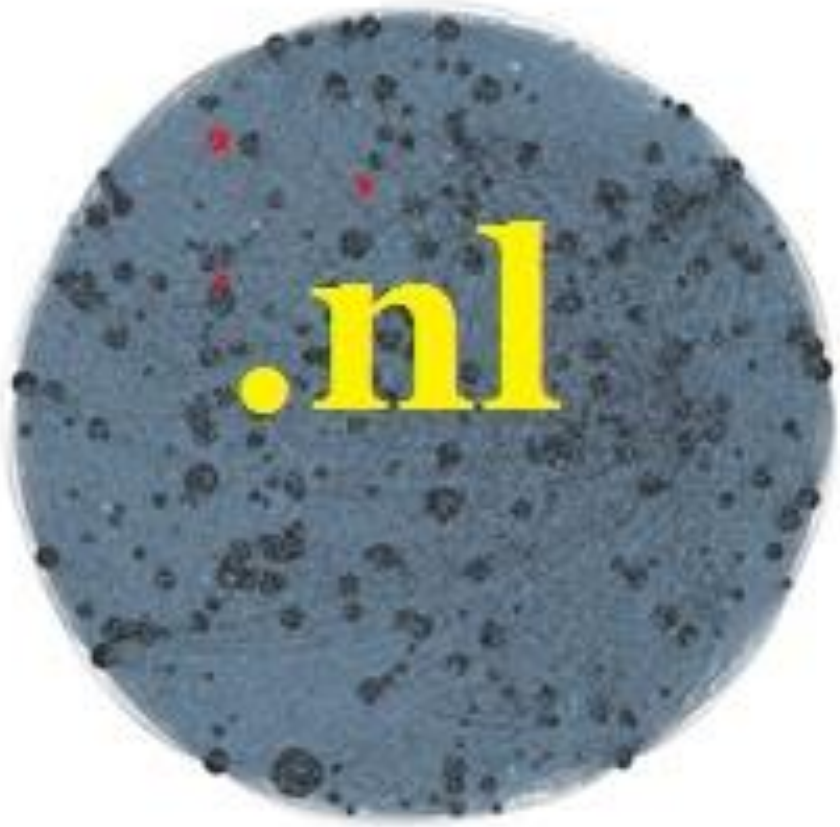
Koninklijke Bibliotheek
Nationale bibliotheek van Nederland



How to harvest born digital conspiracy theories: webarchiving Dutch digital culture in the post-truth era

Dr Kees Teszelszky
Curator digital collections Koninklijke Bibliotheek

www.kb.nl
@keesone

Aim of web archiving @KB-NL

"To select, preserve and make accessible a representative set of websites of the Dutch national web domain"

- Since 2007 15,550 sites selected
- Growth 1,500 sites / year
- Largest born digital web archive in the Netherlands, 35 TB
- No legal deposit in the Netherlands: only on site access

Why this web collection?

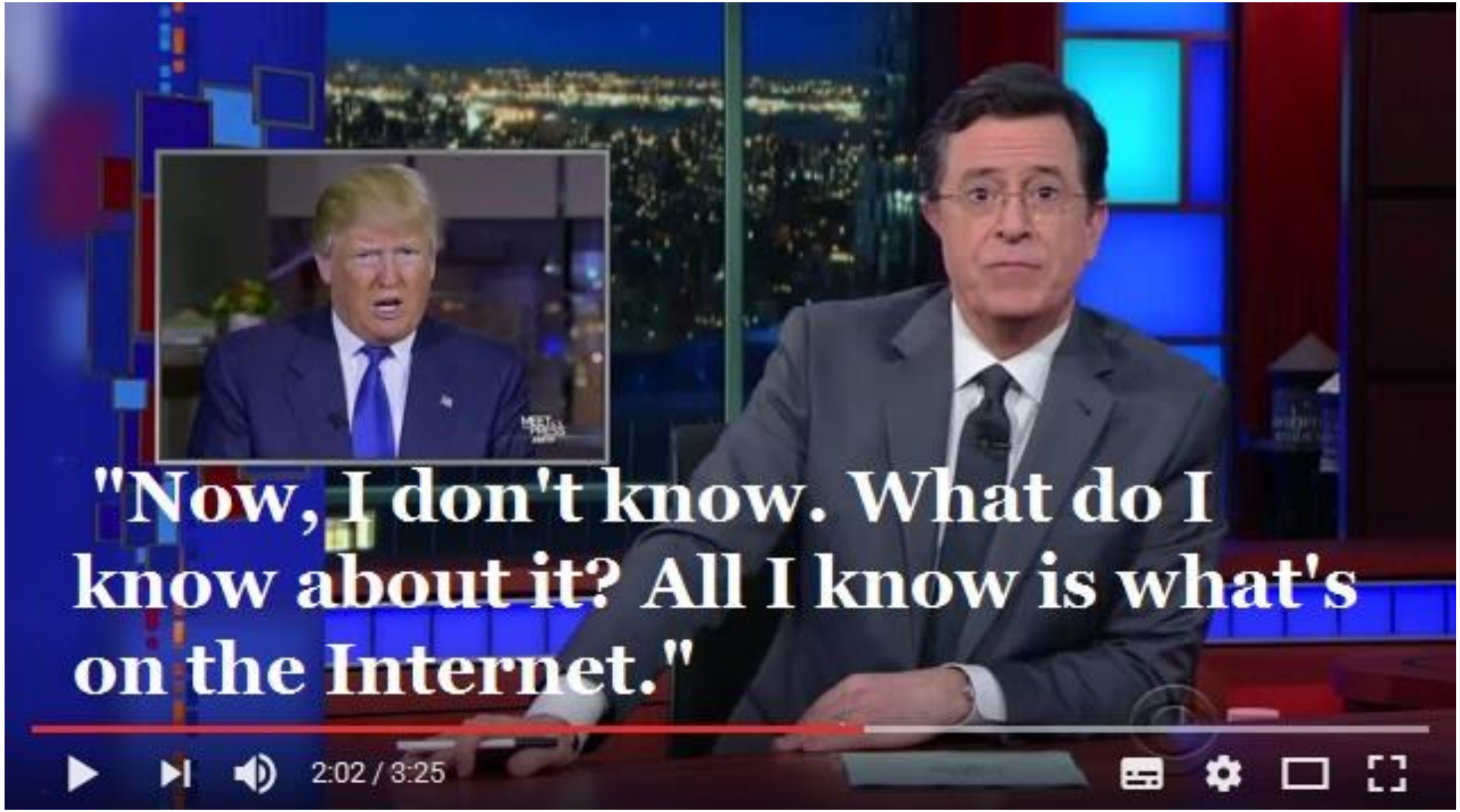
KB-NL: "**web archive of innocence**"

"**Librarian Filter Bubble**" in selective web archiving

- Much web sites on history with limited use as **primary source** for future research of history of our time
- We need a representative set of websites of our time outside of our own "**digital comfort zone**" as librarians, historians and web archivists.

Term (frequentie)	Frequentie in domeinnaam (aantal)
Muse(um, a, ale, etc.)	453
Histor(ie, isch etc.) / Geschie(denis)	198 / 46 = 244
Sticht(ing)	277
Nederland(se etc)	197
Verenig(ing)	158
Kring	89
Heem(kunde, etc.)	75
Boek(en etc.)	67
Fonds	65
(werk, studie)groep	59
Cult(uur, urele, etc.)	46
Genoots(chap)	46
Club	27

"All I know is what's on the internet"



What is actually on the Dutch web:

The Netherlands anno 2018
@Post-Truth era

Follow

Conspiracy theories!!!1!11!!! Fake News!!!1!
Trolling!11!!! Hate speech!1! Extremism!!1!

05:48 PM - 6 June 2018

Selection for the future must be done now

- No time for "**historic distance**" to source material (90 days average)
- Differs from books and other tangible archive sources: only one copy online, almost no "private collections", no offline treasure chests
- Material is sensitive
- Amount of data is huge

Method of making corpus

- PhD thesis Jelle van Buuren, University of Leiden on **online conspiracy theories**, article Jaron Harambam and Stef Aupers, Erasmus University;

  → **Linkextraction** (all hyperlinks in publications);

  → Bring back links to level of **websites**, selection relevant **hyperlinks**;

  → Construction **link cloud** (web sphere!)

  and extraction and selection **Dutch links** from cloud:

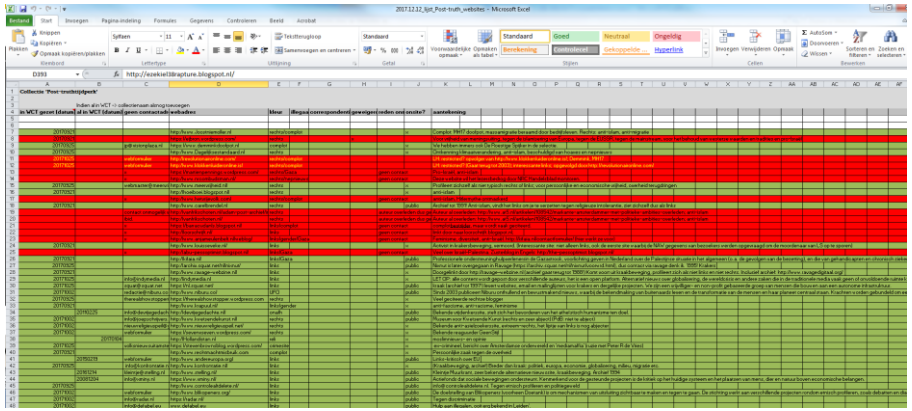
  → **research data set**

- Using suggestions of others (KB, external experts, even conspiracy theorists…)
- Collection of **500 sites**
- **Not all sites will be or can be archived due to legal restrictions, but data remains available for researchers.**

Collections within collection

Themes within collection

- Conspiracy | Anti-Conspiracy
- Spirituality | Flat Earth | End time
- Alternative media | Fake news
- Health | Nature | Gender
- Gaza | Israel
- Anti-Government | Freedom
- Left-wing | Right-wing
- Religion | Sects
- Crime sites
- Climate change denial
- Aspects of fundamentalism



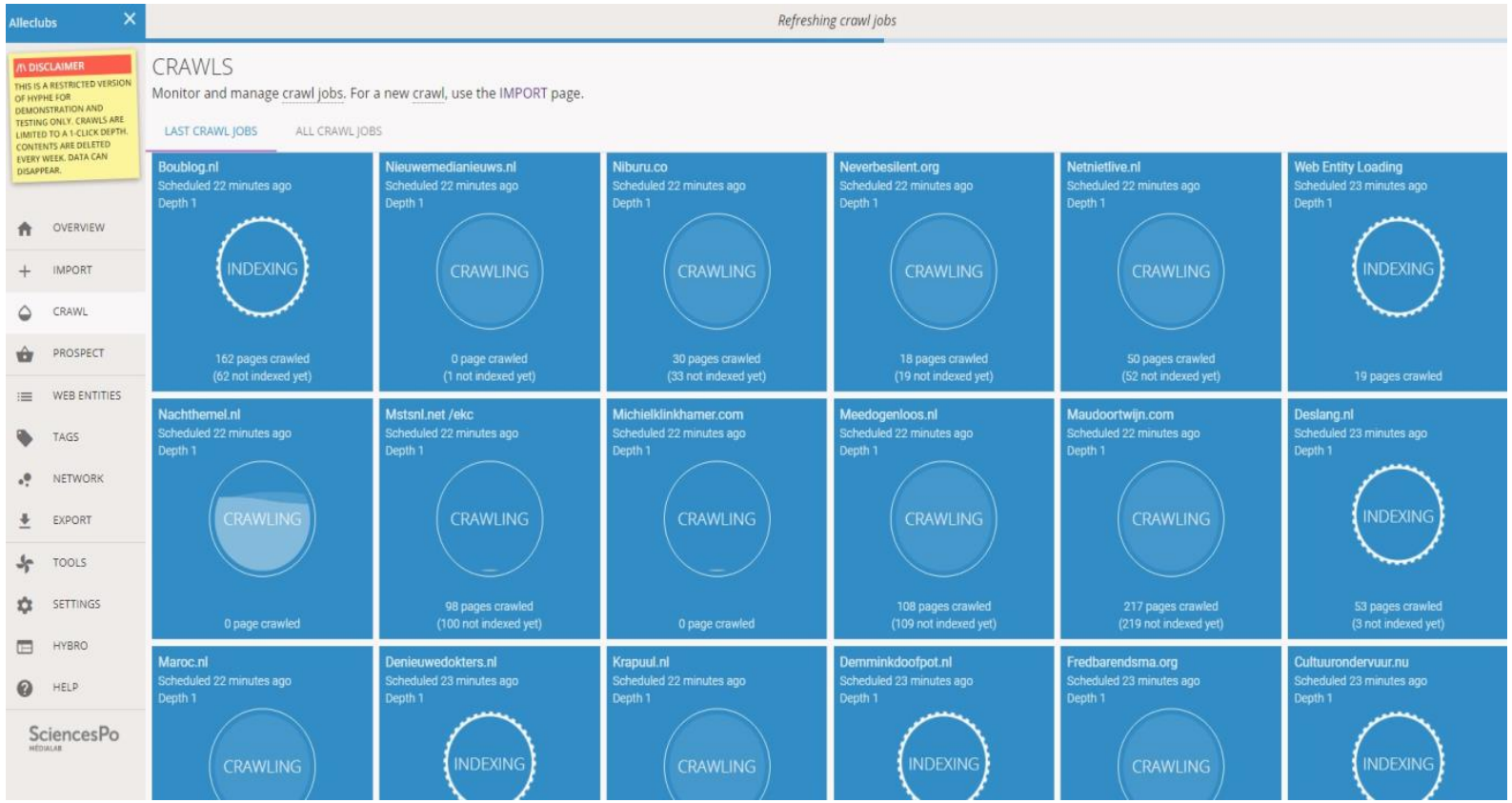
Common theme: mistrust of government

Problems of web archiving Post truth

- Recognition by National Library
- Opt-out message
- Illegal content
- Sites soon offline
- Technical issues
- Refusal

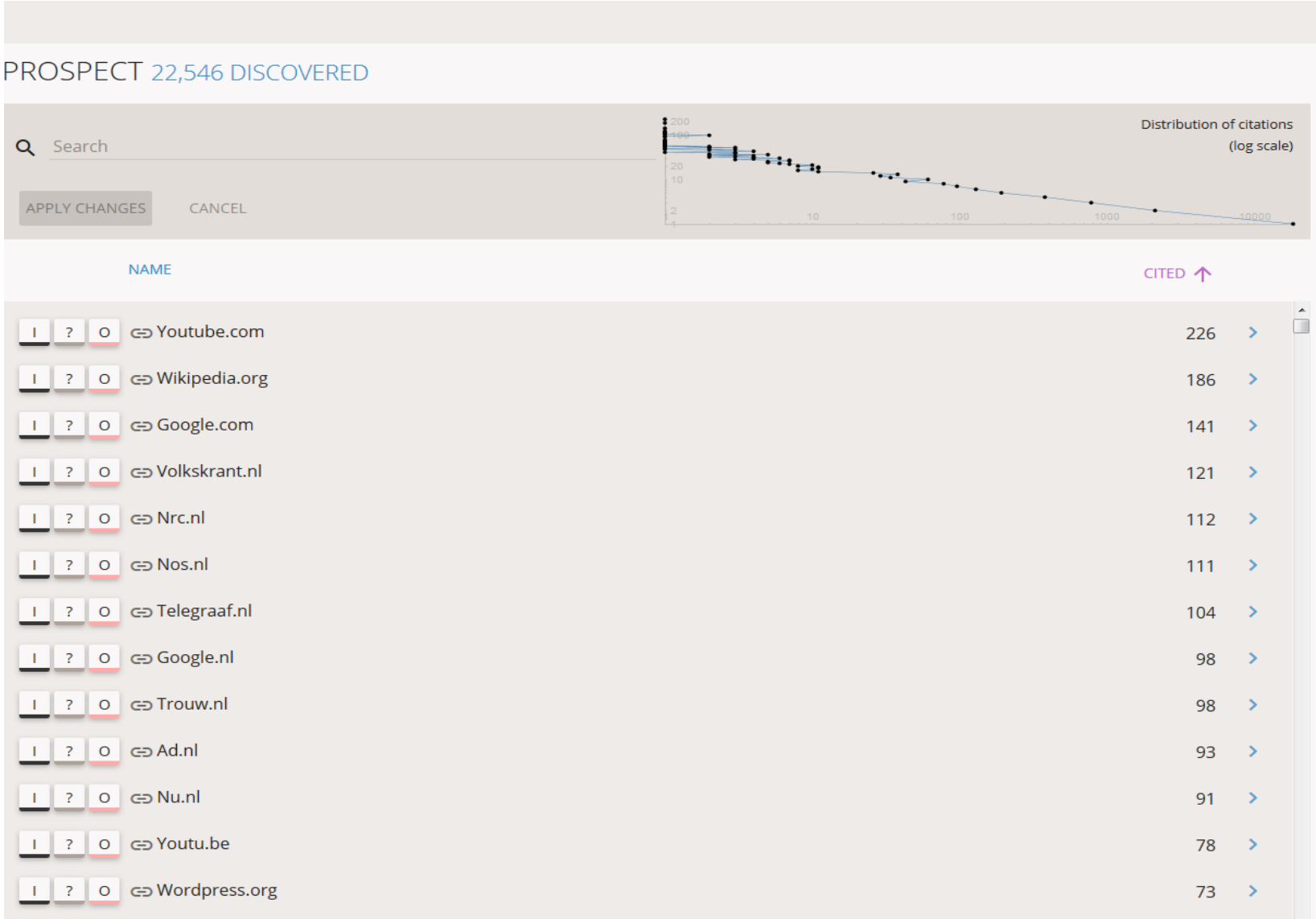
Link analysis

- Hyphe
- Depth: one level
- Overview most linked sites, network



http://hyphe.medialab.sciences-po.fr/demo/#/project/post/network

Most quoted sites



- Corpus: 500 sites
- 5,665 links to Youtube on 226 sites
- Much links to

  ‘Main Stream Media’:

Wikipedia, newspapers, online news

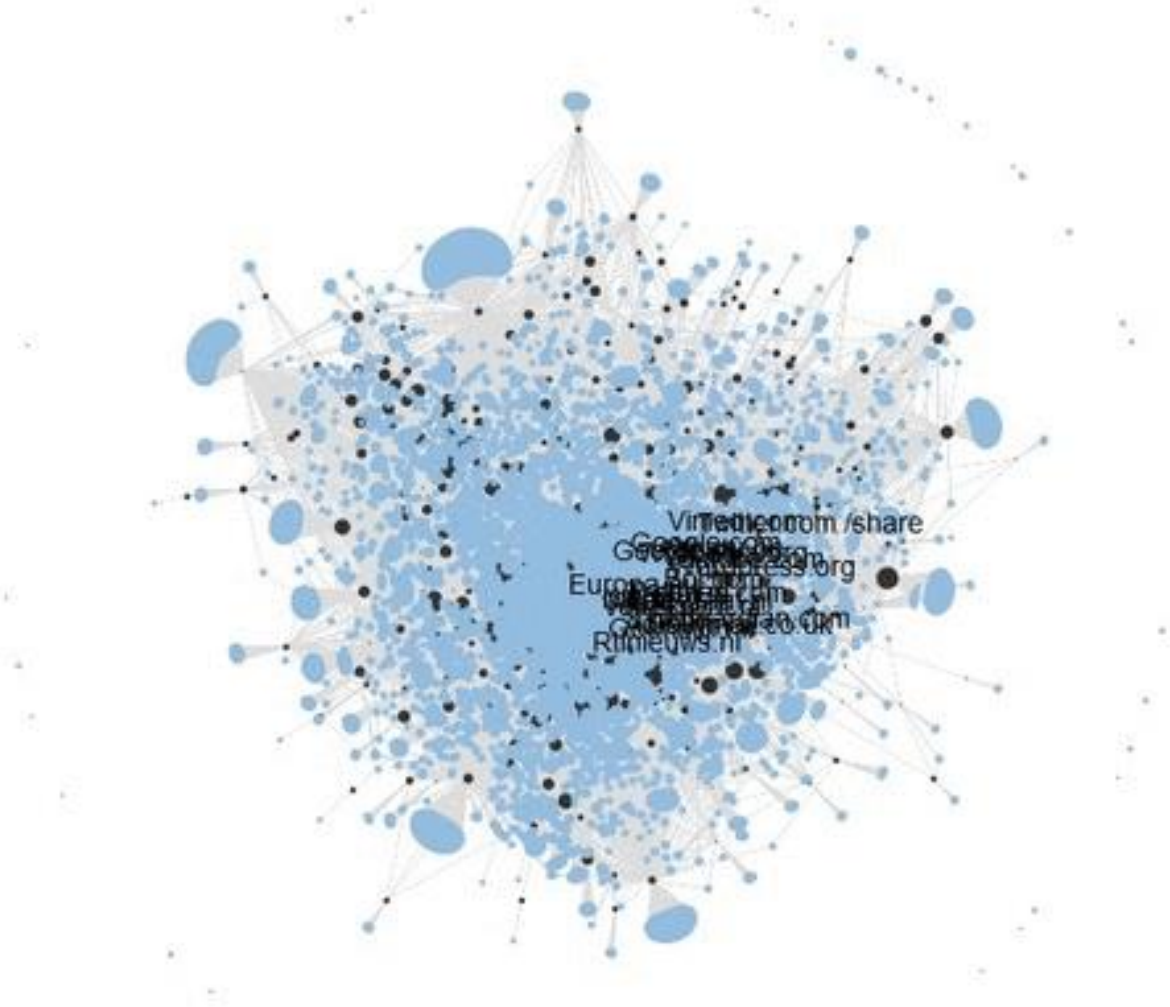
network analysis: internal links

- Shock log Geenstijl.nl most quoted site in collection

network analysis: external links

network analysis: external links

http://hyphe.medialab.sciences-po.fr/demo/#/project/maatschappijkrit/network

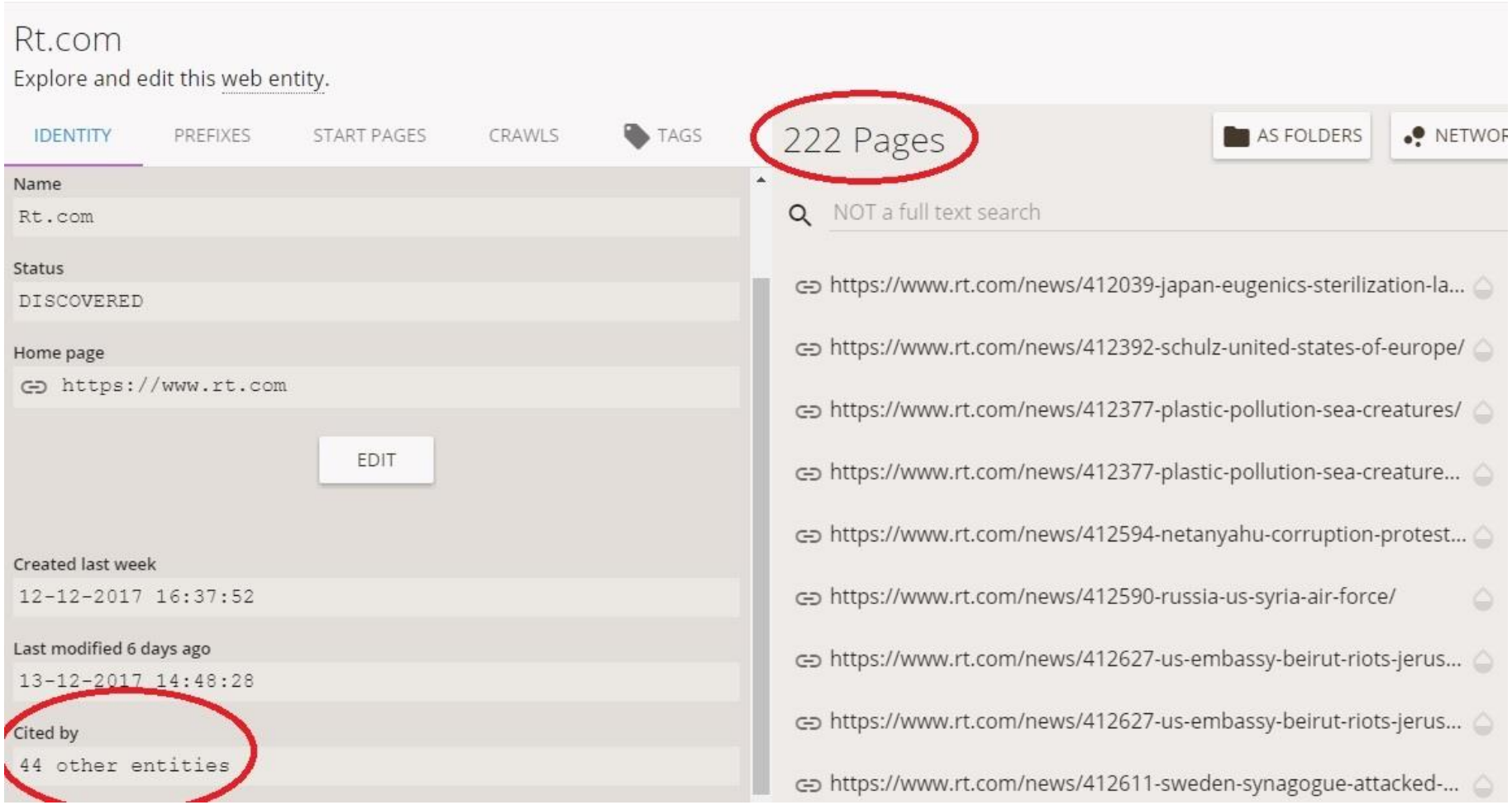


Linkanalyse

Network analysis

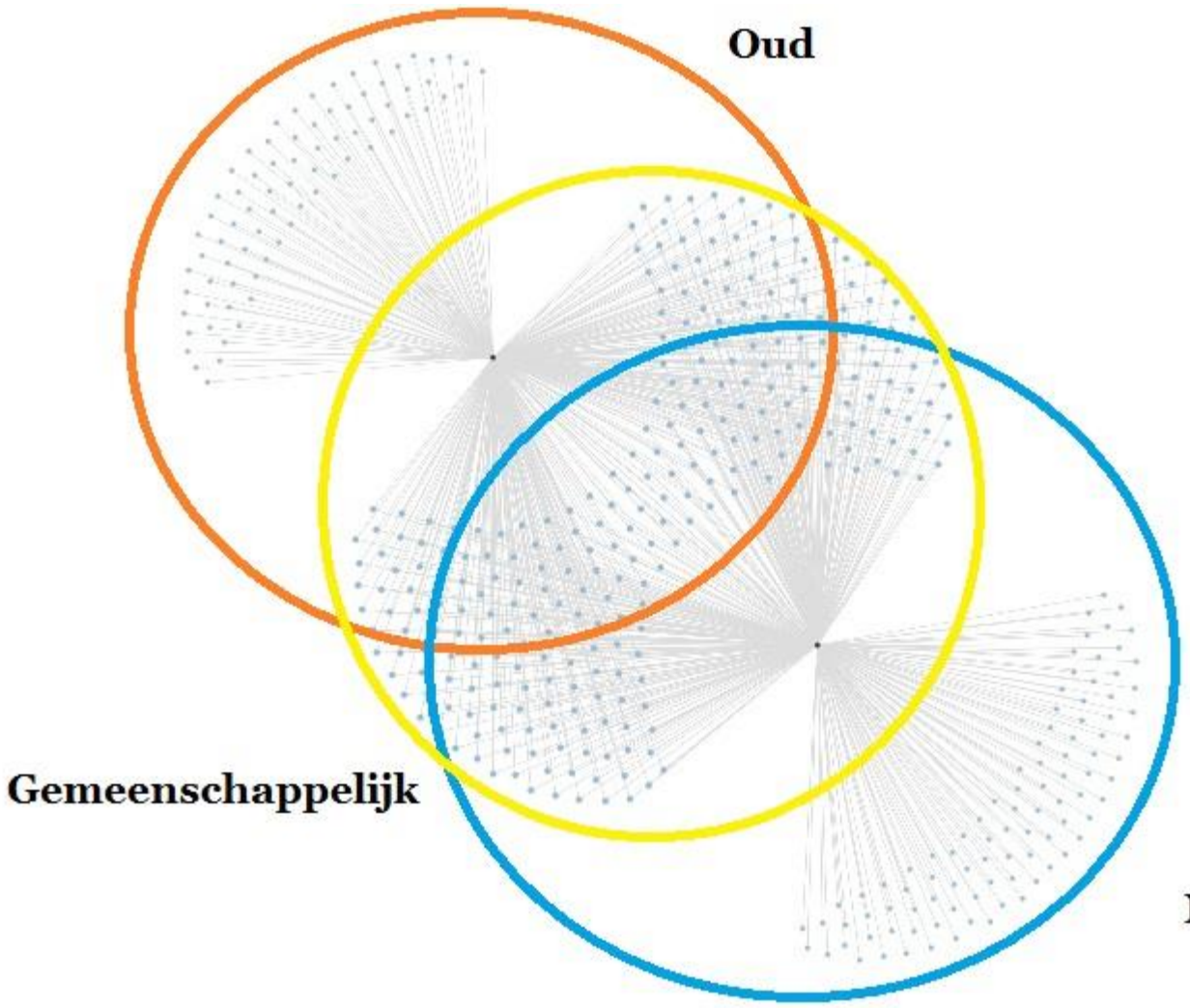
- Digitised library collections become part of the web

Network analysis: Russian links



- Rt.com (Russia Today)
- Russia-Insider.com
- Pravda.ru
- Mid.ru
- Kremlin.ru
- Narod.ru
- Ria.ru
- Government.ru
- Kp.ru
- Orwell.ru
- Posoh.ru

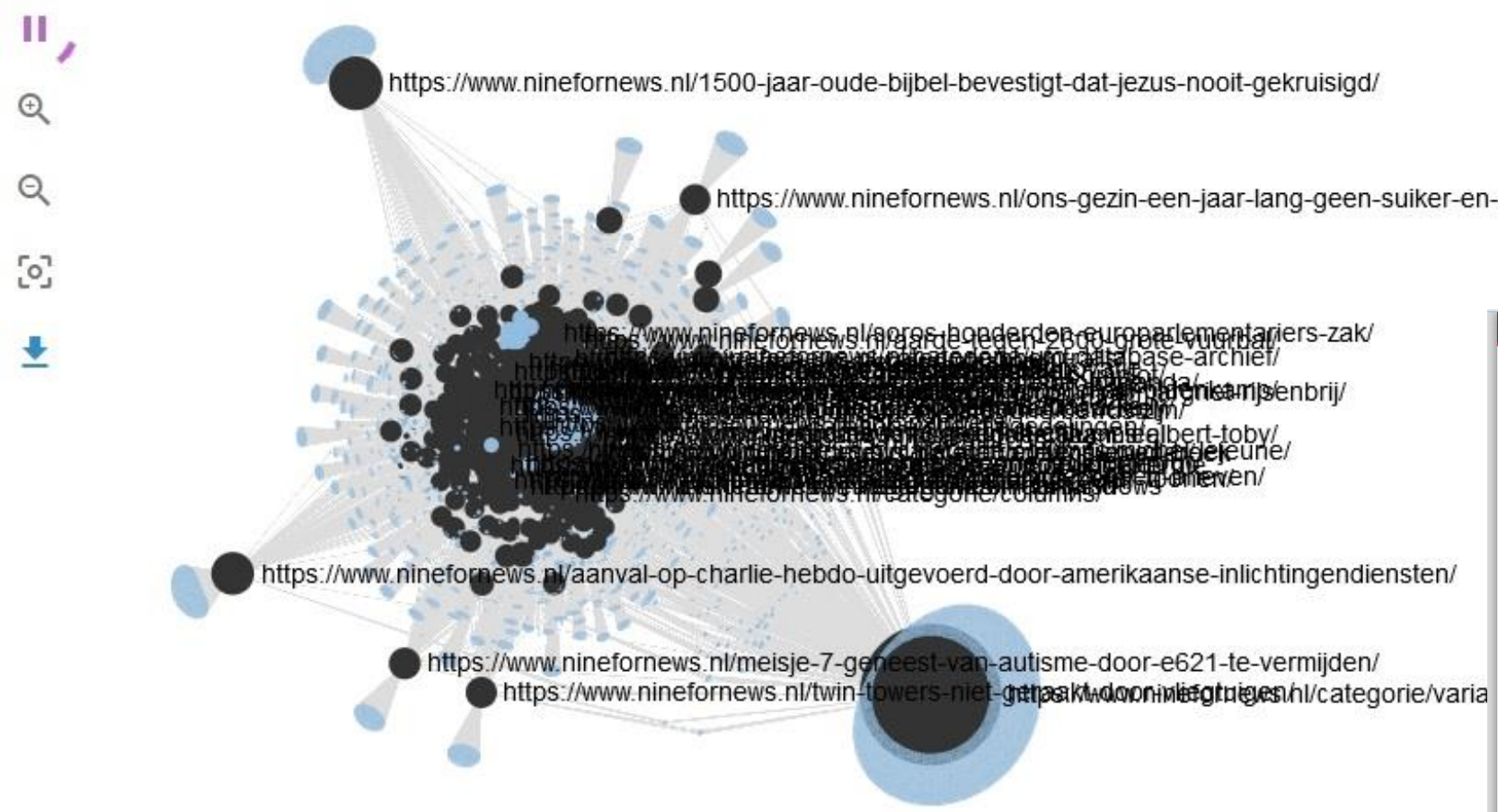
Link analysis



http://members.ziggo.nl/henryv/index.htm (oud) http://www.openbaring.org/ (nieuw)

Link analysis

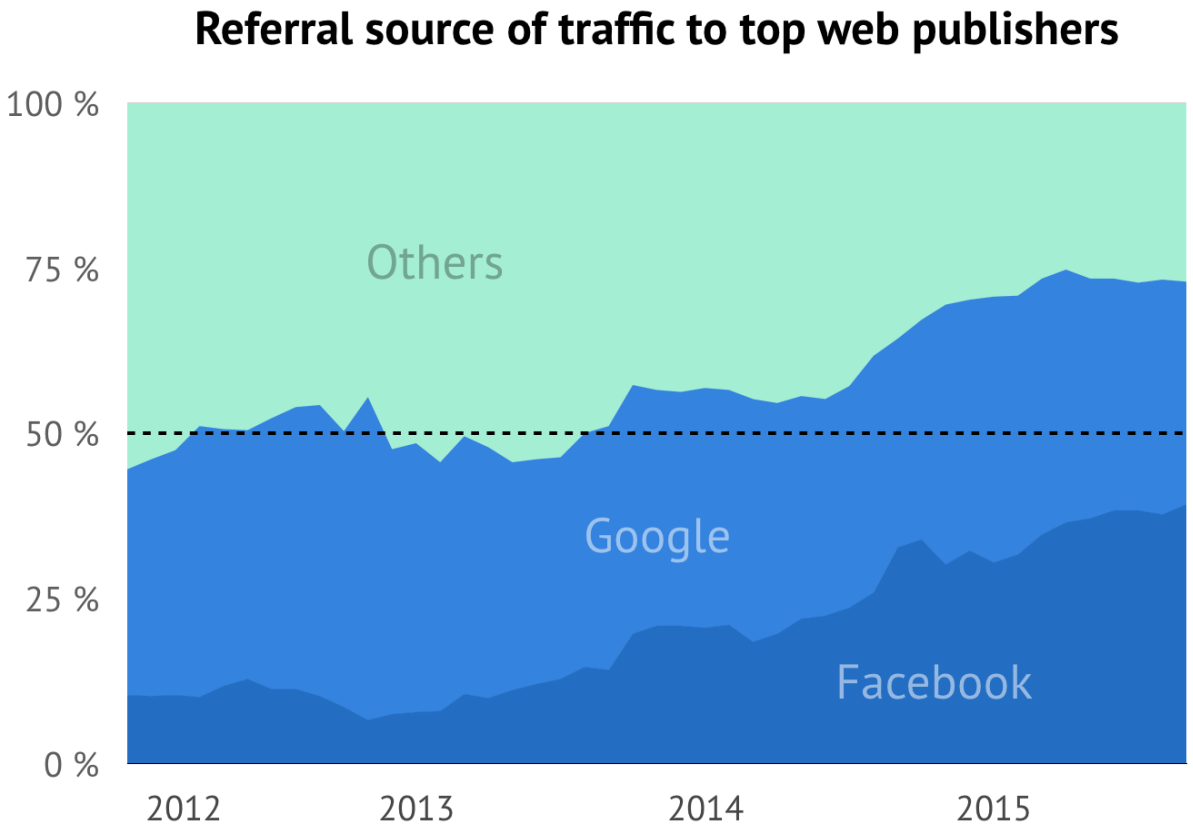
Pages network: Ninefornews.nl



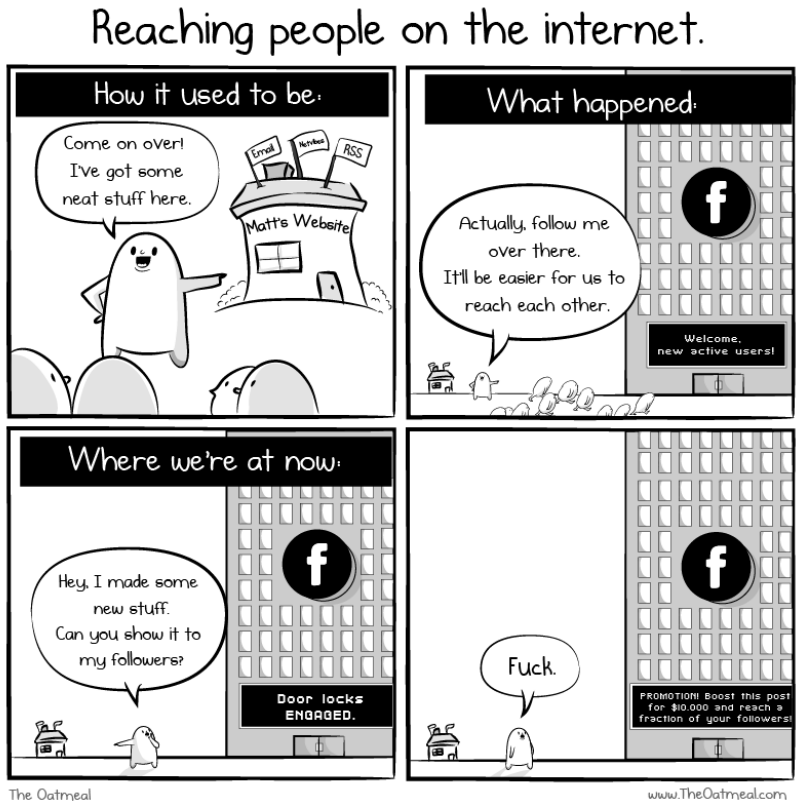
Content analysis on basis of links and web nodes

Trends on the web

Sociale media and apps grow on expense of websites
Content dissapears, cannot be archived by user or archive

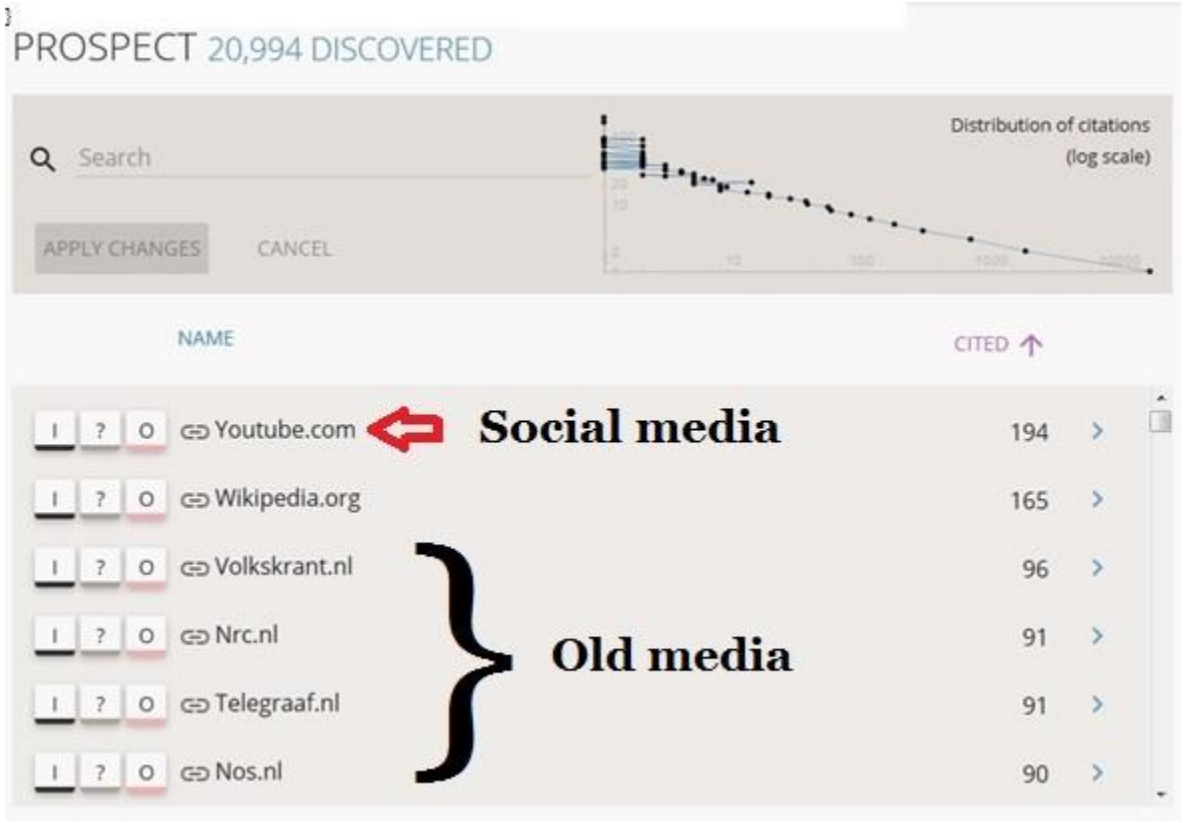


Source: https://blog.parse.ly/post/2855/facebook-continues-to-beat-google-in-sending-traffic-to-top-publishers/



Source: http://theoatmeal.com/comics/reaching_people

Link analysis



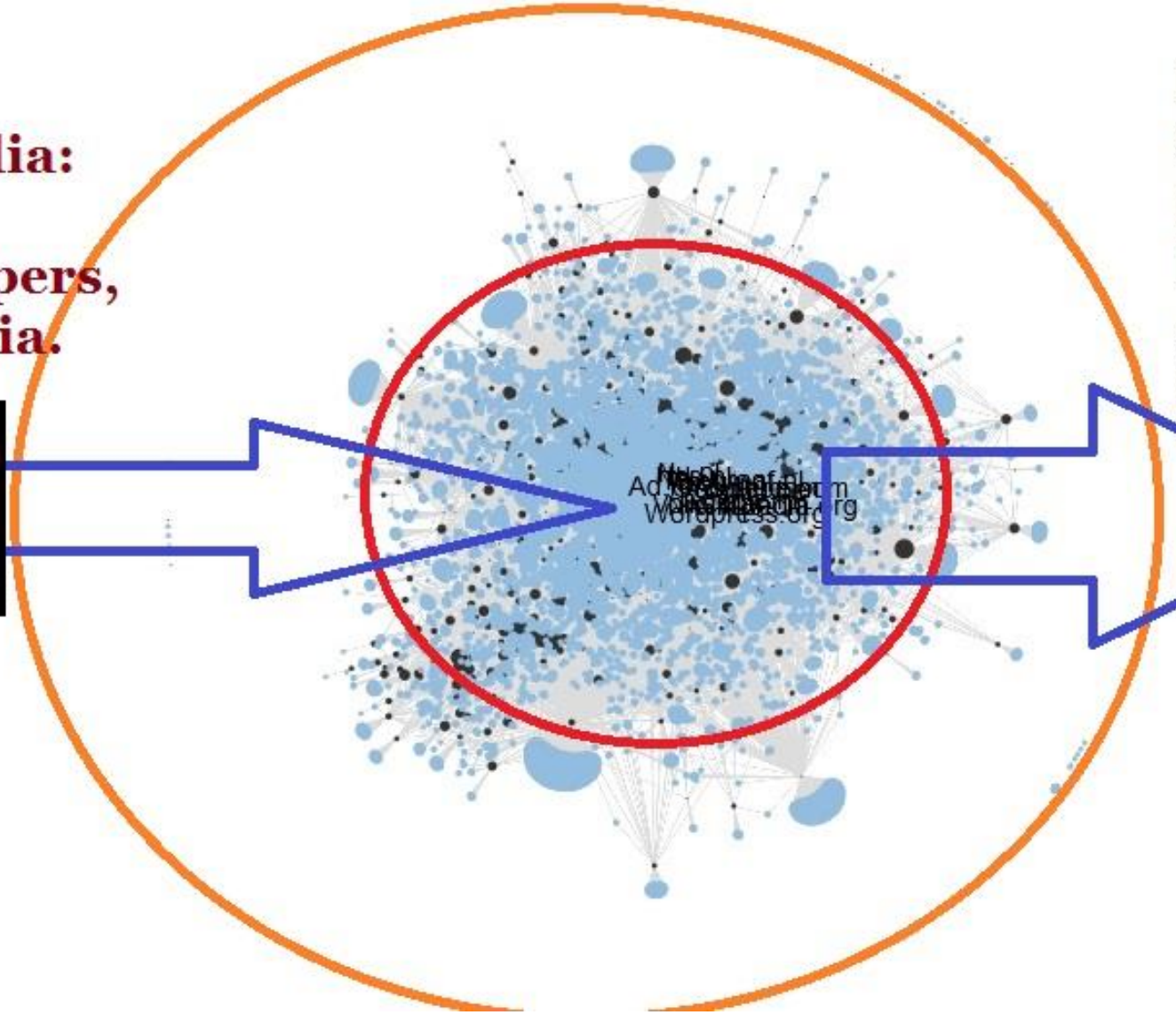
All links, whole collection

All links, leftwing sites

Trends: web in 4D

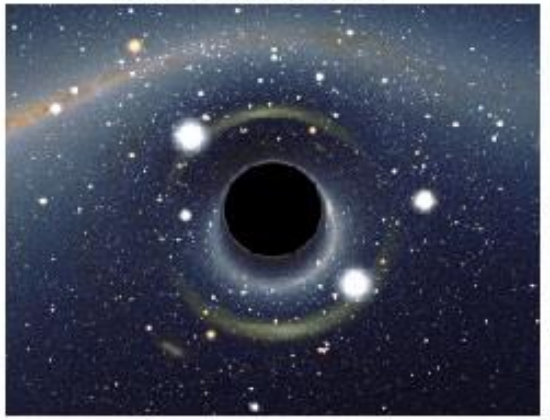
Old media: Books, newspapers, wikipedia.

"Sun"

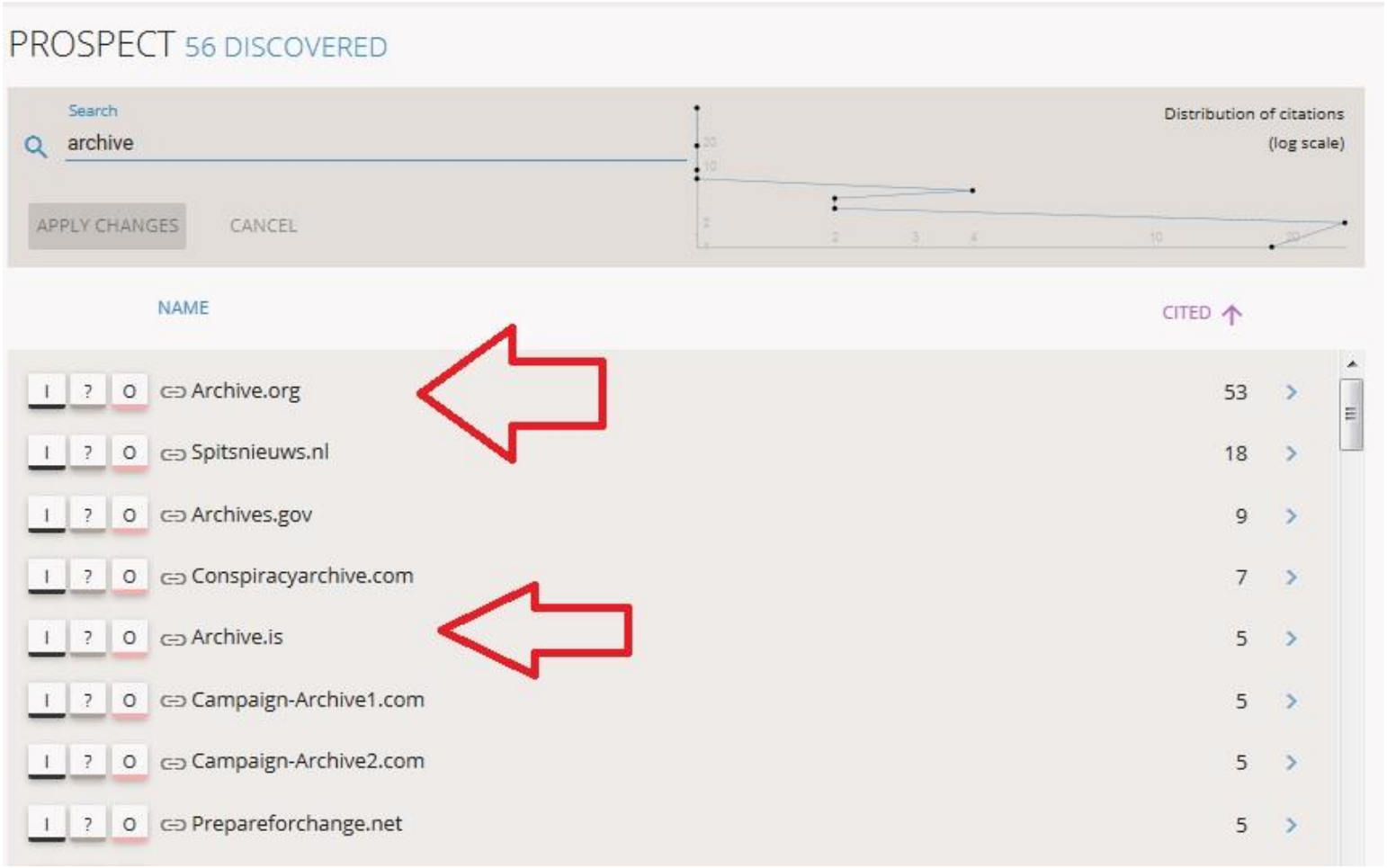


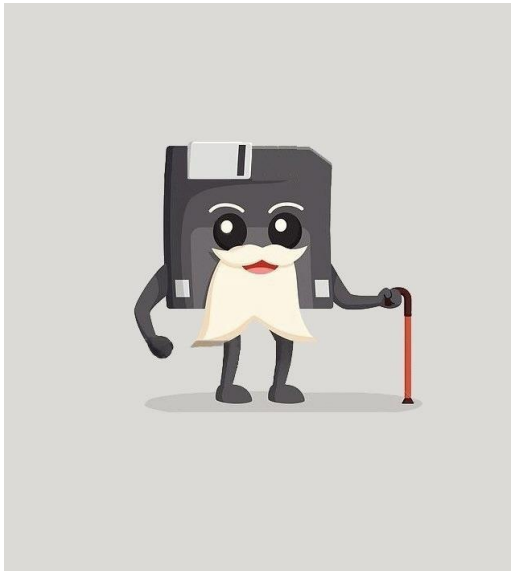
New media: Social media (Youtube, Facebook) Apps (snapchat)

Digital black hole: cannot be archived.



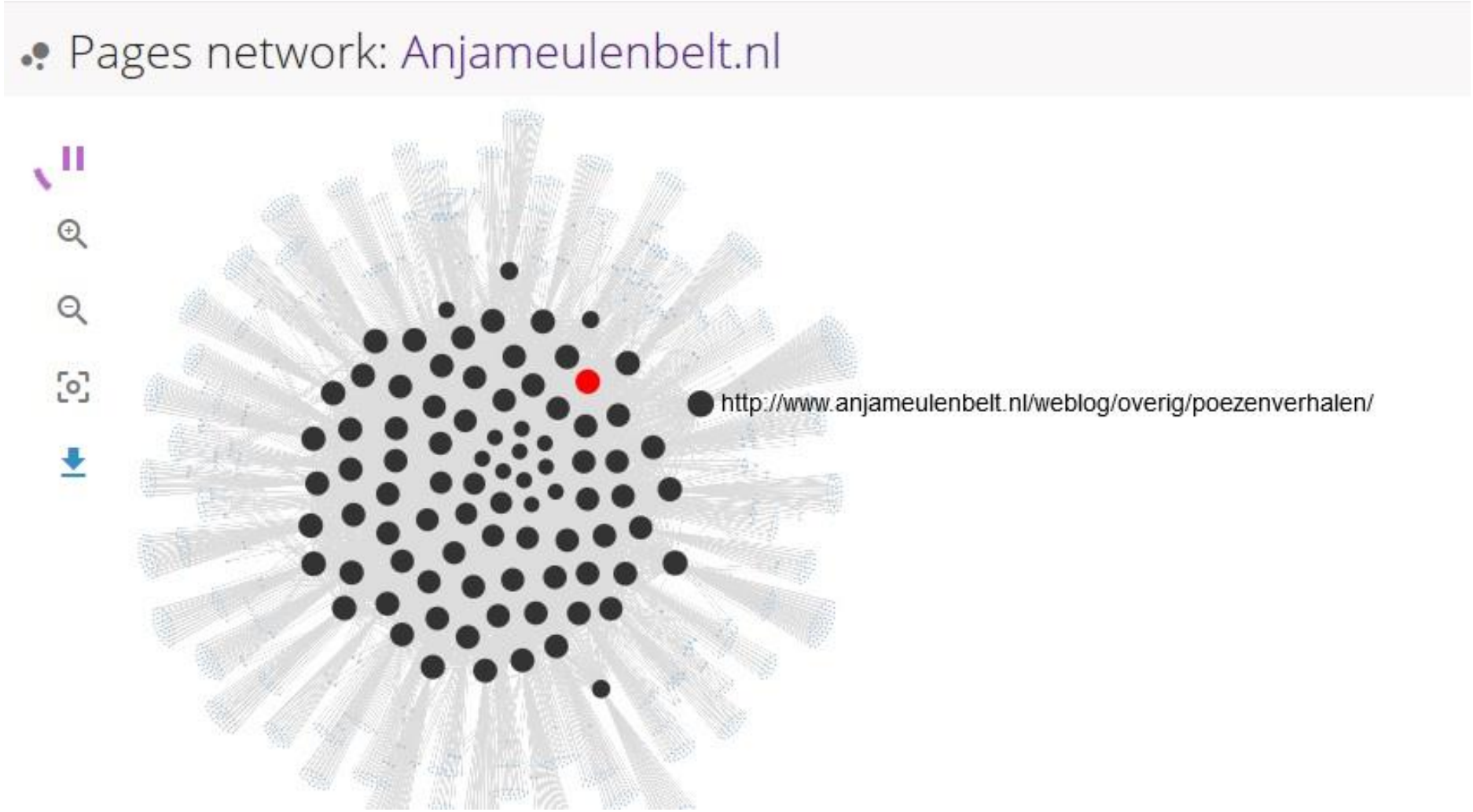
Trend: web archives become part of Post-truth web





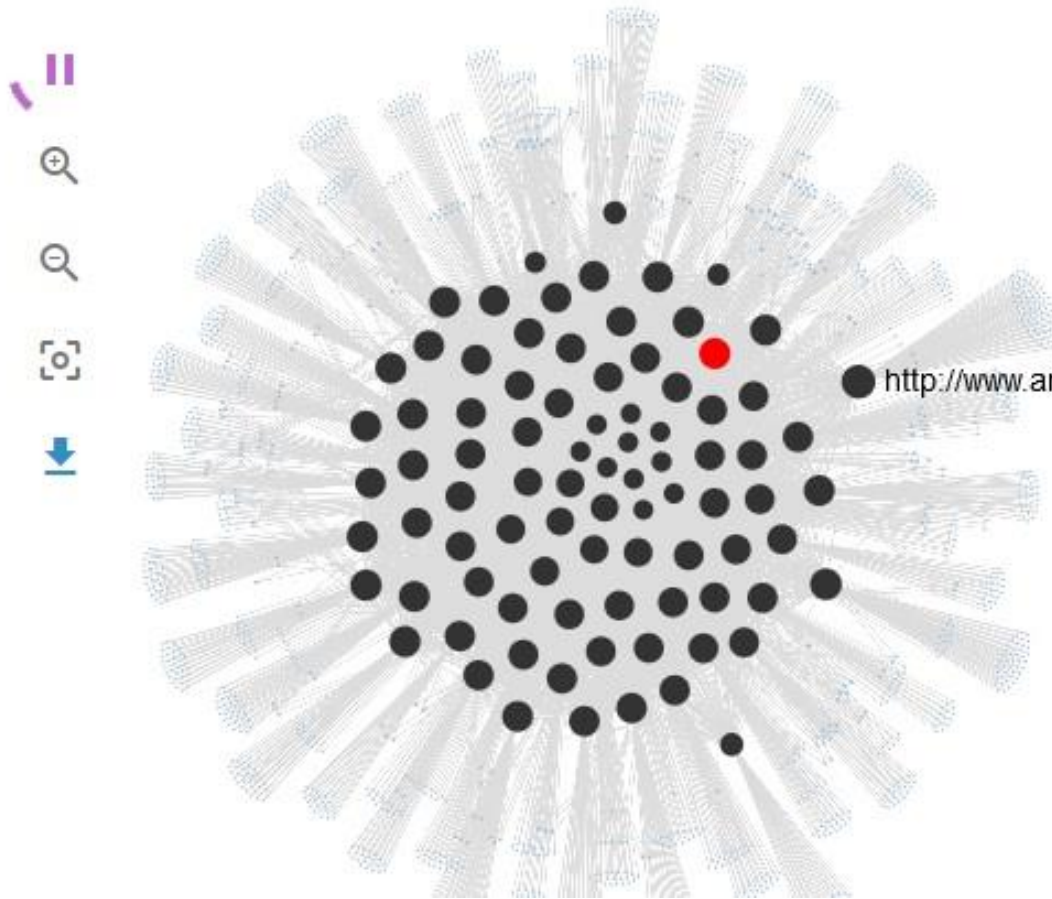
INTERNET ARCHIVE

So what's the web about?



So what's the web about?

Pages network: Anjameulenbelt.nl



Questions?