ual: university of the arts london
london college of communication

Critically engaging with social media research tools

Dr Steven McDermott

s.mcdermott@lcc.arts.ac.uk

@soci

https://snacda.com

- Dr Steven McDermott, London College of Communications, University of the Arts London.

- TAGS; Yourtwapperkeeper; DMI-TCAT; Gephi; Leximancer; for collecting, archiving, analysing, visualising and disseminating social data collected from Twitter, YouTube and text based data.

university of the arts london
london college of communication

Surveillance Capitalism

"The ugly truth here is that much of "big data" is plucked from our lives without our knowledge or informed consent. It is the fruit of a rich array of surveillance practices designed to be invisible and undetectable as we make our way across the virtual and real worlds."(Zuboff, 2016)

The goal of big data analytics is to change people's behaviour at scale.

A Chief Data Scientist of a Silicon Valley company that develops applications to improve students' learning states that…

"The goal of everything we do is to change people's actual behavior at scale. When people use our app, we can capture their behaviors, identify good and bad behaviors, and develop ways to reward the good and punish the bad. We can test how actionable our cues are for them and how profitable for us." (Zuboff, 2016)

ual: university of the arts london
london college of communication

Future Research Plans
Choose Your Targets Carefully

- **Big Data Scientists as Knowledge Creators -** An Ethnography of Data Scientists of 2016 A searchable and interactive network graph of Data Scientists archived in February 2016 - **the digital environment and how it affects individuals and societal groups behaviour in the way they take decisions and seek information**
- **Mapping the 1% -** The Forbes Billionaire List and their Connections – 2015 The Forbes 2015's Billionaires (2015). Forbes ranks more than 1,800 billionaires and their companies, and affiliations with Government Bodies. **Political, cultural and economic conflicts reproduce or extend in digital networks.**
- **Graphs of Wikipedia: Influential Thinkers -** Interactive Graph of Wikipedia: Influential Thinkers **–** assessing **how social inclusion inequalities, class, gender, race, and disability manifest in digital networks**.
- **Members of Parliament in the Digital Environment** - How are members of the political class exchanging information **UK MPs on Twitter August 2014**

university of the arts london
london college of communication

TAGS

By Martin Hawksey

My Blog | Usage Policy

Home | Get TAGS | Support | News | Help

Home › Get TAGS

Get TAGS

To start using TAGS select one of the versions below then follow the steps below:

TAGS v6.0 TAGS v6.1

Which version? If you've not used TAGS before I recommend TAGS v6.1 which has an easy setup. If you've setup TAGS v6.0 you can keep on using that version, your existing archives will keep collecting tweets using your existing authentication. Some background in this post.

1. After your copy has been made open TAGS > Setup Twitter Access and follow the onscreen instructions (when selecting this option you'll be promoted to authorize the script to run several services).
   Important: In the new version of TAGS you only need to run the setup once
2. Enter what data you want to collect on the Readme/Settings sheet and hit TAGS > Run Now!

What's new

Search Forums

Search

Recent Topics

- Last date tweet of every follower
  5 days, 19 hours ago
- Twitter Mentions to GSheet
  6 days, 16 hours ago
- Automating cut and export of data
  2 weeks, 5 days ago
- Date-Time range
  1 month, 1 week ago
- Too many tweets retrieved!
  1 month, 2 weeks ago

News

- Social Media Analytics: Using Data to Understand Public Conversations (course feat. TAGS) #FLsocmed July 18, 2016

Twitter Authorisation

Please be aware that if you are logged into multiple Google accounts your access details will be saved with you default account.

By clicking 'Sign in with Twitter' you agree to abide by the Twitter Policy on 3rd party access

Sign in with Twitter

university of the arts london
london college of communication

Sign in with Twitter

Your TwapperKeeper
archive your own tweets

Archive ID	Keyword / Hashtag	Description	Tags	Screen Name	Count	Create Time	
22	#bigdata	preliminary scrape	research	soci	1040644	Tue, 30 Jun 2015 12:39:44 +0000	
23	#analytics	secondary scrape	research	soci	403748	Tue, 30 Jun 2015 13:17:30 +0000	
24	#iot	secondary scrape	research	soci	1051596	Tue, 30 Jun 2015 13:17:47 +0000	
25	#datascience	secondary scrape	research	soci	179621	Tue, 30 Jun 2015 13:18:14 +0000	
26	#deeplearning	secondary scrape	research	soci	113215	Tue, 30 Jun 2015 13:18:47 +0000	
27	#data	secondary scrape	research	soci	463725	Tue, 30 Jun 2015 13:19:06 +0000	
28	#machinelearning	secondary scrape	research	soci	139813	Tue, 30 Jun 2015 13:20:25 +0000	
29	#internetofthings	secondary scrape	research	soci	169557	Tue, 30 Jun 2015 17:57:39 +0000	

Bills

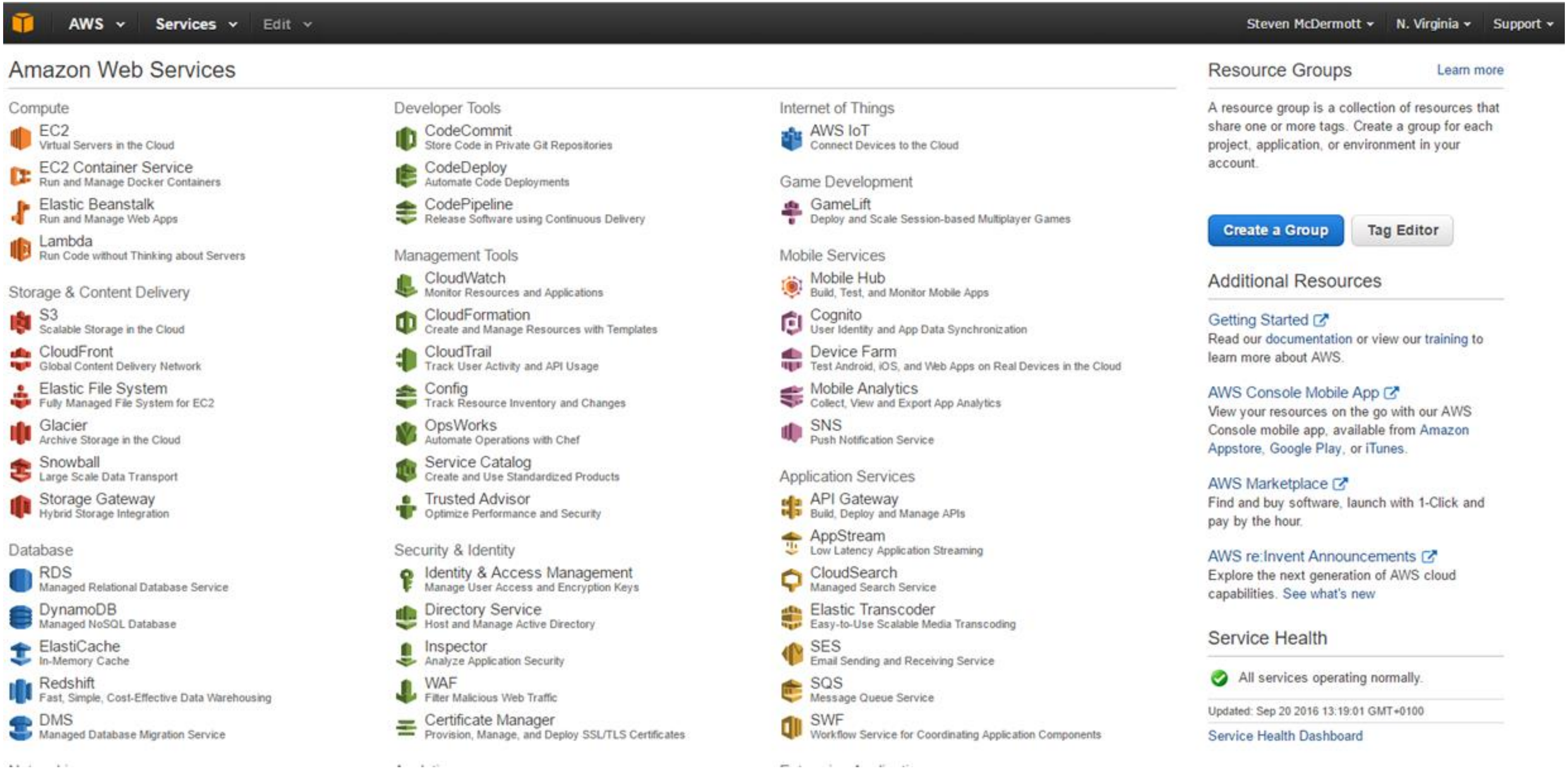
Date: August 2016

Download CSV | Print

Summary	Amount
AWS Service Charges	$55.53
Usage Charges and Recurring Fees View Invoices	$55.53
Other Details	
Payment Summary	$55.53
Tax Invoices View Invoices	
Total	$55.53

+ Expand All

Details	Total
AWS Service Charges	$55.53
Data Transfer	$0.05
Elastic Compute Cloud	$46.24
CT to be collected	$0.00
GST to be collected	$0.00
US Sales Tax to be collected	$0.00
VAT to be collected	$9.24


AWS
Services
Edit
Steven McDermott
N. Virginia
Support
Amazon Web Services
Compute
EC2
Virtual Servers in the Cloud
EC2 Container Service
Run and Manage Docker Containers
Elastic Beanstalk
Run and Manage Web Apps
Lambda
Run Code without Thinking about Servers
Storage & Content Delivery
S3
Scalable Storage in the Cloud
CloudFront
Global Content Delivery Network
Elastic File System
Fully Managed File System for EC2
Glacier
Archive Storage in the Cloud
Snowball
Large Scale Data Transport
Storage Gateway
Hybrid Storage Integration
Database
RDS
Managed Relational Database Service
DynamoDB
Managed NoSQL Database
ElastiCache
In-Memory Cache
Redshift
Fast, Simple, Cost-Effective Data Warehousing
DMS
Managed Database Migration Service
Developer Tools
CodeCommit
Store Code in Private Git Repositories
CodeDeploy
Automate Code Deployments
CodePipeline
Release Software using Continuous Delivery
Management Tools
CloudWatch
Monitor Resources and Applications
CloudFormation
Create and Manage Resources with Templates
CloudTrail
Track User Activity and API Usage
Config
Track Resource Inventory and Changes
OpsWorks
Automate Operations with Chef
Service Catalog
Create and Use Standardized Products
Trusted Advisor
Optimize Performance and Security
Security & Identity
Identity & Access Management
Manage User Access and Encryption Keys
Directory Service
Host and Manage Active Directory
Inspector
Analyze Application Security
WAF
Filter Malicious Web Traffic
Certificate Manager
Provision, Manage, and Deploy SSL/TLS Certificates
Internet of Things
AWS IoT
Connect Devices to the Cloud
Game Development
GameLift
Deploy and Scale Session-based Multiplayer Games
Mobile Services
Mobile Hub
Build, Test, and Monitor Mobile Apps
Cognito
User Identity and App Data Synchronization
Device Farm
Test Android, iOS, and Web Apps on Real Devices in the Cloud
Mobile Analytics
Collect, View and Export App Analytics
SNS
Push Notification Service
Application Services
API Gateway
Build, Deploy and Manage APIs
AppStream
Low Latency Application Streaming
CloudSearch
Managed Search Service
Elastic Transcoder
Easy-to-Use Scalable Media Transcoding
SES
Email Sending and Receiving Service
SQS
Message Queue Service
SWF
Workflow Service for Coordinating Application Components
Resource Groups
Learn more
A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.
Create a Group
Tag Editor
Additional Resources
Getting Started
Read our documentation or view our training to learn more about AWS.
AWS Console Mobile App
View your resources on the go with our AWS Console mobile app, available from Amazon Appstore, Google Play, or iTunes.
AWS Marketplace
Find and buy software, launch with 1-Click and pay by the hour.
AWS re:Invent Announcements
Explore the next generation of AWS cloud capabilities. See what's new
Service Health
All services operating normally.
Updated: Sep 20 2016 13:19:01 GMT+0100
Service Health Dashboard

DMI Twitter Capturing and Analysis Toolset (DMI-TCAT) **By Bernhard Rieder**

» github » issues » FAQ » admin

Data selection

Select the dataset:

digitalhealth --- 149.330 tweets from 2016-06-21 14:00:00 to 2016-08-08 17:32:43 ▼

160.486 tweets archived so far (and counting)

Select parameters:

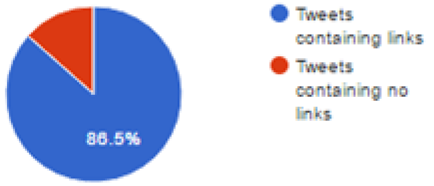
Query:		(empty: containing any text*)
Exclude:		(empty: exclude nothing*)
From user:		(empty: from any user*)
From twitter client:		(empty: from any client*)
(Part of) URL:		(empty: any or all URLs*)
Startdate (UTC):	2016-06-21	(YYYY-MM-DD or YYYY-MM-DD HH:MM:SS)
Enddate (UTC):	2016-08-08	(YYYY-MM-DD or YYYY-MM-DD HH:MM:SS)

update overview

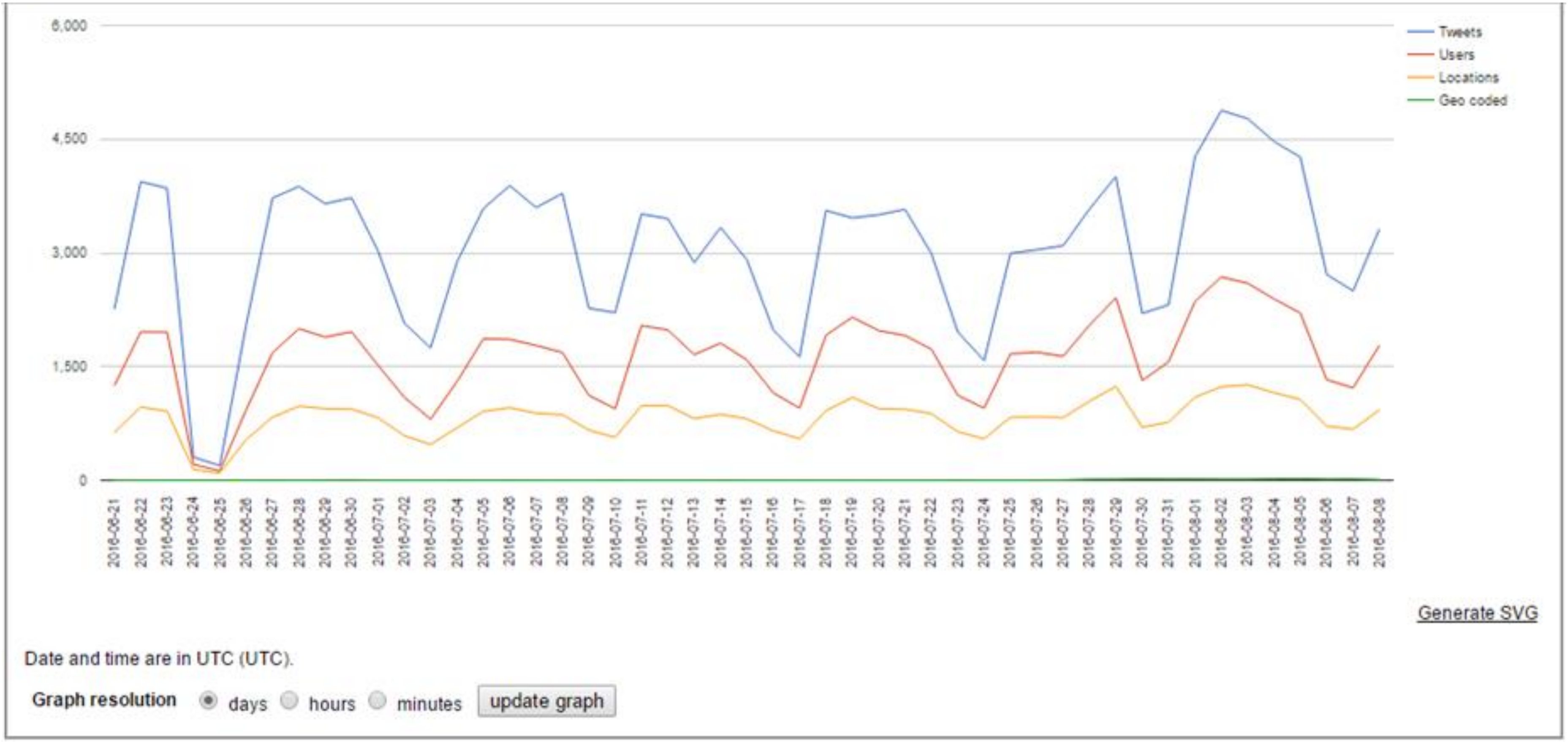
* You can also do AND or OR queries, although you cannot mix AND and OR in the same query.

Overview of your selection

Dataset:	digitalhealth (#digihealthcapital, #digitalhealth, #patient, #patients, 'patient voice', 'patient voices', digitalhealth, mHealth)
Search query:	
Comments:	
Exclude:	
From user:	
From twitter client:	
(Part of) URL:	
Startdate:	2016-06-21
Enddate:	2016-08-08
Number of tweets:	149.330
Number of distinct users:	33.050



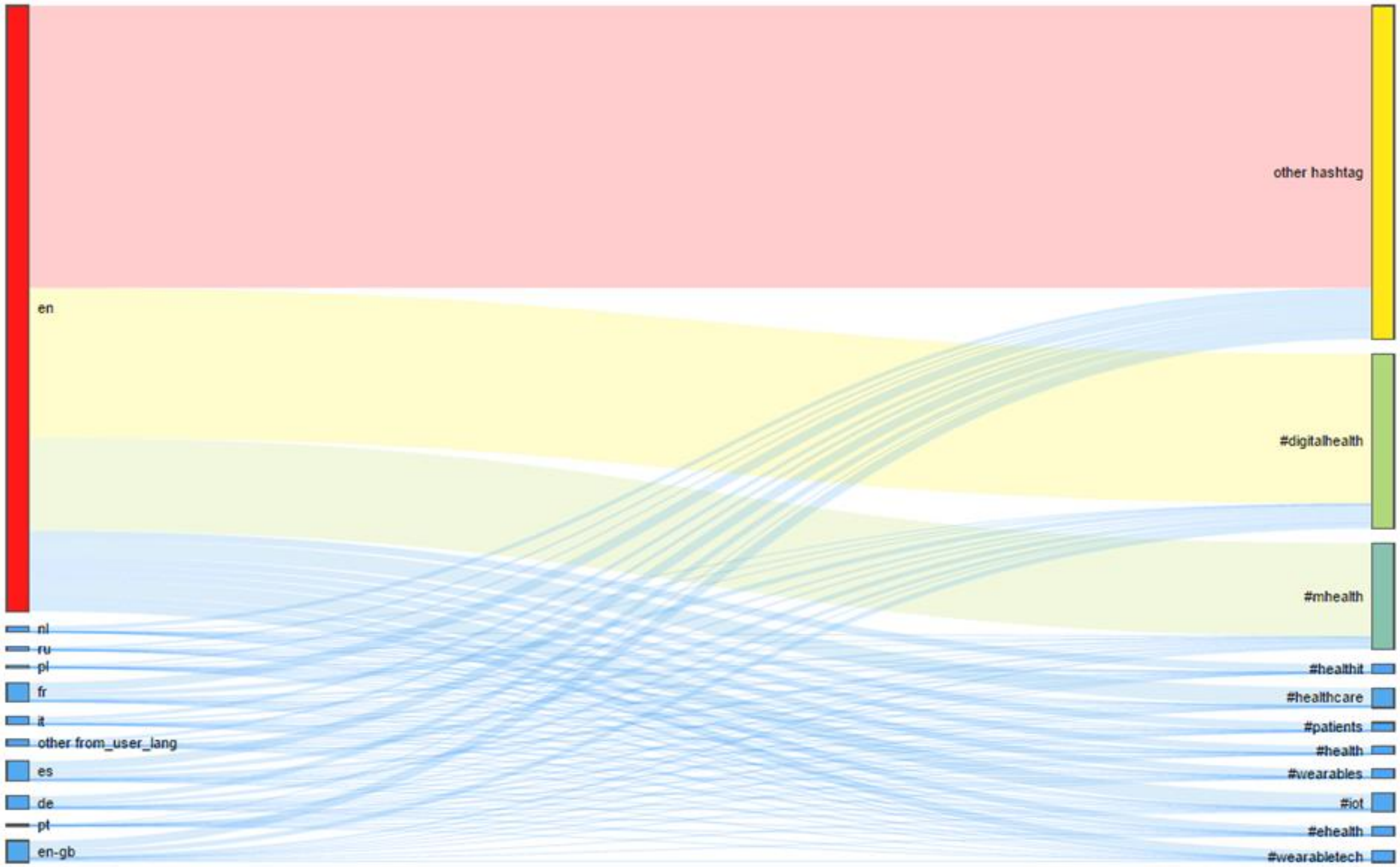
http://54.167.217.25

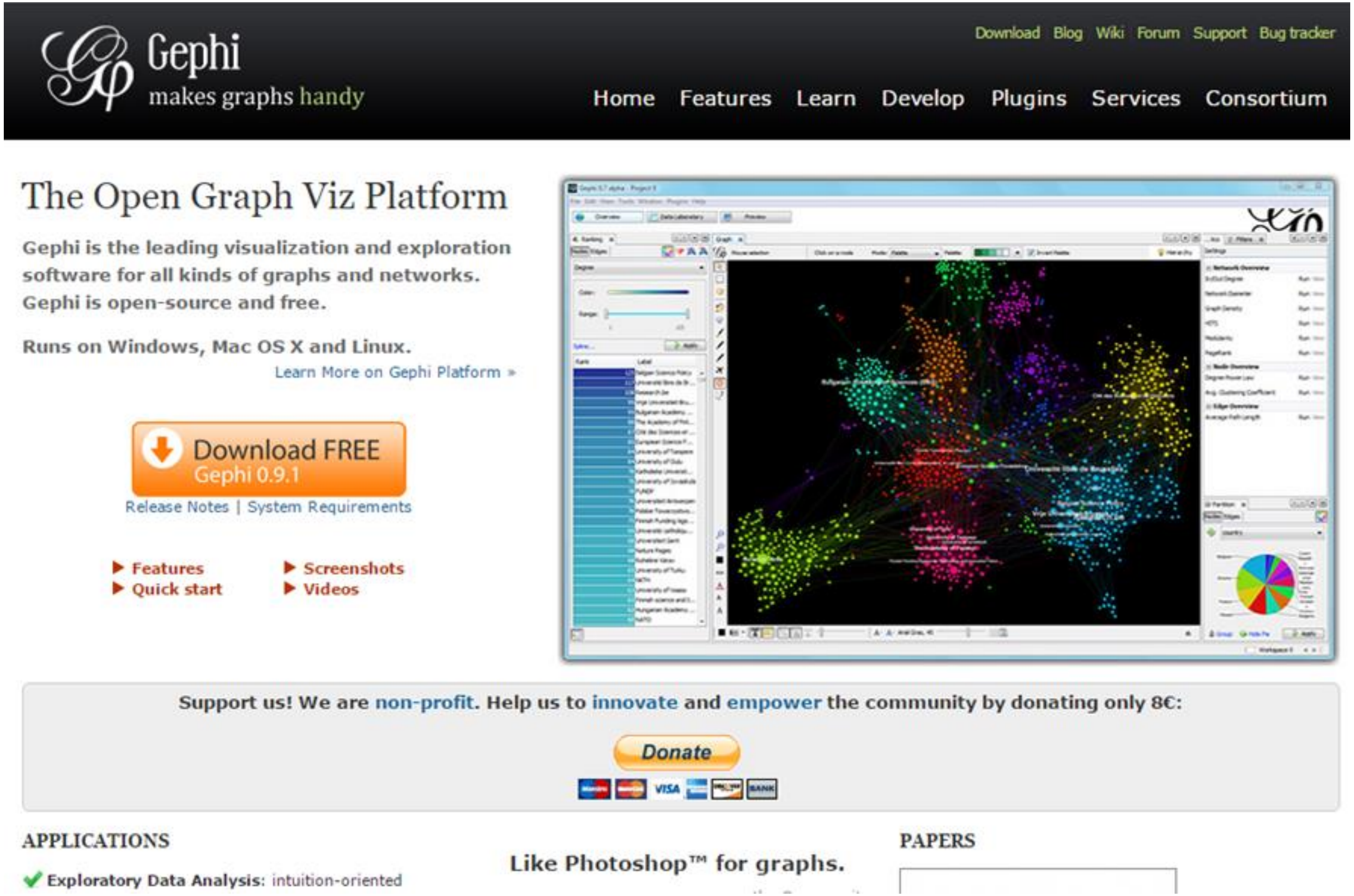

6,000
4,500
3,000
1,500
0
Tweets
Users
Locations
Geo coded
2016-06-21
2016-06-22
2016-06-23
2016-06-24
2016-06-25
2016-06-26
2016-06-27
2016-06-28
2016-06-29
2016-06-30
2016-07-01
2016-07-02
2016-07-03
2016-07-04
2016-07-05
2016-07-06
2016-07-07
2016-07-08
2016-07-09
2016-07-10
2016-07-11
2016-07-12
2016-07-13
2016-07-14
2016-07-15
2016-07-16
2016-07-17
2016-07-18
2016-07-19
2016-07-20
2016-07-21
2016-07-22
2016-07-23
2016-07-24
2016-07-25
2016-07-26
2016-07-27
2016-07-28
2016-07-29
2016-07-30
2016-07-31
2016-08-01
2016-08-02
2016-08-03
2016-08-04
2016-08-05
2016-08-06
2016-08-07
2016-08-08
Generate SVG
Date and time are in UTC (UTC).
Graph resolution
days
hours
minutes
update graph

DMI Twitter Capturing and Analysis Toolset (DMI-TCAT)

Has 36 different options – for extracting or downloading the data that has been collected.

DMI-TCAT is available on Github and is free. However installing it on a server requires money and so does running it.

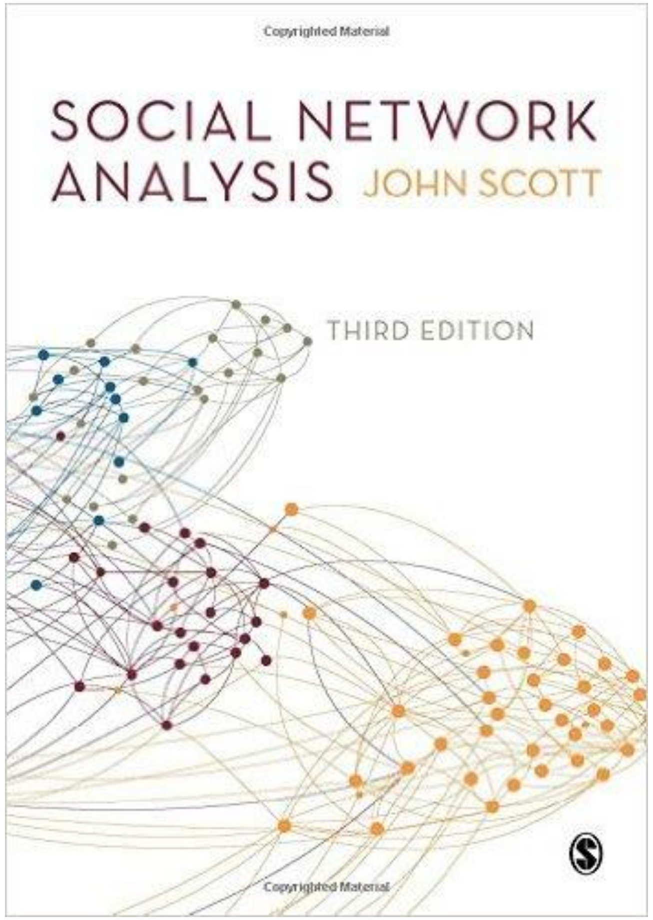

en
nl
ru
pl
fr
it
other from_user_lang
es
de
pt
en-gb
other hashtag
#digitalhealth
#mhealth
#healthit
#healthcare
#patients
#health
#wearables
#iot
#ehealth
#wearabletech


Gephi
makes graphs handy
Download Blog Wiki Forum Support Bug tracker
Home Features Learn Develop Plugins Services Consortium
The Open Graph Viz Platform
Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.
Runs on Windows, Mac OS X and Linux.
Learn More on Gephi Platform »
Download FREE
Gephi 0.9.1
Release Notes | System Requirements
Features
Quick start
Screenshots
Videos
Support us! We are non-profit. Help us to innovate and empower the community by donating only 8€:
Donate
APPLICATIONS
Exploratory Data Analysis: intuition-oriented
Like Photoshop™ for graphs.
PAPERS

The Various Algorithms to Choose From…

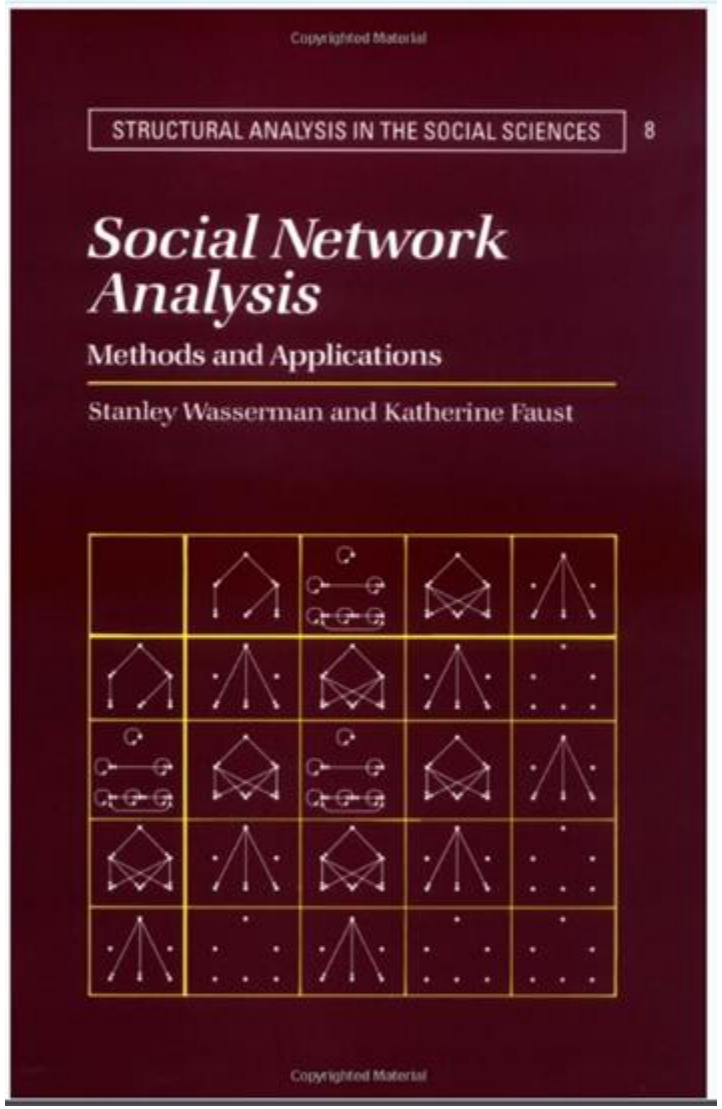
1	id	label	timeset	indegree	outdegree	degree	weighted d	weighted ir	weighted o	eccentricit	closnessc	betweenes	authority	hub	modularity	componen	pageranks	harmonicc	clustering	triangles	eigencentralit
2	897	#healthcare		72	0	72	957	957	0	4	0.408613	0.108482	0.002091	0	6	1	0.001261	0.44089	0.033646	86	0.666603
3	290	#health		47	0	47	373	373	0	4	0.379455	0.072401	0.001375	0	30	1	0.000856	0.41325	0.026827	29	0.473737
4	117	@healthythinker		5	514	519	1708	76	1693	5	0.351963	0.055618	0.000172	0.015915	22	1	0.009896	0.372563	0.001509	202	1
5	8312	@healthranger		1	463	464	613	4	609	5	0.332176	0.054006	5.73E-05	0.005305	1	1	0.011183	0.351481	3.72E-05	4	0.443949
6	32	@		36	0	36	150	150	0	4	0.347737	0.0528	0.00106	0	0	1	0.000665	0.379914	0.02381	15	0.318074
7	4531	@jod_ehealth		1	386	387	1567	22	1567	5	0.335209	0.042413	5.73E-05	0.005305	4	1	0.008545	0.348872	0	0	0.505168
8	7447	@picardonhealth		3	380	383	1205	30	1202	5	0.329081	0.041943	0.000115	0.01061	14	1	0.008787	0.347148	0.000948	69	0.371479
9	333	@meglobalhealth		1	347	348	1095	21	1095	5	0.329585	0.041867	5.73E-05	0.005305	15	1	0.008407	0.343724	0.000183	11	0.305347
10	3491	@healthydebate		1	367	368	1592	28	1592	5	0.334137	0.041493	5.73E-05	0.005305	14	1	0.008557	0.349583	0.00061	41	0.377999
11	9184	@helttihealthy		1	332	333	1290	123	1290	5	0.31885	0.040699	5.73E-05	0.005305	3	1	0.008761	0.336106	0	0	0.265454
12	8728	@dignityhealth		5	344	349	946	103	924	5	0.345461	0.03917	0.000172	0.015915	28	1	0.00767	0.360483	0.001342	81	0.471561
13	14354	@healthdata4all		2	369	371	4765	157	4764	5	0.33654	0.037179	8.59E-05	0.007958	19	1	0.007477	0.351246	0.000557	38	0.568403
14	4059	@dellhealth		5	253	258	1002	32	987	5	0.343485	0.035154	0.000172	0.015915	25	1	0.004849	0.357646	0.003496	115	0.535466
15	2373	@healthloop		5	360	365	2944	299	2898	5	0.342834	0.034383	0.000172	0.015915	19	1	0.006913	0.360259	0.003535	231	0.660325
16	7513	@smart_health_		1	302	303	1323	23	1323	5	0.336961	0.033625	5.73E-05	0.005305	18	1	0.006819	0.349145	0.000682	31	0.416239
17	7733	@healthtech_talk		2	317	319	3176	24	3174	5	0.339506	0.033477	8.59E-05	0.007958	27	1	0.00677	0.356257	0.002156	108	0.539018
18	1468	#digitalhealth		48	0	48	982	982	0	6	0.362891	0.033129	0.001403	0	16	1	0.000826	0.398177	0.044326	50	0.49184
19	1373	@healthspafi		3	290	293	1094	39	1082	5	0.33272	0.032435	0.000115	0.01061	3	1	0.006687	0.346897	0.000565	24	0.369456
20	4015	@healthitcentral		1	330	331	2455	2	2455	5	0.343597	0.030915	5.73E-05	0.005305	33	1	0.006215	0.358078	0.002376	129	0.65792
21	9969	@healthware_intl		3	310	313	1460	99	1417	5	0.336464	0.030884	0.000115	0.01061	30	1	0.00622	0.349157	0.001357	65	0.454601
22	1116	@newhealthfr		1	256	257	1031	31	1031	5	0.316699	0.03034	5.73E-05	0.005305	34	1	0.006387	0.331502	0	0	0.21292
23	10845	@tincturehealth		2	335	337	3038	223	3026	5	0.337874	0.030248	8.59E-05	0.007958	19	1	0.006324	0.35514	0.002038	114	0.57498
24	4169	@digihealthhelp		3	289	292	809	12	797	5	0.33726	0.029908	0.000115	0.01061	16	1	0.005948	0.3553	0.00273	116	0.473344
25	3774	@healthmap		0	269	269	533	0	533	5	0.282092	0.029867	0	0.002653	15	1	0.006627	0.303288	0	0	0.16599
26	3234	@healthbankcoop		1	314	315	1323	118	1323	5	0.329776	0.028812	5.73E-05	0.005305	13	1	0.006186	0.343081	0	0	0.517346
27	12032	@arab_health		2	239	241	860	172	854	5	0.333071	0.028423	8.59E-05	0.007958	15	1	0.005835	0.343165	0.000244	7	0.236781
28	1724	@ibmhealthcare		4	262	266	1218	89	1200	5	0.34099	0.02802	0.000143	0.013263	22	1	0.005266	0.354522	0.002601	91	0.482919
29	11409	@ochsnerhealth		2	225	227	682	65	678	5	0.335969	0.027473	8.59E-05	0.007958	5	1	0.005415	0.345128	0.000551	14	0.212592
30	5668	@barnabas_health		1	224	225	1114	140	1114	5	0.313201	0.027456	5.73E-05	0.005305	21	1	0.0058	0.327501	0	0	0.148858
31	6561	@nexjhealth		1	249	250	945	28	945	5	0.330732	0.026219	5.73E-05	0.005305	26	1	0.005394	0.341336	0	0	0.336692
32	4343	@ehealthscotland		2	241	243	896	17	893	5	0.327902	0.026099	8.59E-05	0.007958	18	1	0.005308	0.341262	0.000754	22	0.315195

Social Network Analysis – By John Scott



Scott, J. (2012). *Social network analysis*. Sage.

Social Network Analysis – Methods and Applications



Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

http://www.semantic-knowledge.com/tropes.htm

TROPES

High Performance Text Analysis for Professional Users

Analysis of written or spoken texts requires that certain questions should be asked with regard to certain objectives. To obtain answers to these questions, texts must be reduced as far as possible to their essentials.

Designed for Information Science, Market Research, Sociological Analysis and Scientific studies, Tropes is a Natural Language Processing and Semantic Classification software that guarantees **pertinence and quality in Text Analysis**.

Extraction of Relevant Information

Tropes can immediately detect contexts, isolate themes and identify principal actors, through the application of three levels of semantic classifications. You can quickly determine who says what to whom; who does what, where and when; and with what purpose.

Qualitative Analysis and Categorization

Tropes identifies the Text Style in order to place it in context and rapidly compare it with other texts. Tropes uses Semantic Meta-Categories to group verbs, adjectives, adverbs, personal pronouns and conjunctions.

Chronological Analysis

Tropes carries out a chronological analysis of a text from which the principal episodes can be isolated, the discussion blocks visualized and the development of an idea followed up.

Products

COMMERCIAL

ACADEMIC

TEXT DOCUMENT TYPES

SYSTEM REQUIREMENTS

LANGUAGE SUPPORT

ACADEMIC PUBLICATIONS TO DATE: 1050

Leximancer School Licensing

Academic Desktop Licence covering all users in a specific School or Faculty. Site-wide licensing options are also available for academic establishments. For more information please contact us.

- Academic School Licence - Please contact us for pricing.

ACADEMIC PURCHASES

Qualified academics and students may purchase a special, reduced cost Leximancer licence for *individual, non-commercial use.* This offer is not valid for commercial customers.

- Desktop Licence: Buy
- 1 Year Annual Desktop Licence: Buy
- Lexi-Portal 1 Month

Leximancer Store

ALL

ANNUAL

PERPETUAL

V4.5 / Windows 12 Month - Academic Desktop

$750.00 AUD

V4.5 / OS X 12 Month Academic Desktop

$750.00 AUD

Leximancer

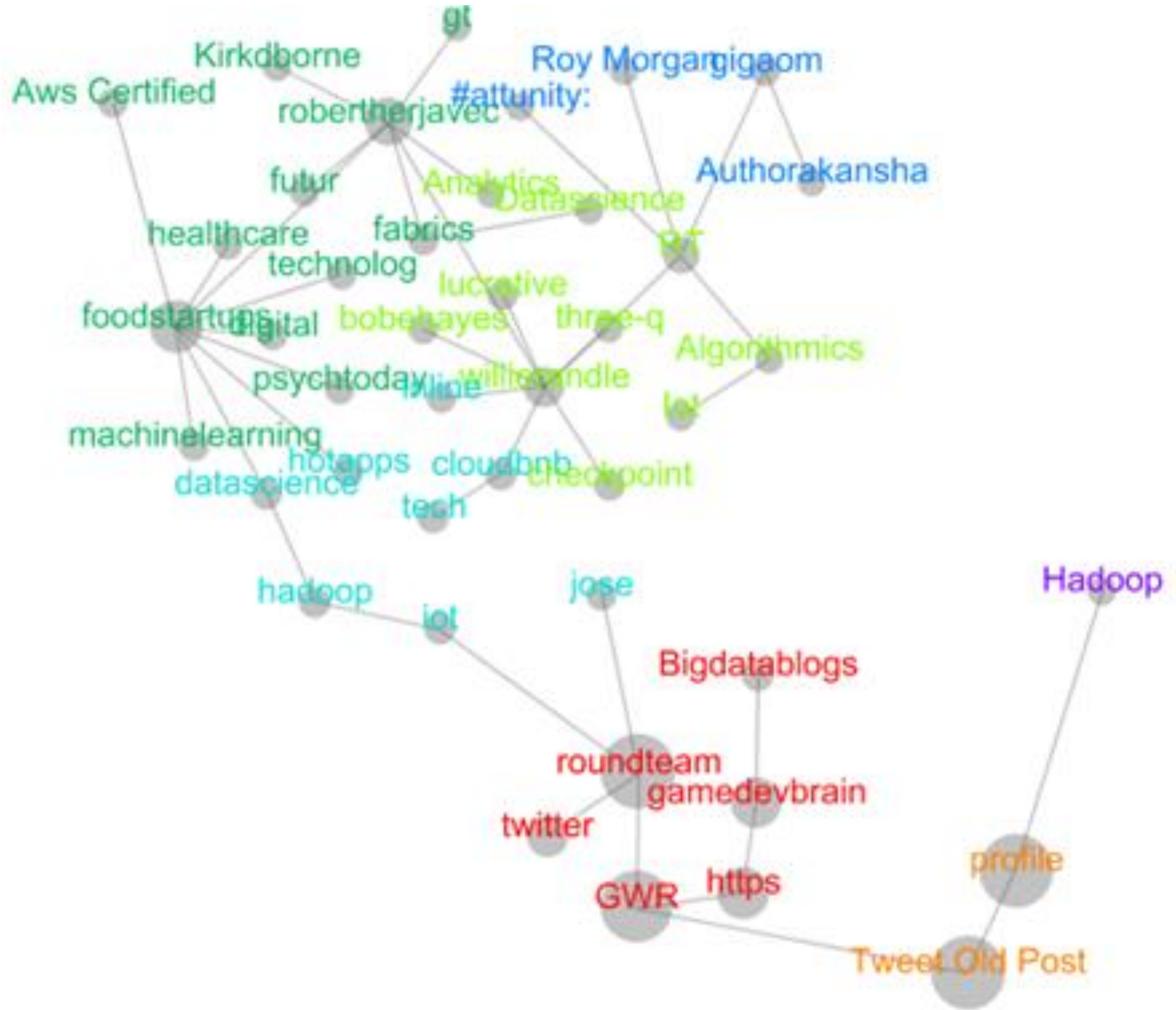
Leximancer is computer software that conducts quantitative content analysis using a machine learning technique. It learns what the main concepts are in a text and how they relate to each other. It conducts a thematic analysis and a relational analysis of the textual data. Leximancer provides word frequency counts and co-occurrence counts of concepts present in the tweets. It is:

[A] Method for transforming lexical co-occurrence information from natural language into semantic patterns in an unsupervised manner. It employs two stages of co-occurrence information extraction— semantic and relational—using a different algorithm for each stage. The algorithms used are statistical, but they employ nonlinear dynamics and machine learning.

Smith and Humphreys, p. 26

Once a concept has been identified by the machine learning process, Leximancer then creates a thesaurus of words that are associated with that concept giving the 'concept its semantic or definitional content'.

1,040,000 #BigData Tweets Analysed using Leximancer



Top Named Individuals

- @authorakansha - I'm 21 & I spent much of my time on Social Media & Photography. I'm also the author of a travel guide series
- @roymorganonline - Recognised leader in Social and Market Research, Melbourne, Australia - roymorgan.com
- @Kirkdborne - The Principal Data Scientist at @BoozAllen, PhD Astrophysicist, ♡ Data Science, Top Big Data Influencer. Ex-Professor http://rocketdatascience.org/ - Booz Allen Hamilton
- @Robertherjavec - Dad. Founder of Herjavec Group. Shark on ABC's Shark Tank. Author of You Don't Have to Be a Shark: Creating Your Own Success. Wherever I need to be. RobertHerjavec.com
- @willierandle - Follow Me, I'll Follow You! Tweet on! El Paso, TX
- @bobehayes - B.O.B. is Chief Research Officer @AnalyticsWeek. PhD in industrial-organizational psychology. Interests in #custexp #bigdata #statistics #analytics Seattle, WA • businessoverbroadway.com

Subscribe | Contact Us   Search in Entire Site   Search

Booz | Allen | Hamilton

WHAT WE DO | WHO WE SERVE | IDEAS & INSIGHTS | ABOUT US | MEDIA CENTER | CAREERS | INVESTORS

Home / Who We Serve / Defense and Intelligence Organizations

FONT SIZE: T T T   PRINT   SIGN UP

- WHO WE SERVE
- CIVILIAN GOVERNMENT AGENCIES
- DEFENSE AND INTELLIGENCE ORGANIZATIONS
  - Army
  - Air Force
  - Navy
  - Marine Corps
  - Defense Agencies
  - Space Agencies
  - Intelligence Community
- INTERNATIONAL
- NONPROFIT ORGANIZATIONS
- COMMERCIAL SOLUTIONS

Defense and Intelligence Organizations

Advanced strategy and technology solutions for our nation's defense

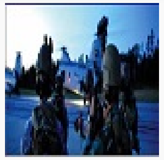
Elusive enemies. Emerging technologies. Unpredictable threats. Now more than ever, mission success depends on agile, adaptive thinking and operations.

From battlefield to boardroom, Booz Allen's global defense experts have the experience, perspective, and know-how to support the Department of Defense's toughest missions.

Since 1940, when founder Edwin Booz helped the Navy gear up for World War II, our focus on national defense has not wavered. Today, as our military services take on new missions, adopt innovative technologies, tackle acquisition and budgeting challenges, and address warfighters' medical needs, our experts are helping the Department of Defense be ready for what's next.

Use of the Department of Defense imagery does not constitute or imply endorsement.

C4ISR



Learn more about Booz Allen's C4ISR solutions.

SMART COMMAND SERIES

Learn more about The Smart Command Town Hall Series that Booz Allen conducted

THE U.S. ARMY WOUNDED WARRIOR PROGRAM

The Army Human Resources Command engaged Booz Allen Hamilton to improve communications and public outreach and ultimately strengthen support for

STAY CONNECTED
- Contact Us
- Sign up to receive Booz Allen's Latest Thinking
- Receive Investor Alerts

FIND YOUR NEXT JOB

Looking for a place to work, learn, and grow? Here are a few opportunities that might be of interest to you.

VIEW ALL CAREERS ›

Tweets by @BoozAllen

Booz Allen Hamilton Holding Corporation

Booz | Allen | Hamilton

Type	Public
Traded as	NYSE: BAH
Industry	Management consulting Government contractor
Founded	1914; 102 years ago
Founder	Edwin G. Booz James L. Allen Carl L. Hamilton
Headquarters	Tysons Corner, Virginia, U.S.[1]
Key people	Horacio D. Rozanski, President & Chief Executive Officer[1] John Michael McConnell, Vice Chairman
Services	Management and Technology Consulting
Revenue	▲ US$ 5.48 billion (*2014*)[2]
Net income	▲ US$ 239.955 million (*FY 2012*)[2]
Number of employees	22,000 (2014)
Website	www.boozallen.com

Edward Snowden was employed by Booz Allen Hamilton and the National Security Agency.

On Twitter the key influencer around the term #bigdata is a contractor who supplies staff to the NSA.

Data Mining, Social Network Analysis and Content Analysis

- What has been presented so far are a few of the techniques and tools of data mining and analytics – with machine learning and automation in Leximancer.
- Such insights that are born from the data and the application of algorithms need to be validated in the light of informed understanding of the **‘never raw data’** position especially for matters related to health.
- The existence of this ‘**data’** is the result of a long chain of requirements, goals and a shift in the wider political economy.
- The ‘insights’ are at the macro level – devoid of context and therefore an immediate sense of meaning.

Calculated Data Patients/Scientists

- Data Science generates crude quantitative knowledge, or “calculated publics” (Gillespie, 2014).
- In order to interpret the visualisations requires human perception.
- Big data and social media analytics can not answer questions about data scientists, or patients on their own.
- Big data and social media analytics requires wider knowledge of context and debates surrounding the topic at hand.

http://ethnographymatters.net/blog/2016/04/12/algorithmic-intelligence-reconstructing-citizenship-through-digital-methods/

Defined, managed, and governed

- Big/Social data does not represent what we think they do.
- Nonetheless it does represent *something*, and this *something* is certainly something worthy of our consideration.
- "That a census or a social survey is a *snapshot of the way our societies are regulated* is rarely remarked on and instead emphasis is given to the presumed objectivity of the categories and their data. This is the ideology of the small data era in action – the claim that it is science and not society that we are seeing through such instruments."
- Even if the referent of population data is not the population itself, we are still dealing with reference and meaning; we are glimpsing not a population in its totality, but the *various ways in which that population is defined, managed, and governed*.

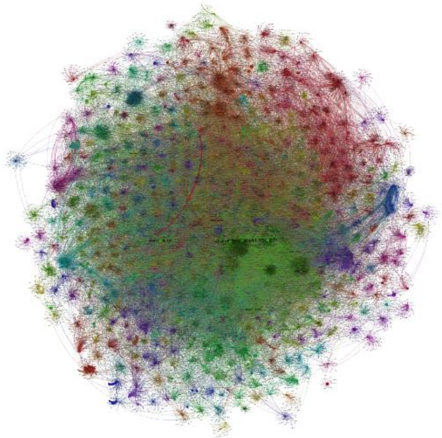
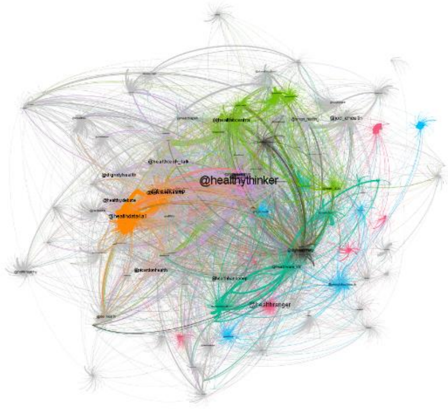
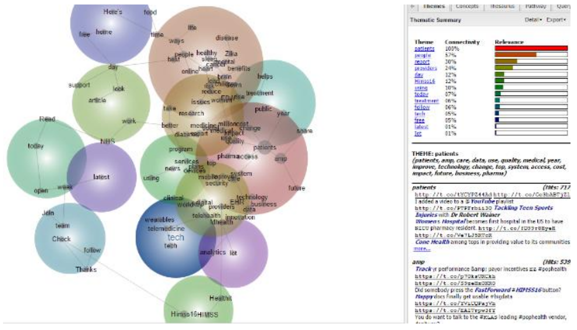
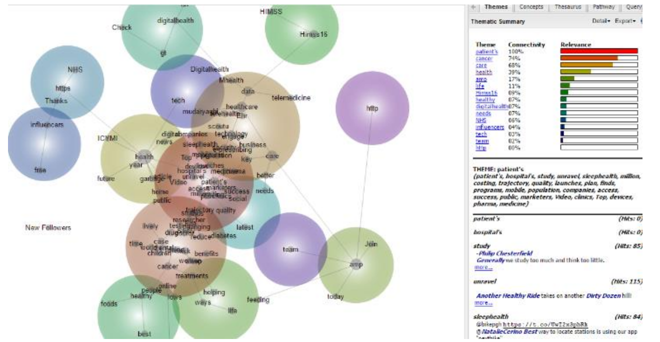
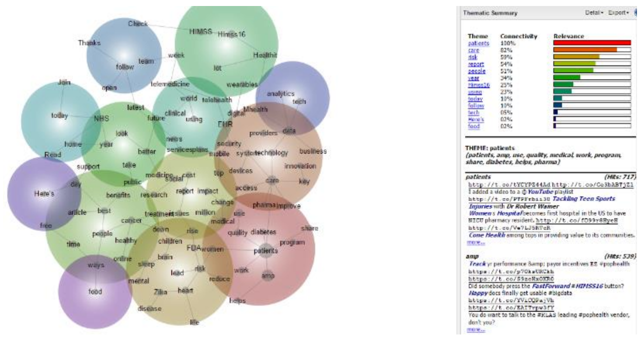
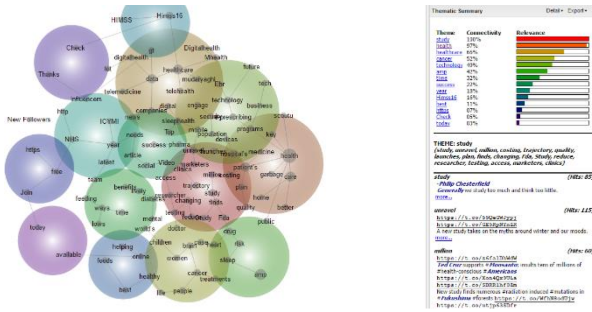
In Order to Interpret Data – Context and Critique are Crucial

- There are those within the Data Science discipline who are prepared to acknowledge the utility of human interpretation of data over algorithmic accounts.
- The Principal Data Scientist at @BoozAllen, (PhD Astrophysicist, Kirk Borne) recently stated that only using computer algorithms for visualisation…

"[…] can miss salient (explanatory) features of the data [therefore] a data analytics approach that combines the best of both worlds (machine algorithms and human perception) will enable efficient and effective exploration of large high-dimensional data".

http://rocketdatascience.org/?p=567

- Unsupervised concept and theme modelling is denaturalising and unfamiliar – but crucially they are not objective, unbiased or neutral.
- They are fictions.
- The modelling algorithm knows nothing about letters, nothing about narrative form, nothing about health.
- All it can tell us is that 1) this string of tokens (in our case, words) co-occur together more than we would expect, all things being equal, and 2) some particular documents (in this case, Tweets) are composed of a certain number of tokens (words) with a relatively high probability of belonging to this topic.

- The algorithm’s lack of knowledge of semantic meaning, and particularly its lack of knowledge of the social media as a form or genre, lets it point us to a very different model of the social.
- Such ‘Reading Machines’ are engaged in datafication of the social.

Datafication

- Datafication as defined by Kenneth Cukier and Viktor Mayor-Schoenberger (2013) refers to: ‘the ability to render into data many aspects of the world that have not been quantified before’.
- This is importantly ‘not the same as digitization, which takes analog content – books, films, photographs -and converts it into digital information, a sequence of ones and zeros that computers can read.
- Datafication is a far broader activity: taking all aspects of life and turn them into data’ (Cukier & Mayor-Schoenberger, 2013). It includes ‘behavioural metadata, such as those automatically derived from smartphones, like time stamps and GPS-inferred locations’ and may be used for a range of purposes ranging from surveillance to citizen empowerment (Kennedy, Poell and van Dijck 2015).

Data is NOT Objective NOT Incontrovertible

- The concern with the notion of **datafication** is that as it attempts to describe a certain state of affairs, as it occurs in one moment, it also flattens human experience, in a way that ethnography always defies, by acknowledging and insisting that whatever we label as ‘data’ is ‘rich’ and ‘lively’, rather than fixed, frozen or representing something ‘true’ about the world.

- Data, particularly that which is derived from huge conglomerate sources, is becoming increasingly a material or source for driving questions for and informing design practice. This is worrying in many ways - metrics derived through Big Data always represent a partial and non-representative sample (Baym, 2013) and thus do not accurately or adequately represent how people engage with and experience the world.
- While some might fall back on the positivist argument that we simply need to improve our measures, people –as everyday designers– will intentionally or unintentionally ensure that data is incomplete, dispersed and unfinished.

- There are growing movements towards data as fiction – embodied within data is an ideology.
- More recently for example the unfitbits online initiative (http://www.unfitbits.com/)
- Yet what is perhaps more worrying is that designers, developers, and policy makers will continue to take Big/Social Data at face value, as an object or representation of a truth that can be extracted from and that reflects the social.
- We are glimpsing the *various ways in which we are defined, managed, and governed.*