile Web"

24-Feb-2017 22:41



DATA NATIVES

moz://a

The problem is that the web is no longer built upon the simple premise of a collection of small static HTML and image files served up with a simple tag structure and readily parsed with a few lines of code. Today's web is richly dynamic, multimedia and increasingly broken into walled gardens and device-specific parallel webs.

Kalev Leetaru "Are Web Archives Failing The Modern Web: Video, Social Media, Dynamic Pages and The Mobile Web"

24-Feb-2017 22:41



DATA NATIVES

moz://a

Challenges to crawl technology

Introducing Firefox Pioneer!⁴



⁴<https://medium.com/firefox-context-graph/make-firefox-better-with-pioneer-10c82d0f9301>

Comparison fo Javascript execution between two people and a crawler for the same page!

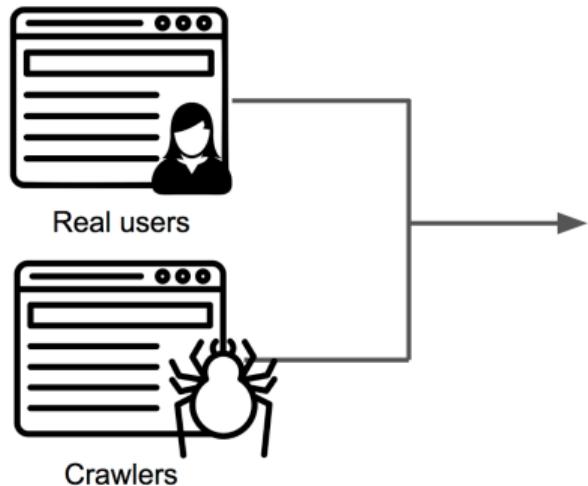


DATA NATIVES

moz://a

Web study stack

Detect invasive scripts



Selective, default-on
Tracking Protection



DATA NATIVES

moz://a

transparency, technology, education



DATA NATIVES

moz://a

Overscripted!

[https://github.com/mozilla/
overscripted/](https://github.com/mozilla/overscripted/)



DATA NATIVES

moz://a