An Analysis of Big Data on Health:

Critique is Not Optional

Dr Steven McDermott

s.mcdermott@lcc.arts.ac.uk

@soci

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 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)

Not the 1ˢᵗ but 2ⁿᵈ Big Data Event

- While we now live in an era marked by the emergence of big data, this is not the first 'big' *information age* but the second.

- The first big data information age occurred in the early 19ᵗʰ century, developed exponentially with a rapid increase in the collection, storage and analysis of information from a growing number of primary sources.

- In particular, traditional and new disciplines emerged that began to adopt strategies of knowledge acquisition. New technologies made new ways of seeing possible and acceptable. New types of data became not only permissible but desirable.

- This was an age of knowledge factories - hospitals - that preceded our now digital computer age but one which set the scene for these later developments.

By Hamish Robertson and Joanne Travaglia - **Engines of Knowledge from the First Information Age: Medicine and the Hospital**
http://sociologicalimagination.org/archives/18555

- The hospital, the university and the laboratory delineated the acceptable boundaries of knowledge production, by directing focus and resources to particular topics and methods, and especially by claiming the power to authorize what constituted acceptable knowledge in specific fields and what did not.

- In short, this period was marked by the development of knowledge 'factories' with all the attendant features of formal organisational structures as they are commonly understood today.

- The locus of the hospital became a valuable site for not only data collection but a site for the production of growing volumes of data and its deployment by an expanding system of experts.

- It seems clear that authority, both then and now, is hugely enhanced by the capacity to establish a knowledge factory, to generate exclusive expert knowledge and to control its application and oversight.

- The current discourse surrounding 'Big Data' is not one to be admired or aspired to…

A *Social Physics*

- In the book *Social Physics,* Pentland (2014) argues that our greatly expanded ability to gather behavioural data will allow scientists to develop "a causal theory of social structure" and ultimately establish "a mathematical explanation for why society reacts as it does" in all manner of circumstances.

- Pentland believes, there is a lack of empirical rigor in the social sciences.

- Unlike physicists, who can measure the movements of objects with great precision, sociologists have had to make do with fuzzy observations.

The Limits of Social Engineering – Carr, 2014

Pentland's Experiments

- Pentland describes a series of experiments that he and his associates have been conducting in the private sector.

- They go into a business and give each employee an electronic ID card, called a "sociometric badge," that hangs from the neck and communicates with the badges worn by colleagues.

- Incorporating microphones, location sensors, and accelerometers, the badges monitor where people go and whom they talk with, taking note of their tone of voice and even their body language.

- The devices are able to measure not only the chains of communication and influence within an organization but also "personal energy levels" and traits such as "extraversion and empathy."

The Limits of Social Engineering – Carr, 2014

Productivity…

- In one such study of a bank's call centre, the 'researchers' discovered that productivity could be increased simply by tweaking the coffee-break schedule.

- "The power of social physics," he writes, "comes from the fact that almost all of our day-to-day actions are habitual, based mostly on what we have learned from observing the behaviour of others."

- Once you map and measure all of a person's social influences, you can develop a statistical model that predicts that person's behaviour, just as you can model the path a billiard ball will take after it strikes other balls.

The Limits of Social Engineering – Carr, 2014

Prediction...

- What really excites Pentland is the prospect of using digital media and related tools to change people's behaviour, to motivate groups and individuals to act in more productive and responsible ways.

- If people react predictably to social influences, then governments and businesses can use computers to develop and deliver carefully tailored incentives, such as messages of praise or small cash payments, to "tune" the flows of influence in a group and thereby modify the habits of its members.

The Limits of Social Engineering – Carr, 2014

Efficiency…

- Beyond improving the efficiency of transit and health-care systems, Pentland suggests, group-based incentive programs can make communities more harmonious and creative.

- "Our main insight," he reports, "is that by targeting [an] individual's peers, peer pressure can amplify the desired effect of a reward on the target individual."

The Limits of Social Engineering – Carr, 2014

Class is out-moded?

- Ultimately, Pentland argues, looking at people's interactions through a mathematical lens will free us of time-worn notions about class and class struggle.

- Political and economic classes, he contends, are "oversimplified stereotypes of a fluid and overlapping matrix of peer groups."

- Peer groups, unlike classes, are defined by "shared norms" rather than just "standard features such as income" or "their relationship to the means of production."

The Limits of Social Engineering – Carr, 2014

We need to talk about Corporate Power, Control and Discipline…

- The 'social' has always been a commercial and scientific resource – now in the digital age the competition regarding claims to which disciplines have justified control over this domain have intensified.

- The social sciences need to defend their subject area in order to preserve it.

- An application of the ethnographic/netnographic approach (Kozinets, 2010), social network analysis, data mining and machine-learning tools to highlight the certainties and uncertainties of Big Data and the Health Industry in order to start the process of uncovering the social and cultural forces that they are appropriating.

- What follows is the application of the tools of Big Data analytics on those that conduct Big Data analytics.

Is It Time to Bring Consumer Data into Health Care?

"Amazon [and other vendors] deliver just about everything but health care — yet probably know more about your habits and behaviours than your doctor does. That may be changing as health care providers begin using the consumer profiling tools that shape advertisements to get to know their patients beyond the examining room."

- *By Martha Hostetter and Sarah Klein 2015*

"There are **just five or six types of patients in the United States**, experts in consumer profiling say."

"Consumers don't realize how this information can be used or even fathom the possibilities of it."
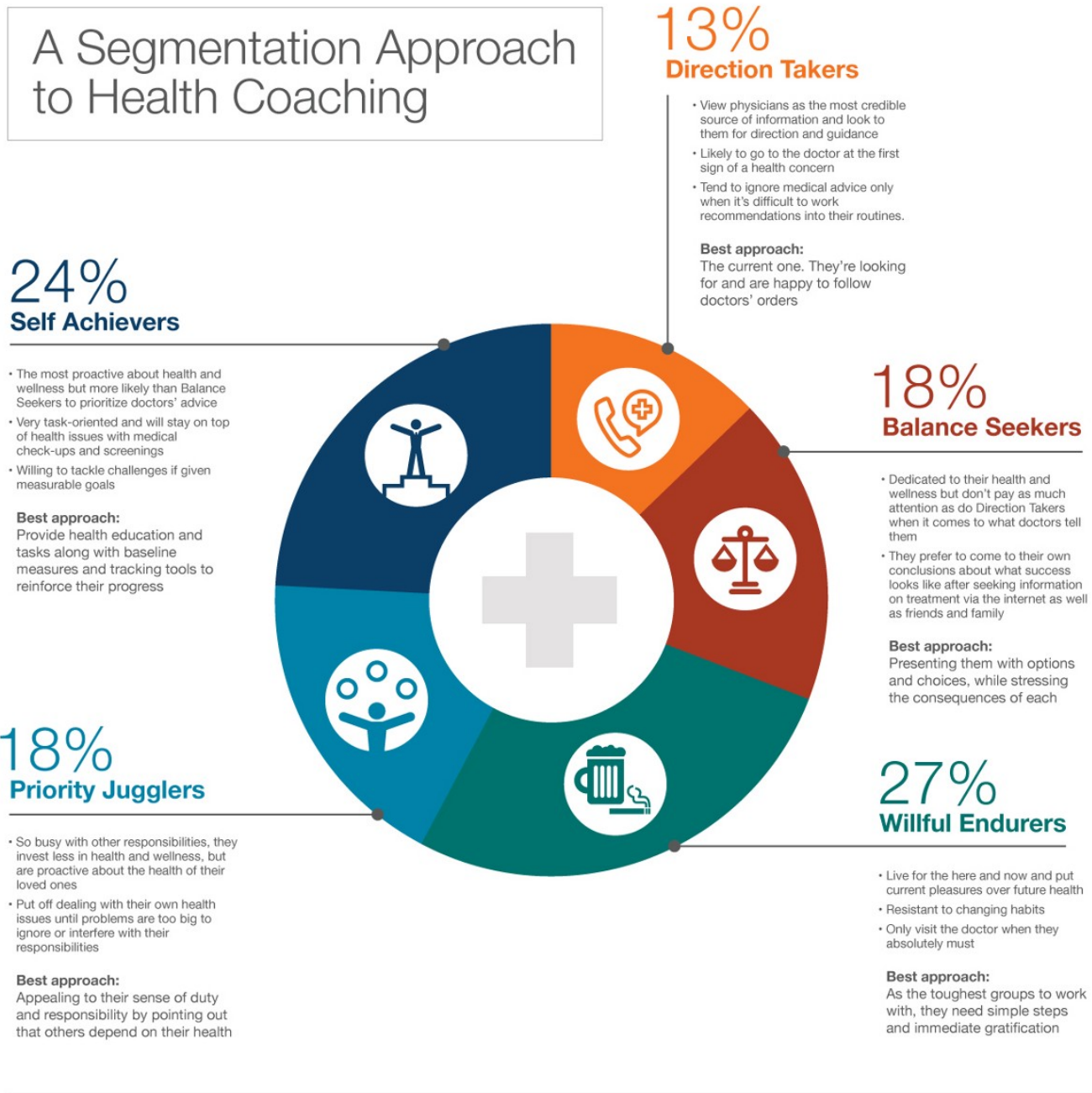
"We get so caught up in technology about how fun it might be to use and how easy," says Linda Hull of consulting firm Mercer.

"Once you go outside of health care institutions that are obligated under federal law to protect your privacy, you are fair game for all kinds of people who are data-mining," including pharmaceutical companies and insurers.

A Segmentation Approach to Health Coaching

13% Direction Takers

- View physicians as the most credible source of information and look to them for direction and guidance
- Likely to go to the doctor at the first sign of a health concern
- Tend to ignore medical advice only when it's difficult to work recommendations into their routines.

Best approach:
The current one. They're looking for and are happy to follow doctors' orders

24% Self Achievers

- The most proactive about health and wellness but more likely than Balance Seekers to prioritize doctors' advice
- Very task-oriented and will stay on top of health issues with medical check-ups and screenings
- Willing to tackle challenges if given measurable goals

Best approach:
Provide health education and tasks along with baseline measures and tracking tools to reinforce their progress

18% Balance Seekers

- Dedicated to their health and wellness but don't pay as much attention as do Direction Takers when it comes to what doctors tell them
- They prefer to come to their own conclusions about what success looks like after seeking information on treatment via the internet as well as friends and family

Best approach:
Presenting them with options and choices, while stressing the consequences of each

18% Priority Jugglers

- So busy with other responsibilities, they invest less in health and wellness, but are proactive about the health of their loved ones
- Put off dealing with their own health issues until problems are too big to ignore or interfere with their responsibilities

Best approach:
Appealing to their sense of duty and responsibility by pointing out that others depend on their health

27% Willful Endurers

- Live for the here and now and put current pleasures over future health
- Resistant to changing habits
- Only visit the doctor when they absolutely must

Best approach:
As the toughest groups to work with, they need simple steps and immediate gratification

Source: c2b Solutions

Network and the Discourse

- There are competing discourses surrounding 'Big Data' and Health. On the one hand business, marketing and advertising interests are promoting Big Data as information that no longer requires theory or the scientific methodologies of old.

- On the other are voices from the academy; digital humanities and computational social sciences that wish to benefit from the volumes of available data. It is these (and other) competing discourses that are the target of this research.

- Here I argue that those engaged in 'data without theory' are generating a relational social mechanism similar to that of self-fulfilling prophesies of Merton, the network effects of Coleman and the bandwagon effects of Granovetter (Donati, 2015:66) and leaving no room for critique.

Emerging Knowledge 'Factories'

- The accumulation of experts, expertise and information strengthened claims as a knowledge factories for the emergence of scientific medicine.
- Are we witnessing the interplay of a gradually emerging technology mixing with new practices and new channels for knowledge production. A very gradual strategy and emergence of new knowledge factories.
- The knowledge producing 'expert' acquires authority and license through a claim to produce certain kinds of knowledge, through processes that were only rarely open to the purview of the layman.
- Expertise becomes closely correlated with authorised knowledge and practice as the knowledge produced, and the engines and engineers of knowledge acquire status.
- For Scientific medicine this was supported by the adoption of the Hippocratic Oath.
- Big Data 'experts' are in dire need of ethics training.

- The connections surrounding big data and health will lead to a mapping of the commercial institutions and people who are engaged in the emergence.
- The textual data will be analysed using software (Leximancer) to determine the tacit knowledge with which the authors refer to 'Big Data' and Health; it will measure levels of nominalisation, modality and generalisation in an attempt to unpack what they claim to know (be certain about) and do not know (be uncertain about) in relation to Big Data and Health.
- By using the tools of Big Data and a netnographic approach the task of unpacking the certainties and uncertainties that surround Big Data and Health will begin.
- Concerned with a lack of "strategic ambiguity" or doubt.
- And when you collect everything – you see nothing.

Insights from the Data?

- Big Data analytics argue that their approach enables an entirely new epistemological approach for making sense of the world and health; rather than testing a theory by analysing relevant data, new data analytics seek to gain insights 'born from the data' (Kitchin, 2014:2).

- 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.

- What follows is a preliminary study of Big Data and health as it appears in social media networks (Twitter initially) and the discursive disciplines. The initial data set consists of Twitter messages containing the phrase Big Data collected July to September 2015.

Overview of the Data Set…

- My analysis covers a number of data sets for hashtags, which are 'keyword hashtags': they are unlikely to serve as the focal point for topical discussion communities.

- Such data sets have a very high percentage of around 90% of the tweets contain URLs or hyperlinks , indicating that they are largely used for sharing links relevant to these topics – and in fact that hashtags in these tweets may be used more as a form of emphasis than with the intent to attract an ad hoc issue public around these topics.

- The different levels of retweeting observed for these hashtags may serve as an indicator of the extent to which *Twitter* users are monitoring these hashtags for information: it is notable that the more topically specific hashtags (such #bigdata) attract considerably greater percentages of retweets (well above 40%), while the more generic terms, such as #job, remain at a much lower level of retweeting.
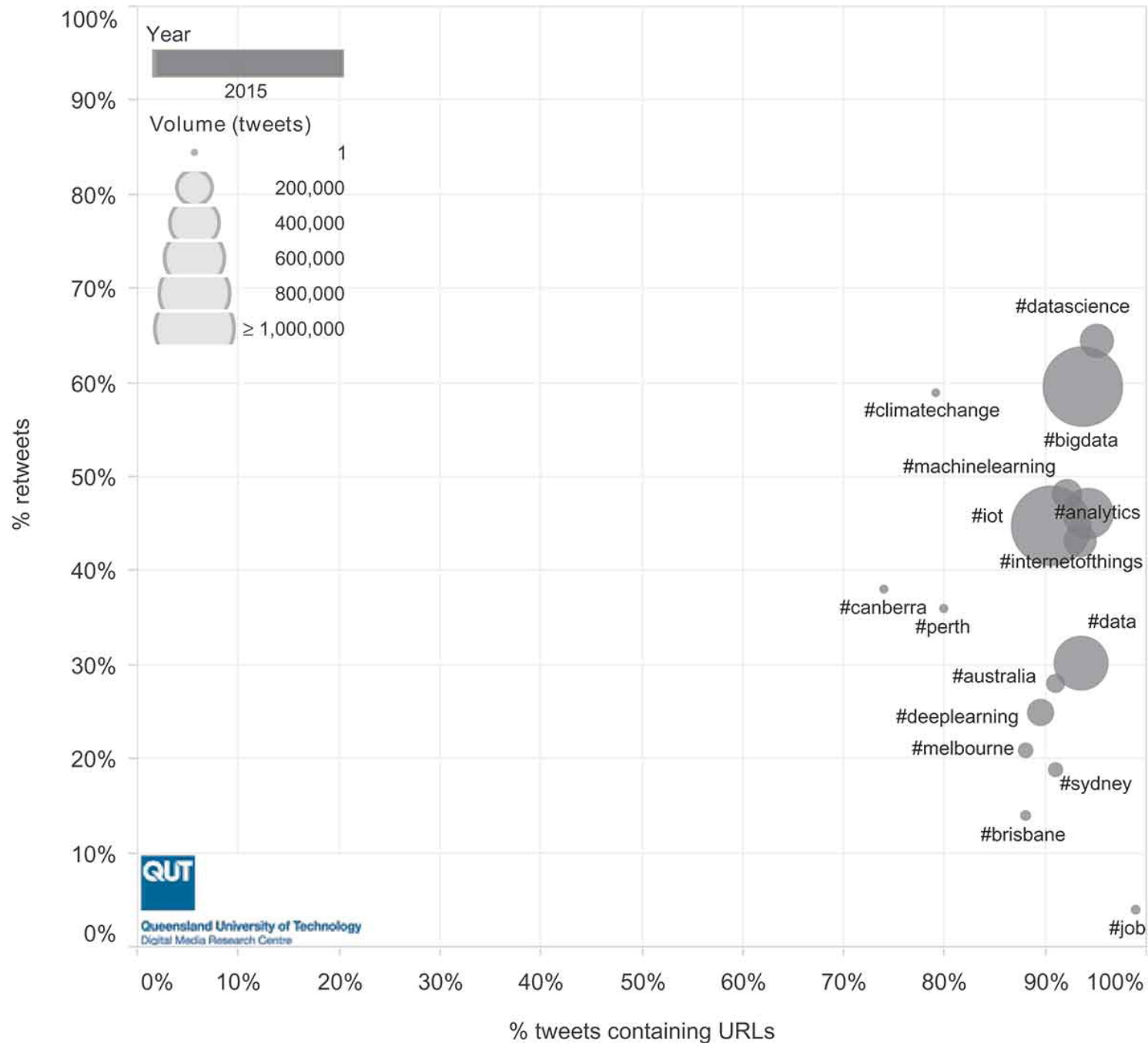
Patterns for keyword hashtag data sets.

"what we have defined as keyword hashtags constitute a very different way of using hashtags – largely for emphasis rather than to institute an issue public –, and the uses and utility of such hashtags remain to be explored in greater detail still."

(Axel Bruns , Brenda Moon , Avijit Paul , Felix Münch 2016)

Total Number of Tweets Archived July to September 2015.

Keyword / Hashtag	Description	Count	Create Time
#bigdata	preliminary scrape	1040644	Tue, 30 Jun 2015 12:39:44 +0000
#analytics	secondary scrape	403748	Tue, 30 Jun 2015 13:17:30 +0000
#iot	secondary scrape	1051596	Tue, 30 Jun 2015 13:17:47 +0000
#datascience	secondary scrape	179621	Tue, 30 Jun 2015 13:18:14 +0000
#deeplearning	secondary scrape	113215	Tue, 30 Jun 2015 13:18:47 +0000
#data	secondary scrape	463725	Tue, 30 Jun 2015 13:19:06 +0000
#machinelearning	secondary scrape	139813	Tue, 30 Jun 2015 13:20:25 +0000
#internetofthings	secondary scrape	169557	Tue, 30 Jun 2015 17:57:39 +0000

How to Read the Maps that Follow...

- In these maps, each dot (or node) represents a Twitter user and the lines joining them represent communication between users, retweets or mentions.

- The size of the dot represents the pagerank of the user. In simple terms, pagerank measures how "important" a node is based on the "importance" of the nodes which connect to it – "importance" denoted by links.

- The width of the line between users indicates the quantity of links. Groups of users with large numbers of links (retweets and mentions) between them stand out as thick and dark lines.

A Map of 30,000 Data Scientists on Twitter 2015.

French

IBM

The Data Lab

American

SASItaly

A Map of 30,000 Data Scientists on Twitter 2015.

French

IBM

The Data Lab

American

SASItaly
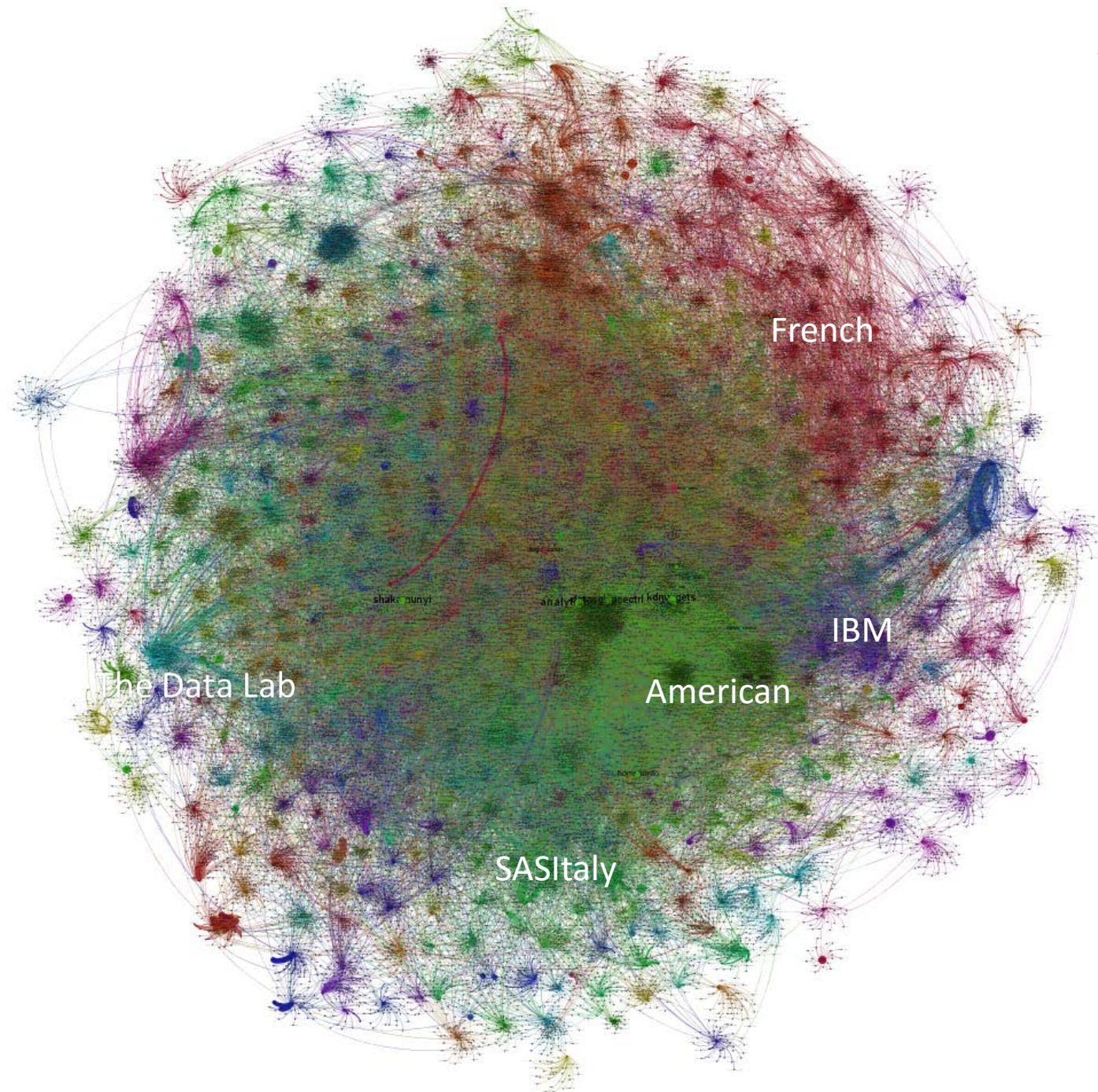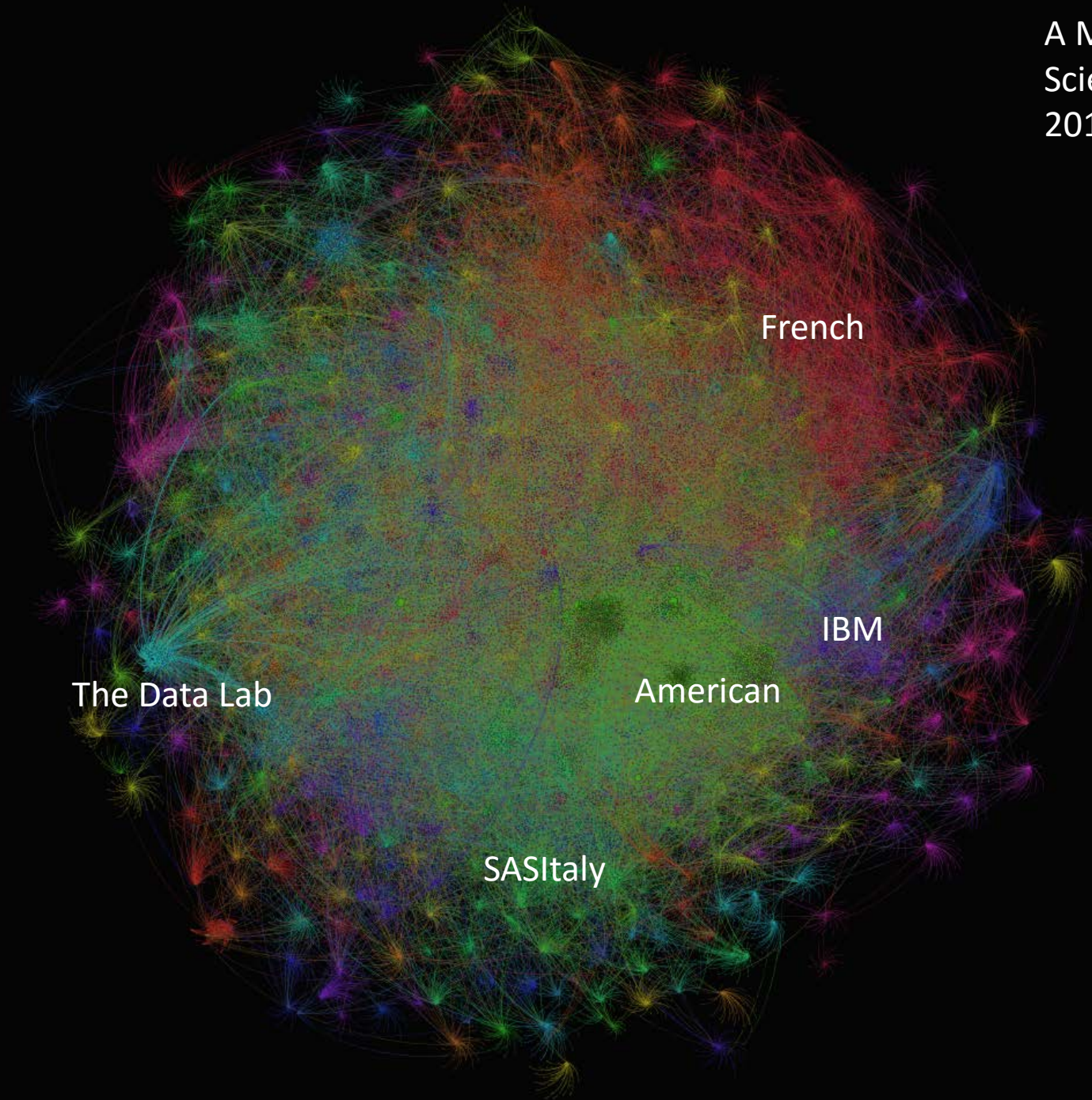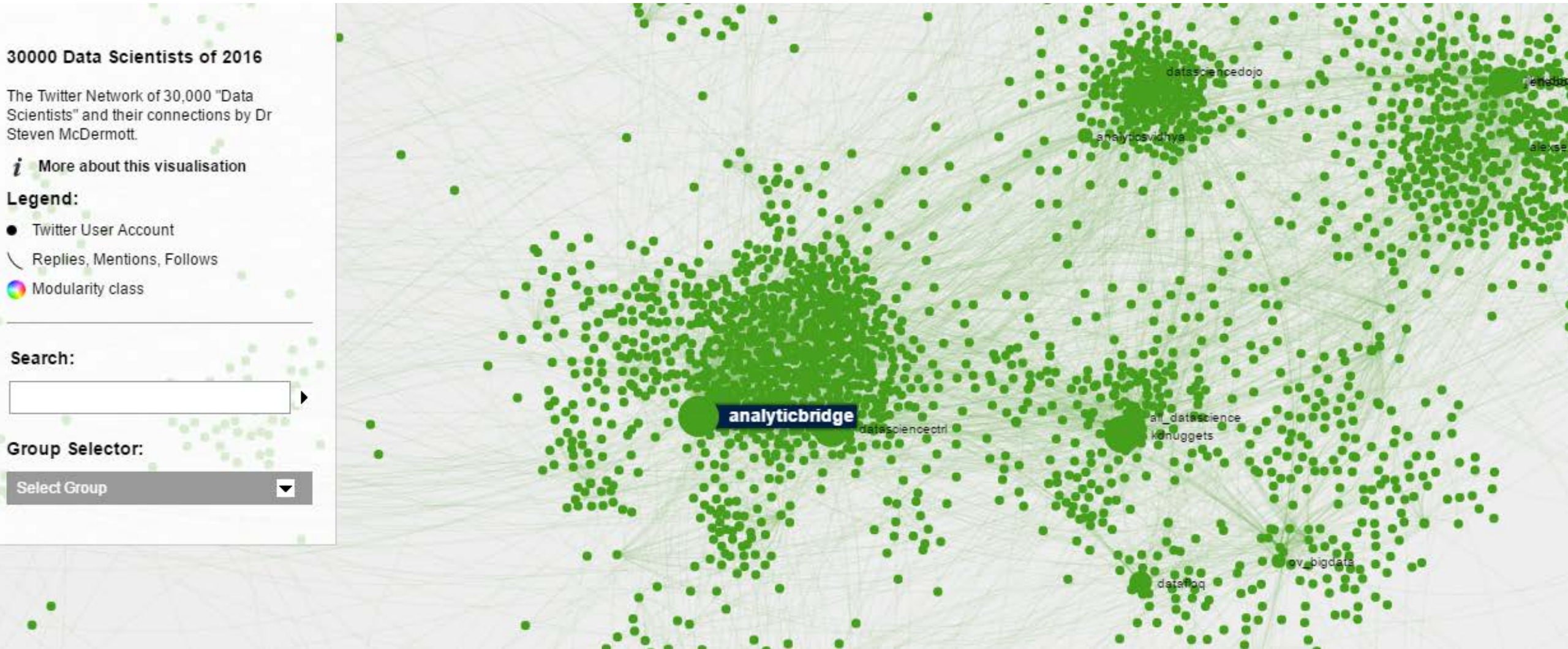
Searchable and interactive version available online - http://snacda.com/2016/02/04/data-scientists-of-2016/
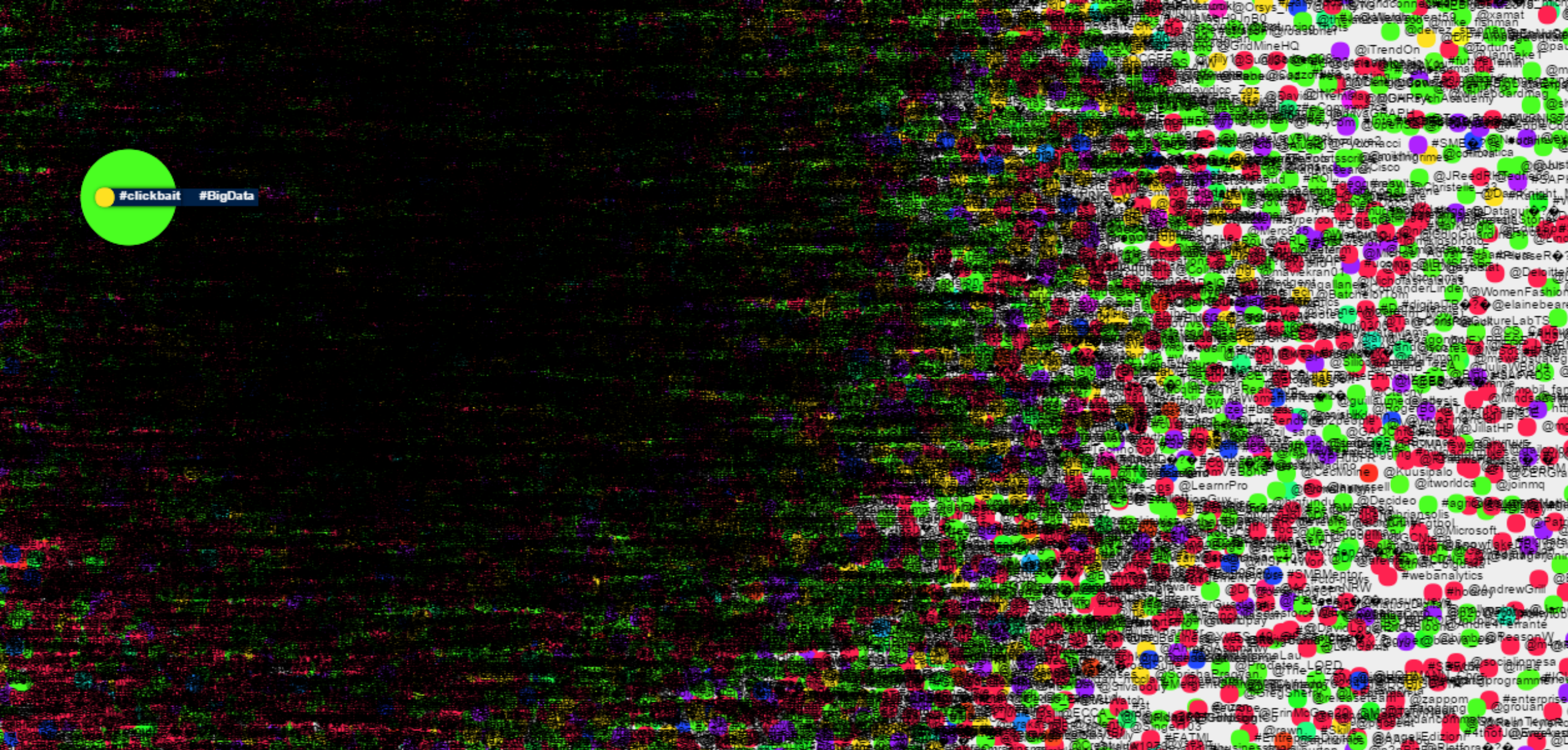
4 million tweets containing the Hashtag #bigdata July 2015
This image is a close up of approx. 50,000 nodes and 130,000 edges from that collection. The two most central hashtags are #bigdata and #clickbait.

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 (or semantic) 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
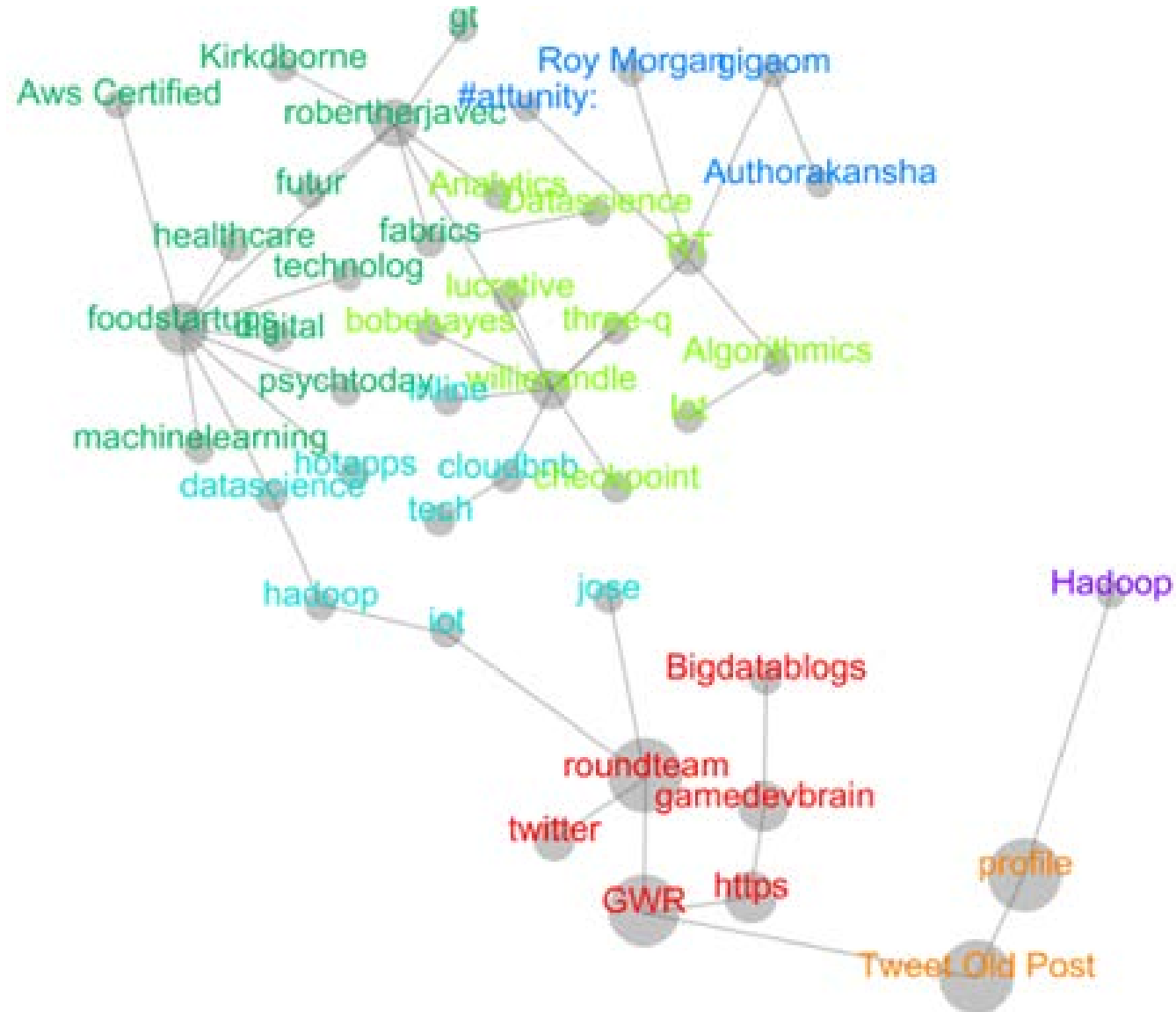
Top Named Individuals

- @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

Booz | Allen | Hamilton

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

FONT SIZE: т 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.

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 ▸

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

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 ⧉

Search...
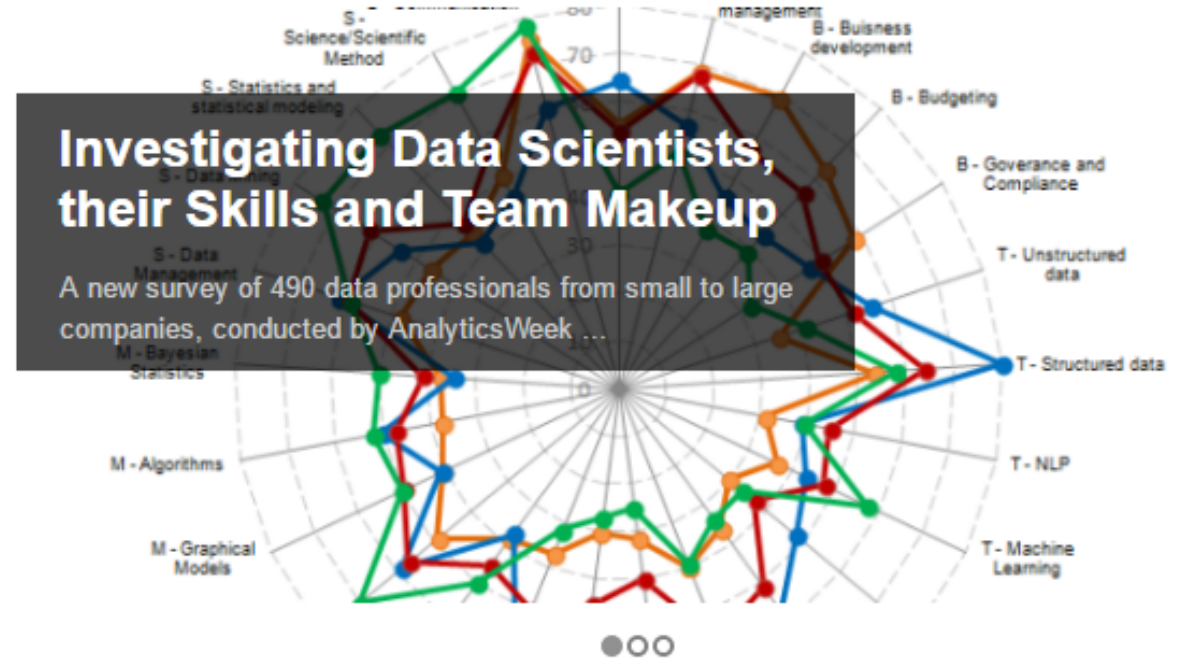
Investigating Data Scientists, their Skills and Team Makeup

A new survey of 490 data professionals from small to large companies, conducted by AnalyticsWeek …

● ○ ○

Understanding Customer Survey Data: Descriptive, Predictive and Prescriptive Analytics to Improve Customer Loyalty

Customer relationship surveys play a major role in helping improve the customer experience and increase customer loyalty. By gathering customer perceptions of their experience as well as their likelihood of

About me

I am Business Over Broadway (B.O.B.). I like to solve problems, primarily business problems,

through the application of the scientific method. I use data and analytics to help make decisions that are based on fact, not hyperbole. My interests are in customer experience, Big Data and analytics. To learn more about me and what I do, click here.

POPULAR   LATEST   COMMENTS   TAGS

Data Science Skills and the Improbable Unicorn
OCTOBER 12, 2015

The Big Picture of Big Data for 2014
JANUARY 29, 2014

Top Named Concepts

- Datascience
- Analytics
- profile
- Roundteam - The Twitter content management platform. roundteam.co
- foodstartups
- Gamedevbrain
- twitter
- gigaom
- iot
- Hadoop
- checkpoint
- Cloudbnb
- datascience
- tech
- machinelearning

- fabrics,
- hotapps
- lucrative
- gt
- healthcare
- three-q
- digital
- Psychtoday - Insight about everybody's favorite subject: Ourselves. New York, NY psychologytoday.com
- futur
- technolog
- inline

Find a Therapist

City or Zip   🔍

Let Your Mind Wander
Happy daydreaming

Most Popular

1 — **The Logic of Falling in Love**

2 — **7 Signs You're Not Mentally Strong, You're Just Acting Tough**

3 — **What Actually Happens When a Narcissist Looks in the Mirror?**

4 — **Are You a Narcissist? 6 Sure Signs of Narcissism**

5 — **Do Smart People Make Good Friends?**
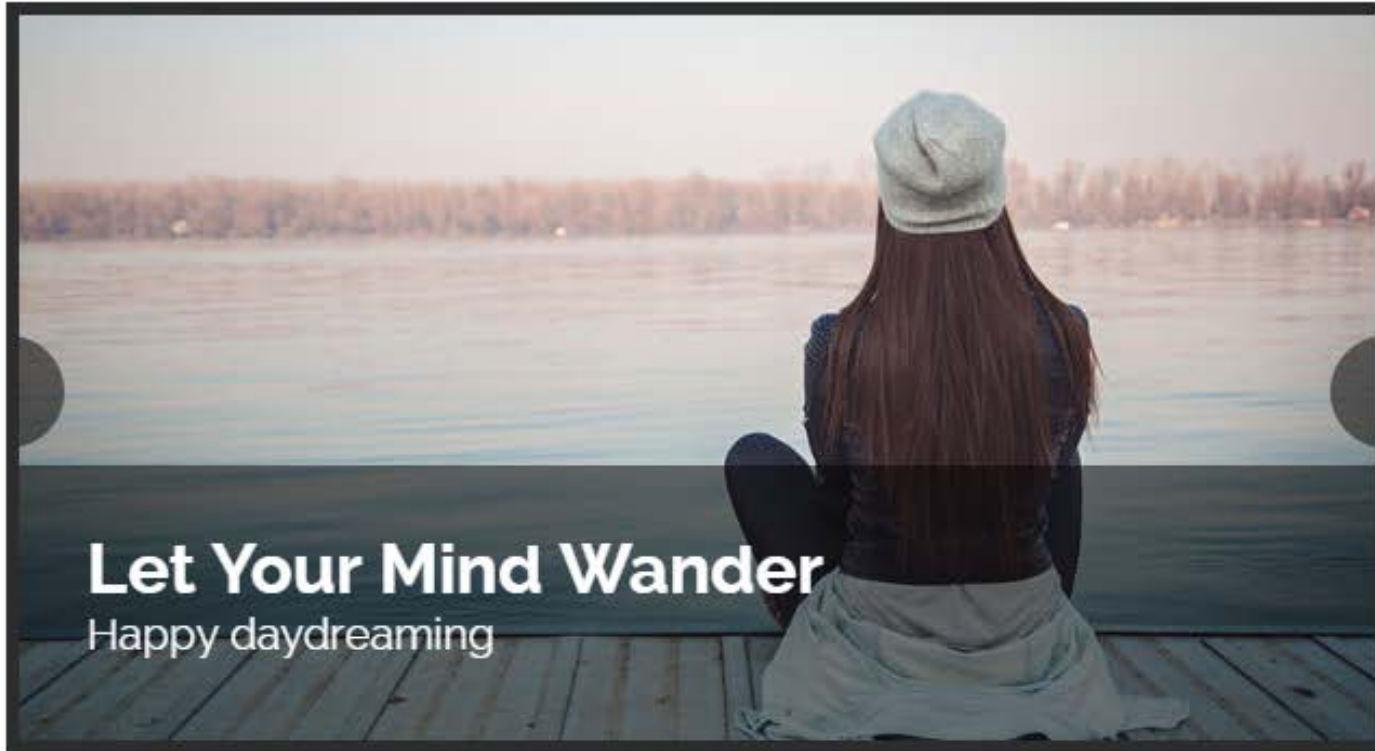
Essential Reads

EVOLUTIONARY PSYCHOLOGY

Living the Rich Life

Big Data and Health Tweets ONLY

- Searching for the Big Data and Health group.
- Or at least those people who sometimes refer to big data and health.
- The graph of 30,000 Big data tweets contained 200 (approx.) Twitter names with the word health appearing in it. (@Ibmhealth for example.)
- The list of 200 entered in Nodexl and it was set to look at the most recent 200 tweets and who they were to.
- It returned a list of 15,000 Twitter accounts and hashtags.

Betweenness Centrality - Influencers
@cdw_healthcare
@digihealthhelp
@mhealthinsight
@dellhealth
@healthtech_talk
@digitalhealth2
@tincturehealth
@picardonhealth
@adihealth
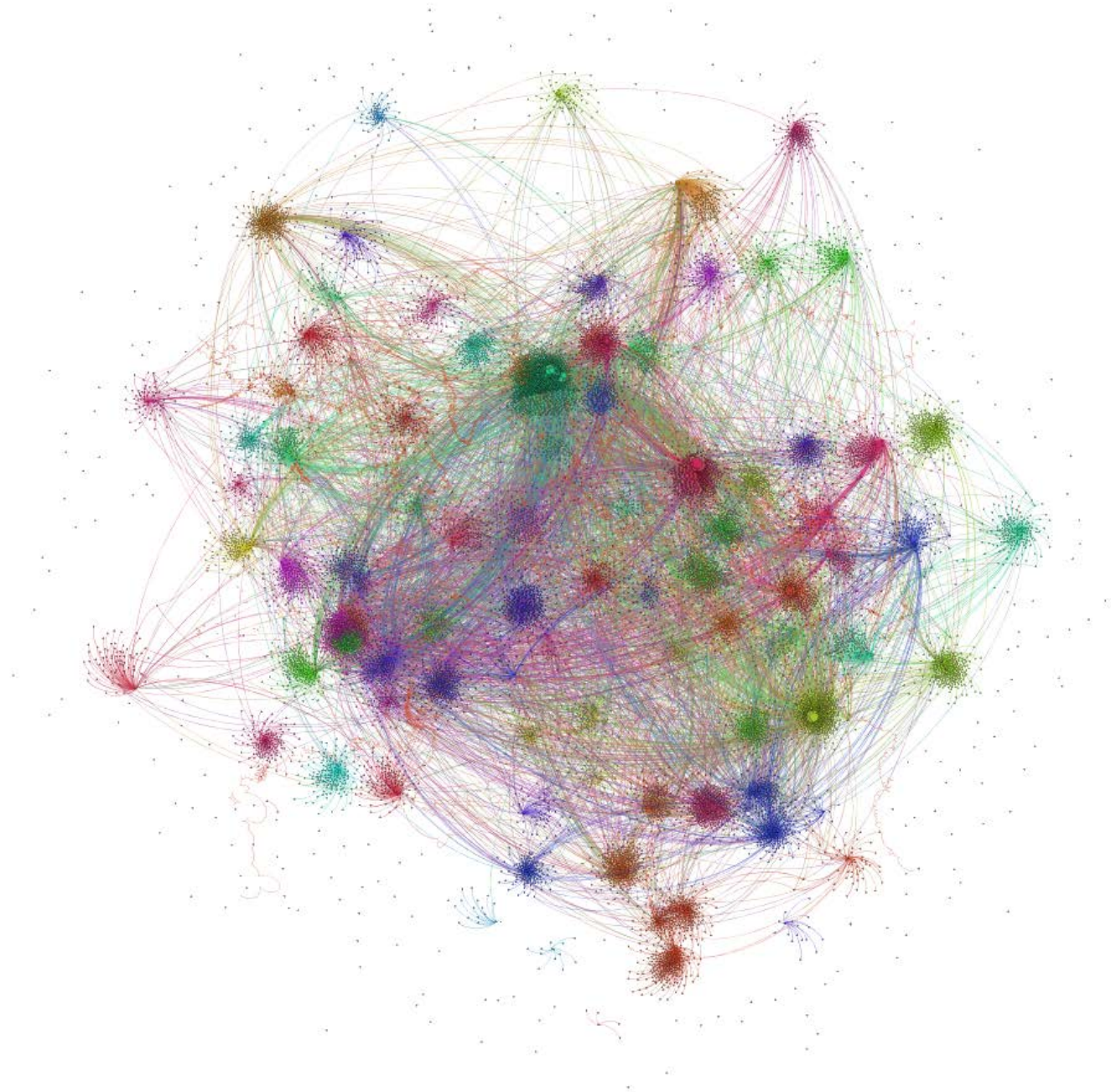
14832 nodes
20016 edges

This image contains the core group and outliers that are loosely connected

HEALTHCARE COMMUNIT

Powering patient care through technology across the continuum of care.

NEWS   TOPICS   PERSPECTIVES   TECH TIPS   RESOURCES

The Empowered Patient

SHARE

THE EMPOWERED PATIENT

Putting Health in the Hands of Patients

Technology is the lynchpin of the patient empowerment revolution. The IT choices you make today can determine how well patients view the care they receive and how much you stand out from your healthcare competition.

OUTPATIENT CLINIC

CDW Corporation

CDW

Type	Public
Traded as	NASDAQ: CDW
Industry	B2B IT products and services
Founded	1984 by Michael Krasny
Headquarters	Vernon Hills, Illinois, United States
Key people	Thomas E. Richards, Chairman & CEO
Products	Desktops, servers, laptops, peripherals, software, telephony products, power, storage
Revenue	▲US$12.9 billion (2015) [1]
Owner	Madison Dearborn Partners Providence Equity Partners
Number of employees	7,200+ [2]
Slogan	People Who Get IT [3]
Website	www.cdw.com

Betweenness Centrality - Influencers
@healthythinker
@healthranger
@jod_ehealth
@picardonhealth
@meglobalhealth
@healthydebate
@helttihealthy
@dignityhealth
@healthdata4all
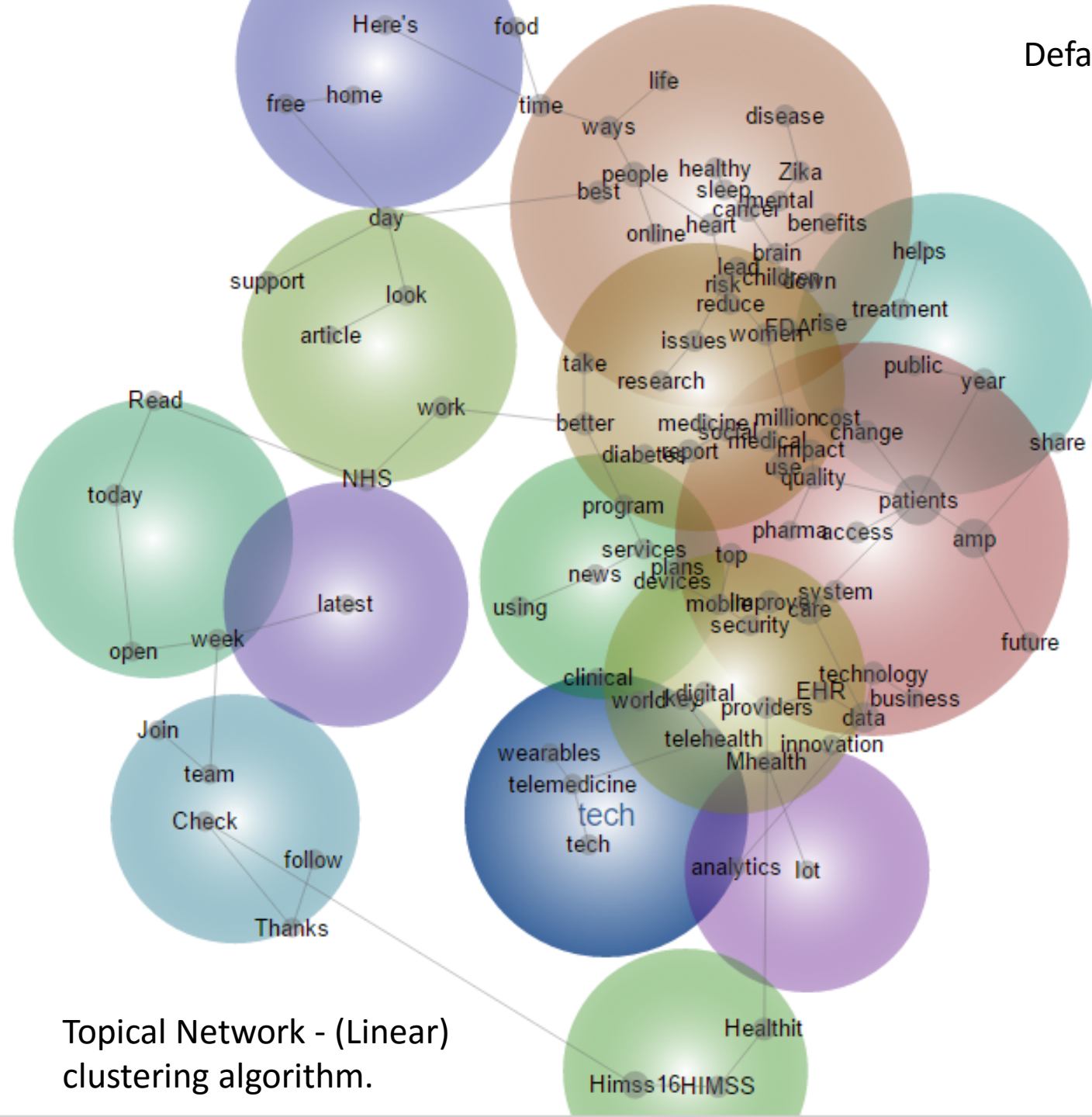@dellhealth
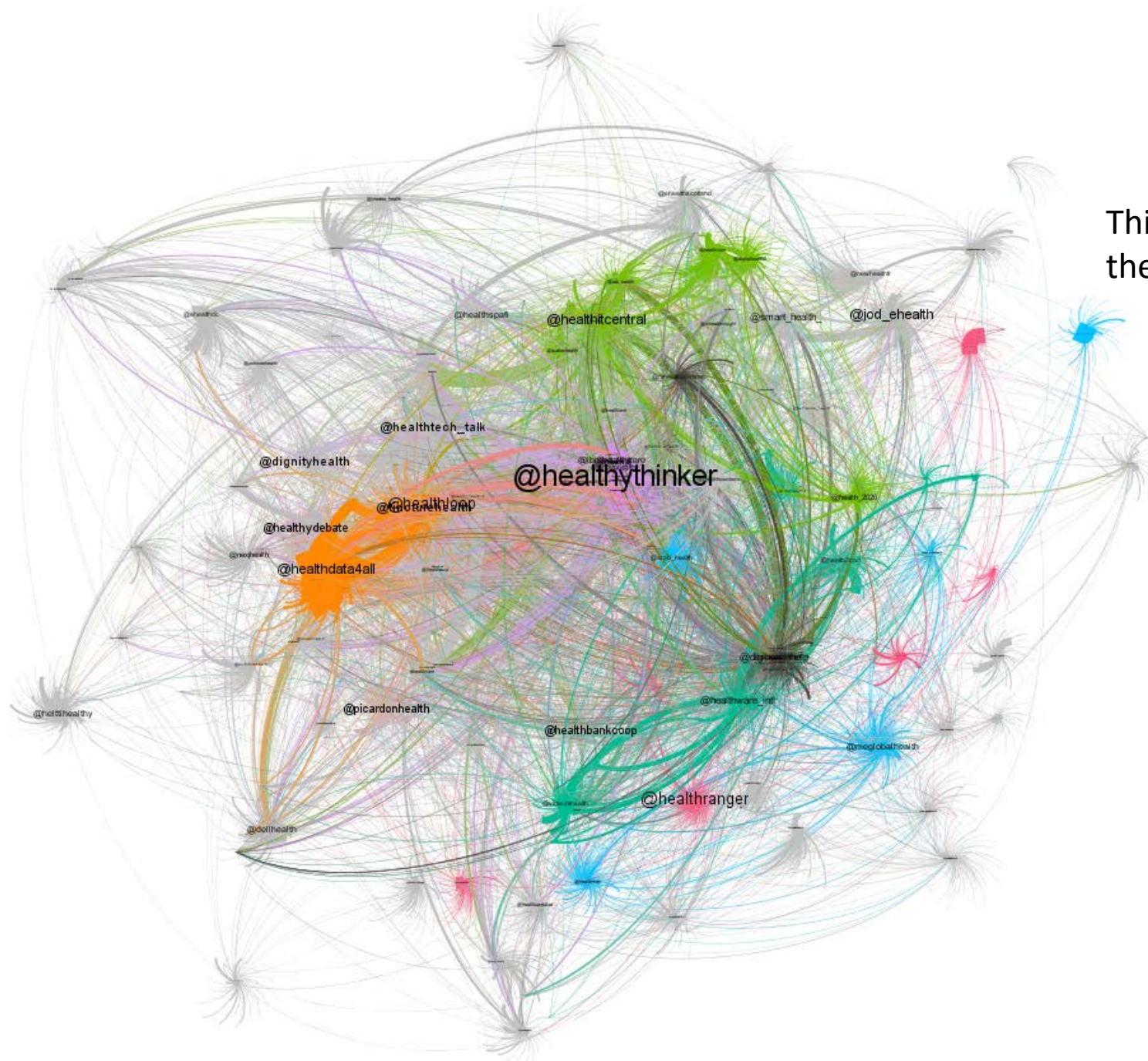
This image contains the core group only

The Various Algorithms to Choose From...

	id	label	timeset	indegree	outdegree	degree	weighted d	weighted i	weighted o	eccentricit	closnessco	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	11	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	0.0008761	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	0.000767	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	0.0008826	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

How to Read a Leximancer Map

- The concepts are clustered into higher-level 'themes' when the map is generated. Concepts that appear together often in the same pieces of text attract one another strongly, and so tend to settle near one another in the map space. The themes aid interpretation by grouping the clusters of concepts, and are shown as coloured circles on the map.

- The themes are heat-mapped to indicate importance. This means that the 'hottest' or most important theme appears in red, and the next hottest in orange, and so on according to the colour wheel.

- Leximancer is not designed to replace the role of the human analyst in the social sciences; rather it is a tool to help analysts perform and draw greater insight from their data.

Default Settings

Topical Network - (Linear) clustering algorithm.

Default Settings

Social Network - clustered using
a Gaussian algorithm

Thematic Summary   Detail▾ Export▾

Theme	Connectivity	Relevance
patients	100%	
care	82%	
risk	59%	
report	54%	
people	51%	
year	34%	
Himss16	25%	
using	23%	
today	10%	
follow	10%	
tech	05%	
Here's	02%	
food	02%	

THEME: patients
(patients, amp, use, quality, medical, work, program,
share, diabetes, helps, pharma)

patients   (Hits: 717)
http://t.co/tYCYPZ44Ad http://t.co/Ge3bABTjZ1
I added a video to a @YouTube playlist
http://t.co/PTPFrbii30 Tackling Teen Sports
Injuries with Dr Robert Wainer
Women s Hospital becomes first hospital in the US to have
NICU pharmacy resident. http://t.co/fO99r8HyeH
http://t.co/Ve7LJ5RTcR
Cone Health among tops in providing value to its communities.
more...

amp   (Hits: 539)
Track yr performance & payor incentives EZ #pophealth
https://t.co/p7OksURCkh
https://t.co/S9zeHxOKRO
Did somebody press the FastForward #HIMSS16 button?
Happy docs finally get usable #bigdata
https://t.co/YViCQPajVh
https://t.co/EAITrpw3fY
You do want to talk to the #KLAS leading #pophealth vendor,
don't you?
more...
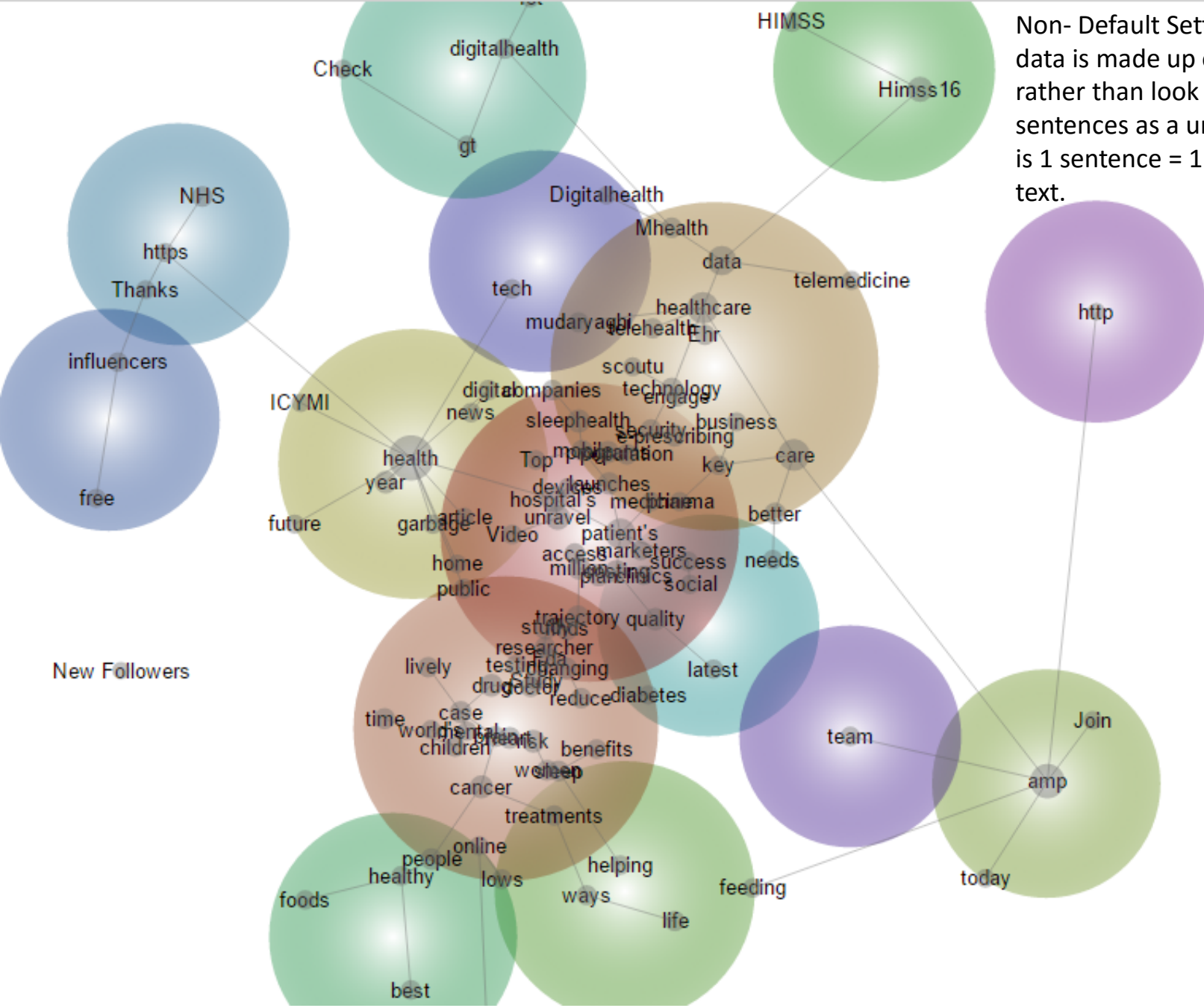
Topical Network - (Linear) clustering algorithm.

Non- Default Settings – the data is made up of tweets so rather than look at 2 sentences as a unit – here is is 1 sentence = 1 unit of text.

Thematic Summary   Detail▾ Export▾

Theme	Connectivity	Relevance
patient's	100%	
cancer	74%	
care	68%	
health	39%	
amp	17%	
life	11%	
Himss16	09%	
healthy	07%	
digitalhealth	07%	
needs	07%	
NHS	06%	
influencers	04%	
tech	03%	
team	02%	
http	00%	

THEME: patient's
(patient's, hospital's, study, unravel, sleephealth, million, costing, trajectory, quality, launches, plan, finds, programs, mobile, population, companies, access, success, public, marketers, Video, clinics, Top, devices, pharma, medicine)

patient's   (Hits: 0)

hospital's   (Hits: 0)

study   (Hits: 85)
  -Philip Chesterfield
  Generally we study too much and think too little.
more...

unravel   (Hits: 115)

  Another Healthy Ride takes on another Dirty Dozen hill!
more...

sleephealth   (Hits: 84)
  @bikepgh https://t.co/UwI2x3pbRk
  @NatalieCerino Best way to locate stations is using our app
  "nexthike"

Social Network - clustered using a Gaussian algorithm

Non- Default Settings – the data is made up of tweets so rather than look at 2 sentences as a unit – here is is 1 sentence = 1 unit of text.

Thematic Summary   Detail▾ Export▾

Theme	Connectivity	Relevance
study	100%	
health	97%	
healthcare	66%	
cancer	52%	
technology	49%	
amp	42%	
time	32%	
success	22%	
year	18%	
Himss16	16%	
best	11%	
https	07%	
Check	05%	
today	03%	

THEME: study
(study, unravel, million, costing, trajectory, quality, launches, plan, finds, changing, Fda, Study, reduce, researcher, testing, access, marketers, clinics)

study (Hits: 85)
- *Philip Chesterfield*
  Generally we study too much and think too little.
  more...

unravel (Hits: 115)
https://t.co/55QeGWJypj
https://t.co/GE5RpNYnZR
A new study takes on the myths around winter and our moods.
more...

million (Hits: 60)
https://t.co/z6falDhWdW
Ted Cruz supports #*Monsanto*; insults tens of millions of #health-conscious #*Americans*
https://t.co/Xon4Qx9ULa
https://t.co/SDRR1hf0Zm
New study finds numerous #radiation induced #mutations in #*Fukushima* #forests https://t.co/WfhN8odUjw
https://t.co/utjp63SDfr

Data Mining, Social Network Analysis and Content Analysis

- What has been presented so far are 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 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 health tweets 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.

- We are glimpsing the *various ways in which we are defined, managed, and governed*.

- 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 Data at face value, as an object or representation of a truth that can be extracted from and that reflects the social.

The Demise of Social Agency

- It would be a 'social stripped of social agency in which the multiple interactions on multiple and hierarchical levels over time create and re-create power laws of social interaction'.

- Akin to August Comte's *laws* of 'social physics'. (Sociology – according to Comte the 6th and greatest of sciences of man)

- Structuralist relational elements and therefore algorithms are in danger of providing justification of supranational systems – (states and multinational corporations) (Wolfe, 2010: 3).

- The indirectly observable force (of social structures) that has gone un-measured is now visualised by mathematical equations.

Mathematical Reductivism

- Mathematical reductivism presenting analysts with a limited social physics devoid of all that is not capable of reduction to logic is a poor representation.

- Social network analysis and Quantified Big Data analysis has developed tools of measurement not dominance. To facilitate a more nuanced understanding, these tools need to allow equal weighting to social, cultural, and agency focused explanations of the social phenomena under observation.

- Comte later argued that there was a 7th science, superior to all sciences.
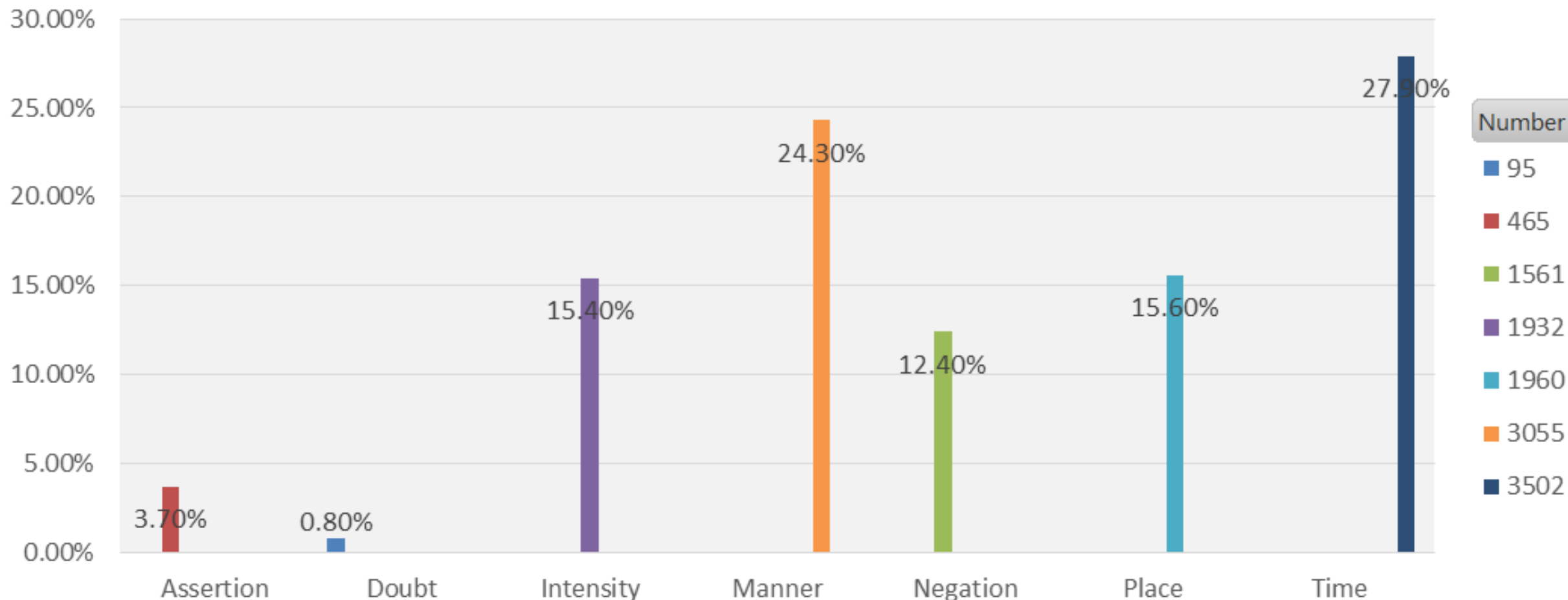
Anthropology – The 7th Science of Humanity

- Anthropology - the last gradation in the Grand Hierarchy of Abstract Science.

- Comte called for the quantitative analysis of society, he also saw a limit in its ability to help explain social phenomena.

Types of modality

- Assertion – I am sure they will revert. It will take them awhile to get the preferred info.
- Doubt – Perhaps time for an encore of these tips, given recent concerns about device safety in this article.
- Intensity – I think Professor John Bachman MD explains the technical side of how it works very well.
- Manner – Mental Elf I think it desperately needs treatment by a popular entertainment show.
- Negation – Patient engagement in population health not one-size-fits-all
- Place – The #refugeecrisis in the middle East and #publichealth
- Time – We're giving away a @Mybasis tracker each day on Twitter.

Occurrance of Modality in Big Data Health Tweets

The Uncertainty Principle Versus the Observer Effect

- The contention being aimed at Big Data, Health and the analytics that they practice are that they have the potential to undermine "freewill".
- The main antidote we have to this is the "certainty of uncertainty".
- To sow doubt - and thereby open up the potential for critique to enter.
- **The uncertainty principle** introduced to quantum mechanics by Heisenberg states that the more precisely a particles position can be known then the less accurate the measurement of its momentum and vice versa.
- Physics refers to a phrase more commonly associated by me with the social sciences – the **observer effect** or the **Hawthorn effect**.
- The Hawthorne effect (also referred to as the observer effect) is a type of reactivity in which individuals modify their behaviour in response to their awareness of being observed.

- The idea that the big data companies like Google and Facebook will hand us back our privacy, free will and right to make decisions is similar to asking Henry Ford to produce the Model T Ford by hand.

- Asking or pleading with Google to give us our autonomy back is asking them to give us their page rank algorithm that takes our surplus behaviour and turns it into predictive analytics. Ask KFC or Coca Cola for their 'secret ingredient'?

- The elephant in the room is BIOMETRIC DATA and whether or not we TRUST Google and Facebook and others to act in the best interests of people or merely continue to extract surplus behavioural data for profit.