SYSTEMS ENGINEERING RESEARCH CENTER

DAU

Digital Readiness Series

Data analytics

Data for the upcoming world: Horizon scanning

By
Dr. Carlo Lipizzi
August 13, 2020

www.sercuarc.org

SYSTEMS ENGINEERING RESEARCH CENTER

Data for the upcoming world: Horizon scanning

- Future cannot be predicted, but in science there is a high level of consistency over time. Data Science today is a steppingstone for an even more informed and complex way of living and doing business, with a continuous integration of sources and media, creating semantic synergies, pushing the boundaries of convenience, value and privacy.
- In this seminar, we scan the major trends in Data Science, starting from the current emerging trends, extrapolating scenarios and presenting live examples of emerging applications

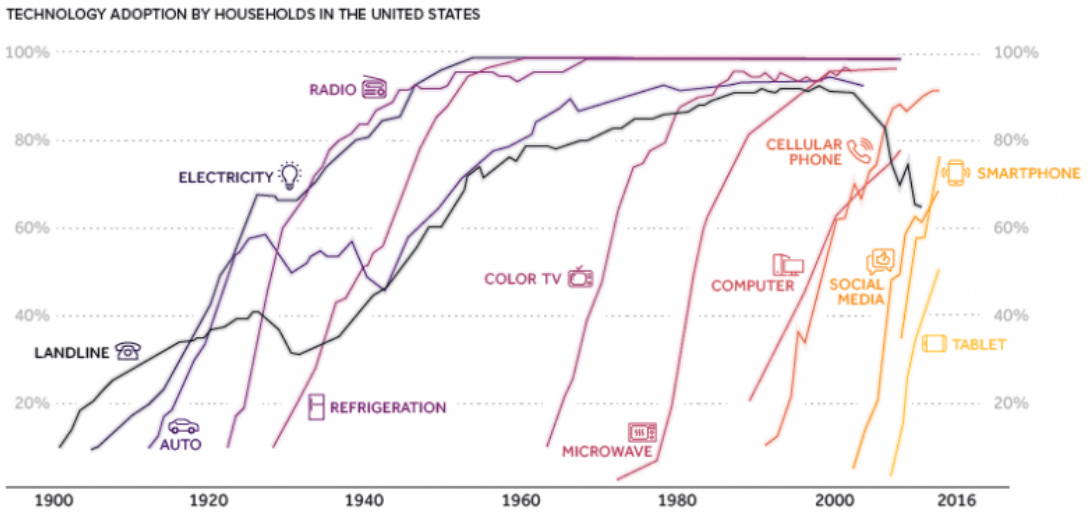
SYSTEMS ENGINEERING RESEARCH CENTER

Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

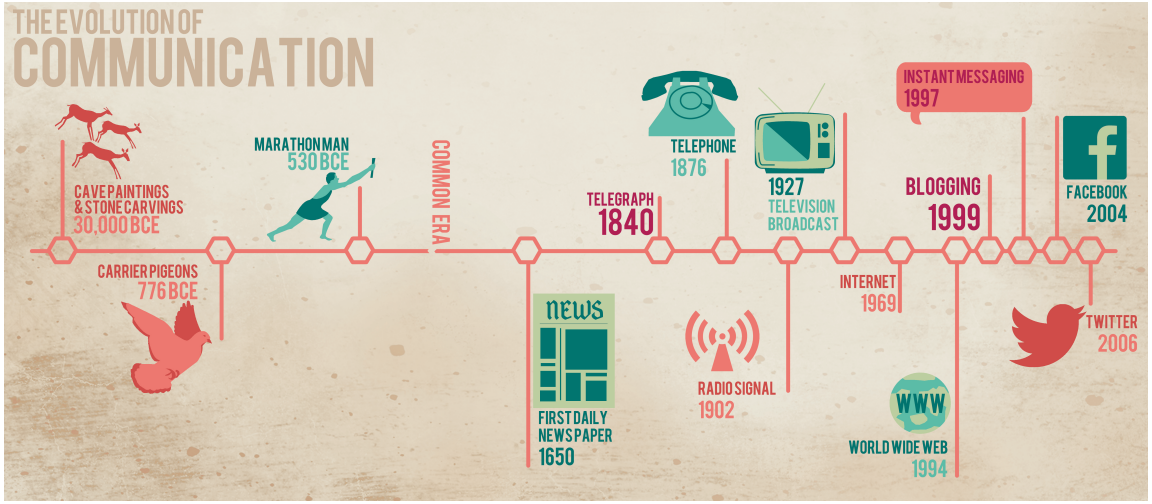
SYSTEMS ENGINEERING RESEARCH CENTER

Faster pacing

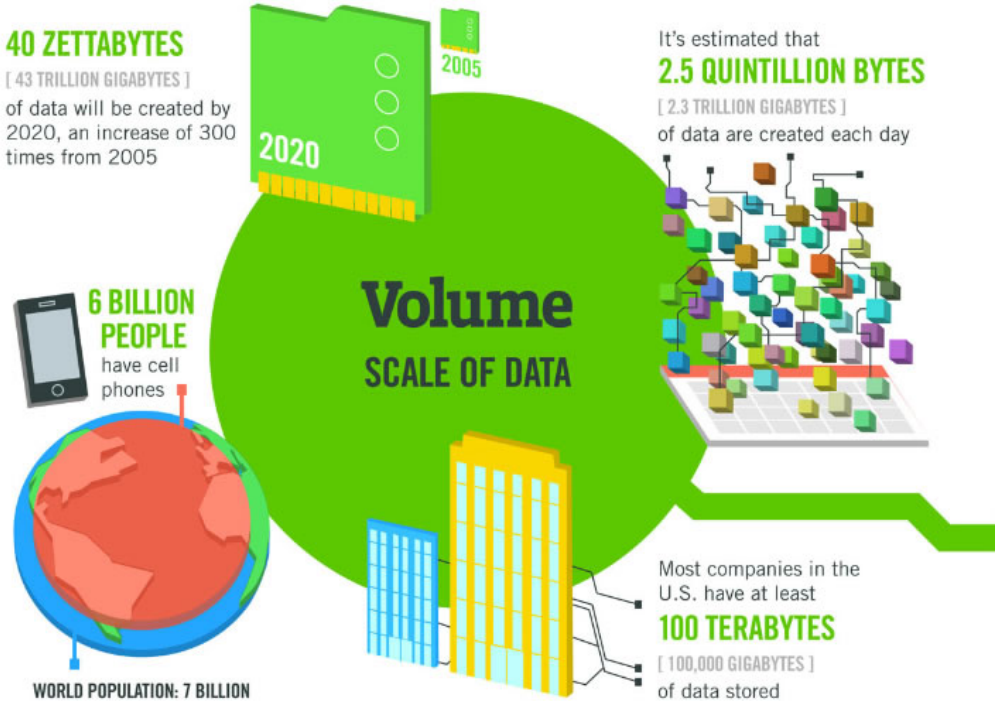


- The rate of adoption of new technologies with high impact on the society is increasing dramatically

- The driving force for the acceleration is communication/data sharing, that is making the world "smaller" for the good and the bad



Data growth: what we share



- The digital tools we are using every day are creating data from everything we do at an unprecedented rate: every day, 2.5 quintillion (1018) bytes of data are created and 90% of the data in the world today was created within the past two years.
- The increasing rate of change, the highly integrated world and the growing massive amount of data we generate and share are the driving forces for the upcoming changes in society and business

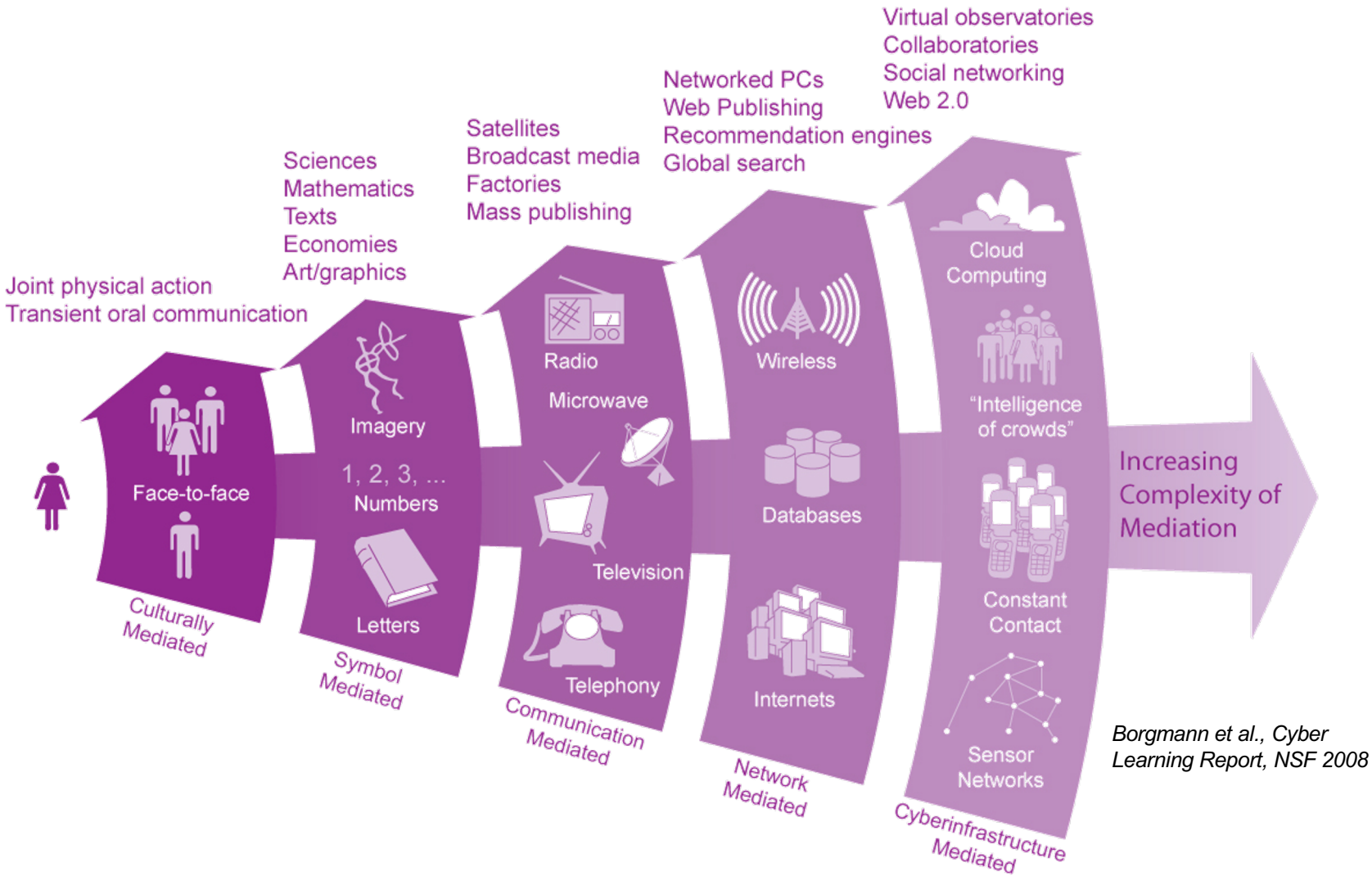
SYSTEMS ENGINEERING RESEARCH CENTER

Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

SYSTEMS ENGINEERING RESEARCH CENTER

The complexity of Learning Mediation



Borgmann et al., Cyber Learning Report, NSF 2008

Mediated Learning Experience refers to the way in which stimuli experienced in the environment are transformed by a mediating agent, usually a parent, teacher, sibling, or other decisive element in the life of the learner

SYSTEMS ENGINEERING RESEARCH CENTER

The “Always-on Internet” society

Focused on

Sharing

Cooperation

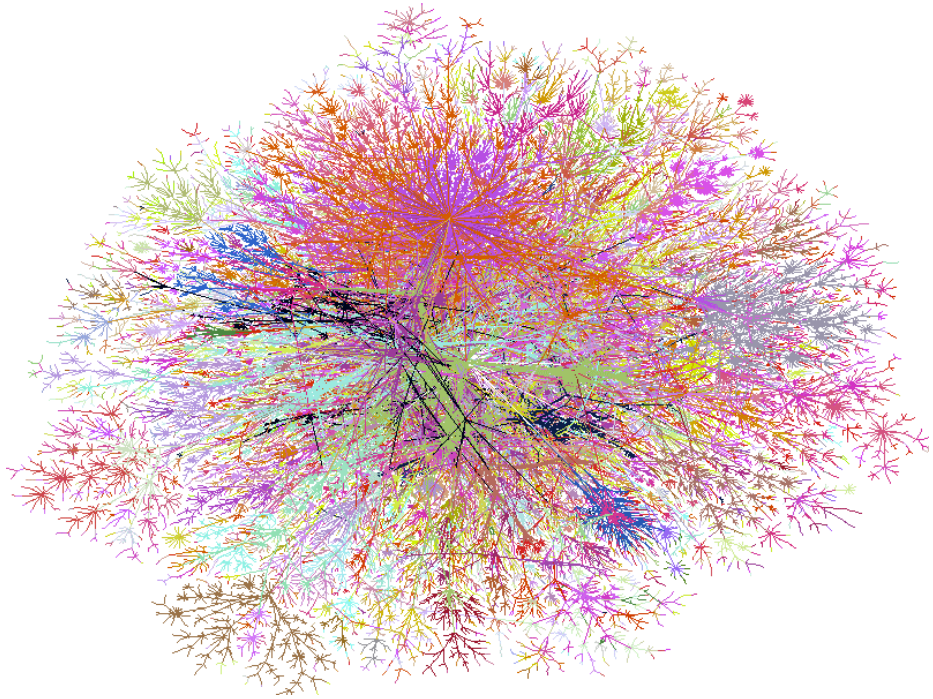
Collective Actions

Reduced Privacy

Source uncertainty

Information overload

“Living in the cloud”



SYSTEMS ENGINEERING RESEARCH CENTER

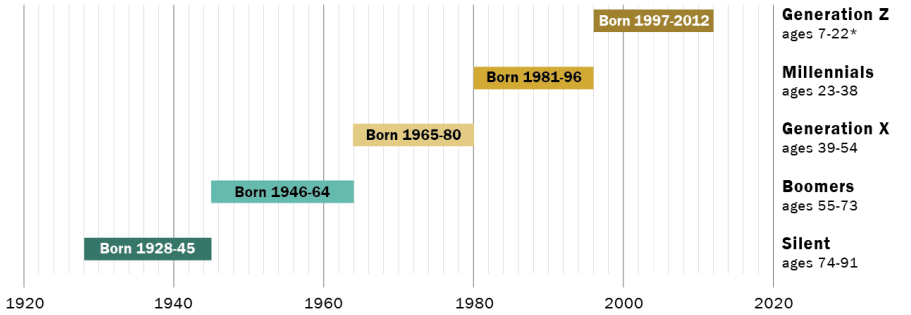
"Living in the cloud" effects: the overstimulation

- The average attention span in 2015 - 8 seconds
- The average attention span in 2000 - 12 seconds
- The average attention span of a goldfish - 9 seconds
- Percent of teens who forget major details of close friends and relatives - 25 %
- Percent of people who forget their own birthdays from time to time - 7 %
- Average number of times per hour an office worker checks their email inbox - 30
- Average length watched of a single internet video - 2.7 minutes

SYSTEMS ENGINEERING RESEARCH CENTER

A snapshot of the generations

The generations defined

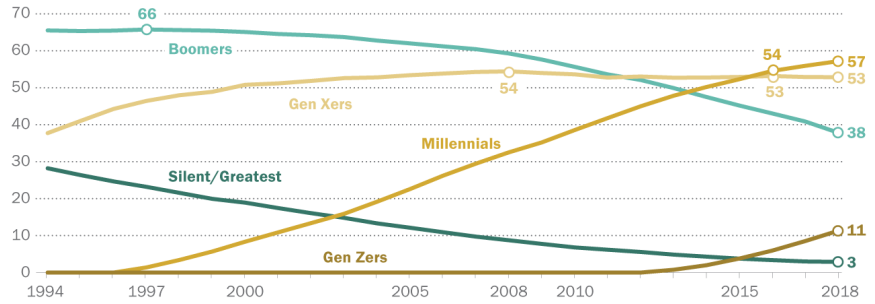


*No chronological endpoint has been set for this group. For this analysis, Generation Z is defined as those ages 7 to 22 in 2019.

PEW RESEARCH CENTER

Millennials became the largest generation in the labor force in 2016

U.S. labor force, in millions

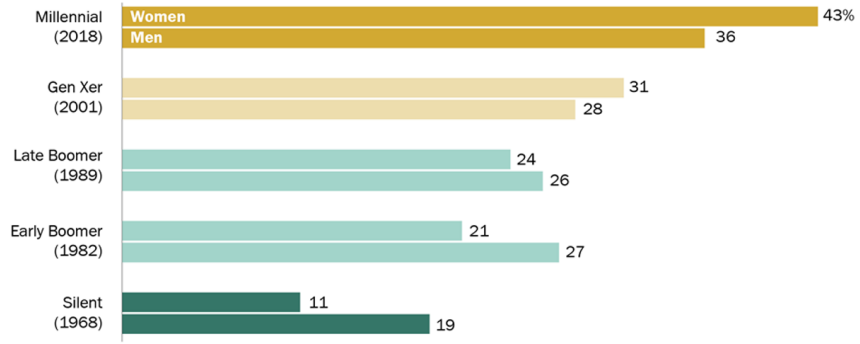


Note: Labor force includes those ages 16 and older who are working or looking for work. Annual averages shown.
Source: Pew Research Center analysis of monthly 1994-2018 Current Population Survey (IPUMS).

PEW RESEARCH CENTER

Among Millennials, women outpacing men in college completion

% of 25- to 37-year-olds who have completed at least a bachelor's degree

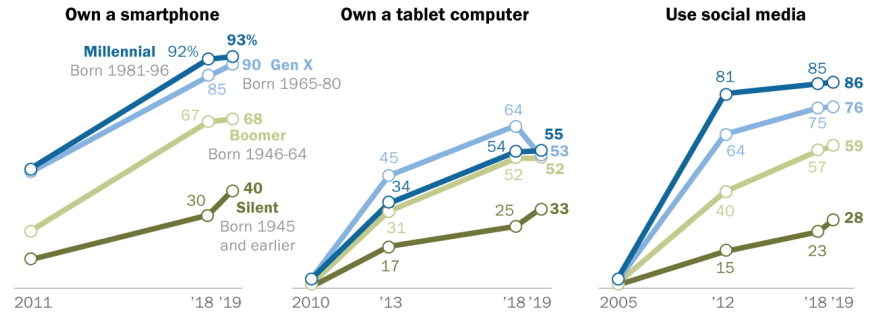


Note: The educational attainment question was changed in 1992. For Boomers and Silents, the shares shown include those who completed at least four years of college (regardless of degree status).
Source: Pew Research Center analysis of 1968, 1982, 1989, 2001 and 2018 Current Population Survey Annual Social and Economic Supplements (IPUMS).

PEW RESEARCH CENTER

Millennials lead on some technology adoption measures, but Boomers and Gen Xers are also heavy adopters

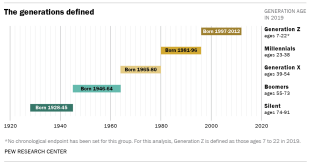
% of U.S. adults in each generation who say they ...

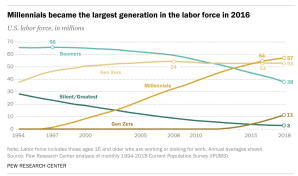


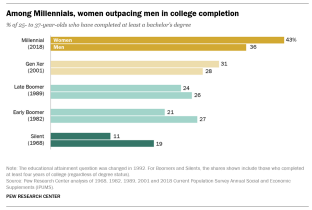
Note: Those who did not give an answer are not shown.
Source: Survey conducted Jan. 8 - Feb. 7, 2019.

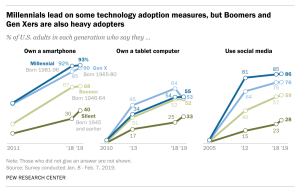
PEW RESEARCH CENTER

A snapshot of the generations – the take







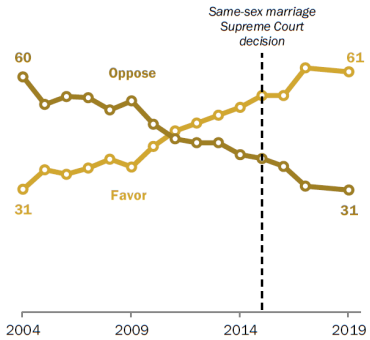


- Using more technology in everyday activities
- More “sophisticated” use of devices and data
- More data will be generated
- Pushing upward the automation
- More “embedded analytics”

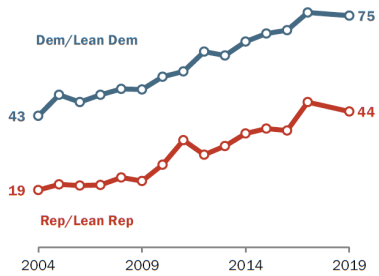
Social Issues

Public remains supportive of same-sex marriage; wide partisan gap persists

% who ___ allowing gays and lesbians to marry legally



% who favor allowing gays and lesbians to marry legally

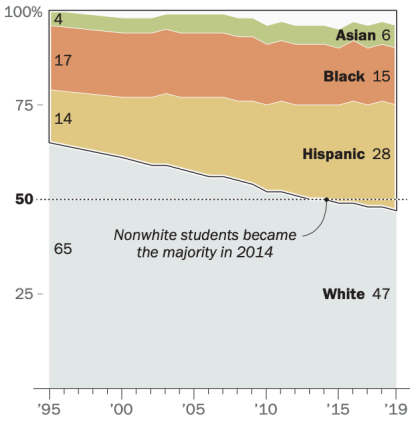


Notes: Data for 2007-2015 based on yearly averages.
Don't know responses not shown.
Source: Survey of U.S. adults conducted March 20-25, 2019.

PEW RESEARCH CENTER

Hispanics are growing as a share of K-12 public school students

% of public school students, by race/ethnicity

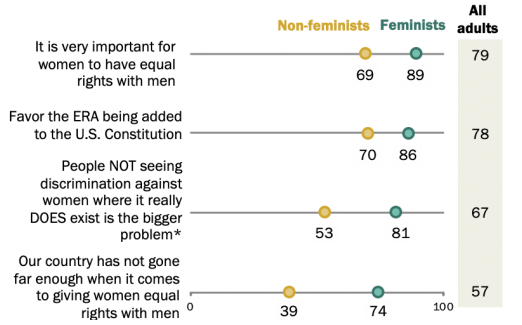


Notes: Among those enrolled in elementary and secondary public schools. Data for 2019 is projected. White, black and Asian are single race, non-Hispanic. Hispanics are of any race. Asian includes Pacific Islander. Multiracial and other single-race groups not shown.
Source: National Center of Education Statistics, U.S. Department of Education.

PEW RESEARCH CENTER

Majority of Americans say gender equality is very important, whether they identify as feminists or not

% of ___ who say ...

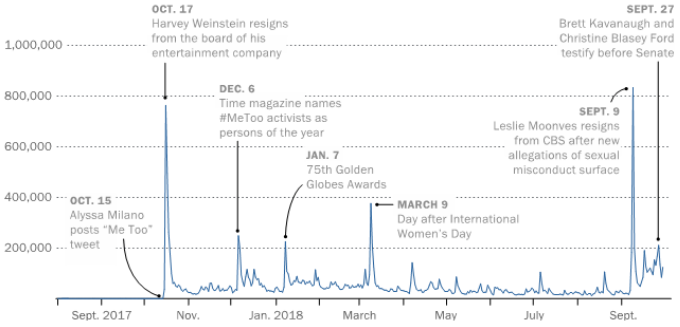


*Compared with people seeing discrimination against women where it really does NOT exist.
Note: Feminists include those who say "feminist" describes them very or somewhat well, and non-feminists include those who say it describes them not too or not at all well.
Source: Survey of U.S. adults conducted March 18-April 1, 2020.

PEW RESEARCH CENTER

The #MeToo hashtag has been used roughly 19 million times on Twitter in the past year, and usage often surges around news events

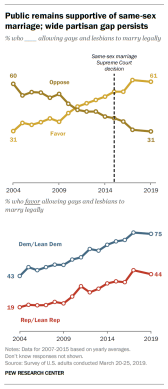
Number of Twitter posts mentioning the #MeToo hashtag, Oct. 15, 2017-Sept. 30, 2018

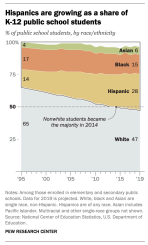


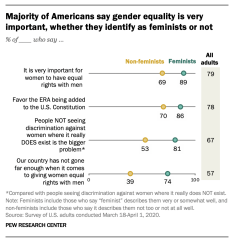
Source: Pew Research Center analysis of publicly available tweets using Crimson Hexagon.

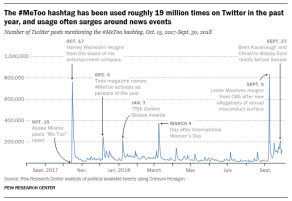
PEW RESEARCH CENTER

Social Issues – the take









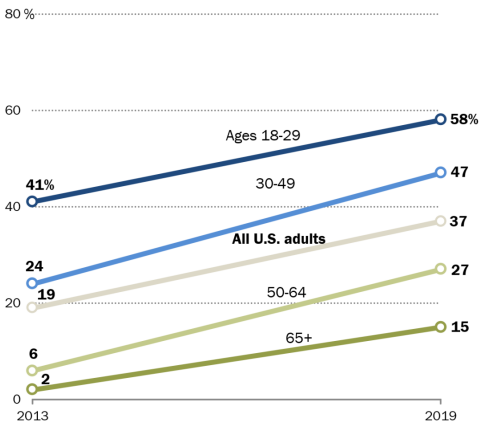
- More diversified cultural predominant elements
- Stronger formal ethical principles
- More polarizing conditions
- More need for customizations/individual-level targeting

SYSTEMS ENGINEERING RESEARCH CENTER

Technology and the Society

Americans of all ages are increasingly likely to say they mostly go online using their smartphone

% of U.S. adults who say they mostly go online using a cellphone

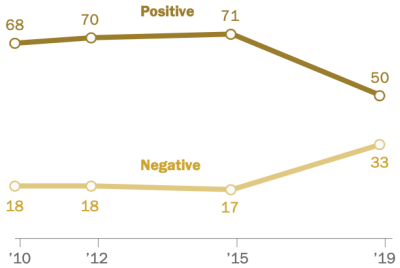


Note: Respondents who did not give an answer or gave other responses are not shown.
Source: Survey of U.S. adults conducted Jan. 8-Feb. 7, 2019. Trend data from previous Center surveys.
"Mobile Technology and Home Broadband 2019"

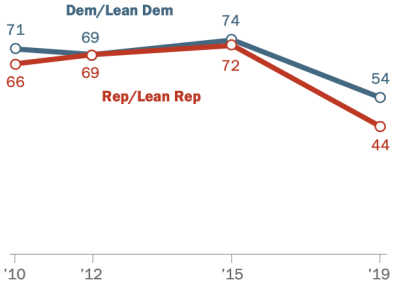
PEW RESEARCH CENTER

Members of both parties are much less positive on impact of tech companies

% who say technology companies have a ____ effect on the way things are going in the country



*% who say technology companies have a **positive** effect on the way things are going in the country*

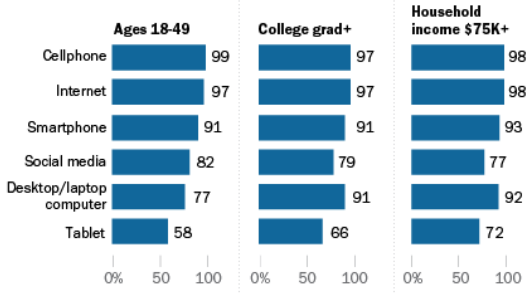


Note: Don't know, other responses not shown.
Source: Survey of U.S. adults conducted July 10-15, 2019.

PEW RESEARCH CENTER

Some groups have reached near-saturation levels for adoption of basic technologies

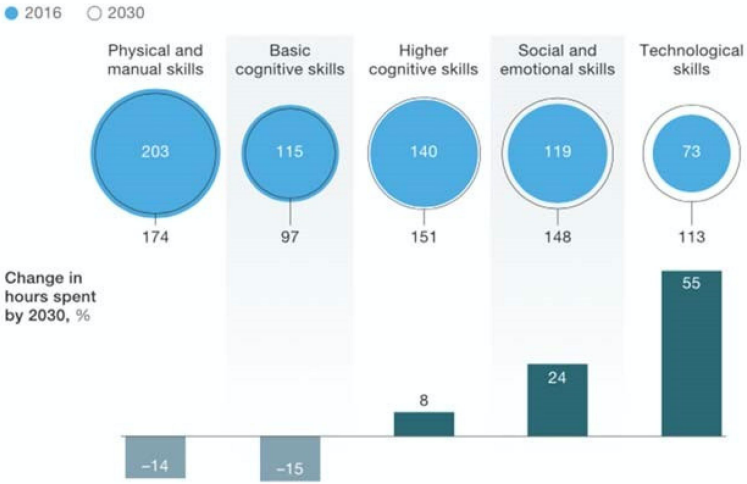
% of U.S. adults who say they own or use this technology



Source: Survey conducted Jan. 3-10, 2018.

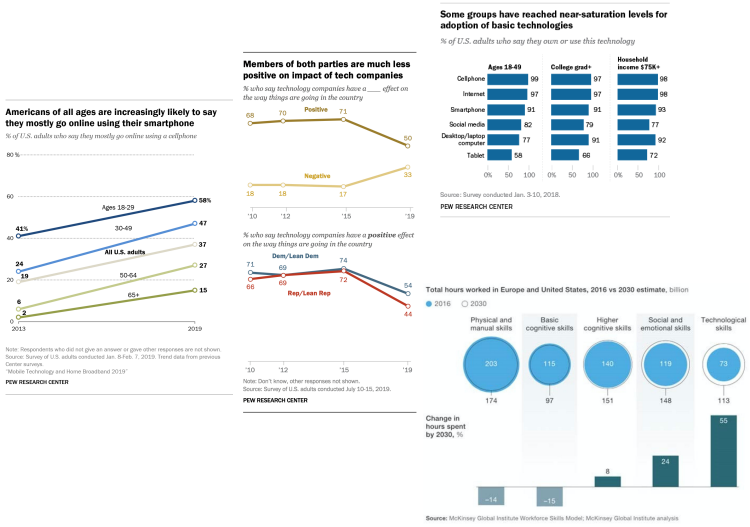
PEW RESEARCH CENTER

Total hours worked in Europe and United States, 2016 vs 2030 estimate, billion



Source: McKinsey Global Institute Workforce Skills Model; McKinsey Global Institute analysis

Technology and the Society – the take



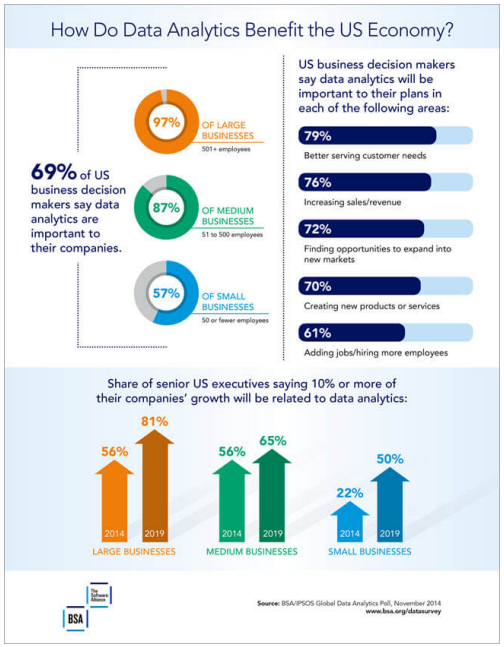
- More mobile access to the net/info, but Covid-19 may change this for good
- More control over technologies
- Plateau effect for leading technology -> need to analytics to find niches
- Continuous shift toward STEM/high added value competencies -> potential increase of the social divide on top jobs; potential increase of ”easy” user interfaces for complex systems
- The gap in skill demand and availability will drive up the use of AI-like solutions and global re-skill (when applicable)

SYSTEMS ENGINEERING RESEARCH CENTER

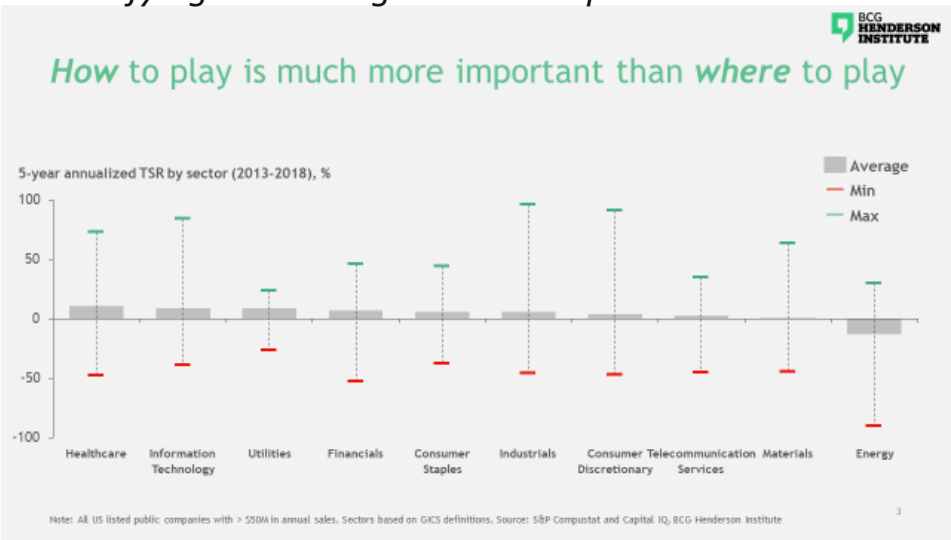
Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

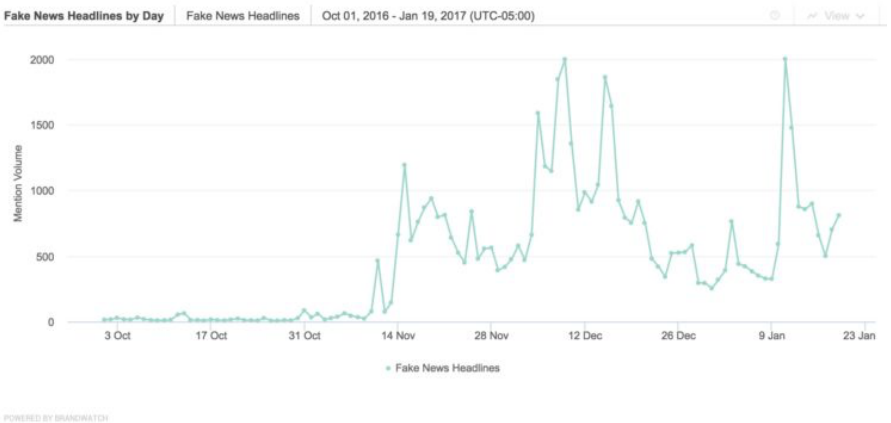
The Business environment

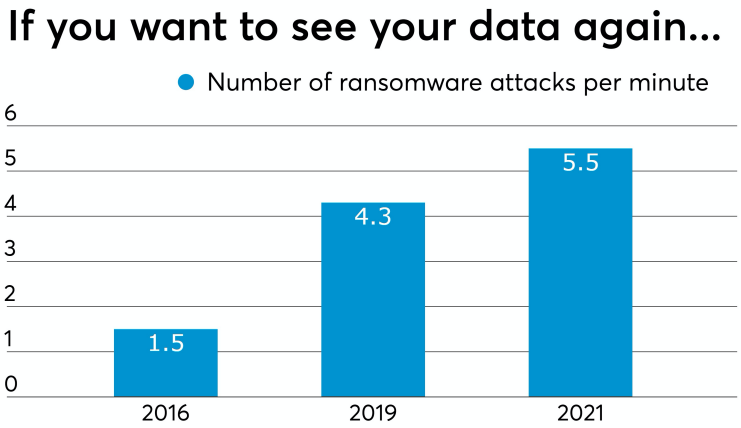


Defying the average is more important than ever



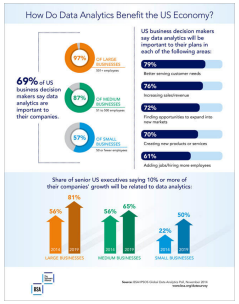
Source: World Economic Forum

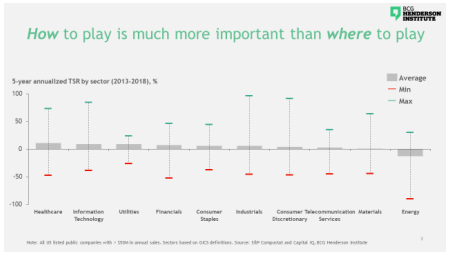


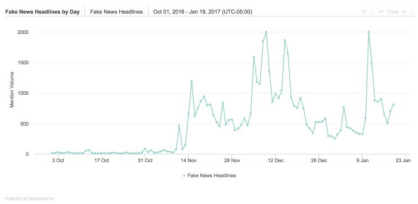


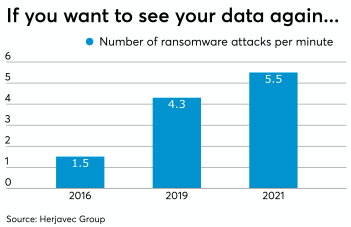
Source: Herjavec Group

The Business environment – the take





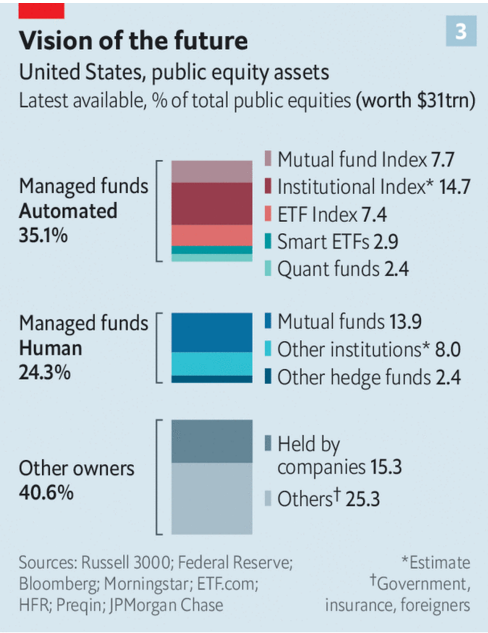




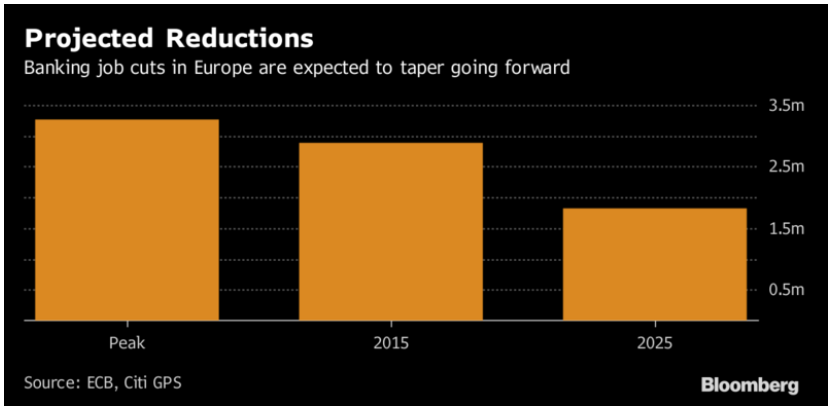
- Security and privacy are soaring in importance: 38% of corporates say improving security and securing the enterprise is a top improvement priority, on par with driving innovation (39%) and improving customer experience (40%)
- Businesses are pivoting around data. If cloud is the digital foundation, then data are the building blocks of the new world technology. Only 4% of corporates claim to have a "highly sophisticated" approach to leveraging data
- “How you play is more relevant than where you play” -> data analytics play a critical role
- There will be an increasing demand for AI/analytics driven solutions to contain cybercrime and vetting information sources

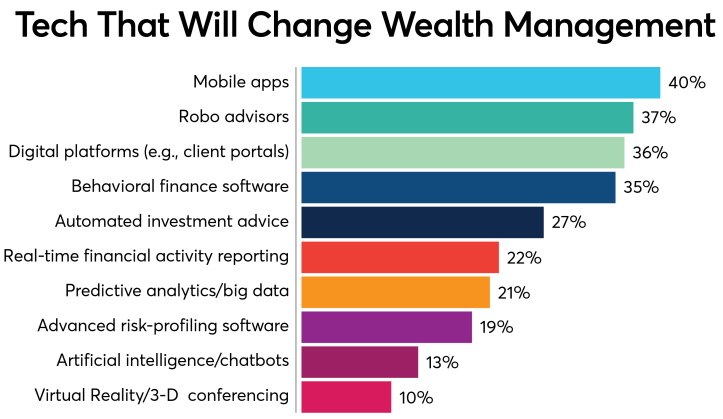
SYSTEMS ENGINEERING RESEARCH CENTER

Impact on Finance

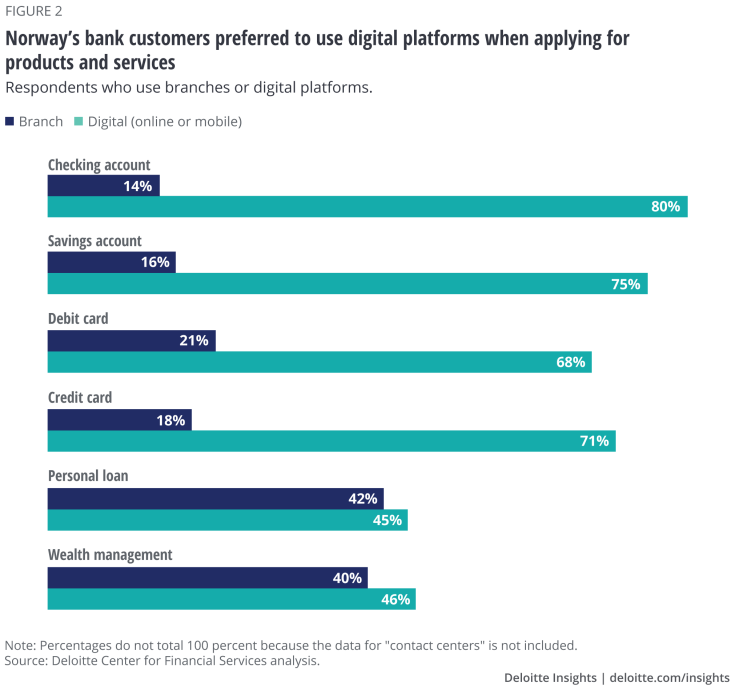


The Economist

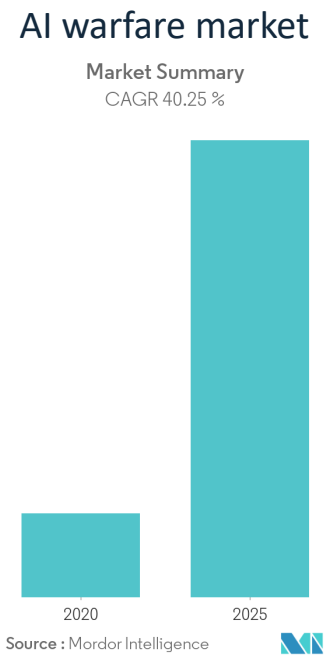


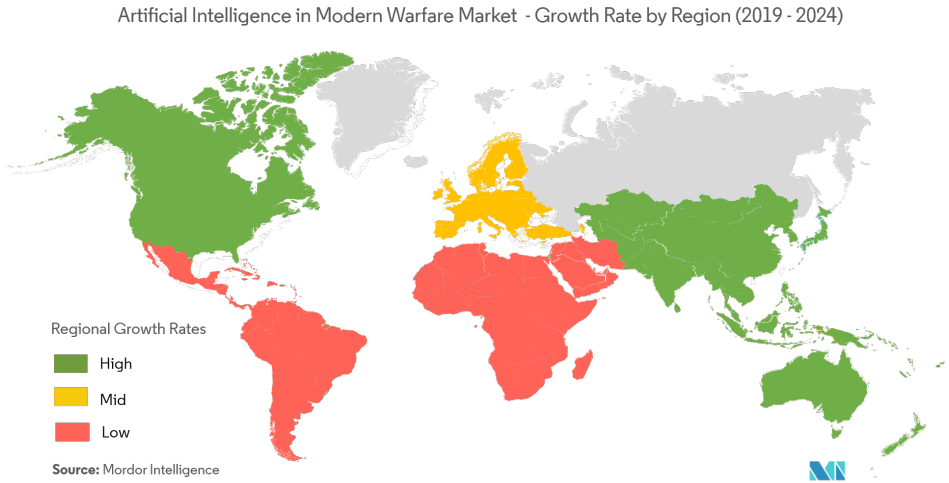


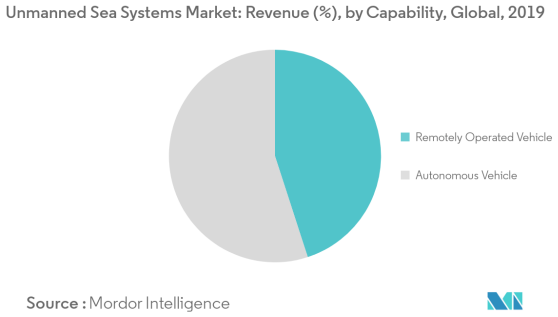
Source: *Financial Planning* Tech Survey, 2017

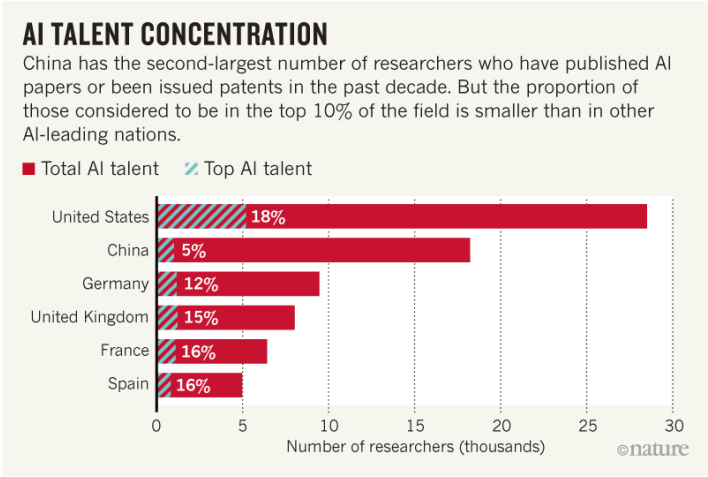


Impact on Military

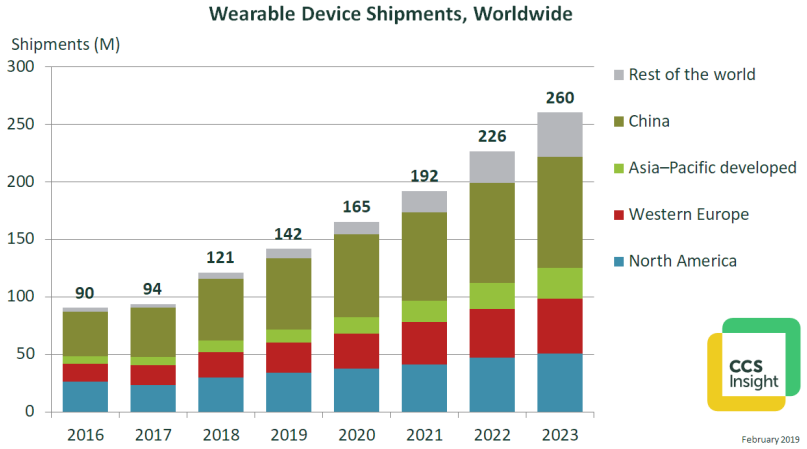


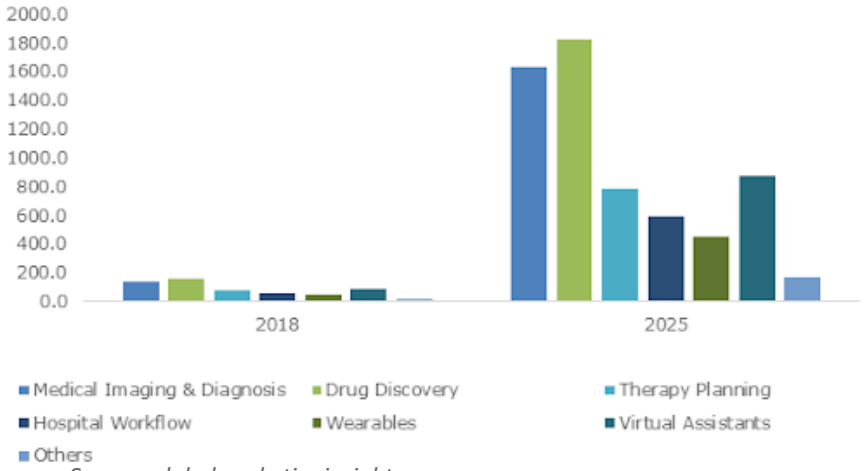




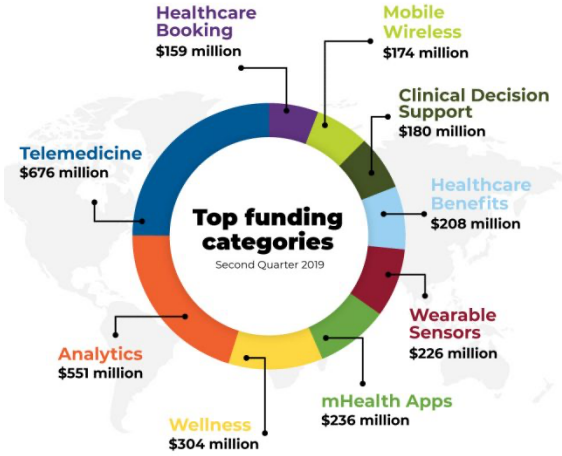


Impact on Wearable - Healthcare





Source: globalmarketinginsights.com



Source: Intersog

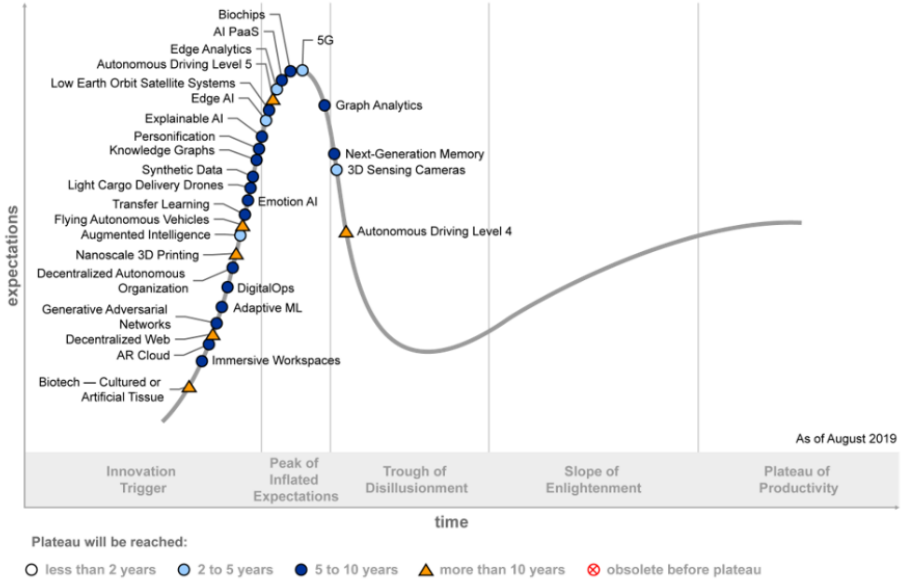
SYSTEMS ENGINEERING RESEARCH CENTER

Agenda

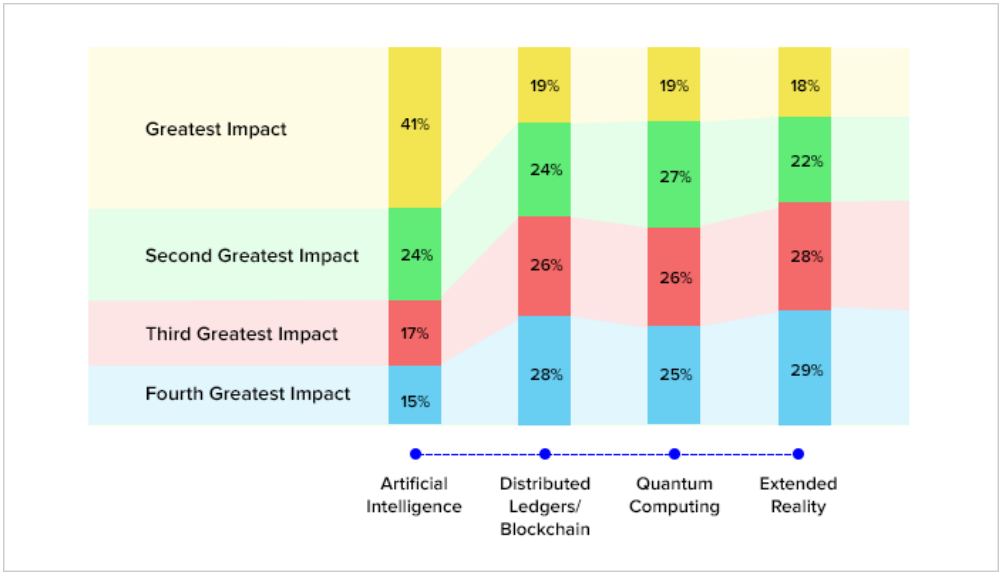
- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

SYSTEMS ENGINEERING RESEARCH CENTER

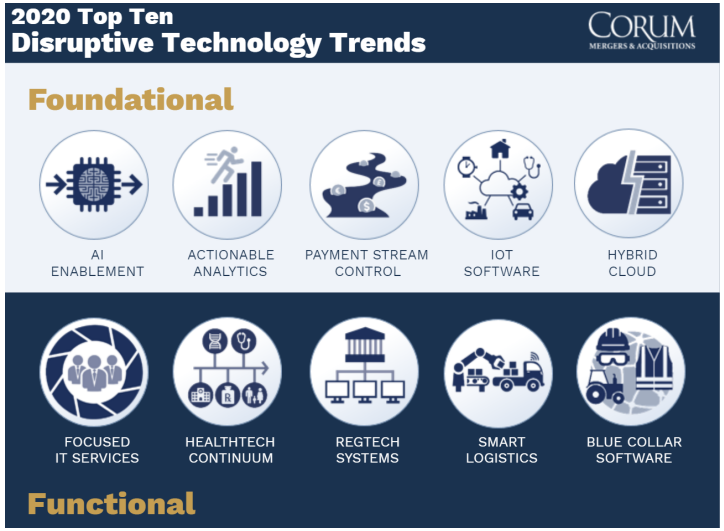
Technology trends

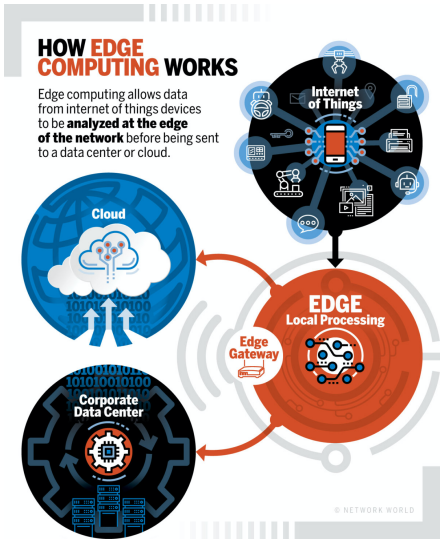


Source: Gartner

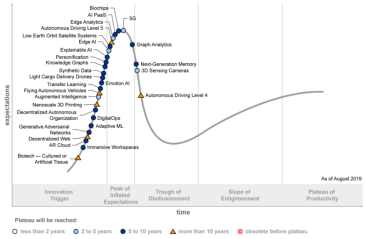


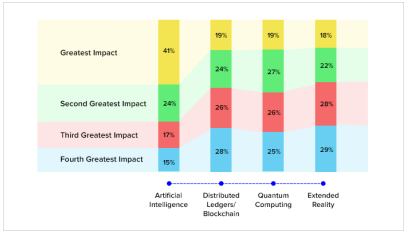
Source: Mobile App Daily

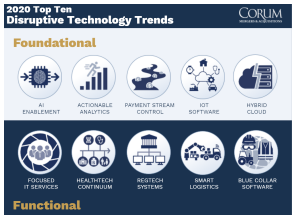


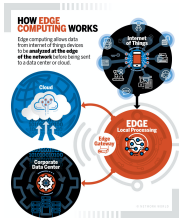


Technology trends – the take









- Technology innovation is data driven: increasing the complexity of the systems requires advanced analytics for management and optimization
- Human Augmentation - Enhancing Cognitive Abilities – Hybrid Systems: combining the speed and pattern-matching capabilities of algorithms with the higher-order reasoning and imagination of humans
- Digital ecosystems: leveraging on an interdependent group of actors (enterprises, people and things) sharing digital platforms to achieve a mutually beneficial purpose
- Sensing and Mobility: combining sensor technologies with AI, systems get a better understanding of the world around them, enabling mobility and manipulation of objects

SYSTEMS ENGINEERING RESEARCH CENTER

Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

SYSTEMS ENGINEERING RESEARCH CENTER

Top data trends happening now

Key contributing factors already happening

- *Decline of the Dashboard* - Data stories will be the most widespread way of consuming analytics, and stories will be automatically generated using augmented analytics techniques
- *X Analytics* (e.g., text analytics, video analytics, audio analytics, etc.) - AI-supported content analytics for video, audio, vibration, text, emotion will trigger major innovations and transformations
- *Augmented Data Management*: Metadata Is "the New Black" - Organizations will utilize active metadata, machine learning and data fabrics to dynamically connect, optimizing and automating most of the data management processes
- *Data Marketplaces and Exchanges* - Large organizations will be either sellers or buyers of data via formal online data marketplaces
- *Relationships Form the Foundation of Data and Analytics Value* - Graph technologies will facilitate rapid contextualization for decision making

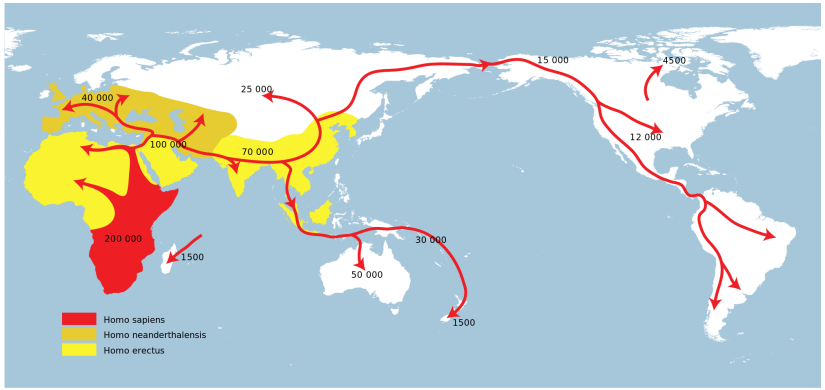
Source: processed Gartner info

Data & NLP trends

Top 10 Data and Analytics Trends That Will Change Your Business

Scaling Business Impact	• Smarter, faster, more responsible AI • Decline of the dashboard • Decision intelligence • X analytics
Transforming Deployment	• Augmented data management: Metadata is the new black • Cloud is a given • Data and analytics worlds collide
Increasing Data and Analytics Value	• Data marketplaces/exchanges • Practical blockchain (for data and analytics) • Relationships form the foundation of data and analytics value

Source: Gartner



SYSTEMS ENGINEERING RESEARCH CENTER

Data/NLP trends – the take

- Augmented Analytics: autonomous or semiautonomous examination of data reducing the data preparation phase, based on AI/ML
- Real-time intelligence, combining analytics with back-end systems able to capture "momentary markets"
- Blockchain - enabling data analytics applications
- Data Fabric to enable seamless data sharing across a distributed network
- NLP technology: end of fully supervised systems, rise of semi supervised, able to leverage on domain-specific knowledge
- NLP technology : data/text convergence via computational representations of knowledge
- NLP technology : semantic data preparation; bias and context detection
- NLP applications: automatic content generation; smart chatbots; primary UI for analytics; semantic search; semantic real time translators

SYSTEMS ENGINEERING RESEARCH CENTER

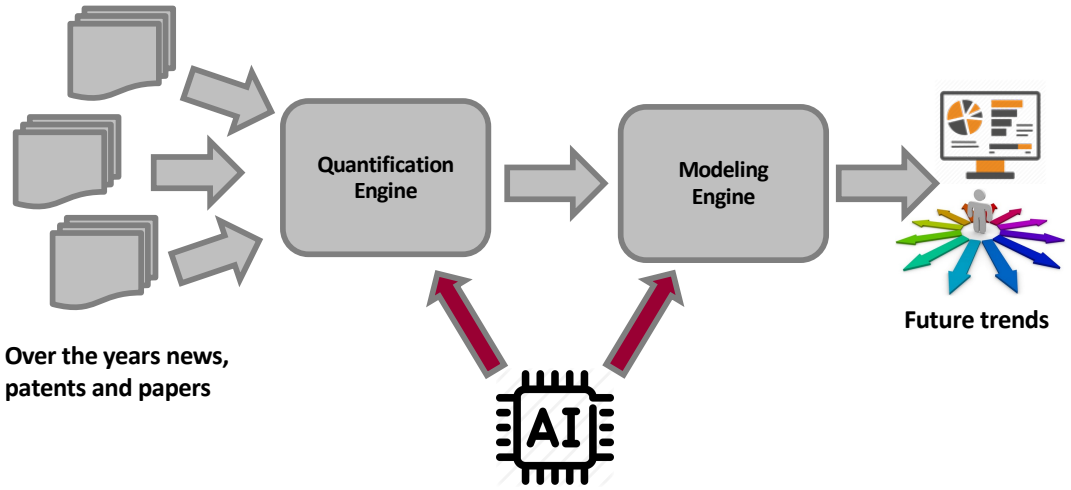
Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

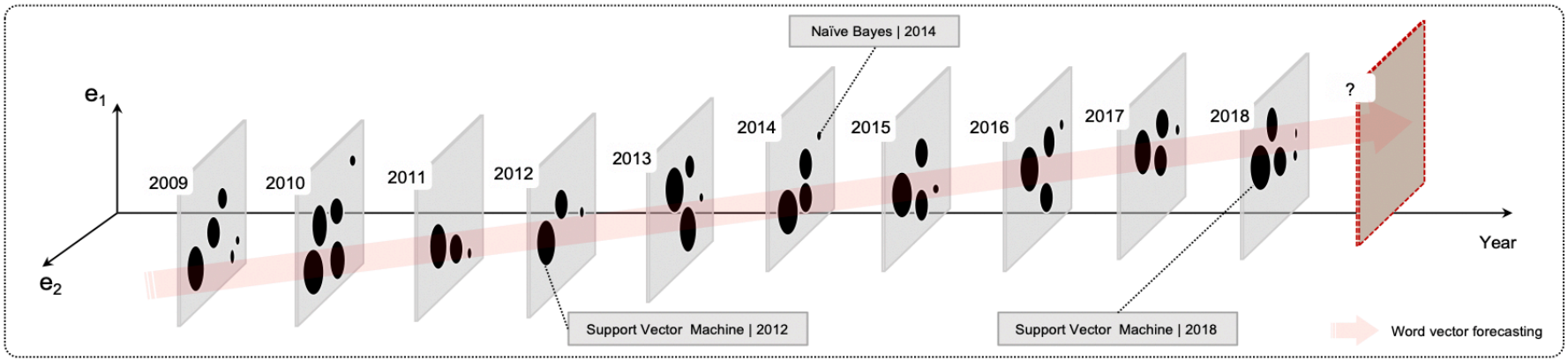
SYSTEMS ENGINEERING RESEARCH CENTER

The “crystal ball”

Future cannot be predicted, but 1. can be estimated; 2. rarely is not heavily rooted in the past

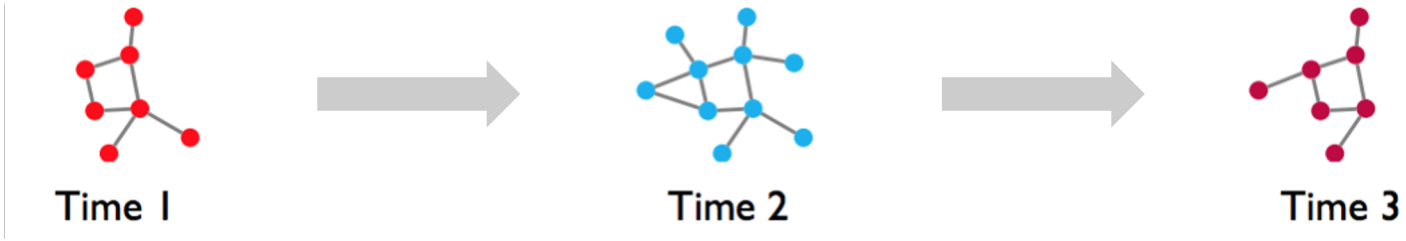


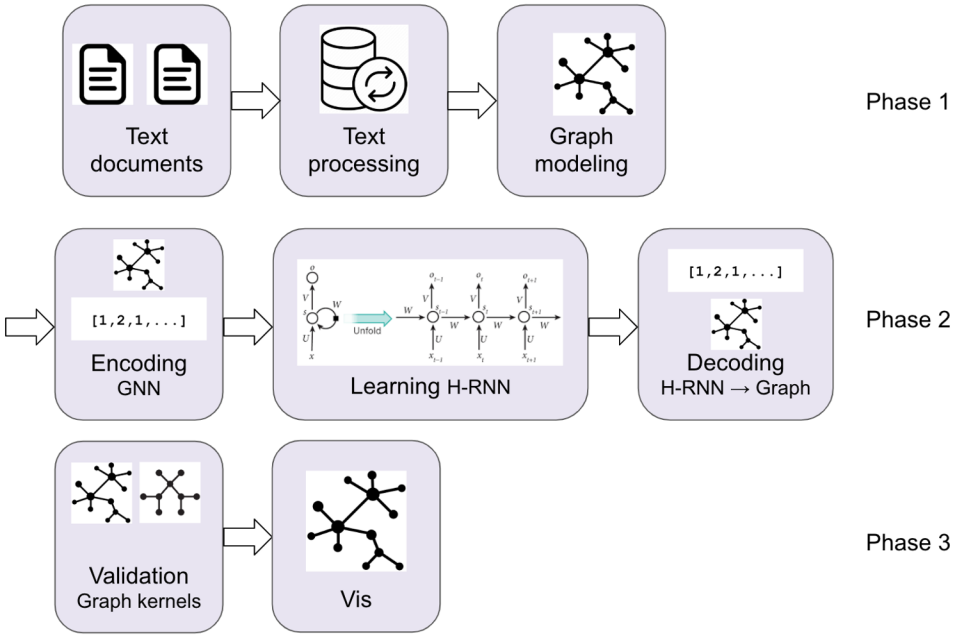
- We take several million documents from the past 10 years
- Transform them into semantic numerical forms (“embeddings”)
- Interpolate the evolution of the elements that are more coherent in time
- The interpolated element is the “predicted” value
- We are experimenting on technology evolution



The “crystal ball” – the network approach

- We create a network of words (“semantic network”) per each time slot, “learn” how they morph over time, predict the next morphed network





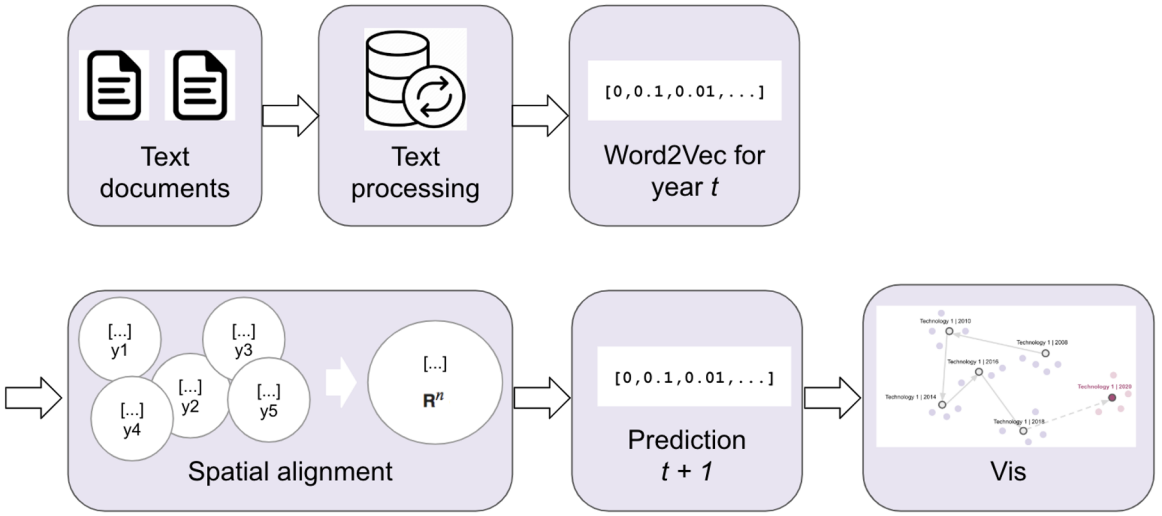
- We use a computational and visual analysis of changes in technology related semantic elements over the years
- In particular, we use Graph Neural Network and Recurrent Neural Network to evaluate and predict the morphing

The “crystal ball” – the linear algebra approach

- We create one computational representation of the text (“embeddings”) per each time slot, consolidate the embeddings keeping their time stamps, predict the next morphed embedding and provide a graphical representation

options_handlers|2009|	-0.000932	-0.001407	0.000226	0.001633	-0.000777	0.000755	-0.000286	0.001624	0.000675	0.001358	...
saver_interfaces|2009|	-0.000596	-0.000535	-0.000562	-0.000451	-0.000196	-0.001652	-0.000134	-0.000058	-0.000462	-0.001643	...
the_corresponding|2009|	0.001436	0.001424	-0.001556	-0.000644	0.000711	0.000534	0.000358	0.000786	0.000078	0.000623	...
their_specification|2009|	0.000222	0.000055	0.000099	0.000759	-0.001104	0.001146	0.001070	-0.000616	-0.000032	-0.000932	...
the_specification|2009|	-0.000213	0.001137	0.000392	-0.001567	-0.000906	-0.000118	0.001648	0.000316	0.001084	0.001122	...
of_the_return_signature|2009|	0.001208	-0.000271	-0.001013	-0.001147	0.000294	-0.000738	0.000322	0.000326	0.001365	0.001450	...
of_the_interface_functions|2009|	-0.000339	0.001518	-0.001098	-0.001328	-0.000350	0.000819	0.000224	-0.001495	0.000424	-0.001567	...
dispat_ching|2009|	-0.001623	-0.000575	-0.000007	0.001236	0.001511	0.000499	-0.000081	-0.001518	0.001245	-0.000205	...
s_methods|2009|	-0.001202	-0.001450	-0.000786	-0.001622	0.000777	0.001346	-0.000725	-0.001443	-0.001479	0.001174	...
one_example|2009|	0.000827	-0.001201	-0.001264	0.001345	-0.000715	-0.001255	-0.001604	-0.001646	0.000730	0.000454	...
a_cus|2009|	0.001183	-0.000796	-0.000373	-0.000408	-0.001534	-0.001096	0.000224	-0.000150	-0.000418	0.000476	...
tomized|2009|	-0.000884	-0.001573	-0.000344	0.000381	0.001275	0.001184	0.001062	-0.001189	-0.001338	0.001104	...
...	...	...	...	...	...	...	...	...	...	...	...
includes_curving|2018|	0.001314	-0.000602	-0.000422	0.001574	0.001549	-0.000776	-0.000711	-0.001506	-0.001033	-0.000914	...
out_the_highest|2018|	-0.001143	0.001263	-0.000859	-0.001494	-0.000820	-0.001288	-0.000580	0.000820	-0.000171	-0.000069	...
and_lowest|2018|	-0.000026	-0.000250	0.000499	-0.001109	-0.000239	0.000018	0.000471	0.000699	0.001169	-0.000150	...
a_second_line|2018|	-0.000248	0.000798	0.000126	0.001276	0.000175	0.001086	0.000707	0.001039	-0.000264	-0.000992	...
a_third_line|2018|	-0.001661	-0.000339	-0.000325	-0.001138	-0.000100	0.000576	-0.001537	-0.000226	0.000220	-0.001397	...
the_slop|2018|	-0.000172	0.001121	-0.000555	0.000125	0.001139	-0.001474	-0.000189	-0.000640	0.000327	-0.001605	...
the_mono|2018|	-0.001561	-0.000819	0.000834	-0.001187	-0.001460	-0.001338	0.001228	0.001022	0.001373	-0.000902	...
polygonal_line_curve|2018|	0.000802	0.001545	-0.000314	0.000163	-0.000074	0.000017	-0.001581	0.000771	-0.000855	-0.001561	...
devia_the_learner|2018|	-0.001343	-0.000226	-0.001513	0.000805	-0.001378	-0.000987	0.000249	-0.000387	-0.001170	-0.000820	...
comparison_syllabic|2018|	0.000584	0.000339	0.001343	0.001360	-0.000717	-0.000492	-0.000374	0.000501	0.000905	-0.001548	...

Example of embeddings: a matrix with the unique “words” as 1st column, each word with numbers defining it in the given text



Steps to create the system

SYSTEMS ENGINEERING RESEARCH CENTER

The “crystal ball” – where we are

- The system has been developed as part of a project sponsored by the U.S. Department of Defense through the Office of the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)) under Contract [HQ0034-19-D-0003, TO#0150]
- At the current stage, the prototype can predict
  - The “neighborhoods” where a technology can be in the future
  - The elements/keywords that will characterize the predicted technology
- We are testing the system “predicting” the present, based on the past. Results are improving, fine tuning the different steps
- One of the critical element is the way the “embeddings” are created. We currently use a leading vectorization method (Word2Vec), but we are evaluating a method based on non-Euclidian spaces/”quantum spaces”

SYSTEMS ENGINEERING RESEARCH CENTER

Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

SYSTEMS ENGINEERING RESEARCH CENTER

Samples of application

Pouria Babvey

Pouria Babvey is a PhD candidate in School of Systems and Enterprises in Stevens Institute of Technology. He holds degrees in Electrical Engineering from University of Sharif (MSc) and the University of Tehran (BSc).

His research efforts focus on the development of artificial intelligence architectures for natural language processing and social network analysis.

During the time at Stevens he has been participating in projects for UNICEF, Wiley, and NSF on scholar networks, social media conversational patterns, evolutionary trends in the tech-products, and human-bot collaboration. He also has been teaching Computer Algorithms for several years and published some books.

Fernanda Capela

Fernanda Capela is a PhD candidate in School of Systems and Enterprises in Stevens Institute of Technology. She holds a a degree in Industrial Engineering with a minor in Financial Management from the Federal University of Rio de Janeiro (UFRJ, Brazil).

She is a data scientist and her research explores the impacts of big and crowdsourced online data in the modern society.

Her areas of interest are Data Analytics, Data Mining, Text Mining, Emotion Recognition, Sentiment and Opinion Analysis, Machine Learning, Social Media and Social Network Analysis, Data Visualization, Data Story-telling, and Computational Social Sciences.

SYSTEMS ENGINEERING RESEARCH CENTER

Agenda

- The datatification
- How Society is changing
- How Business environment is changing
- Technology trends
- Data & NLP trends
- Building the crystal ball
- Sample Applications
- Data Science Trilogy: key take-aways

SYSTEMS ENGINEERING RESEARCH CENTER

Data Science Trilogy: key take-aways

- We are in a digital transformation process that is creating a new kind of economy based on the "datafication" of virtually any aspect of human social, political and economic activity as a result of the information generated by the digitally connected individuals, companies, institutions and machines
- Data Science projects are processes, not just an application of algorithms and tools
- You need to know and interact with your data to get meanings out of it
- Data Science is part science, part craftmanship. The science is getting better, but the craftmanship will stay for a while
- Intelligent systems are formal representations of someone's intelligence: focus first on the intelligence you are bringing to the process, not on the technology
- Data Science develops models of the reality based on the data you can collect: data has to be representative of that reality. You need to know the reality and be sure it is properly represented by the data you have
- Don't disregard results if they look different from what you expect: check the model first but then have an open mind on what the data is telling you

SYSTEMS ENGINEERING RESEARCH CENTER

DAU

STEVENS INSTITUTE of TECHNOLOGY

THE INNOVATION UNIVERSITY®

Thank you!

School of Systems&Enterprises

Dr. Carlo Lipizzi

clipizzi@stevens.edu

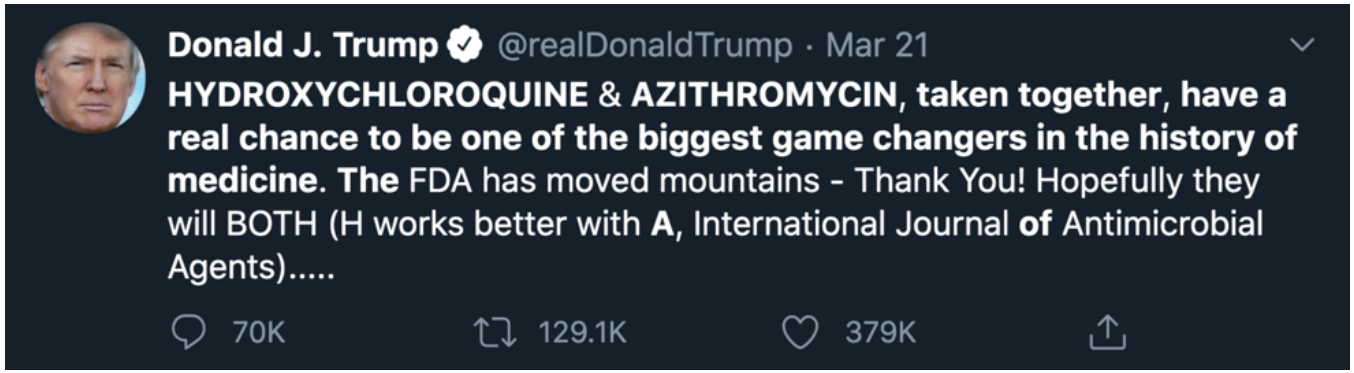
NLP LAB

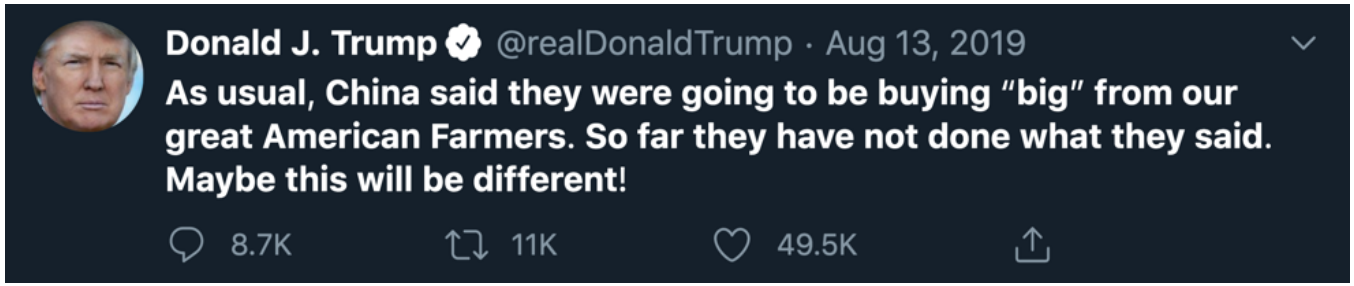
STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

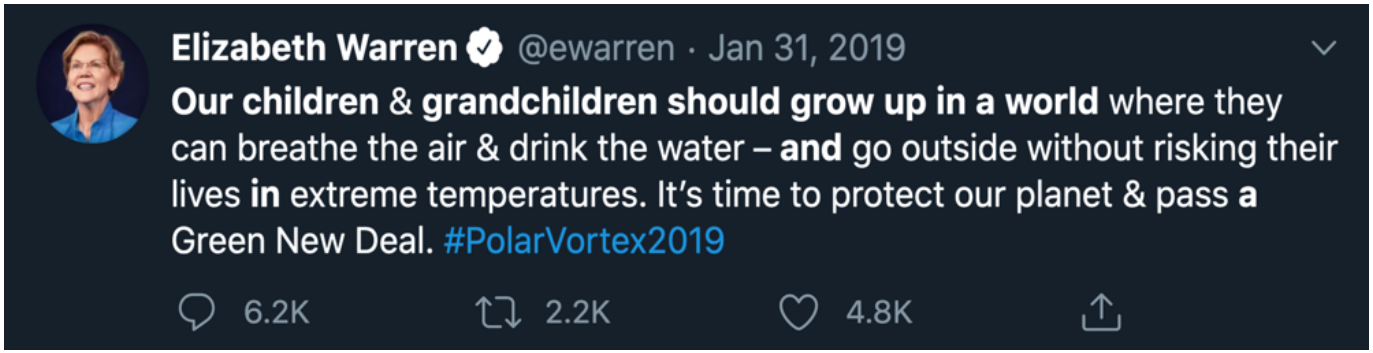
Content-aware Galaxies: Digital Fingerprints of Discussions on Social Media

Stevens Institute of Technology
School of Systems and Enterprises

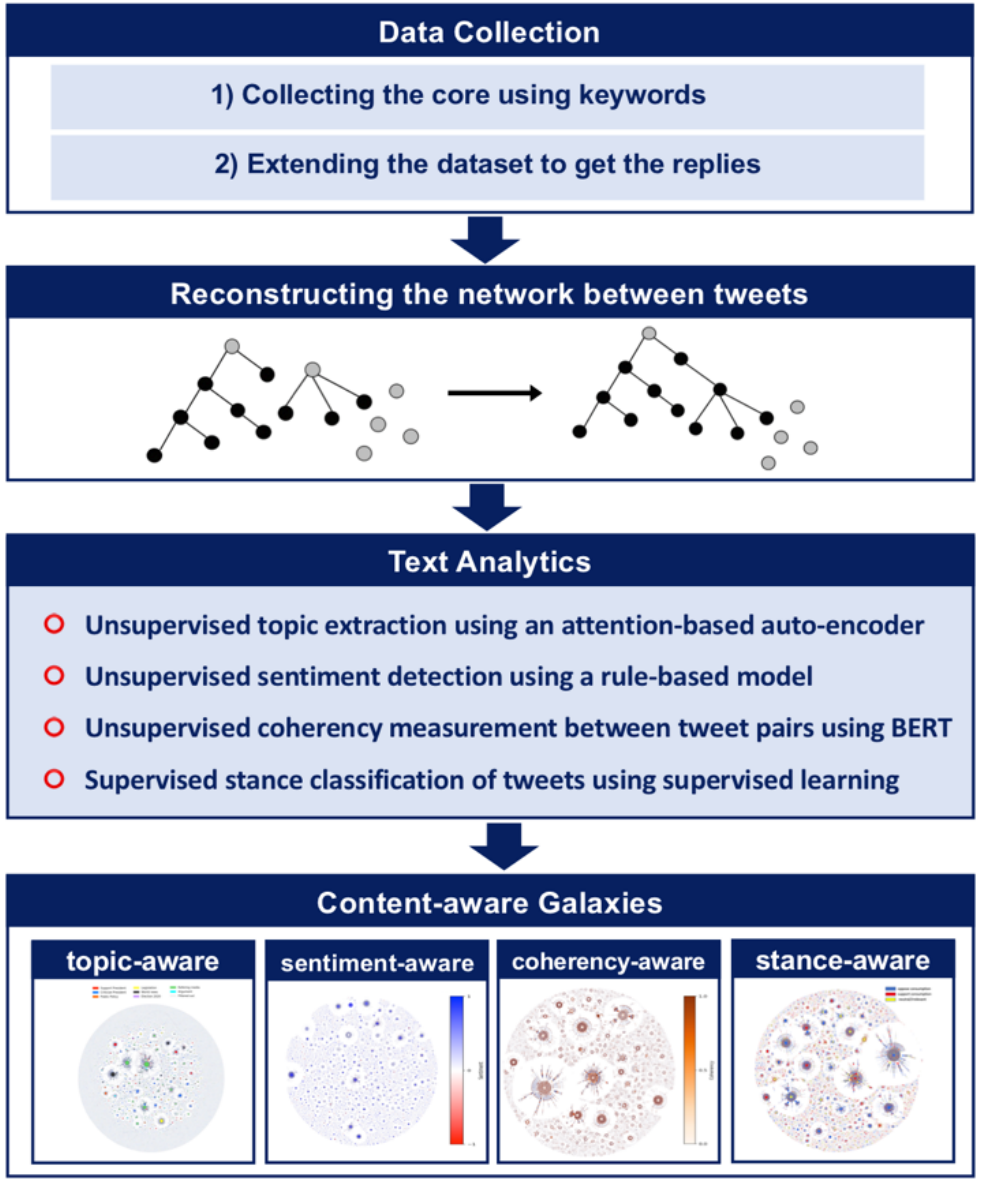
August 13, 2020


Donald J. Trump @realDonaldTrump · Mar 21
HYDROXYCHLOROQUINE & AZITHROMYCIN, taken together, have a real chance to be one of the biggest game changers in the history of medicine. The FDA has moved mountains - Thank You! Hopefully they will BOTH (H works better with A, International Journal of Antimicrobial Agents).....
70K
129.1K
379K

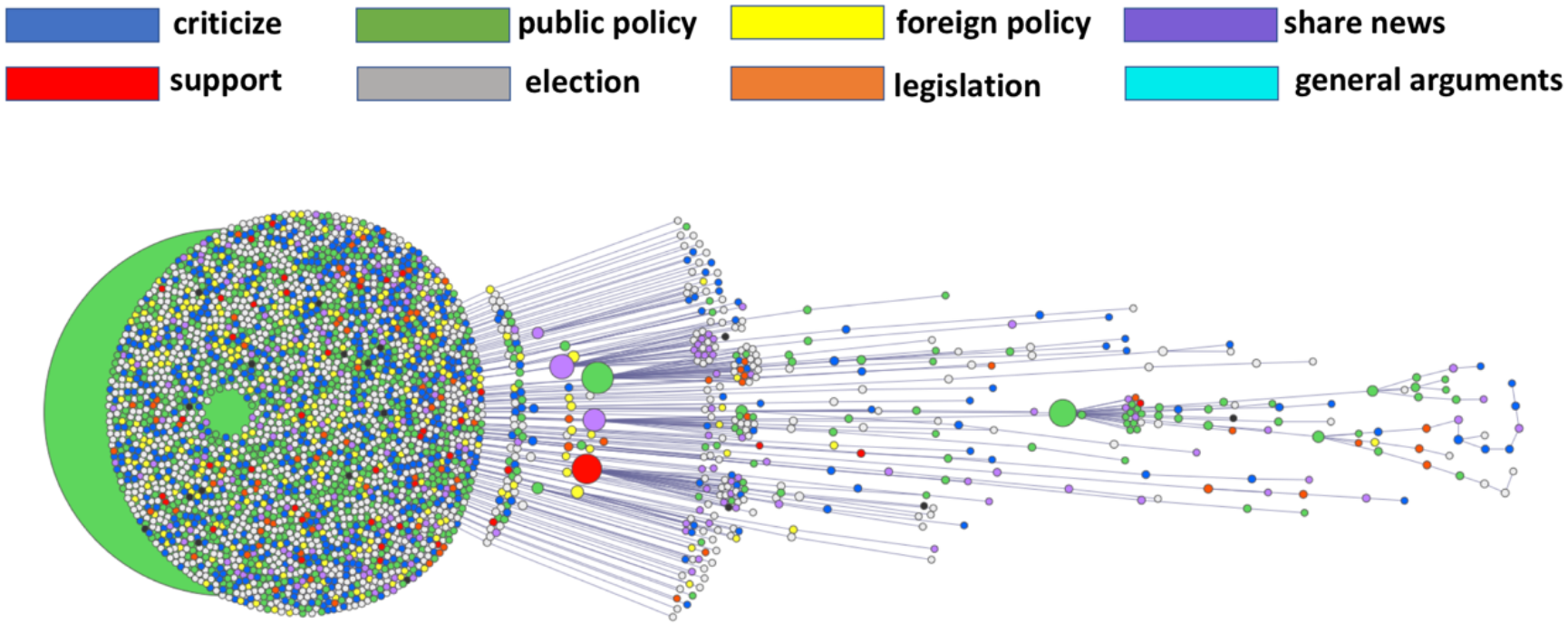

Donald J. Trump @realDonaldTrump · Aug 13, 2019
As usual, China said they were going to be buying "big" from our great American Farmers. So far they have not done what they said. Maybe this will be different!
8.7K
11K
49.5K


Elizabeth Warren @ewarren · Jan 31, 2019
Our children & grandchildren should grow up in a world where they can breathe the air & drink the water – and go outside without risking their lives in extreme temperatures. It's time to protect our planet & pass a Green New Deal. #PolarVortex2019
6.2K
2.2K
4.8K

Framework


Data Collection
1) Collecting the core using keywords
2) Extending the dataset to get the replies
Reconstructing the network between tweets
Text Analytics
Unsupervised topic extraction using an attention-based auto-encoder
Unsupervised sentiment detection using a rule-based model
Unsupervised coherency measurement between tweet pairs using BERT
Supervised stance classification of tweets using supervised learning
Content-aware Galaxies
topic-aware
sentiment-aware
coherency-aware
stance-aware

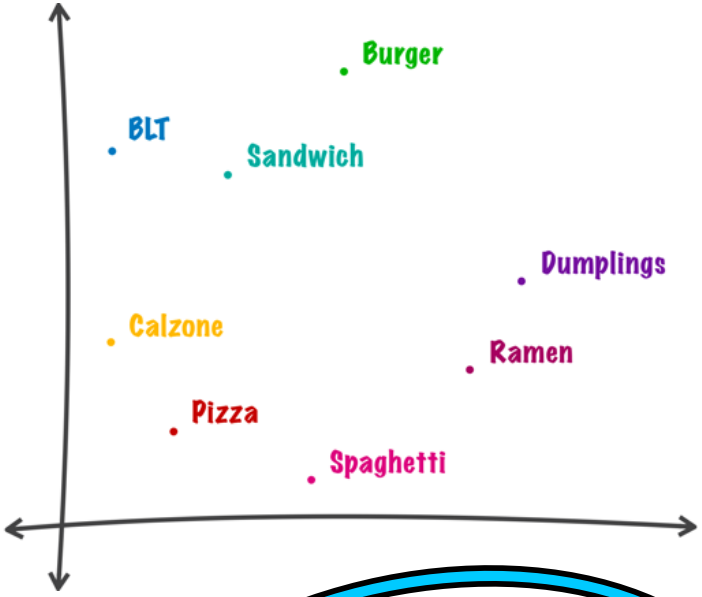
Network visualization



Topic-aware Galaxies

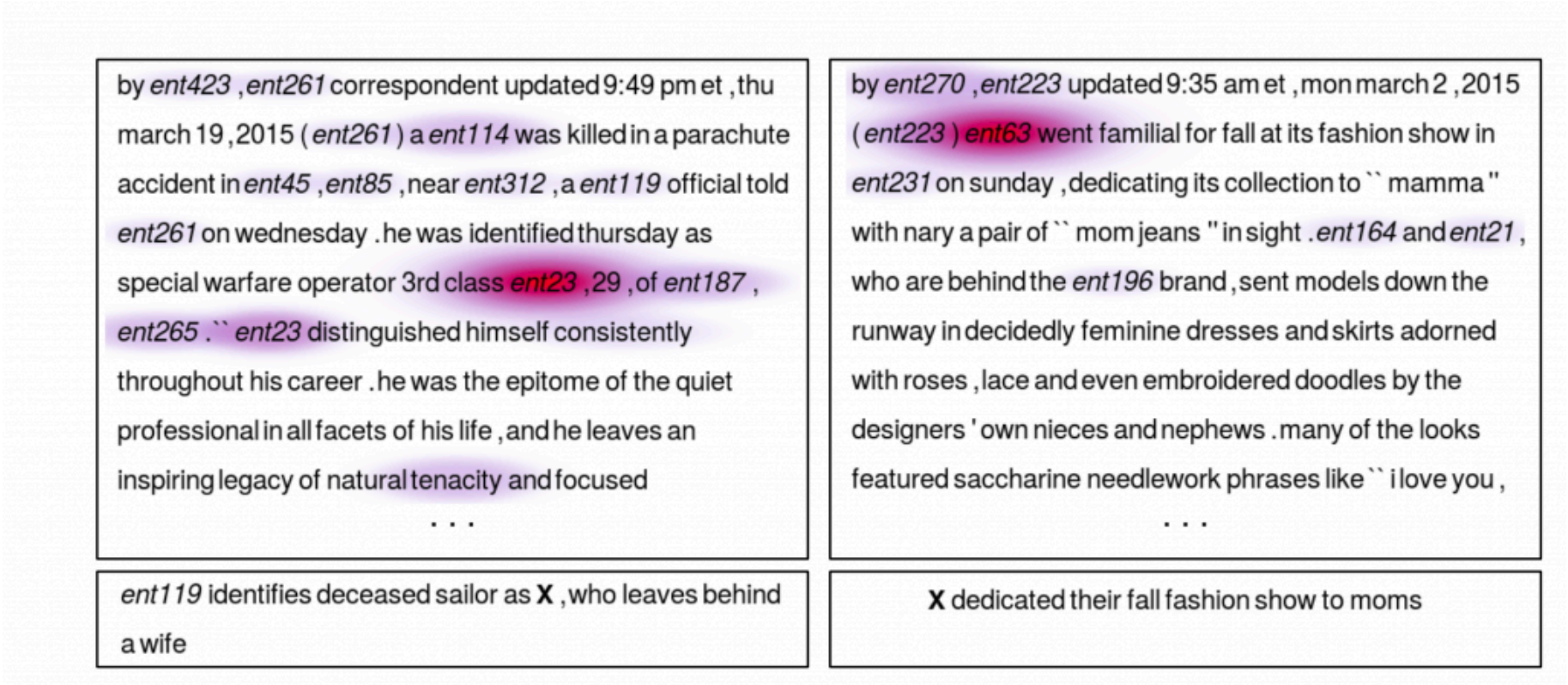
Word2Vec

is used to convert words to vectors

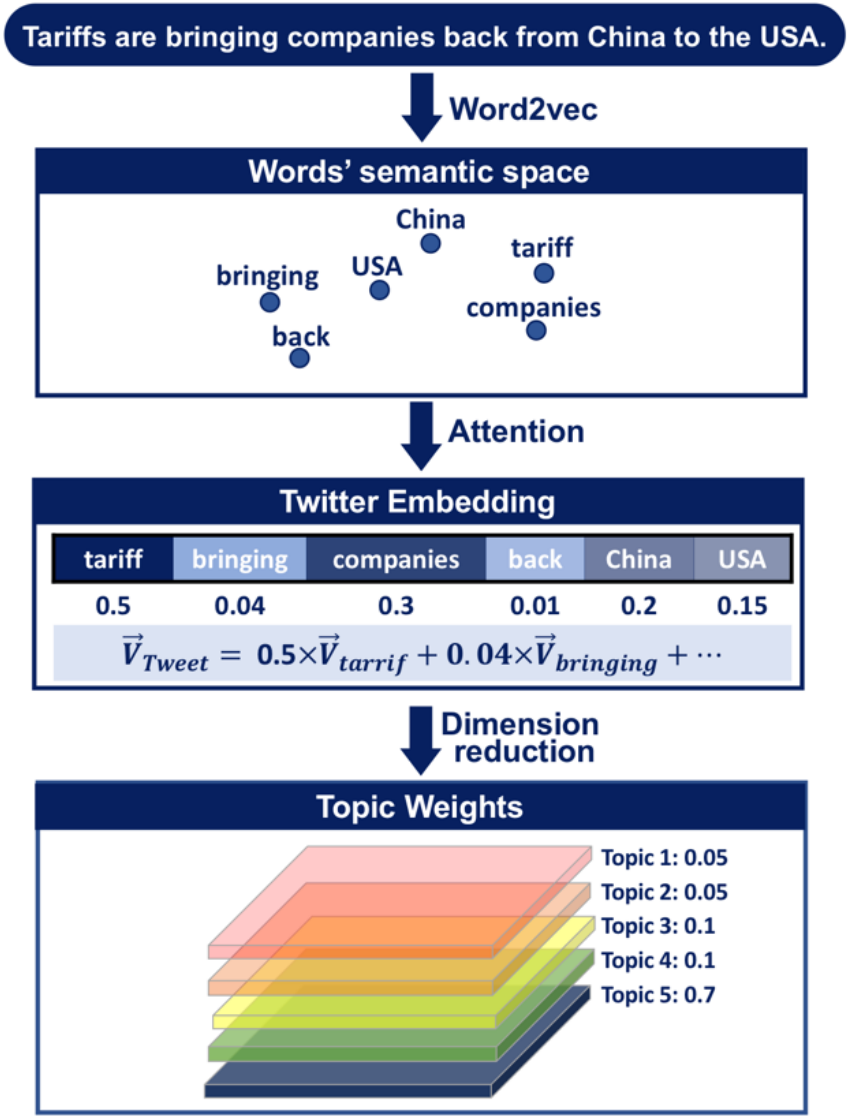


"He is subsidizing farmers with tariff money from China, not borrowed money from China.

Attention Mechanism



Model Architecture

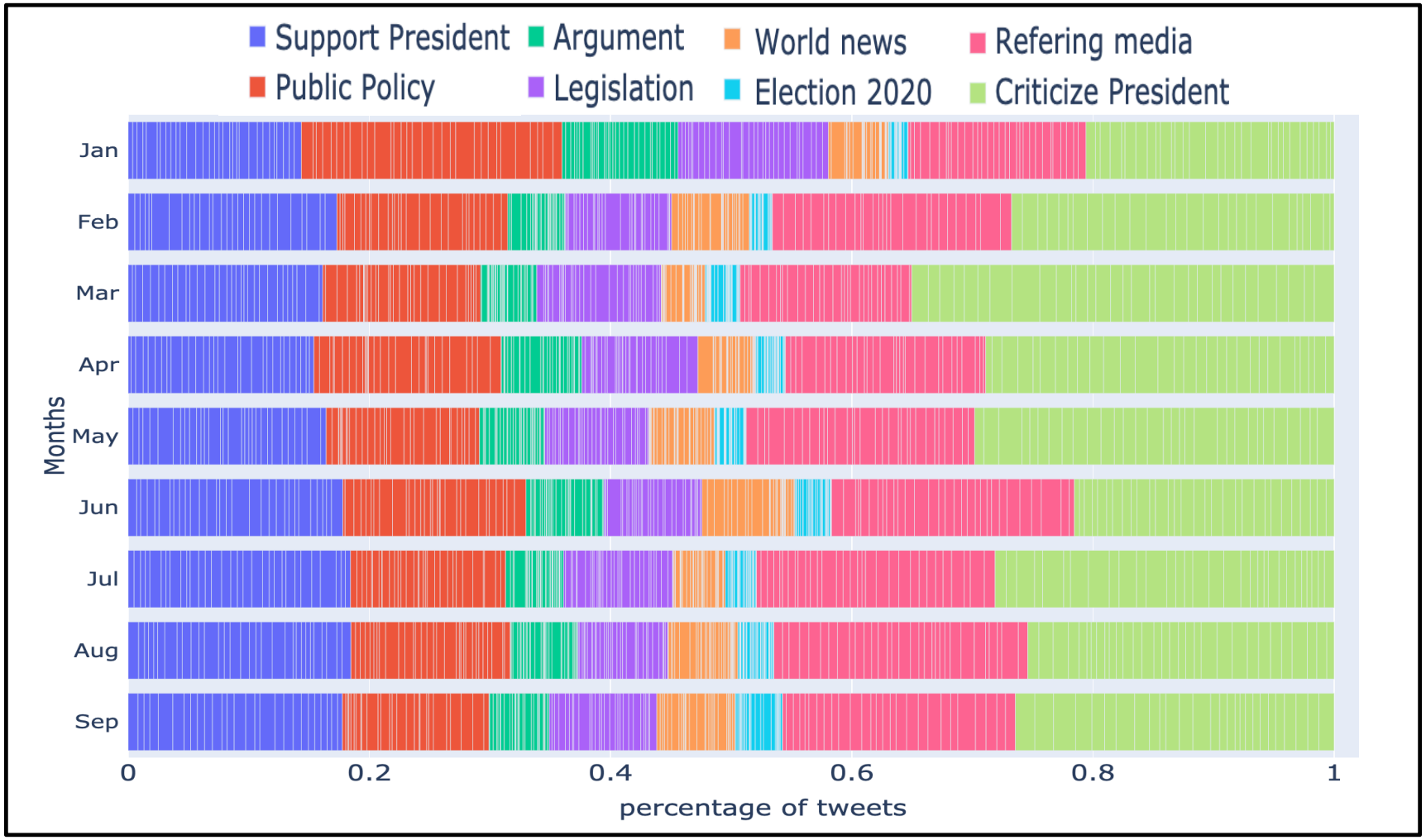


Case Study

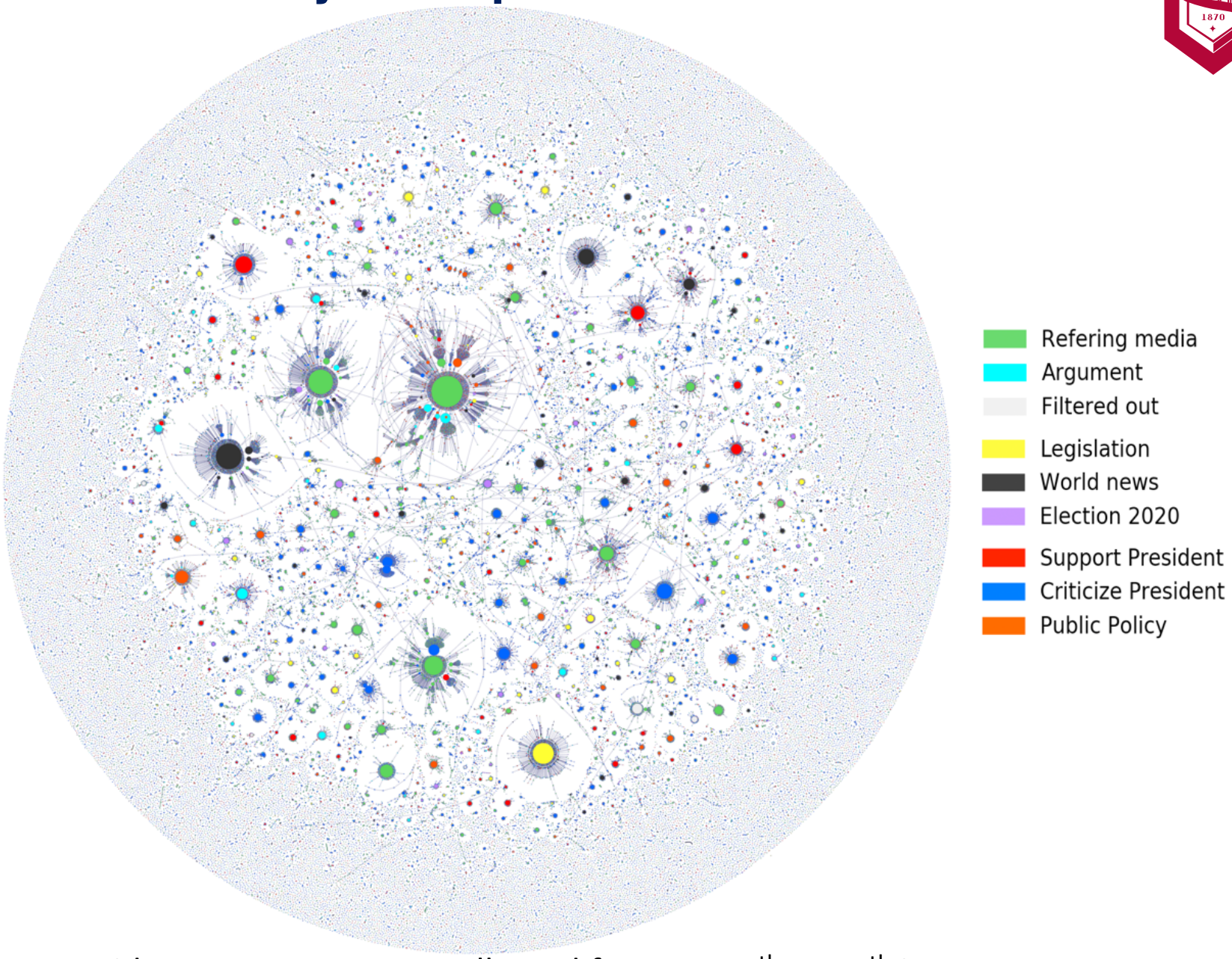
President Trump motivated
Tweets: ~ 30 million tweets
Jan 2019-Sep 2019

Topic Label	Sub-topic Label	Top words
Support President	**affirming President**	georgesoros, do_your_job, trump_war_room, mitchellvii
	admiring President	blessed, grateful, spirit, best_president_ever, patriot, wonderful, sacrifice
Criticize President	**impeach President**	resignnow, jilevin, impeachpence, impeachnow
	racism	divisive, spewed, xenophobia, bigotry, demagogue, sexism, racial
	satire	eating, cheese, hamberders, juice, drink, snort, cheeseburger, pizza
	sexual assault	assaulted, teenager, sexually, pregnant, incest, infant, raping, lesbian, daughter
	President investigations	cover-up, perjury, prosecution, McCabe, investigation, russiagate, robertmuller, indictment
Public policy	**financial struggles**	contractor, charity, loan, mortgage, stiffed, property, reimbursement, owes, salary,
	illegal immigration	illegal_aliens, coyote, smuggling, sanctuary_cities, Yuma, Guatemala, Honduras
	economy	deregulation, wage, industry, layoff, stagnant, growth, gdp, manufacturing, export, economy, inflation, income, taxation, recession, subsidy, revenue, farm
Next election	**election 2020**	2020election, beat, gerrymandering, challenger, candidate, senatorial, ballot, berniesanders, voter, ewarren, andrewgillum, incumbent
World news	**foreign policy**	Erdogan, Pakistan, Turkey, Maduro, jzarif, india, Syria, Kurd, Afghan, Cuba, Tehran, Mullah, Japenese
Arguments	**arguments**	although, confused, obviously, honestly, opposite, impressed, saying, neither, unfortunately, genuinely, complicated, admit, besides, weird
Legislation	**Congress news**	fake_emergency, veto, override, resolution, rand_paul, senator_collins, unanimously, senate_majldr, legislating, senate, government_shutdown
	bill process	require, function, procedure, legislative, statute, limiting, institution, mandatory, enact, administer, obligation, governmental, banning
referring media	**Democrats' media**	huffpost, columnist, newsweek, editorial, newyorker, latimes, buzzfeednews
	Republicans' media	gov_mike_dewine, Shulkin, Toomey, elizabeth_pipko, james_aydelott
	diplomatic visits	guest, hosted, palace, Buckingham, Orlando, dinner, ceremony, Manchester, visiting, scheduled, Miami, invitation, delegation, Fayetteville, trumpukvisit
	talking media	shadow_banning, retweets, algorithm, censored, suspended, user, content, unfollowed, Parler, trollbots, satire, feedback, facebook, pinned, fakenews, Koppel
	reporting events	December, 10th, October, alert, date, decodes, percent, estimated, margin, consecutive, neverforgotten, posted, apprehended

Share of topics over time



A Topic-aware Galaxy Example

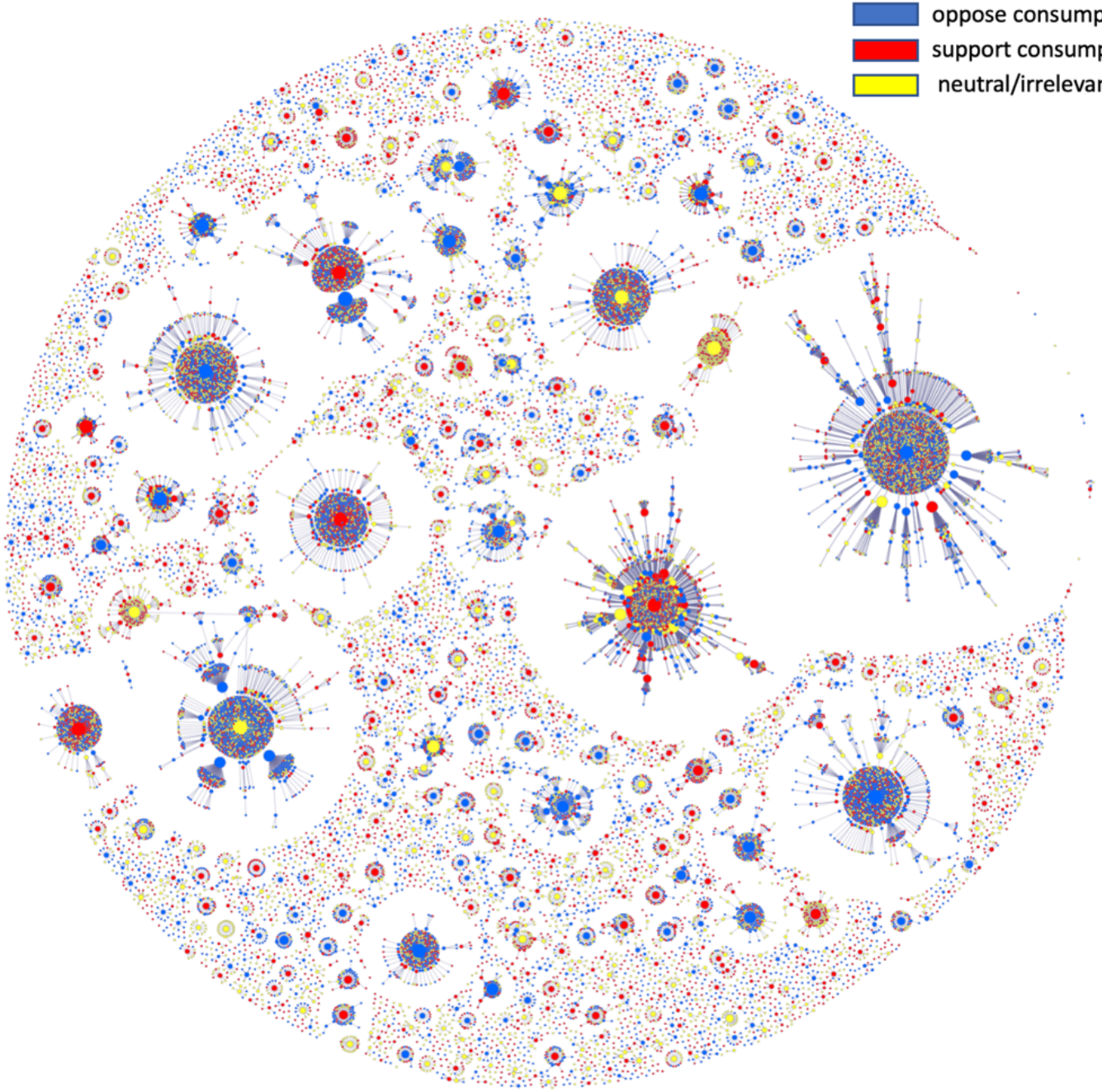


President Trump, Tweets collected from Sep 9th to 11th in 2019

Stance-aware Galaxies

Case Study: Hydroxychloroquine

- First, collected all the tweets that contained “Hydroxychloroquine” in the first 5 months of 2020 as the core (~400k).
- Next, using some fields we extended our dataset to collect all the replies (5 millions)
- Our results show **29% support, 38% oppose, and 33% neutral** messages
- Criticism increased over time
- **Neutral language** did not generate any **cross-cutting discussion** with **significant majority** of one side in the generated discussions



Tweets posted from March 20th to April 4th in 2020 containing ~47000 tweets.

STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

SYSTEMS
ENGINEERING
RESEARCH CENTER

Text Mining and NLP Applications for Decision-Making

Fernanda Capela
fcapela@stevens.edu

August 2020

Risk Panel System – Problem Scope

- Assess the risk of a company to **lose competitiveness** in its market segments, based on investments and acquisitions.
- **Decision support system** that recommends technological applications for companies in the security industry.

  - *Who are my major competitors?*
  - *What is my position compared to my competition?*
  - *What technologies do they have that I don't?*
  - *What new technologies from my market segment are being commercialized? Which ones are being developed?*
  - *Which technologies will give me the best competitive advantage return?*

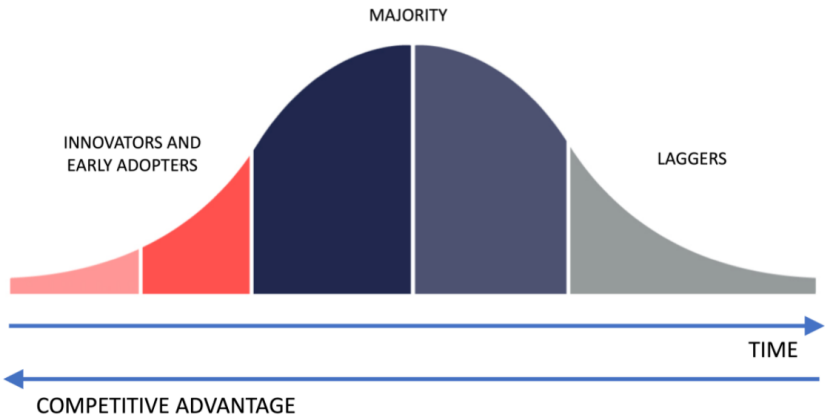
Risk Panel System – Metric Definitions

Who are my major competitors?

- Companies with similar technology portfolio that compete with me for market share.
- Technologies that can overcome mine – **Technology Kill Chain**

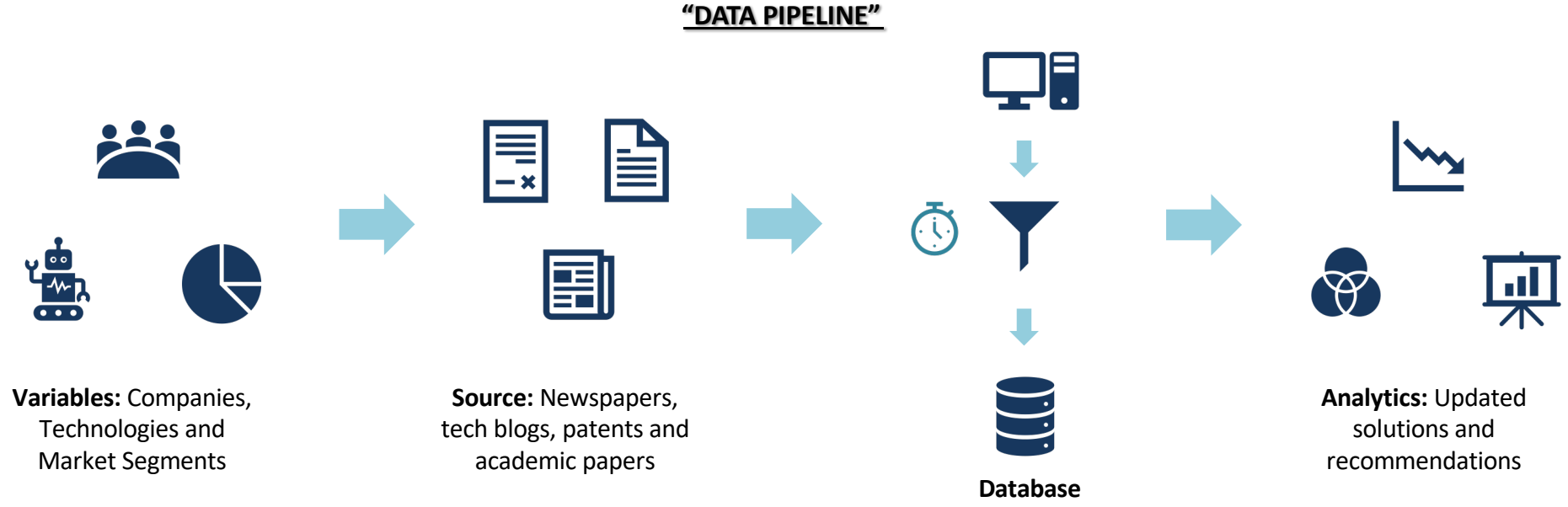
Which technologies will give me best competitive advantage return?

- Define competitive advantage.
- Create an index to represent the concept quantitatively - **MOAT**



Risk Panel System – Data Gathering

- What type of data do we need to answer each one of the problem's questions?
- Consider time period! (Tech industry changes fast!)
- What is the data source? Is it enough? Do we need to combine different datasets? – Web scrapping, APIs, etc.



Risk Panel - Analysis and Recommendations

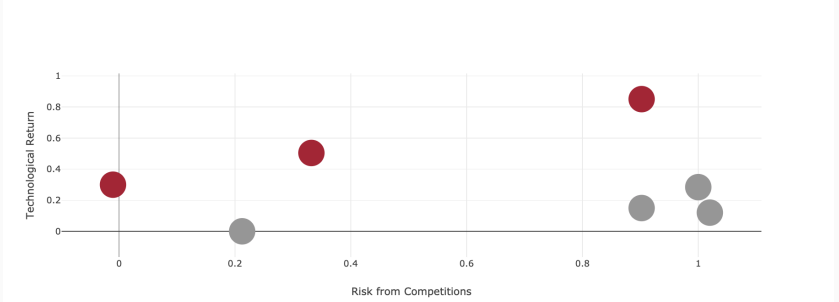
Meshing Capability and Threat-based Science & Technology Resource Allocation

SYSTEMS ENGINEERING RESEARCH CENTER

Recommendation | Comparative Analysis | Risk Assessment | Preference Inputs | Metrics Details | Events/Acquisitions | Implementation Guide | Help

Sector Assessment

Evaluating a tech-driven sector can be similar to optimizing an investment portfolio as it is a decision and trade-off regarding expected return and risk.



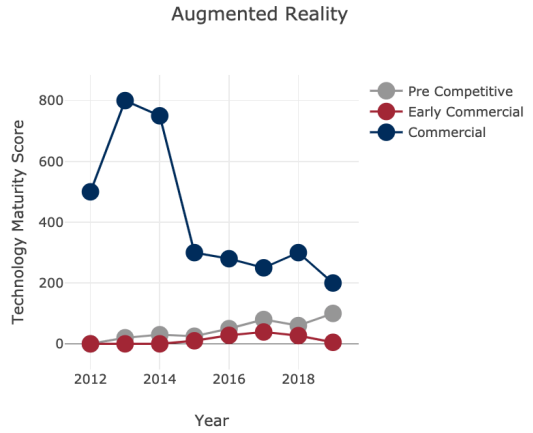
Notes: Sector(s) in red are recommended as they offer the highest expected return for a defined level of risk, or the lowest risk for a given level of expected return. Sector(s) in grey are other sectors. You may adjust your preference in the **Preference Inputs** page.

Sector Recommendation

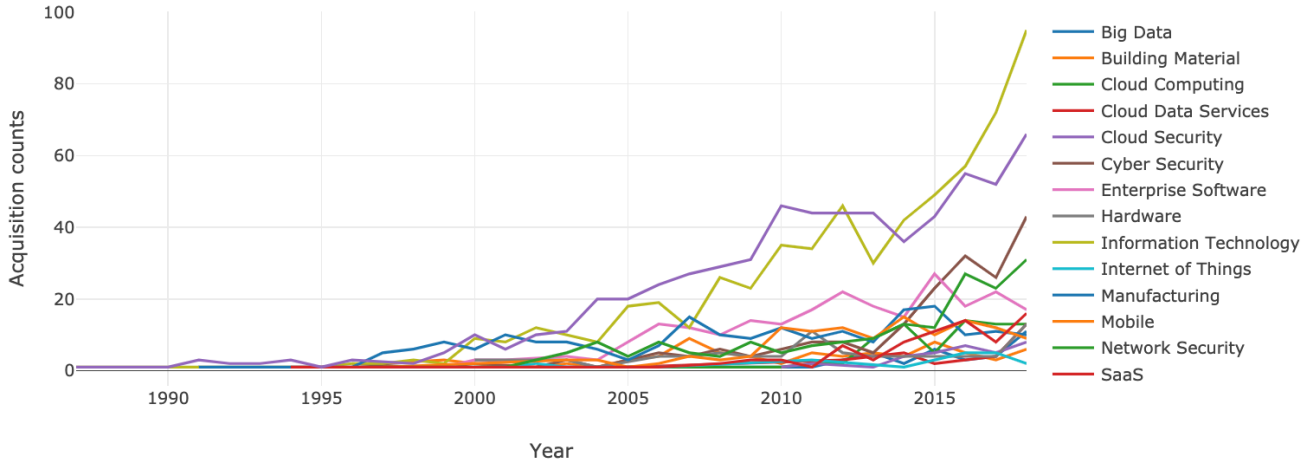
Based on the technological assessment scores on the left and your risk preferences in the second tab, we suggest general action for the following sector:

Enabling Technology

Explore the major enabling technology maturity for the selected sector. Technology maturity score is currently defined as Room Theory based distance.

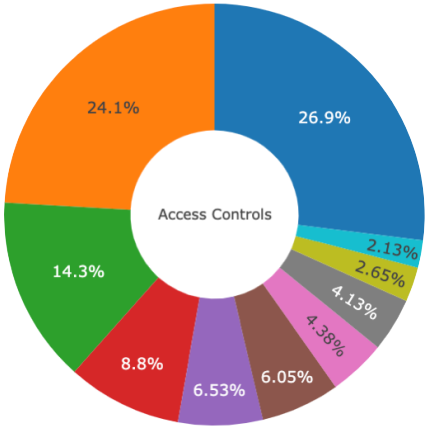


Tracking and monitoring of data trends, giving insights about the market.



Market Environment

Gather competitors information. Colors represents companies, and the number currently is derived from revenue.



Risk Panel - Analysis and Recommendations

Recommendation | Comparative Analysis | Risk Assessment | Preference Inputs | Metrics Details | Events/Acquisitions

Players and Strategy

Make selections to interact with the system.

Select Your Company:

The system will make recommendations for the selected company.

SAAB AB-B

Select Competitors:

Add or remove competitors.

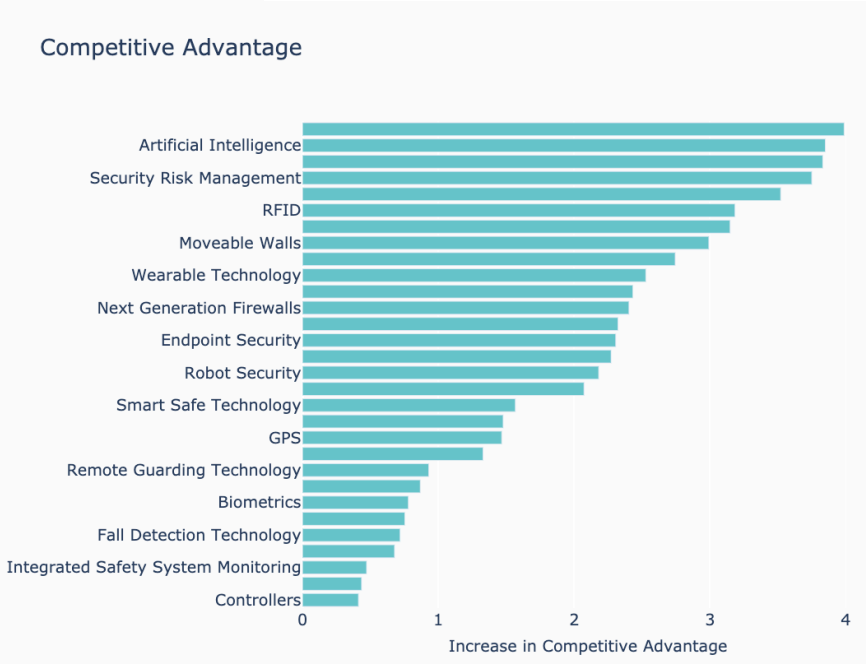
ADT INC | ALLEGION PLC | ASSA ABLOY AB-B | BRINK'S CO | DORMAKABA HOLDING | G4S PLC | GEMALTO | GEO GROUP INC/TH | LOOMIS AB-B | PALO ALTO NETWORK | S-1 CORPORATION | SAAB AB-B | SECURITAS AB-B | STANLEY BLACK & DECKER | SYMANTEC CORP | UNITED TECH CORP

Select Strategy:

- Technology Portfolio evaluates technology ownership and recommends new technologies based on major competitors' technology portfolio.
- Competitive Advantage considers time of ownership as key pilar of competitive advantage and recommends technologies to optimize increase in competitive advantage.

(•) Technology Portfolio
( ) Competitive Advantage

Recommendation Ranking



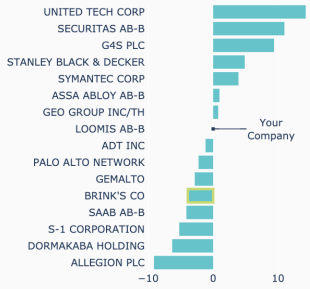
Interactive: User can make selections and play with different scenarios.

Risk Panel - Analysis and Recommendations

Recommendation | Comparative Analysis | Risk Assessment | Preference Inputs | Metrics Details | Events/Acquisitions | Implementation Guide | Help

Comparative Strenghts and Weaknesses with Competition

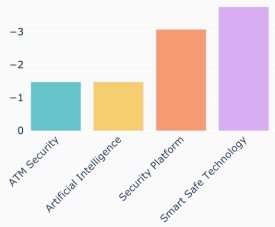
Your strategic position compared to competitors.
Your major competitors are highlighted.



Select a Technology to Divest:

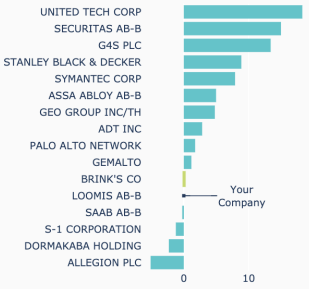
Smart Safe Technology

Loss in Competitive Advantage if Divest in Technology



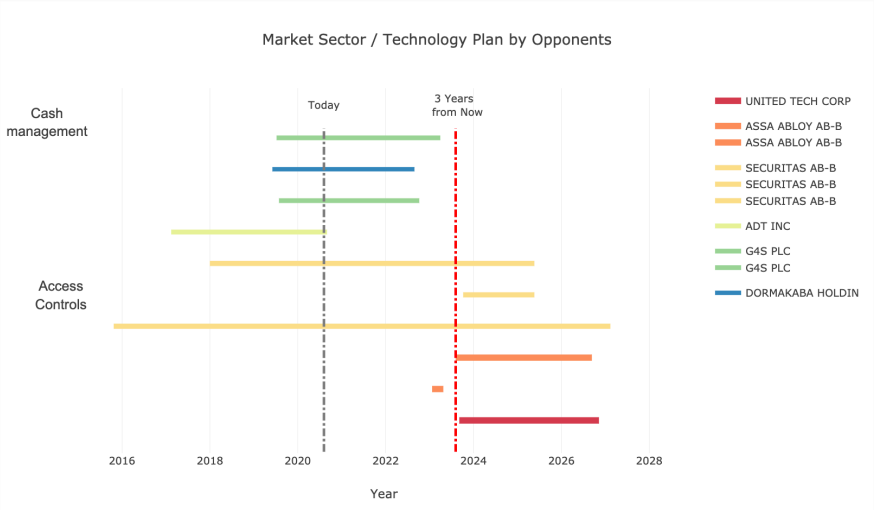
Ranking Loss for Divesting in Technology

Change in strategic position after dropping investment on selected technology.
Your major competitors are highlighted.



Estimation of Impact of Investing and Divesting in Technologies.

Market forecasting



STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

stevens.edu

Thank You!

fcapela@stevens.edu