Software-based data collection - Praxis

HIIG Methodenworkshop, 30. Mai 2013.
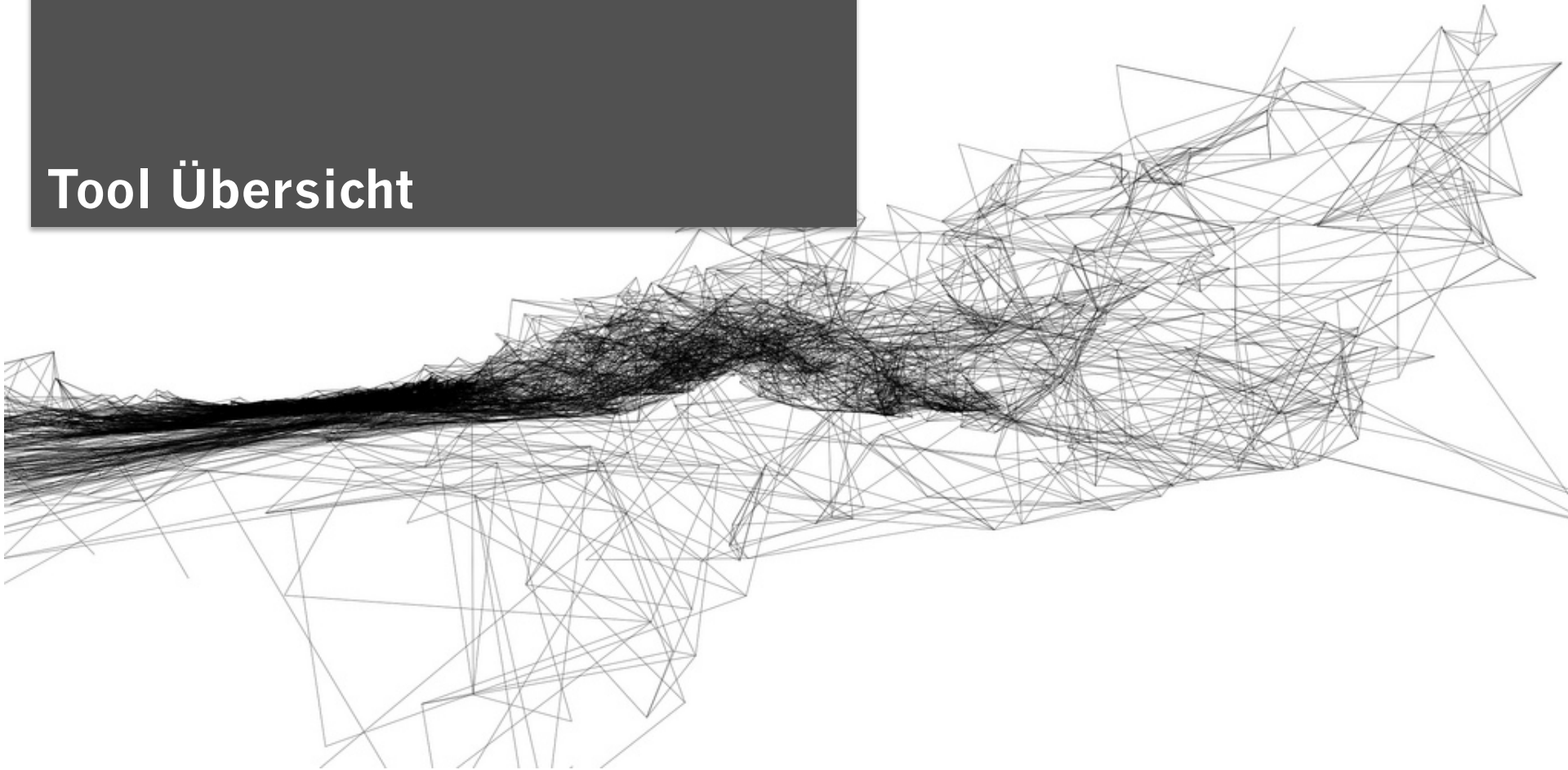
Dr. Carolin Gerlitz
University of Amsterdam/Goldsmiths, University of London

Schedule

- Tool Übersicht
- Mini-Projekt
- Forschungsbrowser
- Schritt 1: Cross-Platform Analysis
- Schritt 2: Longitudinal Analysis

Tool Übersicht

Digital Methods Tools

- Tools für **Datenerhebung**, **Datenreinigung**, **Auswertung** & **Visualisierung**.
- Ausgehend von Devices, Plattformen, oder digitale Objekten.
- Entwickelt für akademische und nicht-kommerzielle Nutzung.
- Frei zugänglich unter https://wiki.digitalmethods.net/Dmi/ToolDatabase

DMI Tools

Media Analysis: Media Monitoring | Mapping | Cloudin
Data Treatment: Data Collection | Data Analysis | Infor
Natively Digital: The Link | The URL | The Tag | The D
Device Centric: Google | Google Images | Google New   a |
IssueCrawler | Twitter | Facebook
Spherical: Web Sphere | News Sphere | Blogosphere |

Actor Profiler

Launch tool   Instructions & Scenarios of Use

The Actor Profiler works in tandem with the Issue Crawler. It calculates the top ten actors on an Issue Crawler map, and profiles each of them. The profile consists of each actor's inlinks a...

Censorship Explorer

Launch tool   Instructions & Scenarios of Use

Check whether a URL is censored in a particular country by using proxies located around the world.

Compare Lists

Launch tool   Instructions & Scenarios of Use

Compare two lists of URLs for their commonalities and differences.

Convert Issuecrawler to Navicrawler

Launch tool   Instructions & Scenarios of Use

Convert an Issuecrawler XML file into the WXSF format of the Navicrawler file

Tools als Assemblage

- Keine Ein-Klick-Option sondern **Kombination** von Tools.
- Je nach Forschungsfrage müssen Tools neu kombiniert und aufeinander abgestimmt werden.

DMI Tools pertaining to Mapping

Media Analysis: Media Monitoring | Mapping | Clouding | Comparative Media Analysis
Data Treatment: Data Collection | Data Analysis | Information Visualization
Natively Digital: The Link | The URL | The Tag | The Domain | The PageRank | The Robots
Device Centric: Google | Google Images | Google News | Google Blog Search | Yahoo | Yo
Facebook | Amazon
Spherical: Web Sphere | News Sphere | Blogosphere | Tag Sphere | Video Sphere | Image Sphere

Amazon Book Explorer

Launch tool | Instructions & Scenarios of Use
Provides different analytics for Amazon.com's book search

Dorling Map Generator

Launch tool | Instructions & Scenarios of Use
Input tags and values to produce a Dorling Map (i.e. bubbles). Output is an svg.

Extract URLs

Launch tool | Instructions & Scenarios of Use
Extracts URLs from an Issuecrawler result file (.xml). Useful for retrieving starting network.

FlickrPhotoPoolNetwork

flickr
Group Pool

Launch tool | Instructions & Scenarios of Use
Outputs the contact network of Flickr users from a Photopool in .gdf format

Issuecrawler

Launch tool | Instructions & Scenarios of Use
Enter URLs and the Issue Crawler performs co-link analysis in one, two or three

Der Link

- Links als **assoziative oder Beziehungs-Marker** (Rogers 2002).
- Link zwischen Websites deutet auf Beziehung hin.
- Links als 'cornerstone of the web' (Shields), 'essence' des Webs (Foot & Schneider), 'foundation of the web' (O'Reilly).
- Wie kann man Links nutzen, um Assoziations- und Beziehungsnetzwerke zu erstellen?

Issuecrawler

- Issuecrawler untersucht Link-Beziehungen zwischen Websites für die Analyse von Netzwerken.
- Input: Liste von thematisch verknüpften Websites.
- Analyse: einfache, gegenseitige und Schneeball Links.

Instructions of use | Scenarios of use | Français | Italiano | 한국어 | FAQ | Allied tools

issuecrawler

| the Lobby | **Issue Crawler** | Network Manager | Archive |

Friday, September 16, 2011

@405.40

Launch Crawl

Results from the Harvester:

All the urls will be crawled, and their network will be located. For best results, use the linkpages, e.g., www.site.org/links

```
http://acampada-aviles.org
http://acampadaalm.es
http://acampadabcn.wordpress.com
http://acampadalugo.blogspot.com
http://alcala.tomalaplaza.net
http://alicante.tomalaplaza.net
http://benicarlo.tomalaplaza.net
http://bilbao.tomalaplaza.net
```

View   Edit   Update   Remove   Save Results

Crawling Methods

⊙ Co-link
Crawl seeds and retain URLs with at least two links from seeds. Co-linkees become seeds for subsequent iteration.

○ Snowball
Crawl seeds and retain URLs with at least one link from seeds.

○ Inter-actor
Crawl seeds and return inter-linkings among seeds.

Current and Queued Crawls   rss

23 Aug 2011 new influencers
14 Sep 2011 Religious Blogosphere in the US 2011
15 Sep 2011 PartidosEspañaElecciones20N2011
7 Sep 2011   VisitAalborg
7 Sep 2011   Outdoor Network
8 Sep 2011   unibrennt snowball 4
8 Sep 2011   Spoilers Lost
8 Sep 2011   Redie
9 Sep 2011   Spoilers Lost-2
9 Sep 2011   Jonge NL Bloggers
9 Sep 2011   Lost_2007_1
9 Sep 2011   Lost_2007_2
9 Sep 2011   Lost_2007_3
10 Sep 2011 ehealth
10 Sep 2011 Politikblogs_Austria_Snowball
14 Sep 2011 medicinal marijuana controversy snowball
14 Sep 2011 medicinal marijuana controversy inter-actor
14 Sep 2011 Prion 14.11.2011
14 Sep 2011 Aid worker blogs (test 5)
16 Sep 2011 Electronic Health Records in Denmark. 16 sept 2011

Issuecrawler

- Beziehungs-Terminologie auf Grundlage von Linkverhalten (Rogers 2002).

Linking Styles

totem	kinship	extra-curricular	third-party
links to no domains	only links to its own domain	links to a domain that is not its own	links to two domains but not to its own

non-reflexive	missionary	transdiscursive	fully transdiscursive
links to three domains but not to its own	links to its own domain and to one other	links to its own domain and to two others	links to all domains including its own

Figure 2. Actor Hyperlink Language, Govcom.org, Design and Media Research Fellowship, Jan van Eyck Academy, Maastricht, 1999. ©Govcom.org Foundation, Amsterdam, 1999. Reproduced with permission.

Issuecrawler

- Issuecrawler Visualisierung von Websites zum Thema "Green Home" (Marres 2009).

Figure 1. Issue network disclosed by green home blogs, Issuecrawler, March 2008
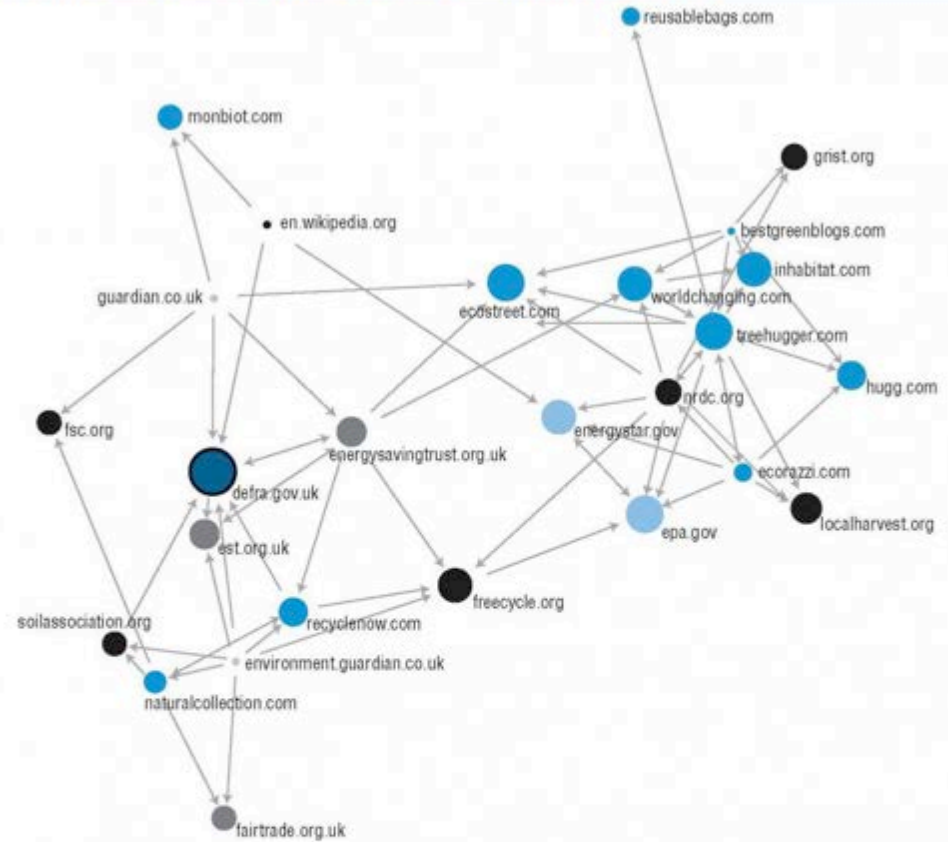
Map details

Author:	Noortje Marres
Email:	marres@dda.nl
Crawl start:	4 Mar 2008 - 05:10
Crawl end:	4 Mar 2008 - 08:05
Privilege starting points:	off
Analysis mode:	site
Iteration:	2
Depth:	2
Node count:	26

Map generated from Issuecrawler.net by the Govcom.org Foundation, Amsterdam.

Legend

- (.com)
- (.gov.uk)
- (.org)
- (.org.uk)
- (.gov)
- (.co.uk)

Issuecrawler

- Issuecrawler Visualisierung zu Beziehungen zwischen den Top 1000 Websites nach Alexa.

Issuecrawler

- Issuecrawler Visualisierung zu Beziehungen in der historischen holländischen Blogosphere (Weltevrede & Helmond 2012).

The Dutch Blogosphere in Transition
The rise and evolution of the Dutch blogosphere 1999-2009.

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009

Dutch Blogosphere in Transition   201
Where do bloggers blog? The rise and fall of TLDs and blog platform software.

Research and Visualization aided by tools from
digitalmethods.net from the Digital Methods Initiative,
Gephi and E-diaspora Atlas Platform.

Research_ Erik Borra, Anne Helmond and Esther Weltevrede

digital
methods
initiative

Dutch
Blogosphere

Links & Facebook

- Hyperlinks als ein, soziale Medien als anderer Beziehungsmarker (Gerlitz & Helmond 2013).
- Plattformen erlauben das liken, teilen und tweeten externer URLs.
- Wie oft wurde eine URL gelikt, geteilt oder mit Facebook kommentiert?

facebook DEVELOPERS   **Documentation   Forum   Showcase**

Social plugins
Home › Documentation › Social plugins

The easiest way to add Facebook to your site.

Social plugins enable you to provide engaging social experiences to yo
users with just a line of HTML. Because they are hosted by Facebook,
plugins are personalized for all users who are logged into Facebook — even if
the users haven't yet signed up for your site.

Like Button
The Like button lets users share pages from your site back to their Facebook profile with one click.

Recommendations
The Recommendations plugin gives users personalized suggestions for pages on your site they might like.

Login with Faces
The Login with Faces plugin shows profile pictures of the user's friends who have already signed up for your site in addition to a login button.

Comments
The Comments plugin lets users comment on any piece of content on your site.

Activity Feed
The Activity Feed plugin shows users what

'Mermaids: The New Evidence': Never-Before-Seen Footage Of Mermaids Allegedly (VIDEO)

Posted: 05/27/2013 8:17 pm EDT

| 1,371 | 197 | 186 | 255 |

f share   tweet   email   comment

GET TV ALERTS:
Enter email   SIGN UP

FOLLOW: TV Replay, Video, Television, Animal Planet, Mermaids, Mermaids Animal Planet, Mermaids The New Evidence, Mermaids The New Evidence Animal Planet, TV News

Like Scraper

- Die Facebook API ermöglicht Anfragen wie oft URLs gelikt, geteilt oder kommentiert wurden.
- Like Scraper erlaubt systematische Suche und erhebt Facebook Aktivitäten per URL.

Like Scraper

Input

Enter URLs (1 per line):

```
http://www.hiig.de
http://www.golem.de/1110/87312.html
http://www.hiig.de/doktorandenpositionen/
http://www.hiig.de/events/methodentraining-forschungsgegenstand-
werkzeug-internet/
```

[i Like Scraper]

Like Scraper, an Introduction

Batch queries Facebook likes.

For each URL entered, this script queries the Facebook api and retrieves the number of likes, shares, comments and clicks for given URLs. The output is a table with the URLs queried and the numbers retrieved.

url	normalized url	share count	like count	comment count	click count	total count
http://www.hiig.de	http://www.hiig.de/	25	7	0	0	32
http://www.golem.de/1110/87312.html	http://www.golem.de/1110/87312.html	2	4	0	0	6
http://www.hiig.de/doktorandenpositionen/	http://www.hiig.de/doktorandenpositionen/	0	0	0	0	0
http://www.hiig.de/events/methodentraining-forschungsgegenstand-werkzeug-internet/	http://www.hiig.de/events/methodentraining-forschungsgegenstand-werkzeug-internet/	5	0	0	0	5

netvizz v0.72

This application allows you to extract data from different sections of the Facebook platform for research purposes. It creates network files in the **gdf format** (a simple text format that specifies a graph) as well as statistical files using a **tab-separated format**.

These files can then be analyzed and visualized using graph visualization software such as the powerful and very easy to use **gephi** platform or statistical tools such the interactive visualization software **Mondrian**.

Big networks may take some time to process. **Be patient!**

Privacy policy and credits are **here**.

___

your personal friend network:

Creates a network file with all the friendship connetions in your personal network.

Step 1 – Select user data to include in the file (sex, interface language, and profile age ranking are standard):

☐ friends' like and post count (public and visible to logged user), includes counts for received likes and comments on posts, adds an additional ±4 seconds of waiting time per friend

Step 2 – create a gdf file from your personal network by clicking **here**

___

your like network:

Create a bipartite network (gdf file) from your friends and their likes (both users and liked objects are nodes) **here**. Only liked pages are provided, not external objects. Count on waiting about a second per friend.

___

groups:

You are a member of the following groups. Netvizz can extract two types of social networks (both gdf files):

– friendship connections (API limits for group data are changing regularly, current version should be able to get up to 5000 group members. This may take a very long time, i.e. hours).
– interactions (if a user comments or likes another user's post, a directed link is created – currently the last 200 posts are take into account)

Netviz

- Netzwerkanalyse in Facebook fokussiert sich oft auf Freundschaftsbeziehungen.
- Welche Netzwerke entstehen, wann man sich auf Inhalte/gemeinsam gelikte Seiten konzentriert?

Freundes-Netzwerk visualisiert mit Gephi.
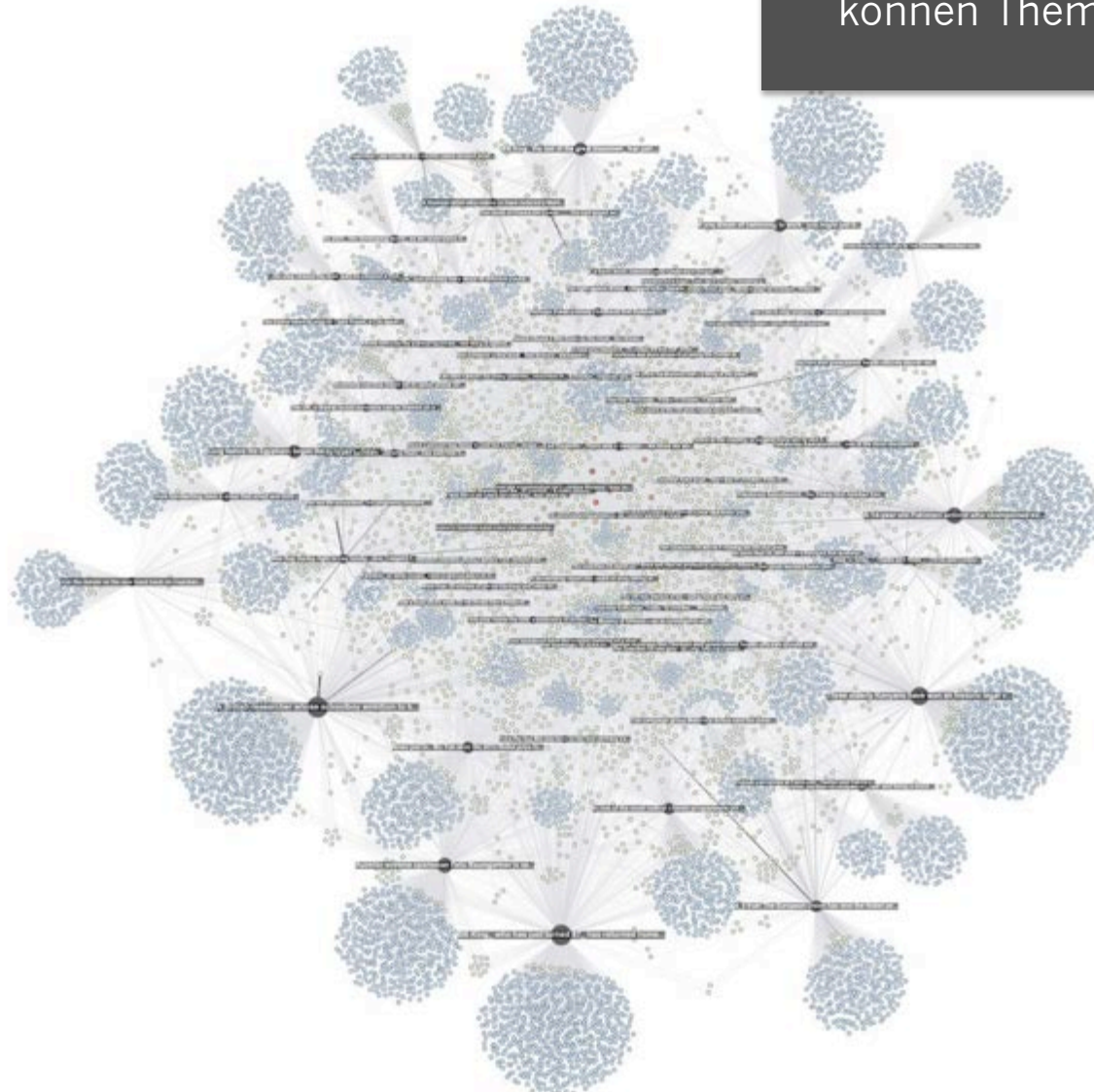
Netviz

Aktivitäts-Netzwerk der Guardian Seite.

Netviz

- Vergleichende Aktivitäts-Netzwerke New York Times vs. Wall-Street Journal.
- Haben beide Magazine gemeinsame Liker? Und wenn ja, sind sie auf beiden Seiten aktiv?

Twitter

- Digital Methods Twitter Capture and Analysis Toolset (TCAT).
- Nutzt die Twitter Streaming API, fast-Echtzeit Zugang zu Tweets.
- Erhebung, Archivierung & Analyse.
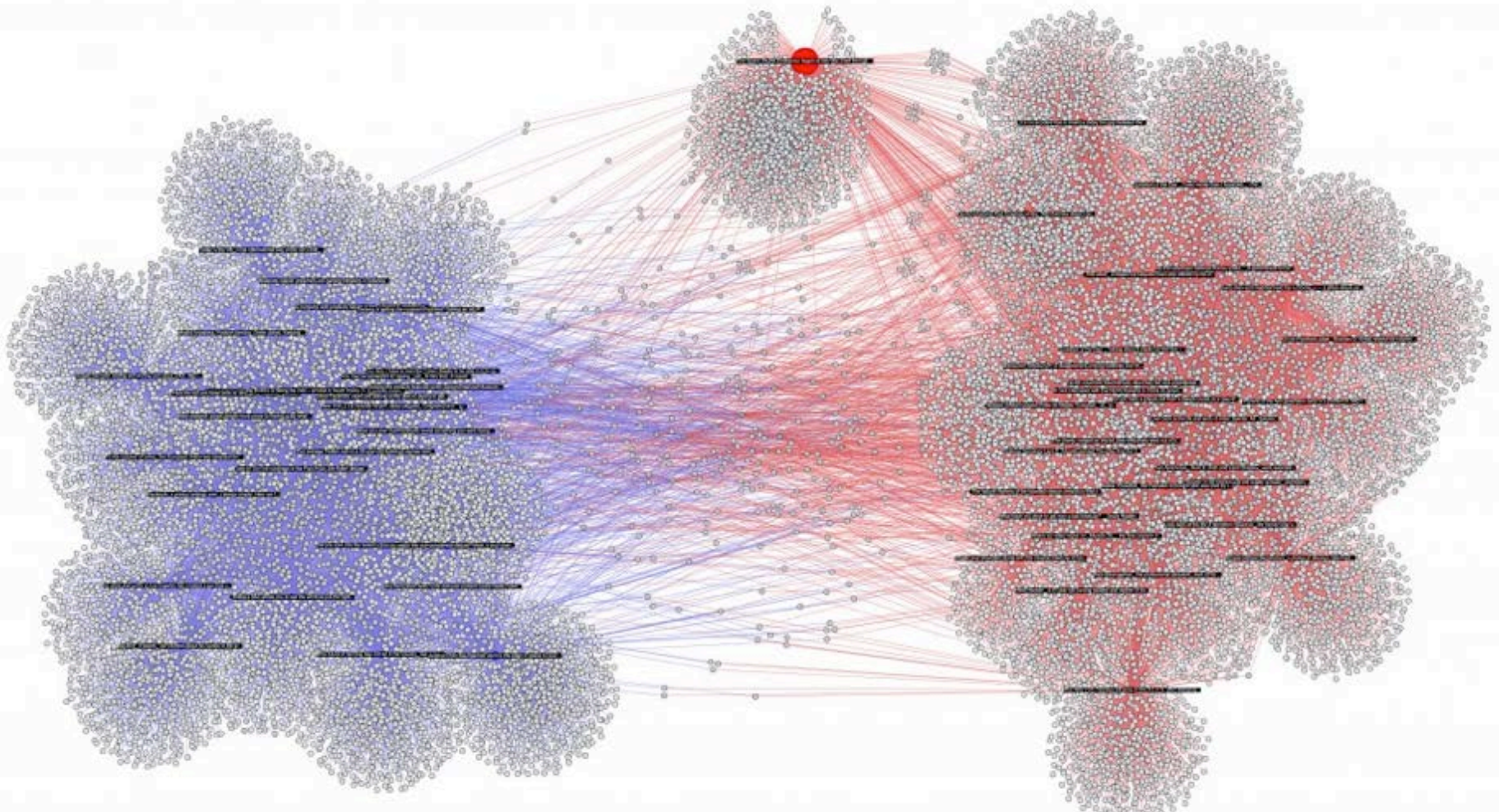
Twitter Analytics

Data Selection

Select the dataset:

copyright --- 3.429.099 tweets from 2012-11-23 15:53:48 to 2013-05-29 10:14:49

Select parameters:

Query:		(empty: containing any text*)
Exclude:		(empty: exclude nothing*)
From user:		(empty: from any user*)
URL (or part of URL):		(empty: any or all URLs*)
Startdate:	2012-11-23	(YYYY-MM-DD)
Enddate:	2013-05-28	(YYYY-MM-DD)

update overview

* You can also do AND or OR queries, although you cannot mix AND and OR in the same query.

Overview of your selection

Dataset:	copyright (#tu, acta, #freeandopen, #netfreedom, #ETNO, #WCIT12, WCIT, #sopa, sopablackout, IPRED, TPPA, TPP)
Search query:	
Exclude:	
From user:	
(Part of) URL:	
Startdate:	2012-11-23
Enddate:	2013-05-28
Number of tweets:	3.423.139
Number of distinct users:	867.756

- Tweets containing links (38.5%)
- Tweets containing no links (61.5%)

160,000

120,000

80,000

40,000

0

- Tweets
- Users
- Locations
- Geo coded

Twitter

Frequencies

Select how the frequencies should be calculated:

○ overall ○ per hour ◉ per day ○ per week ○ per month ○ per year ○ custom: `YYYY-MM-DD;YYYY-MM-DD;...;YYYY-MM-DD`

Hashtag frequency

Creates a .csv file (open in Excel or similar) that contains hashtag (#hashtag) frequencies.

Use: find out which hashtags are most often associated with your subject.

» launch

User mention frequency

Creates a .csv file (open in Excel or similar) that lists usernames and the number of times they were mentioned by others.

Use: find out which users are "influentials".

» launch

User tweet frequency

Creates a .csv file (open in Excel or similar) that lists usernames and how many tweets they posted.

Use: find the most active tweeters, see if the dataset is dominated by certain twitterati.

» launch

User tweet+mention frequency

Creates a .csv file (open in Excel or similar) that lists usernames and both tweet and mention frequencies.

Use: see wether the users mentioned are also those who tweet a lot.

» launch

Url frequency

Creates a .csv file (open in Excel or similar) that contains the frequencies of tweeted URLs.

Use: find out which contents (articles, videos, etc.) are referenced most often.

» launch

Host name frequency

Creates a .csv file (open in Excel or similar) that contains the frequencies of tweeted domain names.

Use: find out which sources (media, platforms, etc.) are referenced most ofter.

» launch

Identical tweet frequency

Creates a .csv file (open in Excel or similar) that contains tweets and the number of times they have been retweeted indenticaly, per day (date range > 2 days) or per hour (date range 2 days or smaller).

Use: get a grasp of the most "popular" content.

» launch

Twitter

Network

Social graph by mentions

Produces a directed graph (.gdf, open in gephi) based on interactions between users. If a users mentions another one, a directed link is created. The more often a user mentions another, the stronger the link ("link weight"). The "count" value contains the number of tweets for each user in the specified period.

Use: analyze patterns in communication, find "hubs" and "communities", categorize user accounts.

» launch

Co-hashtag analysis

Produces an undirected graph (.gdf, open in gephi) based on co-word analysis of hashtags. If two hashtags appear in the same tweet, they are linked. The more often they appear together, the stronger the link ("link weight").

Use: explore the relations between hashtags, find and analyze sub-issues, distinguish between different types of hashtags (event related, qualifiers, etc.).

» launch

Experimental

Associational profile

Produces an associational profile as well as a time-encoded co-hashtag network.

Use: explore shifts in hashtags associations.

» launch

Bipartite hashtag-user graph

Produces a bipartite graph (.gexf, open in gephi) based on co-occurence of hashtags and users. If a user wrote a tweet with a certain hashtag, there will be a link between that user and the hashtag. The more often they appear together, the stronger the link ("link weight").

Use: explore the relations between users and hashtags, find and analyze which users group around which topics.

» launch

Bipartite hashtag-mention graph

Produces a bipartite graph (.gexf, open in gephi) based on co-occurence of hashtags and @replies. If an @reply co-occurs in a tweet with a certain hashtag, there will be a link between that @reply and the hashtag. The more often they appear together, the stronger the link ("link weight").

Use: explore the relational *activity* between mentioned users and hashtags, find and analyze which users are considered experts around which topics.

» launch

Bipartite hashtag-URL graph

Creates a .csv file (open in Excel or similar) that contains URLs and the number of times they have co-occured with a particular hashtag.

Creates a .gexf file (open in Gephi) that contains a bipartite graph (.gexf, open in gephi) based on co-occurence of URLs and hashtags. If a URL co-occurs with a certain hashtag, there will be a link between that URL and the hashtag. The more often they appear together, the stronger the link ("link weight").

Use: get a grasp of how urls are qualified.

» launch

Automatisch vs. manuell

- Notwendigkeit von manueller Datenerhebung und/oder Analyse.
- Digitale Sortierung, manuelle Extraktion.
- Manuelle Kategorisierung von Ergebnissen und Nutzertypen.

Wikipedia

- Wikipedia ermöglicht vglw. guten Zugang zo Aktivitäten und Änderungen.
- Erlaubt die Analyse von Kollaborationsprozessen und ihrer Technizität.

nt Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk

Global warming: Revision history

View logs for this page

Browse history

From year (and earlier): 2013   From month (and earlier): all   Tag filter: Go

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary.
External tools: Revision history search · Contributors · User edits · Number of watchers · Page view statistics

- Interaction
  Help
  About Wikipedia
  Community portal
  Recent changes
  Contact Wikipedia
- Toolbox

(cur) = difference from current version, (prev) = difference from preceding version, **m** = minor edit, → = section edit, ← = automatic edit summary
(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) ⊙ 21:36, 28 May 2013 IRWolfie- (talk | contribs) **m** . . (161,793 bytes) (+11) . . *(fix)*
- (cur | prev) ⊙ 04:42, 28 May 2013 Northamerica1000 (talk | contribs) . . (161,782 bytes) (+251) . . *(Reflinks: Converting bare references)*
- (cur | prev)○ 04:37, 28 May 2013 Northamerica1000 (talk | contribs) . . (161,531 bytes) (+8) . . (→See also: +Science portal)
- (cur | prev)○ 00:03, 28 May 2013 Peterl (talk | contribs) . . (161,523 bytes) (-20) . . (→Notes: rmv wikilinks to countries - should either be all (but that would be overlinking) or none.)
- (cur | prev)○ 18:57, 27 May 2013 IRWolfie- (talk | contribs) . . (161,543 bytes) (+1) . . (I forgot to say, I forked content from Surveys_of_scientists'_views_on_climate_change in an earlier edit just there.)
- (cur | prev)○ 18:45, 27 May 2013 IRWolfie- (talk | contribs) . . (161,542 bytes) (-21) . . (→Others: "scientists and non-scientists", or people as we call them.)
- (cur | prev)○ 18:34, 27 May 2013 IRWolfie- (talk | contribs) . . (161,563 bytes) **(+1,137)** . . (→Other views: add and separate)
- (cur | prev)○ 18:07, 27 May 2013 FeydHuxtable (talk | contribs) . . (160,426 bytes) **(+1,572)** . . (updating)
- (cur | prev)○ 17:05, 27 May 2013 IRWolfie- (talk | contribs) . . (158,854 bytes) (+361) . . (more precise)
- (cur | prev)○ 13:32, 27 May 2013 IRWolfie- (talk | contribs) . . (158,493 bytes) **(-1,230)** . . (Undid revision 557014757 by FeydHuxtable (talk) unclear why material was removed, unclear source for material about freezing, WP:RECENTISM)
- (cur | prev)○ 13:25, 27 May 2013 FeydHuxtable (talk | contribs) . . (159,723 bytes) **(+1,230)** . . (updating)
- (cur | prev)○ 07:43, 24 May 2013 Nigelj (talk | contribs) . . (158,493 bytes) (-454) . . (Reverted good faith edits by David.moreno72 (talk): Rm addition based on 1997 research. That is not current news, and more recent work make it much worse than that. ([[WP:...)
- (cur | prev)○ 07:34, 24 May 2013 David.moreno72 (talk | contribs) . . (158,947 bytes) (+454) . . (→Observed temperature changes: Added temperature reconstructions)
- (cur | prev)○ 05:58, 18 May 2013 HiLo48 (talk | contribs) . . (158,493 bytes) (+10) . . (Reverted unsourced change)
- (cur | prev)○ 05:54, 18 May 2013 Daren420c (talk | contribs) . . (158,483 bytes) (-10)

Wikipedia

- Artikelnetzwerk Global Warming.
- Forschungsfrage: Wie werden Kontroversen gelöst und welche Optionen bietet die Plattform an?

Global Warming Wikipedia Network

Method_ Crawl articles related to Global Warming (indicated by bi-directional links), outputting network.

Articles used for further analysis are highlighted.

Digital Methods Summer '09
Analysis_ Michael Stevenson, Carolin Gerlitz, Erik Borra and Richard Rogers
Map_ ReseauLu
Design_ Marieke van Dijk, Michael Stevenson and Richard Rogers
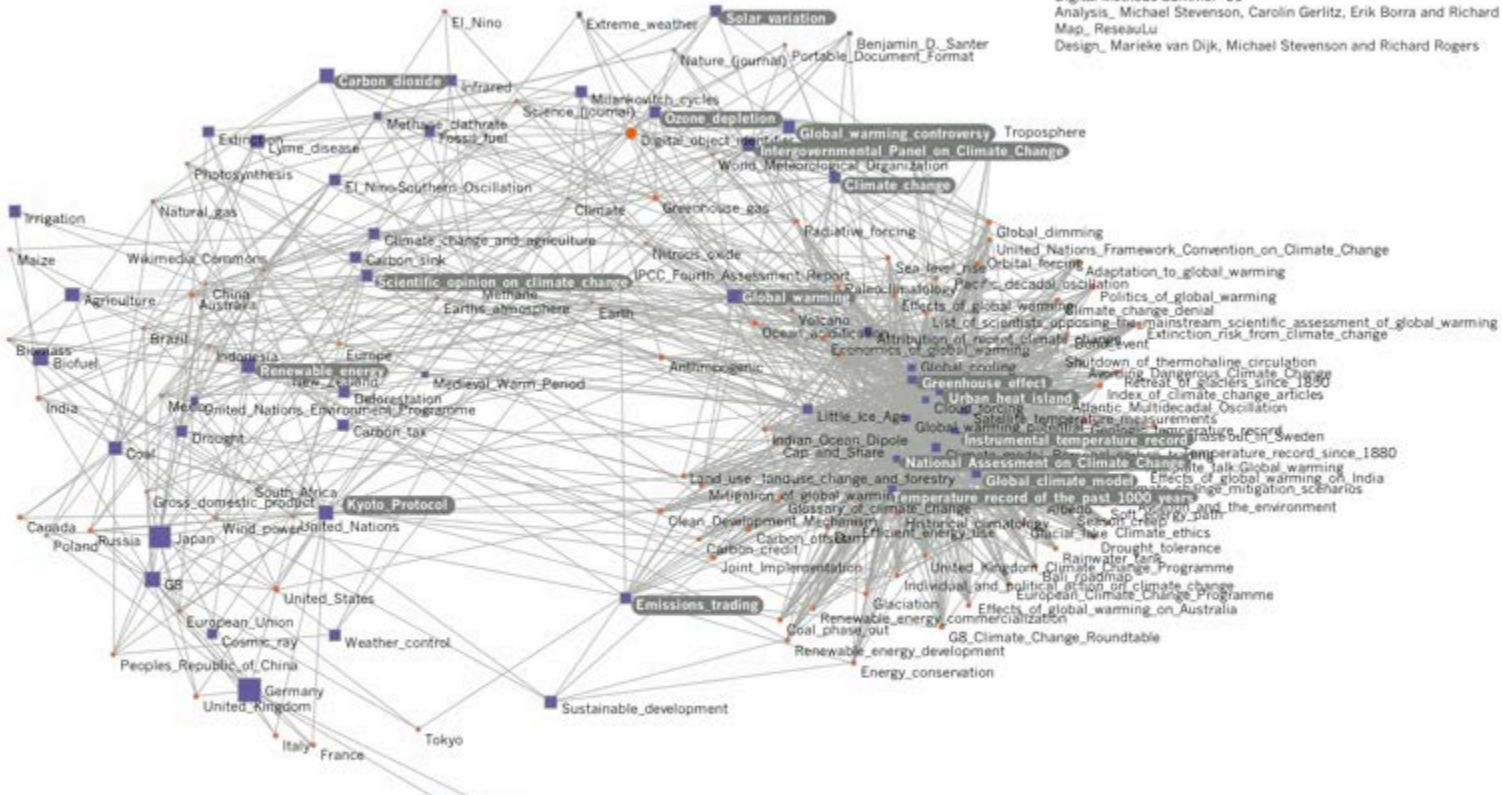
Wikipedia

Activity in the Global Warming Network

Question_ How is the Global Warming network structured by editor activity?

Method_Identify the total number of edits per article and create Dorling maps.

Findings_The activity within the Global Warming network is distributed very unequally. Whereas Global Warming is characterised by large numbers of edits and re-edits, articles like "Global Climate Model" or "National Assessment on Climate Change" hardly change. One sees that "Climate Change" and "Scientific Opinion on Climate Change" form the mid range in relation to their total number of edits.

Digital Methods Summer '09
Analysis_ Carolin Gerlitz and Michael Stevenson
Design_ Carolin Gerlitz, Michael Stevenson and Marieke van Dijk
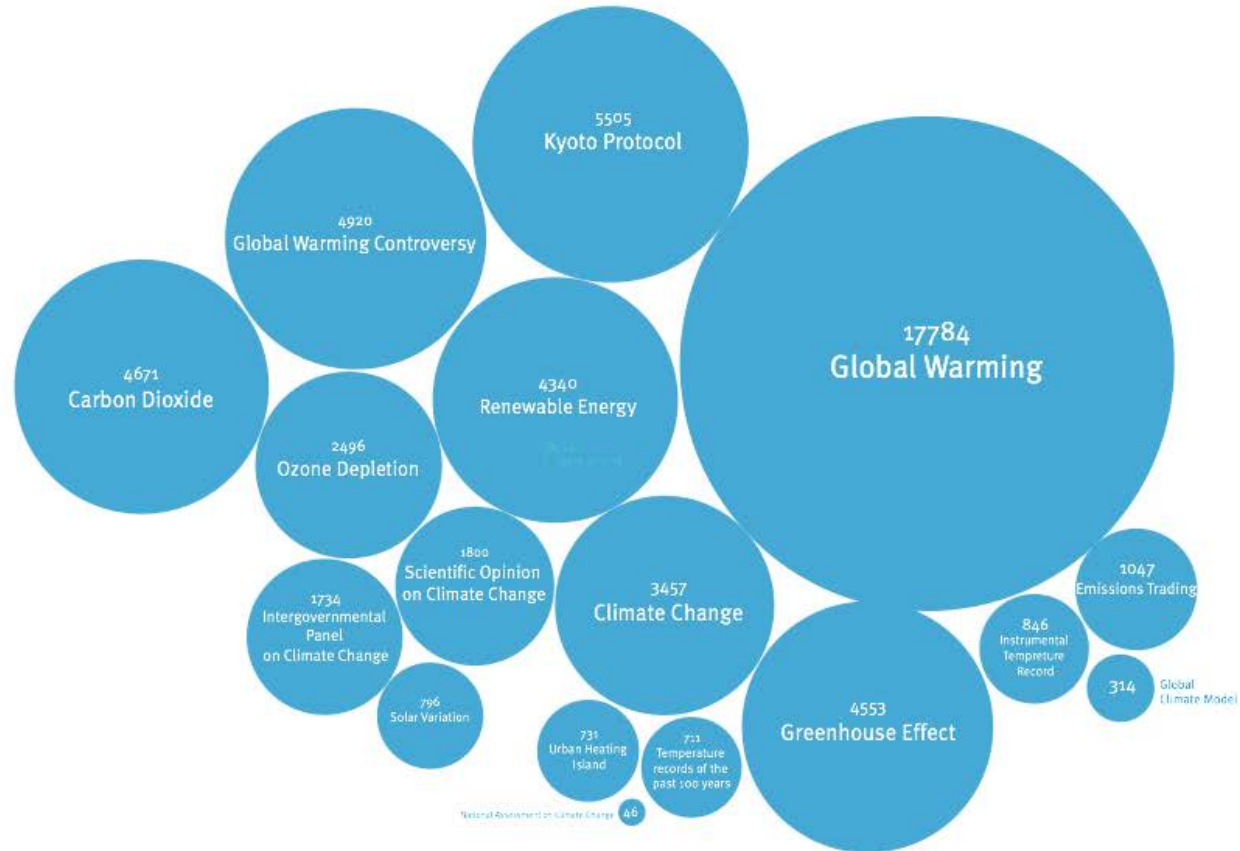
Wikipedia

Debates and their Displacement on Wikipedia

Question: How is controversy managed on Wikipedia, and how does this effect the encyclopedia's composition of an issue?

Findings: After heavy editing and much discussion on the Climate Changeand Global Warming articles regarding scientific consensus, active editors decided to create the article "Scientific Opinion on Global Warming." The Dorling Map shows the number of edits and editors for "Climate Change" in the three months prior to the creation of the "Scientific Opinion" article, and the same for both articles in the following three months. One sees, in a sense, the relocation of the debate and its debaters, as well as a corresponding decline in activity on the article "Climate Change."

Digital Methods Summer '09
Analysis_ Carolin Gerlitz and Michael Stevenson
Design_ Carolin Gerlitz, Michael Stevenson and Marieke van Dijk

Edits to "Climate Change"

Edits to "Scientific Opinion on Climate Change"

24 June to 24 September 2003

24 September to 24 December 2003

Wikipedia

Wikipedia

Where are the Users?
Top User Activity per article

Question_ How is editor activity distributed among the articles of the Global Warming network?

Method_ Query the Revision History Statistics of each article for the top 100 editors. Comparing the total edits of each editor within the Global Warming network, the 20 most active users were identified. The following Bubble Lines show the total number of edits per editor in each article of the network.

Findings_ The Bubble-Lines show that the Global Warming network is maintained and created by a limited number of core editors, led by William M. Connolley followed by Raymond Arrit and VSmith. These three core editors are active above average on almost each article within the network. Among the most active editors UBeR poses an exception as this editor only participates in few articles - but on a large scale. Further, one sees that Renewable Energy and Emission Trading are not maintained by the core users of the network but by a different group. A number of top users like OmegaTron and The-Seven have particular interests and do not participate in the editing of general articles such as Global Warming. Finally, there is one bot among the top editors: ClueBot.

Digital Methods Summer '09
Analysis_ Carolin Gerlitz and Michael Stevenson
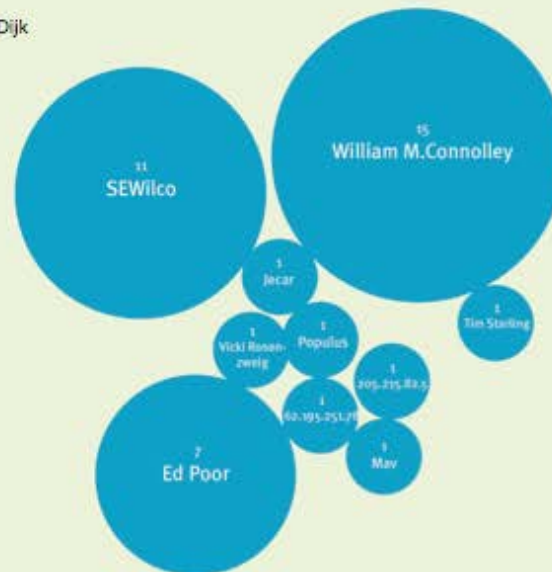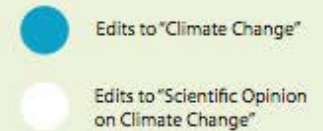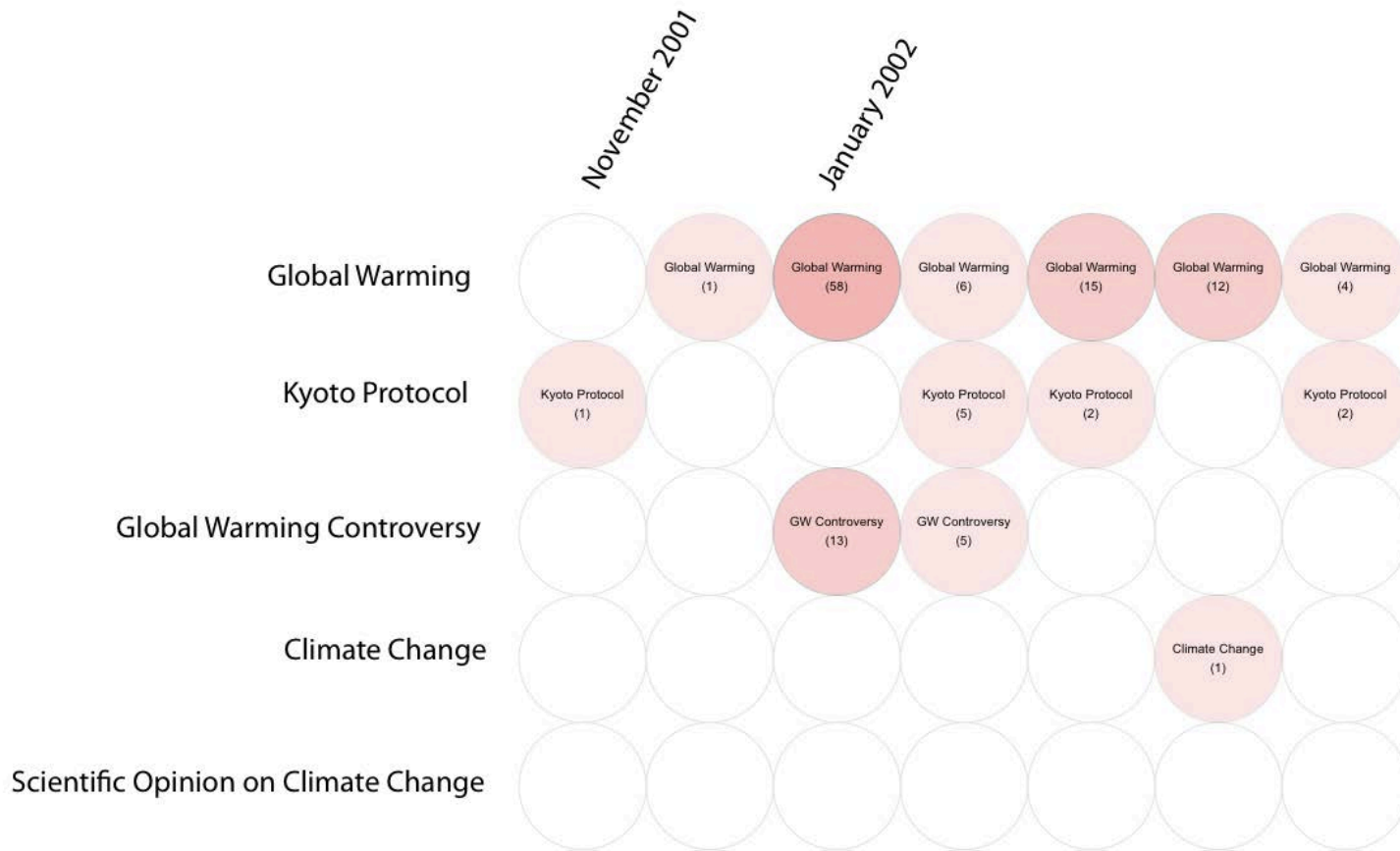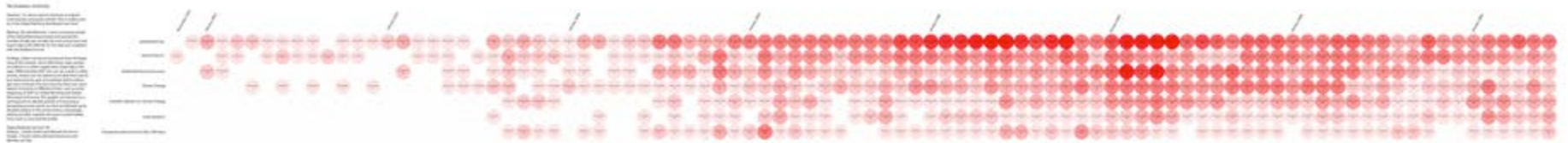Design_ Carolin Gerlitz, Michael Stevenson and Marieke van Dijk

Wikipedia

Where are the bots?
Top bot activity per article

Question_ How is bot activity distributed among the Global Warming network?

Method_Query the Revision History Statistics and get the bot edits for each article of the network. Identify the 15 most active bots by their total number of edits within the network. Create BubbleLine graphic with the activity of each bot in each article.

Findings_As the articles are arranged according to the total number of edits, the graphic shows that bots are more active in more edited articles. As one sees in the upper left corner of the graphic, the top four bots ClueBot, SmackBot, Tawkerbot2, AntiVandalBot are especially active in the top edited articles. Within the Global Warming Network, Cluebot can be identified as the most active bot. Only on Global Warming Tawkerbot2 and AntiVandalbot are more active. The amount of active bots per article decreases in relation to the number of total edits per article. Overall, the number of influential bots is rather limited as the majority of bots just perform up to 10 edits.

Digital Methods Summer '09
Analysis_ Carolin Gerlitz and Michael Stevenson
Design_ Carolin Gerlitz and Michael Stevenson.

Software-based Data collection

Verschiedene Rollen von Software:

- Software extrahiert und archiviert digitale Daten.

- Software struktiert und macht digitale Daten zugänglich.

- Software bietet Analysemethoden basierend auf Aggregation, Statistik, Co-Wort Analyse, Netzwerkanalyse an.

- Software nutzt Algorithmen zur Visualisierung.

- Natively Digital Methods? Oder eher **Interface Methods?**

- Methodische Anleihen und Analyseschritte aus verschiedenen Disziplinen: Soziologie, Ethnographie, Informatik etc.

Mini-Projekt Google Scraper

- Arbeitet in Gruppen von 2-3 Leuten.
- Wählt ein Thema/Kontroverse, das sowohl momentan als auch über die letzten Jahre hinweg relevant war: Euro Crisis, Same Sex Marriage, War, Terror, Internet Regulation etc.
- **Start**: Research Browser.
- **Schritt 1:** Cross-Spherical Analysis: Wie wird das Thema momentan in Web, Blogs und News diskutiert?
- **Schritt 2:** A sort of longitudinal analysis.
- **Schritt 3:** Welche weiteren Methoden?

Research Browser

- Idealer weise einen separaten Forschungsbrowser einrichten.
1. Aus allen Services ausloggen.
2. Browserverlauf & Cookies löschen.
3. Google: Personalisierung von Suchergebnissen löschen: https://www.google.com/history/optout
4. Lokale oder globale Google Version?

www.google.com/ncr

Google Web History

Customization based on signed-out search activity is disabled.

Enable customizations based on search activity.

Sign in or create a Google account to get more personalized search and ad results using your signed-in Web History.

©2013 Google   Google Home   Help   Privacy & Terms   About Google

Schritt 1

- Identifiziere relevanten Suchbegriff.
- Nutze Google Web, Blog & News Scraper.
- Top 100 Ergebnisse.
- Download als xls Datei.

- Kopiere die Artikeltitel der Ergebnisse von Web/Blog/News Scraper und nutze Tagxedo.com um wiederkehrende Begriffe zu visualisieren.

- Identifiziere 3 (kontroverse) Unterthemen.
- Kopiere Top 100 Ergebnisse in URL-Feld und 5 Unterthemen in Query Feld.

autism (17) "sudden infant death syndrome" (5) epilepsy (9) "multiple sclerosis" (4)

Schritt 2

- Identifiziere 3-4 relevante **Zeitintervalle** (1, 3 oder 6 Monate)
- Nutze die Google Web Scraper Intervalsuche.
- Top 100 Ergebnisse.
- Export als .xls Datei.

- Kopiere die Artikeltitel der Ergebnisse und nutze Tagxedo.com um wiederkehrende Begriffe zu visualisieren.
- Erstelle gleich formatierte Taxedo Visualisierungen pro Interval und vergleiche.

Schritt 3

- Was sind die Grenzen und Möglichkeiten der vorgestellten Methode?

- Mit welchen weiteren software-basierten Methoden könnte Euer Thema noch erforscht werden?

- Welche Plattformen oder Devices wären relevant?