Enriching Knowledge Series

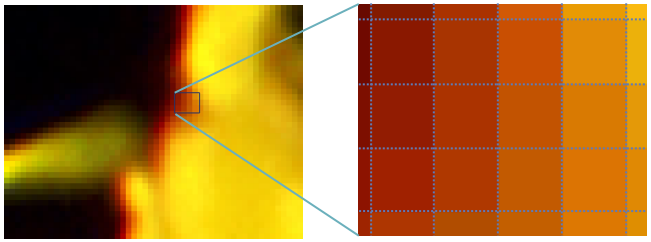
The Uses and Applications of Big Data in Daily Life

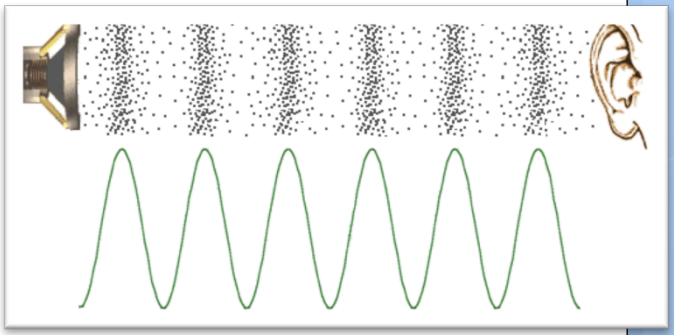
Dr. Reynold Cheng
Associate Professor
HKU Computer Science
ckcheng@cs.hku.hk

Reynold Cheng (鄭振剛)

- **Position:** Associate Professor, Graduate Program Director
- **Education:** HKU (BSc, MPhil 95-00), Purdue (PhD, 00-05), HKPolyU (Asst. Prof, 05-08);
- **Research:** Database; Data mining; social networks;
- **Teaching:** Big Data (UG); Database (UG,PG)

Big Data: Introduction





Data

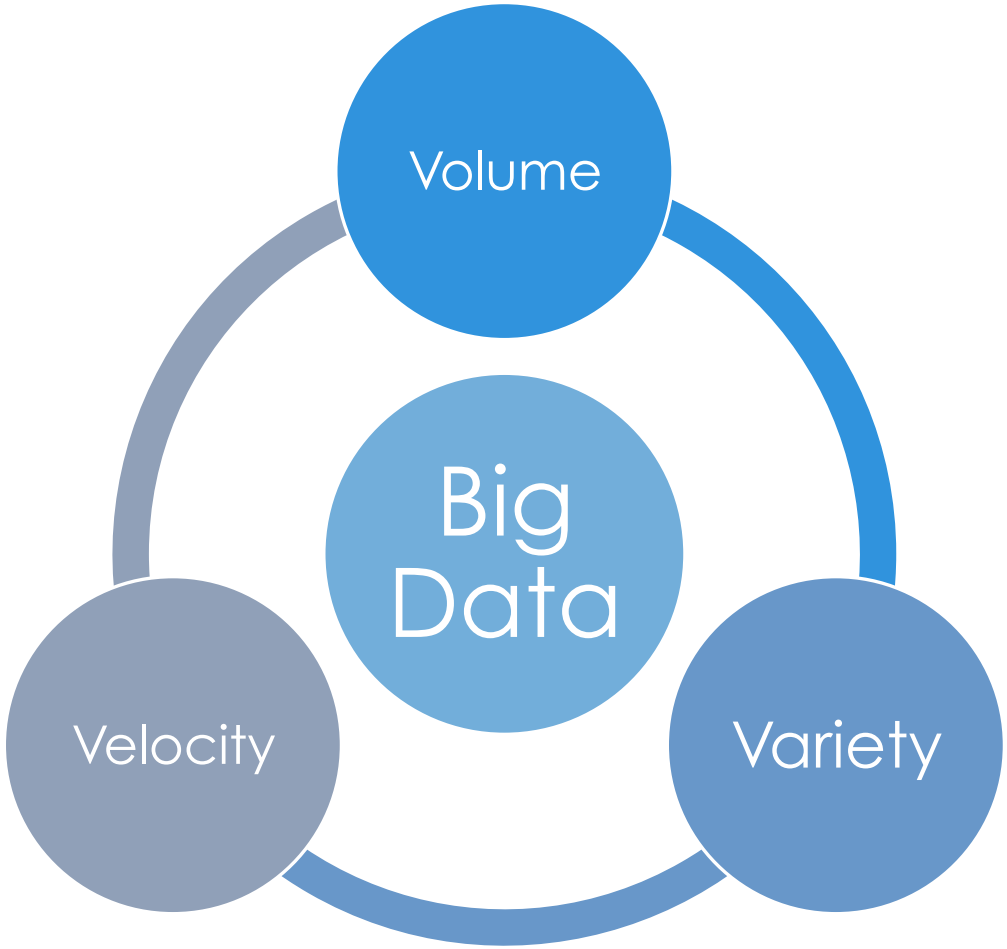
- **Text**
  - stored **word-by-word**, **character-by-character**
- **Image**
  - A 2D array/matrix of pixels (***picture elements)***
- **Audio**
  - Sound wave represented by a sequence of values
- **Video**
  - a sequence of images + audio

Image source: http://www.mediacollege.com/audio/01/sound-waves.html

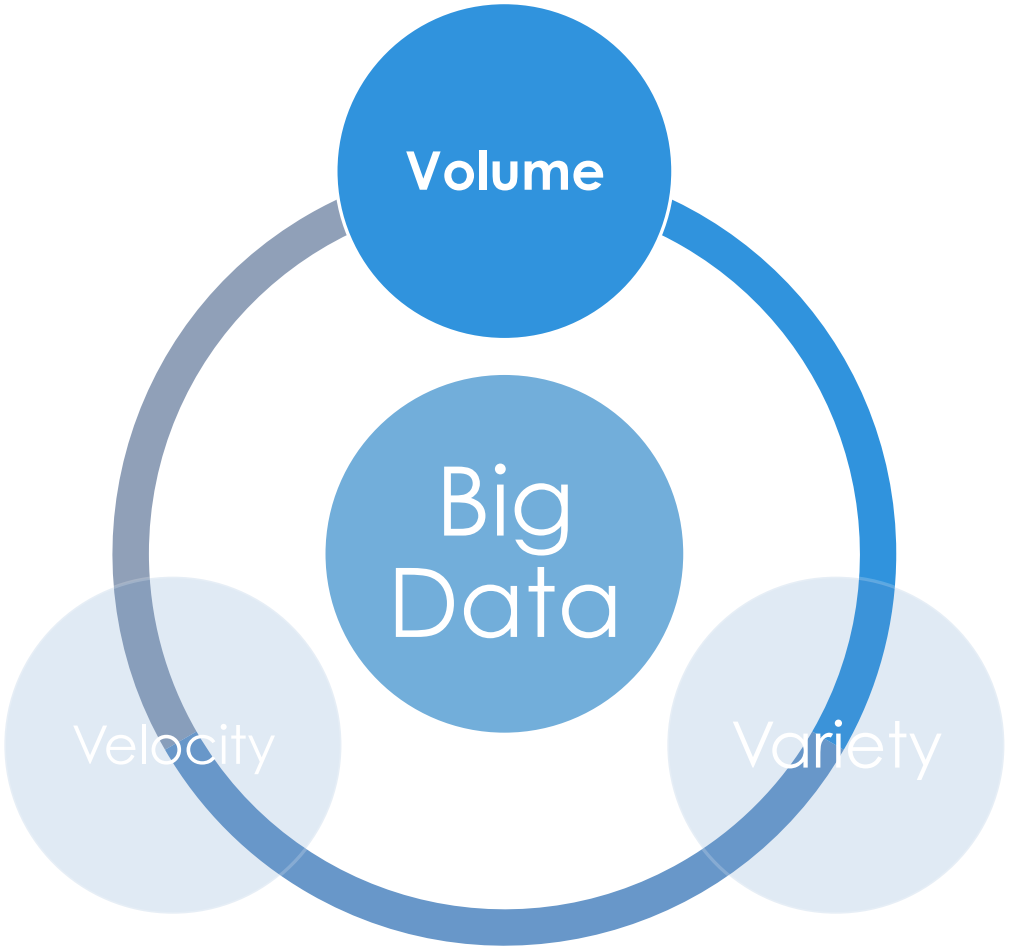
Big Data

- "Big data is data that exceeds the processing capacity of conventional database systems." [4]
- "Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications." [1]

The 3 Vs of Big Data

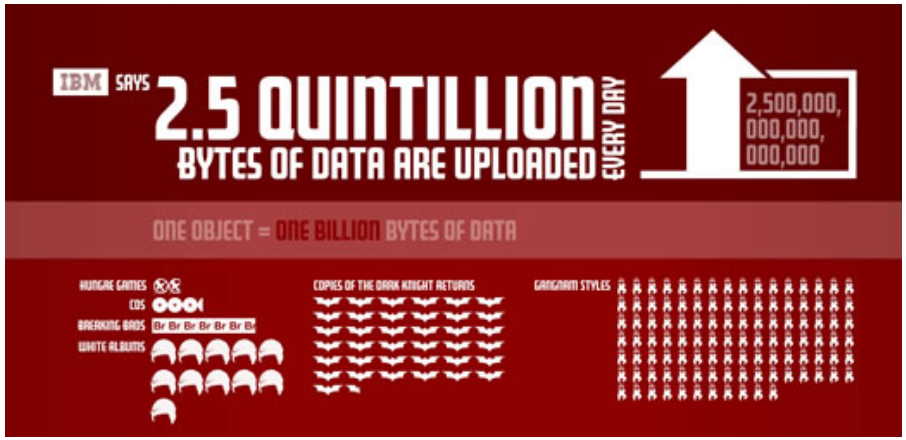


The 3 Vs of Big Data

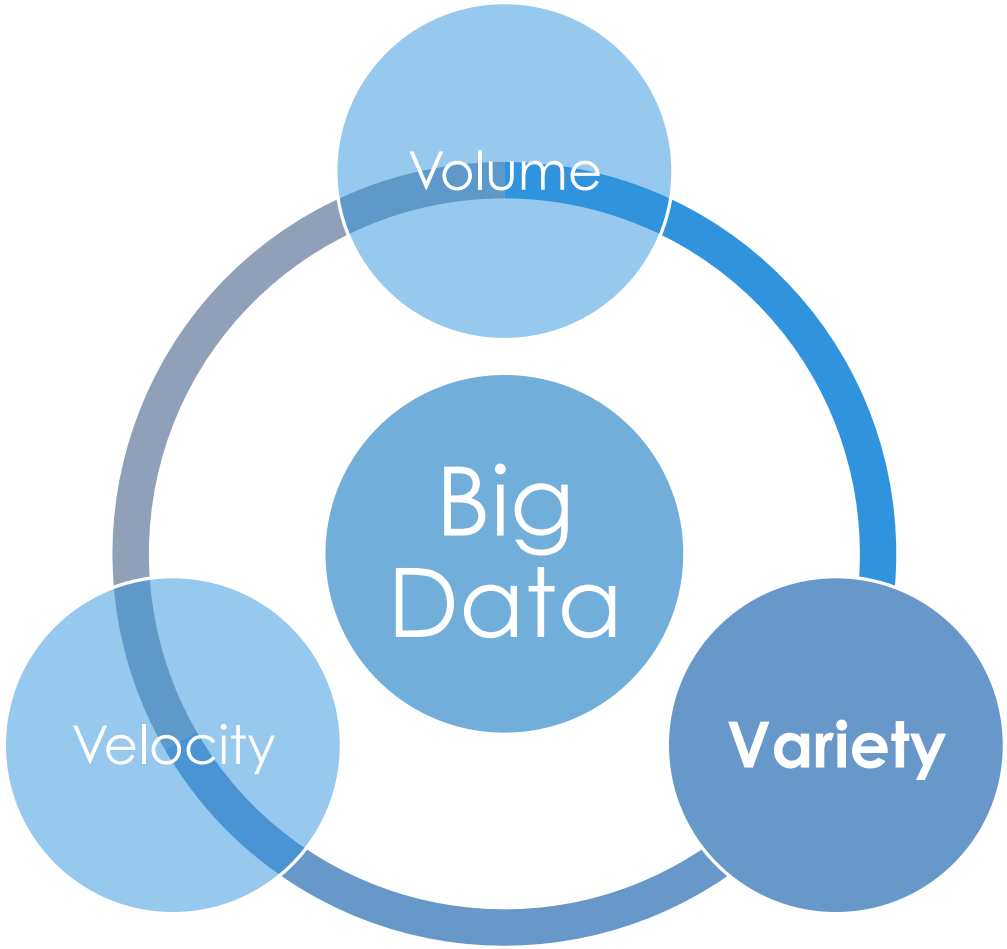


Big Data

- Every day, we create **2.5 quintillion** bytes of data
- **90%** of the data in the world today has been created in the **last two years** alone.



The 3 Vs of Big Data

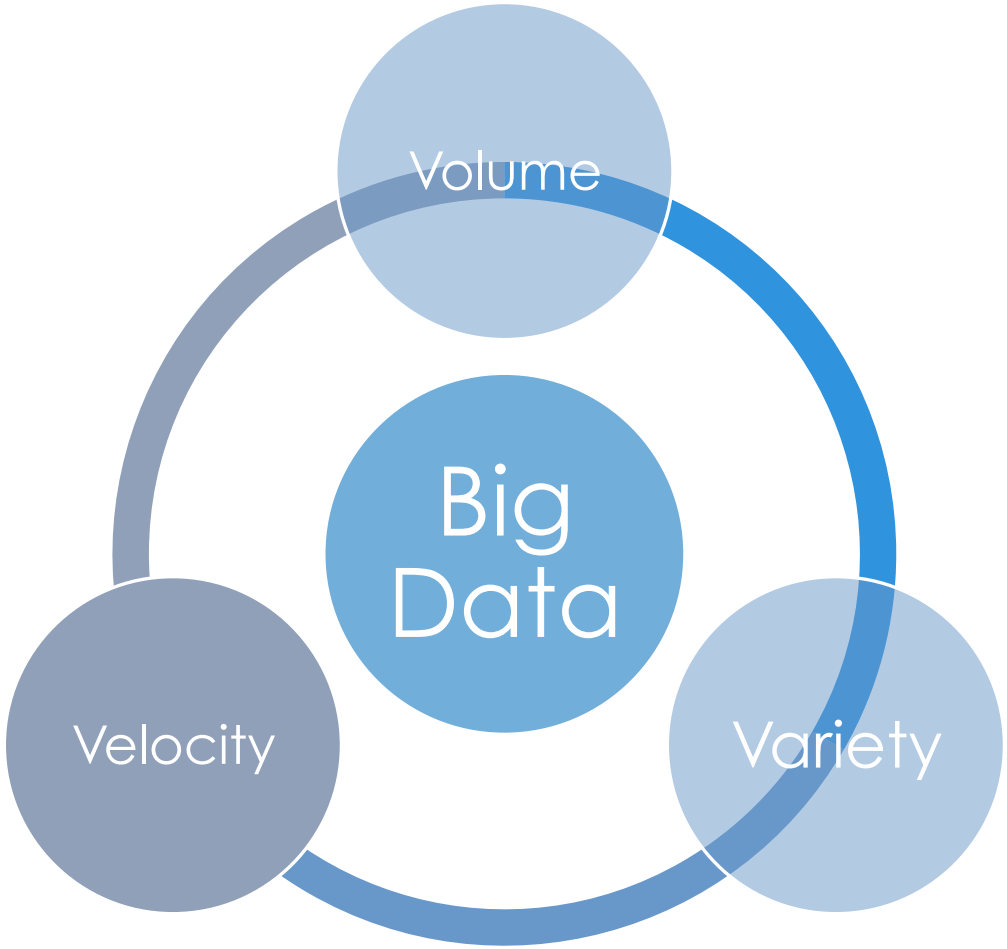


Sources of Big Data

- **Internet**
- **Crowd sourcing**
  - Wikipedia, forums, answers
- **Social networks**
  - Facebook, instagram
- **Microblogs**
  - Twitter
- **Sensors**
  - Mobile phones, GPS, Camera

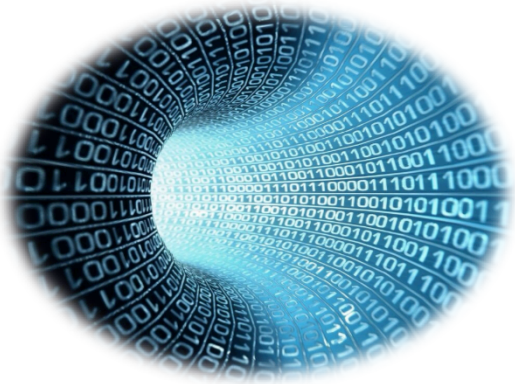
Image source: http://avtecmedia.com/images/blog/social-networking-logos.jpg

The 3 Vs of Big Data



Traffic statistics

- Juniper Research estimated that 14.7 trillion mobile messages would be sent from mobile devices in 2012*
  (i.e. 466,133 messages every second)



* http://www.statista.com/statistics/262005/mobile-message-traffic-worldwide/

"TransDec: A Big-Data Framework for Decision-Making in Transportation Systems" by Prof. Cyrus Shahabi

Road Traffic data

- RIITS (Regional Integration of Intelligent Transportation Systems)http://www.riits.net/
- Collects data from Los Angeles

Traffic data

Variety (GPS, video, loop sensor, events)

Data Ty	Size (in KB)	(in seconds)	Minute (in KB)	Hourly (in KB)	Daily (in KB)	Annual (in KB)	3 Years (in KB)
bus_mta	23	86400	0.96	0.96	23.00	8,395.00	25,185.00
bus_mta	1065	120	532.50	31,950.00	766,800.00	279,882,000.00	839,646,000.00
cctv_inv.x	57	86400	0.04	2.38	57.00	20,805.00	62,415.00
cms_inv.xn	52	86400	0.04	2.17	52.00	18,980.00	56,940.00
cms_rt.xml	48	75	38.40	2,304.00	55,296.00	20,183,040.00	60,549,120.00
event_d7.xml	11	75	8.80	528.00	12,672.00	4,625,280.00	13,875,840.00
rail_mta_inv.xml	1	86400	0.00	0.04	1.00	365.00	1,095.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
vds_art_ladot_inv.xml	2538	86400	1.76	105.75	2,538.		00
vds_art_ladot_rt.xml	969	60	969.00	58,140.00	1,395,360.00	306,400.00	1,527,919,200.00
vds_fr_d7_inv.xml	957	86400	0.66	39.88	957.00	349,305.00	1,047,915.00
vds_fr_d7_rt.xml	361	30	722.00	43,320.00	1,039,680.00	379,483,200.00	1,138,449,600.00
Total KB from XML data	**13980**	**864660**	**6,985.28**	**419,060.38**	**10,057,449.00**	**3,670,968,885.00**	**11,012,906,655.00**

Velocity

Volume

Big Data Applications

Big Data Applications

Discover insights for science, healthcare, government, vehicle traffic control, private sectors, and international development.

- Develop intelligence for
  - making decisions
  - finding relationships
  - solving problems
  - increasing profits
  - increasing productivity
  - improving life quality

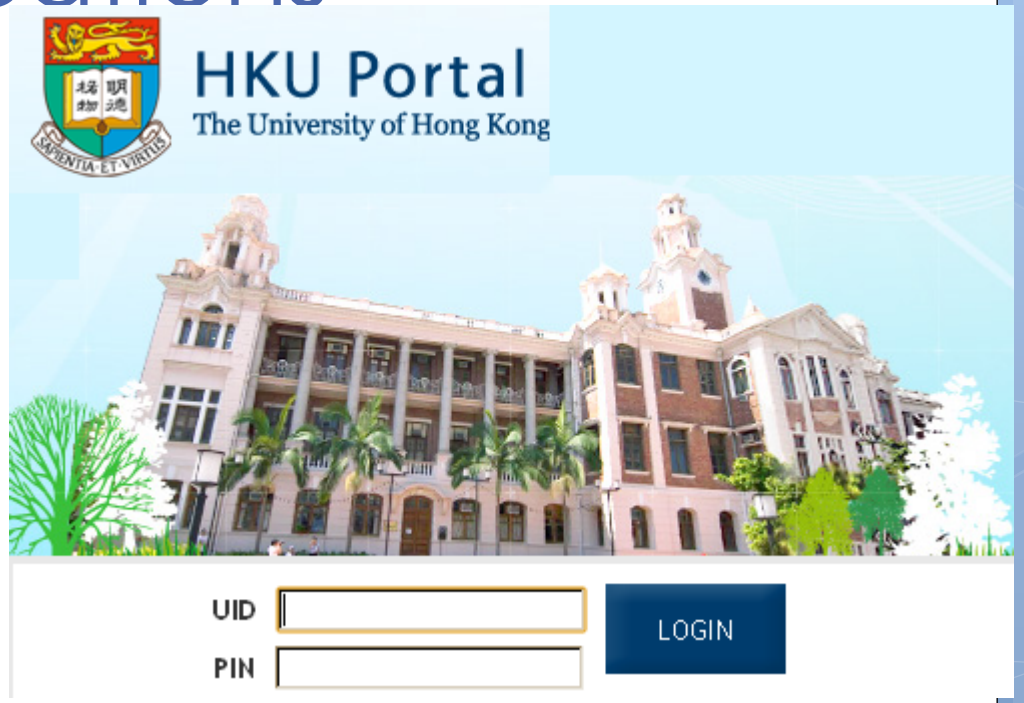
Image source: www.business2community.com

Sources of Big Data

- Social Network data
- Medical data
- Sensor data
- Location data
- Daily applications
- Web data
- Image and video data
- Crowdsourcing data
- Open data

Daily applications

- Banking
- Airline
- **Universities**



Student records

Emails

News and announcement

University fee payment

Course registration

Borrowed books

Daily applications

- Banking
- Airline
- Universities
- **Smart card systems**

Customer information | Balance | Transactions

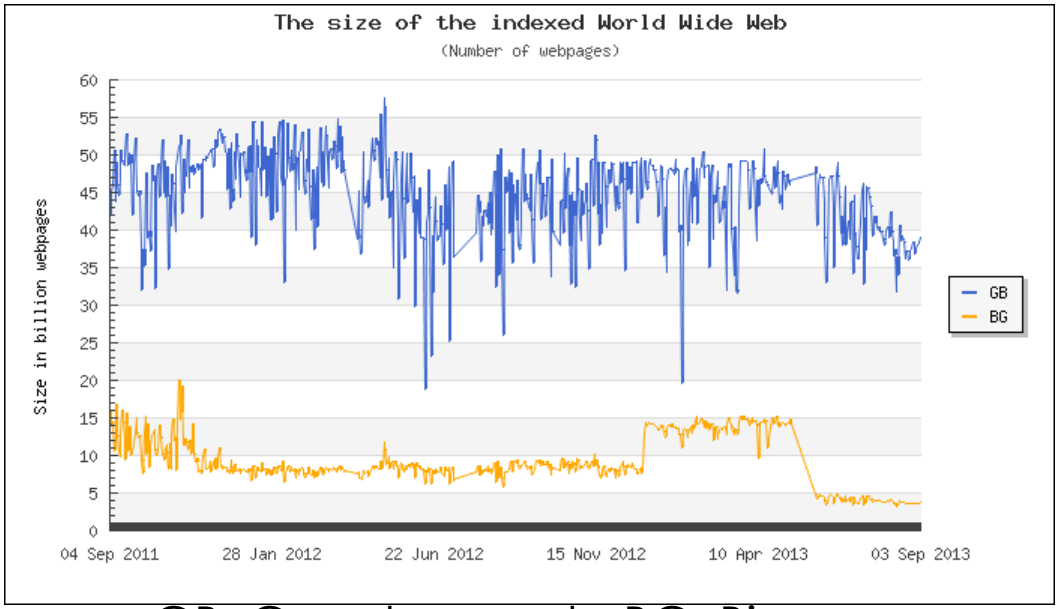
Web data [42]

- Abundance of web data for information discovery!
  - Google
  - Amazon
  - Walmart
  - United States government
- Analyze "big data" to:
  - identify patterns of behavior
  - make correlations
  - Make predictive assessments

Image source: successfulworkplace.com

Web data

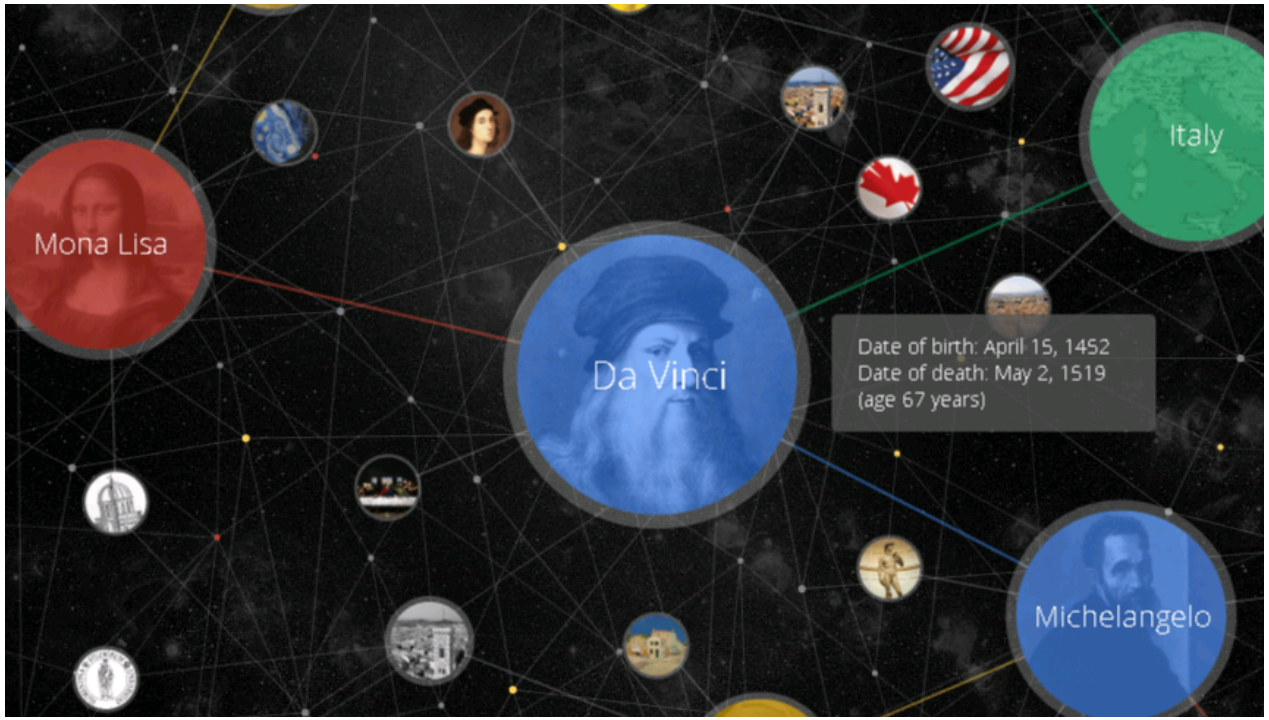
- About 40 billion web pages are indexed by Google search in Sep 2013 [2]



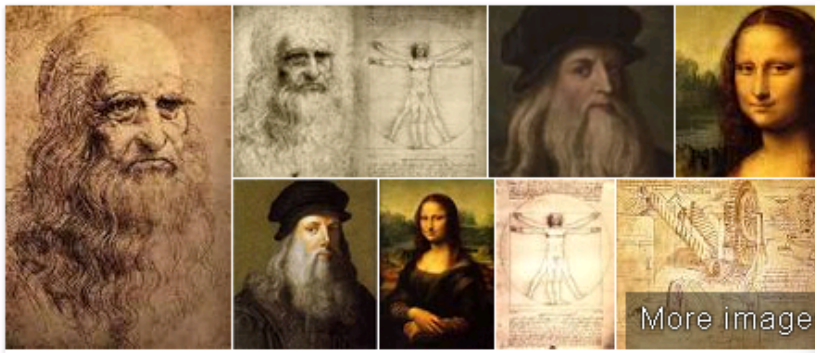
GB: Google search; BG: Bing.

Knowledge Graph: Google

- Represents knowledge as a graph



Knowledge Graph: Google



Leonardo da Vinci

Painter

Leonardo di ser Piero da Vinci was an Italian Renaissance polymath: painter, sculptor, architect, musician, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist, and writer. Wikipedia

Born: April 15, 1452, Vinci, Italy

Died: May 2, 1519, Amboise, France

Full name: Leonardo di ser Piero da Vinci

Period: High Renaissance

Buried: Chapel of Saint-Hubert

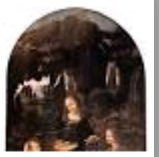
Artwork


Mona Lisa
1517


The Last Supper
1498


Vitruvian Man
1490


Lady with an Ermine
1490


Virgin of the Rocks
1486

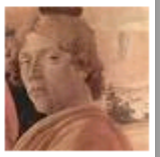
People also search for


Michelan...


Raphael


Pablo Picasso


Vincent van Gogh


Sandro Botticelli

Knowledge Graph: Yago

- Yago is a knowledge database derived from Wikipedia
- It has knowledge of more than 10 million persons, organizations, cities, and 120 million facts

http://www.mpi-inf.mpg.de/yago-naga/yago/

Google search

- By analyzing how web users search on Google, Google is able to automatically complete your search query while you are typing [5]

If some users searched "big fata" and then search "big data" right after that, Google may consider "big fata" a typing mistake of "big data"

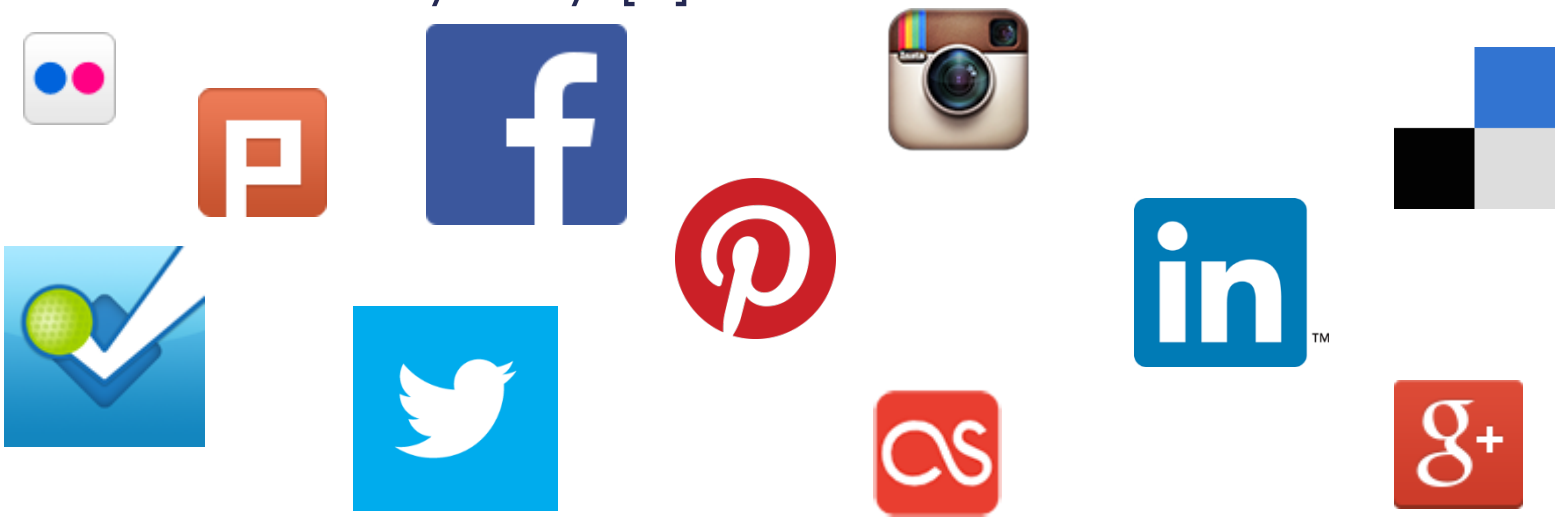
Activities in Amazon.com

- User activity such as Facebook likes can be used to enhance user experience
- By storing each customer's searches and purchases and almost any other piece of information available, Amazon is able to guess what a customer would like to buy

amazon

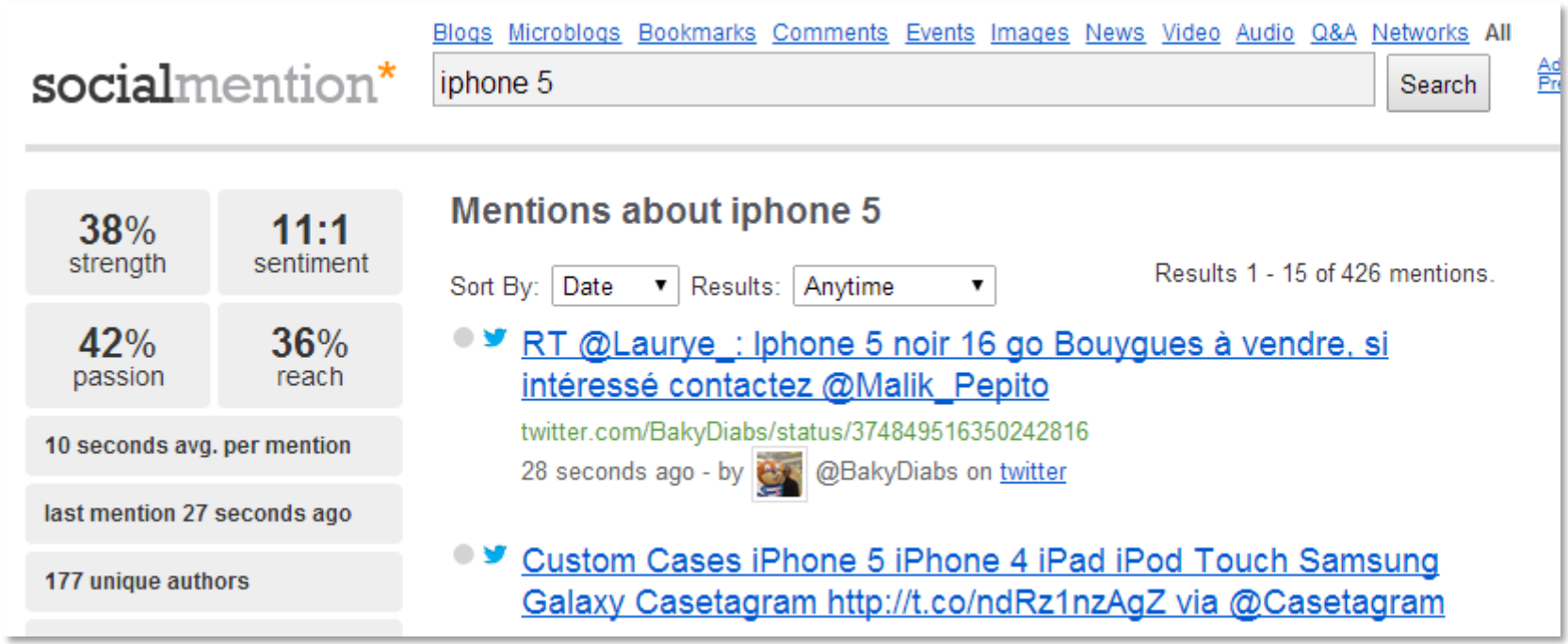
Social Network data

- An enormous number of users are providing data every second
- Facebook and Twitter processes Terabytes of data every day [3]



Brand Monitoring

- A quick search on Twitter allows company to gather feedbacks on new products shortly after release
- There are web services that allow user to monitor keywords on multiple social web sites



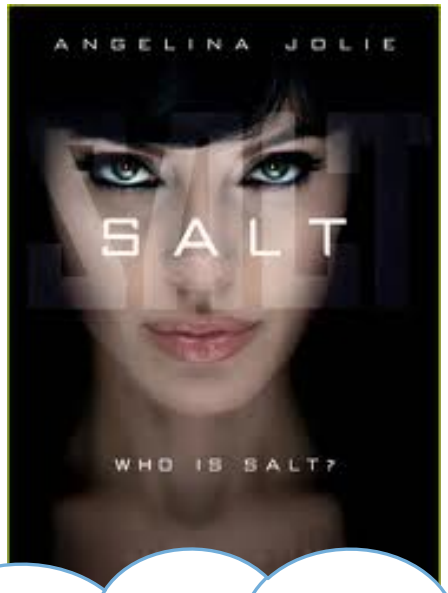
Social Genome

Walmart

Hanna

I love Salt!

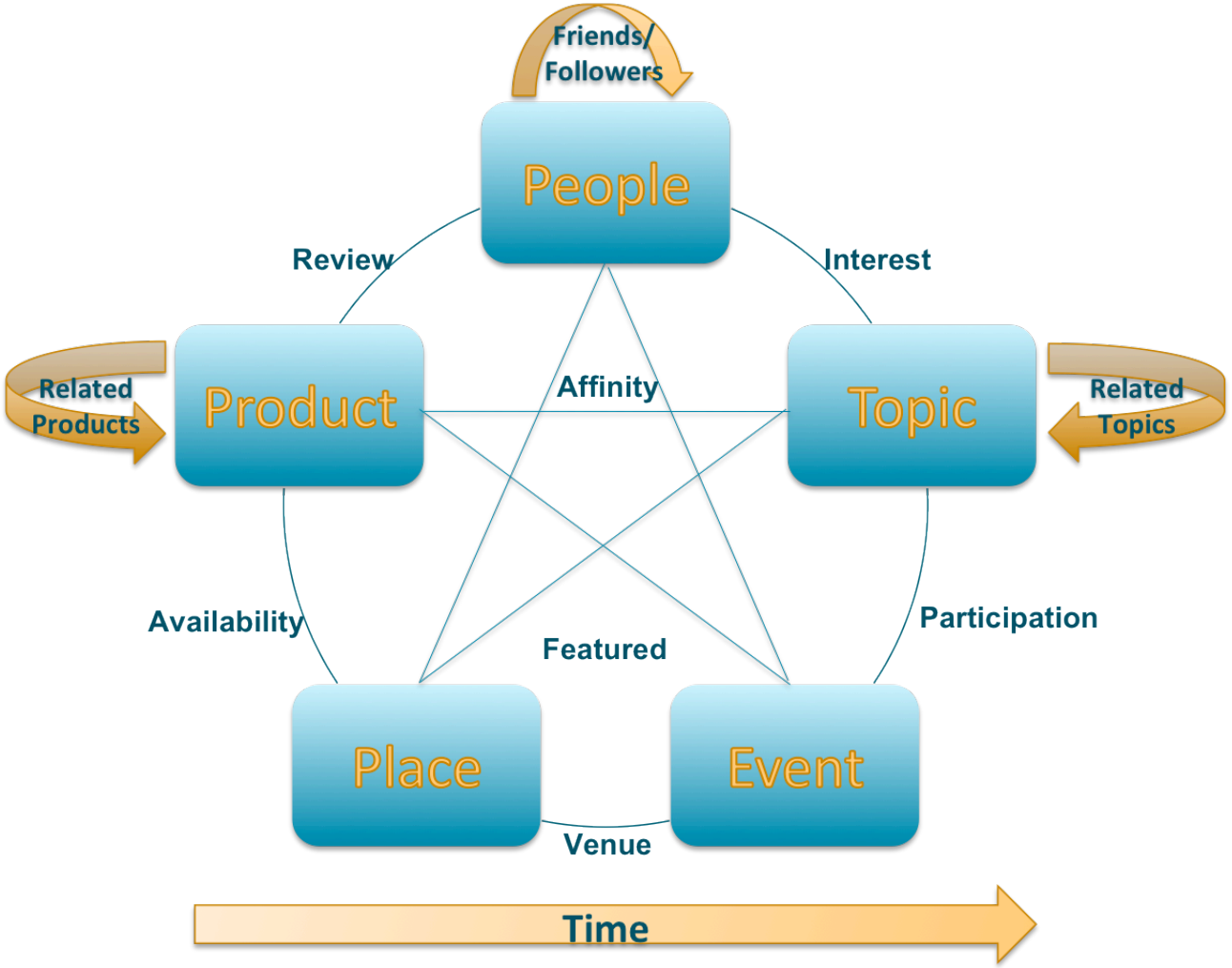
Walmart



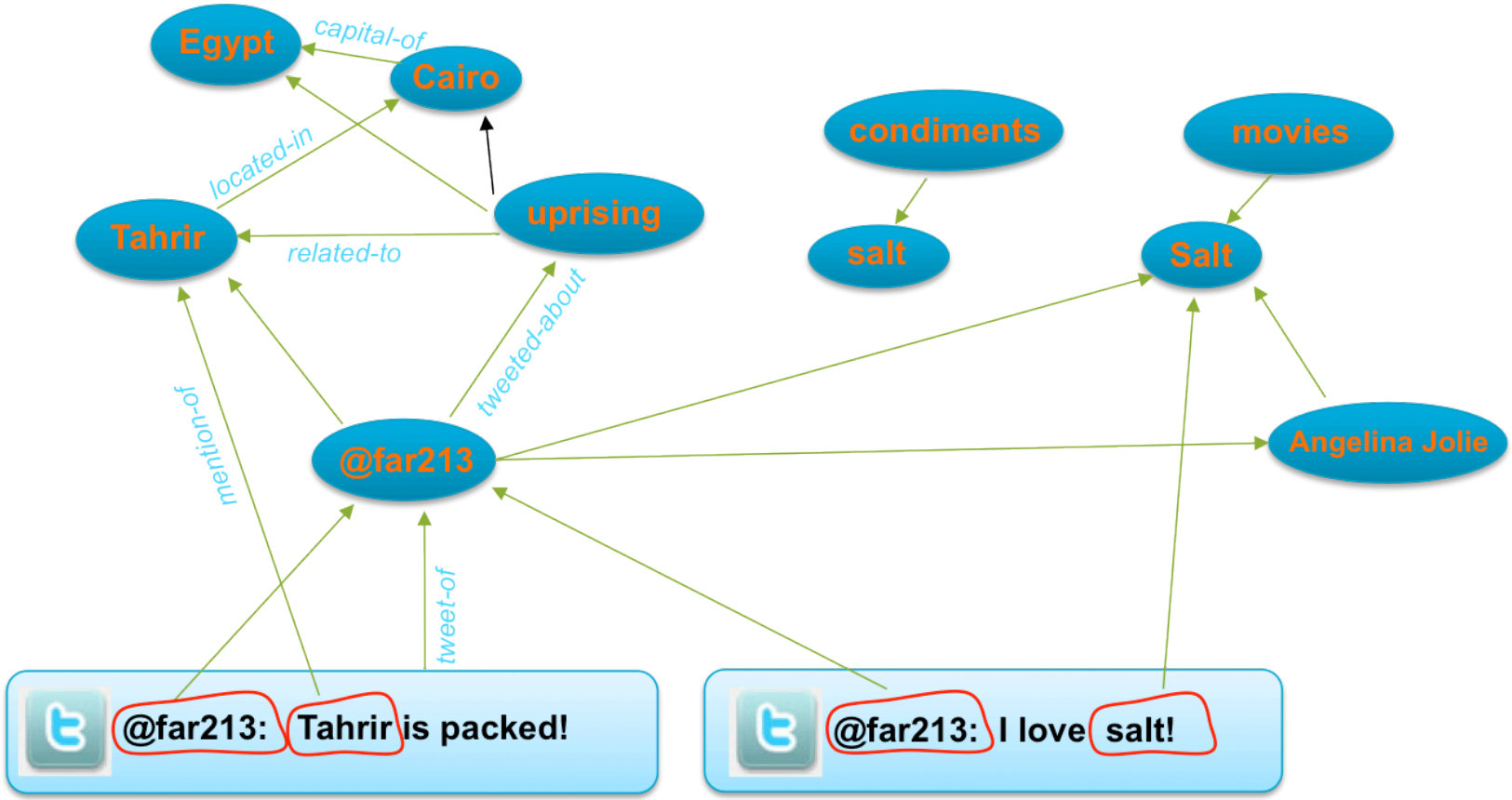
Hanna's friend

You asked us to remind that Hanna's birthday is coming up. She tweeted positively about "Salt". Do you want to buy something for her?

Social Genome: A Knowledge Base



Social Genome: A Knowledge Base



Social Genome: Examples

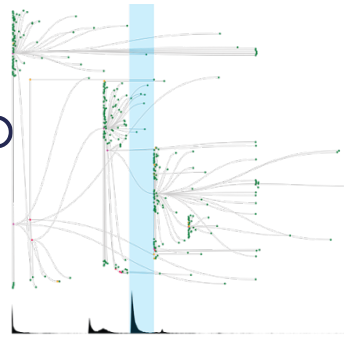
- Quickly detected that **Susan Boyle** is becoming an interesting person in social media
- Monitored social media to collect more information about her

- **Coffee machine recommendation:**
  - Based on Facebook postings on gourmet coffee
  - Recommend a new DeLonghi EC266 coffee, a 45% discount if at least 50 customers sign up

The New York Times

Text data analysis

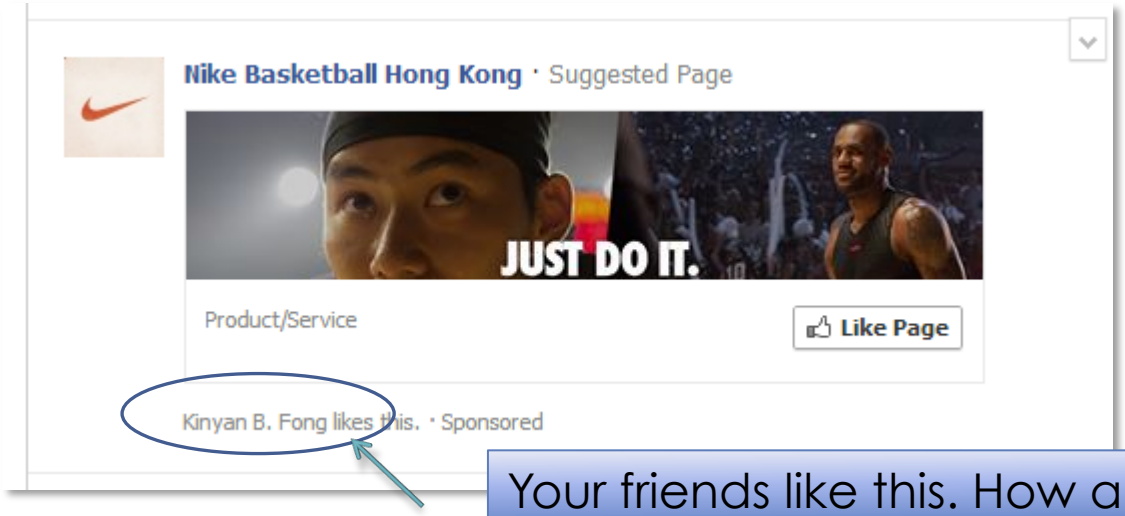
- New York Times used big data for text analysis and web mining [1].
- Cascade project use the browsing logs to understand and predict when an online conversation will result in a tidal wave of content consumption on the Times. [40]



- This helps to understand the users' needs and interests to improve their marketing strategy.

Marketing - Facebook

- Facebook makes use of friend suggestions, targeted ads and other member-focus offers.
- Information such as user's preferences, history, current activity, etc. can be used to create focused recommendations
  - quite accurate for the majority of users. [1]



Your friends like this. How about you?

Medical data

- Size of human genome is about 3GB.
- DNA sequencing is getting faster and cheaper
- Cost to map out an individual's genome: **US$1 Million** 2007
- **$1000-$4000** US dollars 2013 [41]
- More raw genomic data can be kept in repositories for research

Oxford Nanopore's disruptive MinION USB device [7]
The smallest DNA sequencing instrument as reported in 2012

Healthcare with Big Data

- "A better understanding of the relationship between treatments, outcomes, and patients will have a huge impact on the practice of medicine in the United States." [1]
- In healthcare technologies, researchers are developing advanced Big Data analytic and visualization tools to help them to find cure to cancer [8]
- Large amount of data like family histories, clinical test results and genomic data can be collected and analyzed, in a way like how Facebook and other social network analyze data [9]

Healthcare with Big Data [43]

- Personalized treatment based on their own genetic information becomes possible.
- Hospitals can make use of the big data efficiently to avoid preventable complications, e.g. blood clots and hospital re-admissions
- In November 2011
  - PatientsLikeMe: >120, 000 patients in 500 different condition groups (Currently, >220, 000 patients in 2000 different condition groups)
  - ACOR: >100, 000 patients in 127 cancer support groups
  - 23andMe: >100, 000 members in their genomic database
  - SugarStats: a diabetes health social network with >10, 000 members

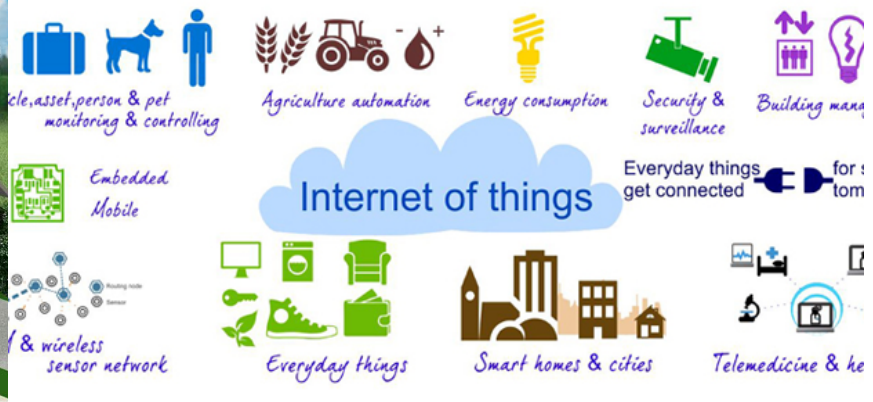
https://www.23andme.com

Image and video data

- A picture is worth a thousand words
- A large number of gigapixel images/videos are uploaded and processed on the web every day [9].
- They become very useful data when analyzed.

Tokyo Tower Gigapixel Panorama by Jeffery Martin
http://360gigapixels.com/tokyo-tower-panorama-photo/

The Internet of Things

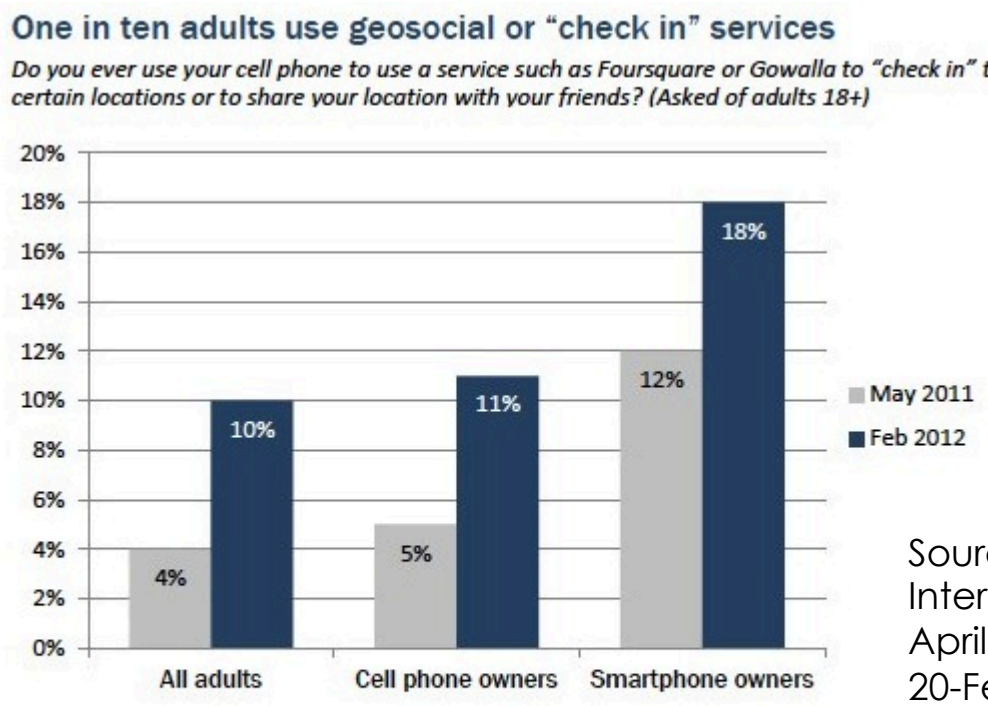


The Internet of Things

- Uniquely identifiable objects and their virtual representations in an Internet-like structure (Wikipedia)
- A sensor-detectable identifier is attached on every object.
- A large amount of related data will become available
- More than 30 billion devices will be wirelessly connected by 2020 [16]

Location data

- Check-in service usage is increasing a lot. The number has been more than doubled in 9 months in May, 2012. [12]



Source: Pew Research Center's Internet & American Life Project April 26-May 22, 2011 and January 20-Febrary 19, 2012 tracking surveys

Image source: http://marketingland.com/check-in-service-usage-has-more-than-doubled-in-past-9-months-study-says-11792

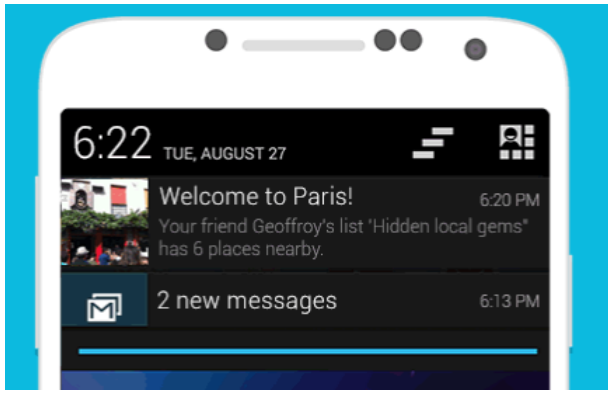
Online Maps

- Google Map allows you to Identify your location and plan routes
- Route information from other users provides suggestions for the best route

Locations | Roads | Images

Location-based services

- Map services on mobile phones are able to navigate users to their destination
- A company can send SMS promotions to subscribers when they are near to their stores [13]
- Application can suggest a dish to you when you are in a restaurant, notify you that your friends are nearby, or suggest you for places that you may want to visit [14]



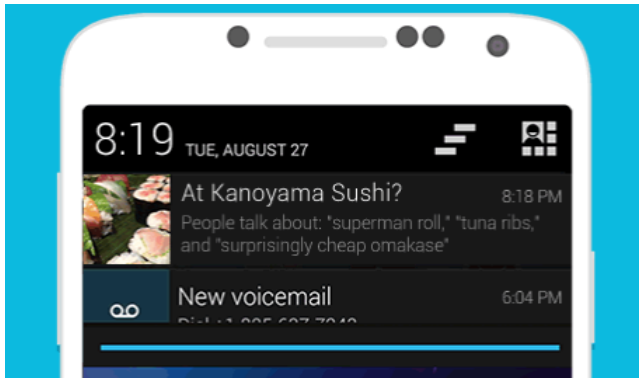


Image source: http://blog.foursquare.com/2013/08/29/a-smarter-foursquare-so-you-dont-miss-a-thing/

Location tracking

- Companies can monitor their delivery truck location with GPS installed on them
- A bus company can also publish such information for passenger on web

Real time bus location in the web site of Enshu Railway Co., Ltd, Hamamatsu, Shizuoka, Japan
http://info.entetsu.co.jp/navi/pc/location.aspx?no=15 (Only available in Japanese)

- Why do you think it is hard to find a taxi in Singapore during rainy days?

Singapore Taxis [37]

- During rainy days, hard to find taxi in Singapore
- **EXPLANATION 1:** Taxis are slow to avoid accident
- **DATA:** GPS location of taxis are often on roadside; few cars are on the road
- **EXPLANATION 2:** Higher customer need
- **DATA:** Taxi income drops significantly
- **FACT:** A Singapore law says that higher penalty is imposed for car accidents that happen in rain

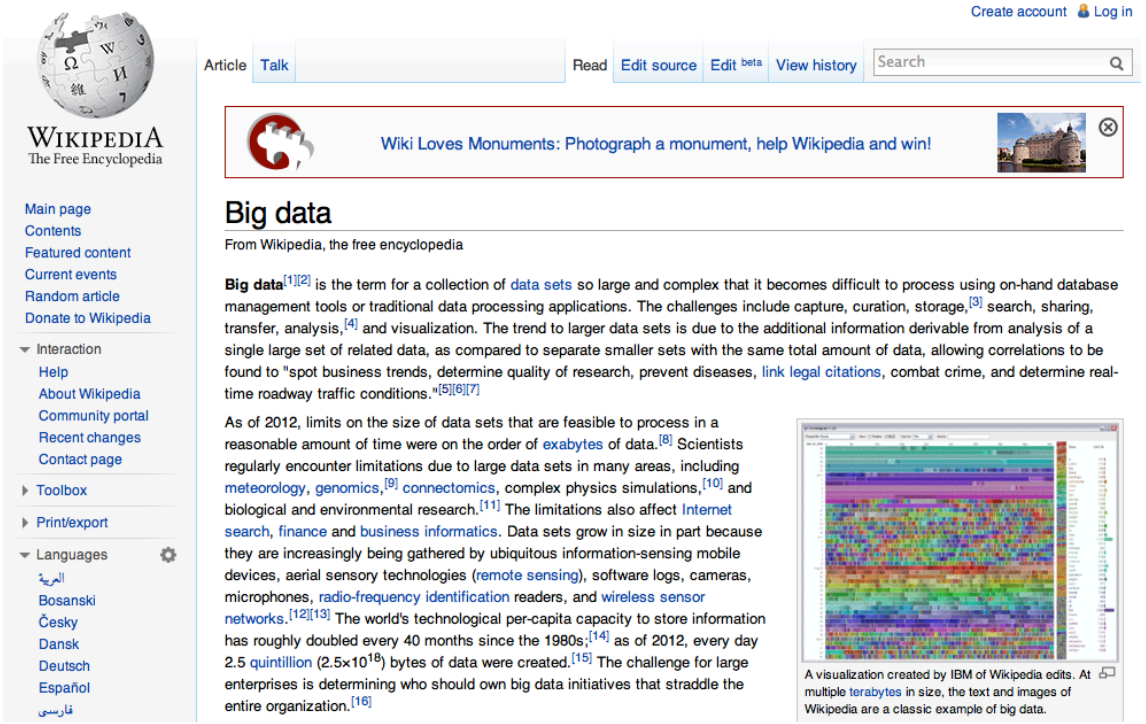
Adapted from "TransDec: A Big-Data Framework for Decision-Making in Transportation Systems" by Prof. Cyrus Shahabi

Customer data analysis

- Walt Disney Company used the big data "to correlate and understand customer behavior in all of its stores, theme parks and web properties." [1]
- Every year, there are about 100 million visitors visited Disney parks.
- Walt Disney collects visitor data by a wireless-tracking wristband "MagicBand":
  - visitors information, purchase history, location, etc.
  - Helps Walt Disney to make better decisions, to improve its offerings and tailor its marketing messages. [38]

Image source: disney.wikia.com

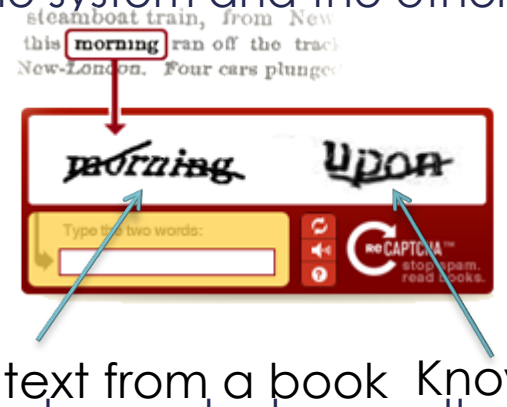
Crowdsourcing data



- Apart from collecting data actively, it is also possible to collect data from the crowd.
- Wikipedia now hosts 4 million pages [17]
- These data need to be carefully validated before used

reCAPTCHA

- reCAPTCHA uses CAPTCHA to help digitizing text of books
- Idea: present images of two words, one is already recognized by the system and the other is the word to be digitized



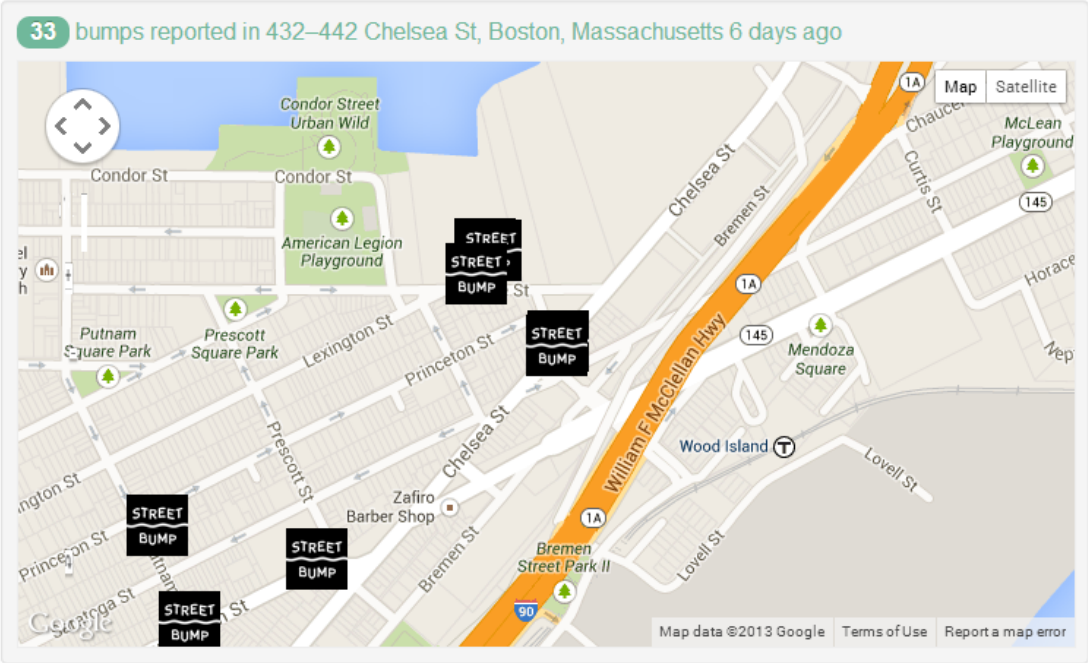
Unknown text from a book Known text

- If user answers the known text correctly, their answers to the unknown text will be kept as reference [18]

Image source: http://www.google.com/recaptcha

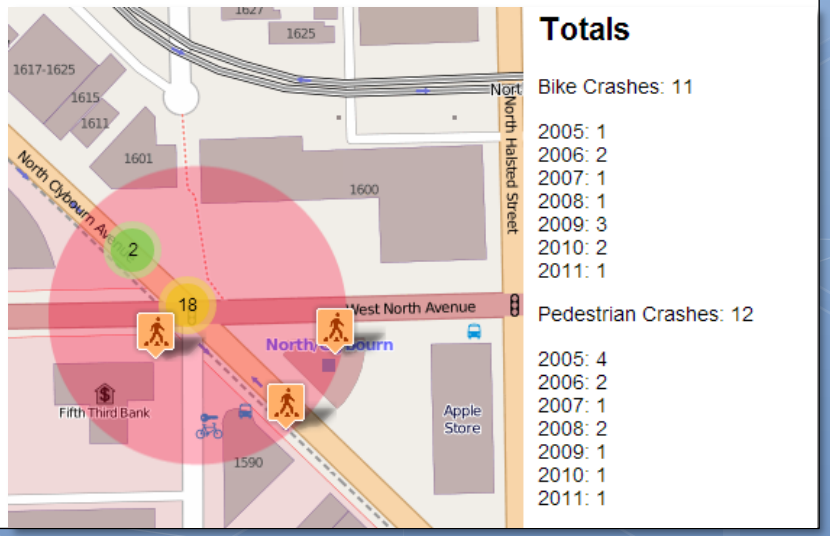
Street Bump

- Street Bump is a crowd-sourcing project that collects data about the smoothness of a user's ride in the city of Boston.
- Street bump is identified and is used to find and fill potholes quickly.



Example: Chicago Crash Browser

- Using data from the transportation department, the website shows a map with the no. of automobile-pedestrian and automobile-bicycle crashes for any point in Chicago [25]



Google Person Finder

- As part of Google Crisis Response, Google Person Finder will be deployed when right after any serious disasters [23]
  - Used in the Boston Marathon bombings on April 15, 2013 and China earthquake on April 20, 2013
- Survivors, family and friends can use the interface to post/search for information about each other
- In 2011 Tohoku earthquake and tsunami, Person Finder has collected 616,300 records [24]

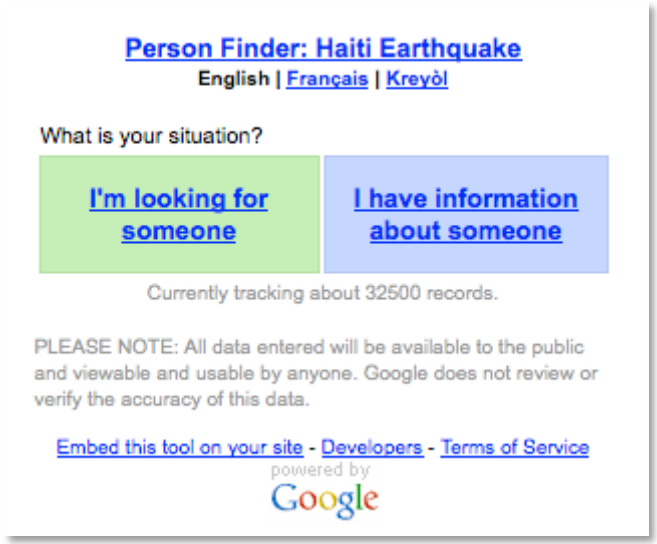


Image source: http://en.wikipedia.org/wiki/Google_Person_Finder

Example: DontEat.at

Social Impact of Big Data

Big Data on Social Network

Image source: sisobproject.wordpress.com

Big Data: Impact on society

Image source: www.ohmygeek.net , www.buzzfeed.com

Big Data: Good or Bad?

- The above network data analysis studies relationship-based data
  - Facebook has 140 billion friendships
- Edward Snowden's story
- Use of social media to support terrorism

Is Big Data Good or Bad?

Is Big Data Good or Bad?

Government officials were having an important meeting…

game. Share pictures of Kenji doing his job. #KenjiGoto #iamkenji #ISIS #IS

http://news.mingpao.com/ins/人大政協開會玩手機睡覺%20網民祝做好中國夢/web_tc/article/20150203/s00004/1422946456103

http://news.mingpao.com/ins/bbc記者：勿分享健二斬首片%20莫被isis牽着走%20促分享工作照-倡傳媒反思應否報道isis/web_tc/article/20150203/s00005/1422949068281

Are people rude when they use Social Media?

- An online survey of 2,698 people by corporate training firm VitalSmarts in February 2013:


3 in 4 have WITNESSED AN ARGUMENT ON SOCIAL MEDIA
4 in 5 report RISING INCIVILITY online
2 in 5 have BLOCKED UNSUBSCRIBED OR UNFRIENDED someone as a result
1 in 5 reduced in-person contact with someone over a cyber argument

Spread of Emotions

- Which of the following emotions do you think spread the fastest and widest in social networks?

Anger

Joy

Sadness

Disgust

Spread of Emotions

- A study on China's Twitter-like micro-blogging site Sina Weibo studies 70 million tweets from 200k users
- Anger spreads faster and wider than joy, sadness, and disgust!

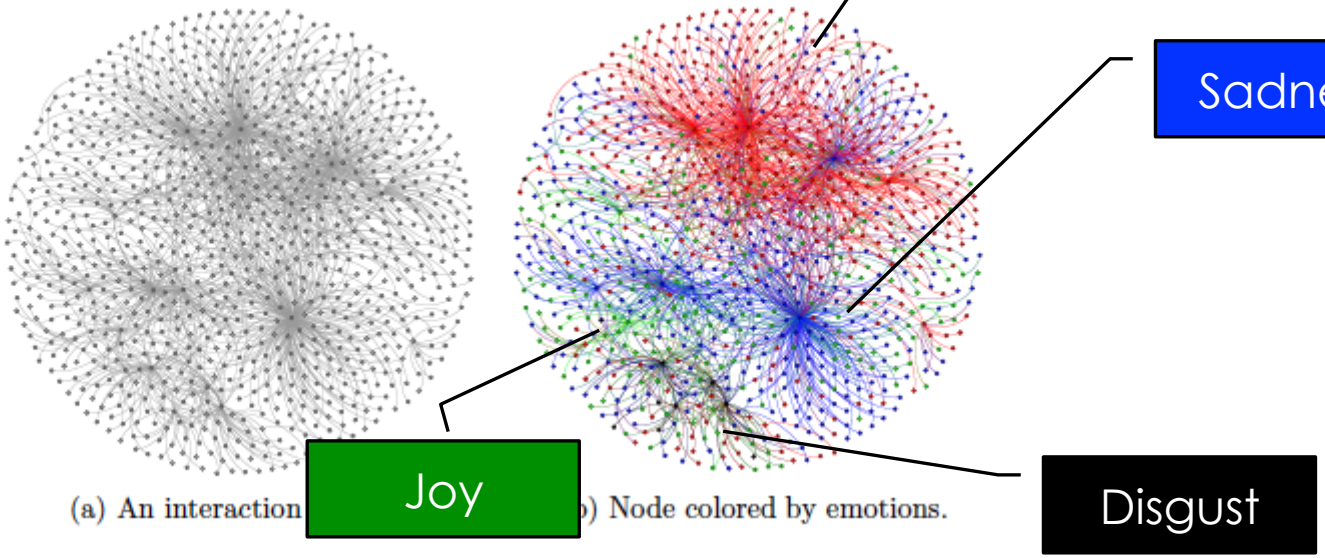
Anger

Joy

Sadness

Disgust

Spread of Emotions



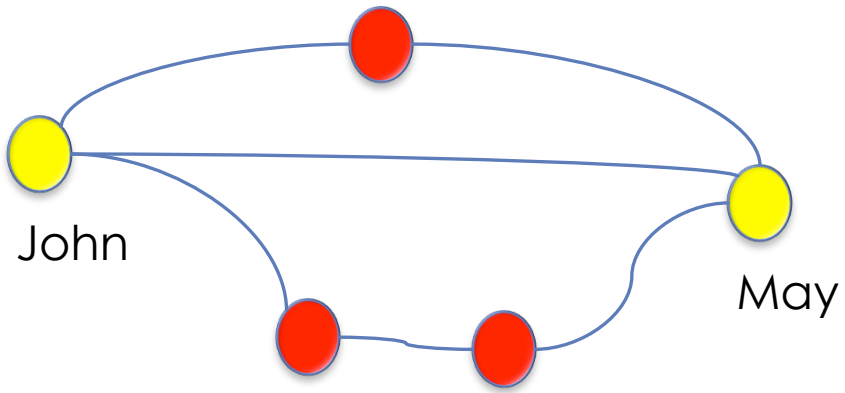
(a) An interaction) Node colored by emotions.

Spread of sentiments

- Angry emotion could spread more quickly and broadly in the network
- The correlation of sadness is surprisingly low and highly fluctuated.

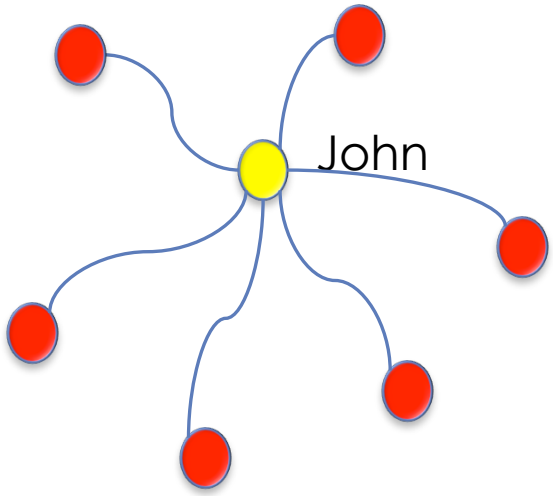
Spread of sentiments

- There is a stronger "sentiment correlation" between two users if they share more interactions.



Spread of sentiments

- Users with larger number of friends posses more significant sentiment influence to their neighborhoods."



Are people rude when they use Social Media?

- Do you think people are less Polite on social media than in person?
- How often did your emotional conversations held on social media have been resolved?


88%
believe that PEOPLE ARE LESS POLITE
on social media than in person
81%
say emotional conversations held on social media are
most often
UNRESOLVED

VitalSmarts Survey Observations

- Social media platforms have become the default forums for:
  - holding high-stakes conversations
  - blasting polarizing opinions

VitalSmarts Survey Observations

- The public forum that allows no immediate feedback or the opportunity to see how our words will affect others
- People using social media platforms can cause a degradation of dialogue that has potential to destroy our most meaningful personal relationships

Fast Spreading of Messages

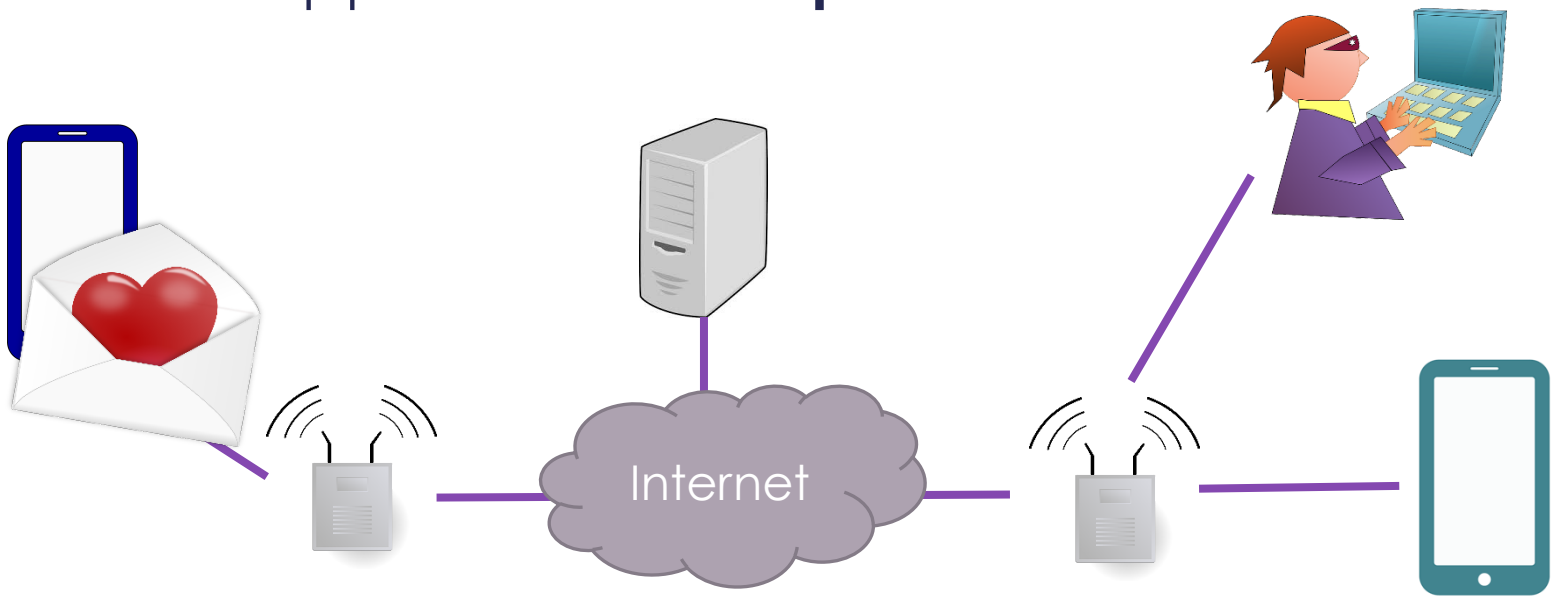
- "Using Twitter and Facebook could harm moral values, as they don't allow time for compassion or admiration."

- "If things are happening too fast, you may not ever fully experience emotions about other people's psychological states and that would have implications for your morality."

  -- Mary Helen Immordino-Yang, the University of Southern California

Security & Data Protection

- Until August 2012, messages sent via Whatsapp were sent as **plaintext**



Privacy

Image source: www.makemeentertain.com

Forbes / Investing

The Little Black Book of Billionaire Secrets

JAN 10, 2014 @ 08:56 AM 37,472 VIEWS

Target Data Breach Spilled Info On As Many As 70 Million Customers

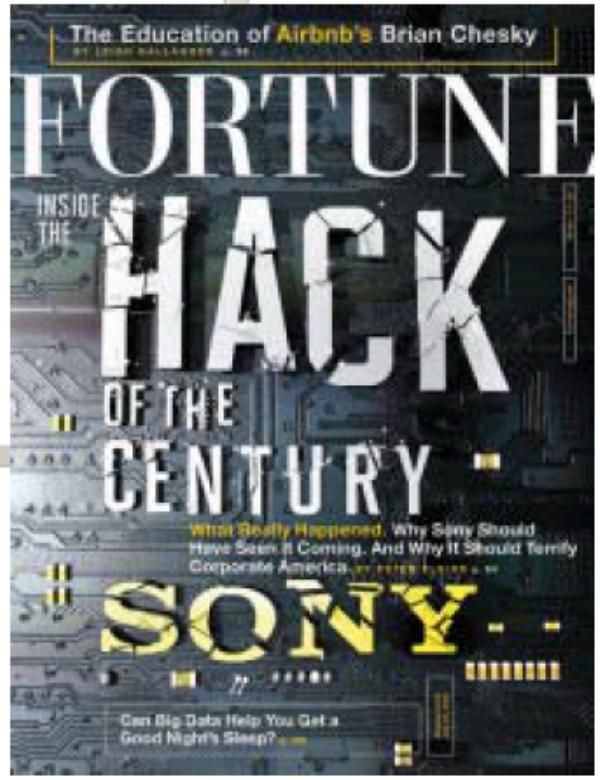
Maggie McGrath
FORBES STAFF

Got one eye on the markets, the other on Gen Y's pressing $$ issues

FOLLOW ON FORBES (744)

FULL BIO >

The data breach th
nightmare before Christmas for Target TGT -1.71% and
millions of customers just got a little bit worse: the
said Friday morning that the information stolen between
November 27 and December 15, 2013 included personal
information of as many as 70 million people — more than the
40 million the company originally estimated.



Acknowledgement: Lily Chan, Director, Privacy, Data Protection & Cybersecurity, Microsoft

The surprise of big data analytics – Target's pregnancy prediction

If it works in this way…

Acknowledgement: Henry Chang, Law and Tech Centre, HKU

Privacy

- Is it **ethical** for Google to analyze your web activity, including your email, to know that you have bought a ticket?
- **Leakage** of private data, due to the use of the Foxy software, and the loss of USB drives that contain thousands of patients' records, have also raised serious legal and social concerns.
- Can **sensitive** data be released to public?

Image source: http://www.tnooz.com/2013/04/27/news/for-opprtunities-as-a-startup-look-no-further-than-big-data/

Search can be offensive

- A man in Japan sued Google of autocomplete for suggesting a connection to crimes when his name is entered.
  - Court ordered Google to pay 300,000 yen to the plaintiff
- A German federal court asked Google to ensure terms generated by auto-complete are not offensive or defamatory
- A HK tycoon is suing Google after search results linked him to web contents that he claims is defamatory

Search can breach privacy

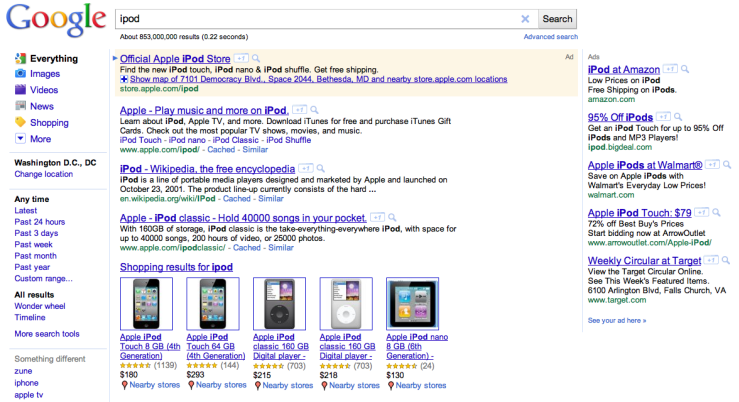
- Is it ethical to search over a person, and announce his/her name?
- Hong Kong Bride-to-be refuses gift money less than $ 500
  - Netizens have **digged out** and publicized the girl's photos, job, phone numbers, and pictures of her and her ex-boyfriends.
  - http://hk-magazine.com/city-living/news/bride-be-refuses-gift-money-lower-500

Paid Inclusion

- In 2004, many search engines (e.g., MSN, Yahoo) display links obtained through **paid inclusion**
- Web publishers pay to have their sites indexed and frequently refreshed
- Microsoft has removed this feature from their search engines
- http://www.wired.com/techbiz/media/news/2004/07/64092

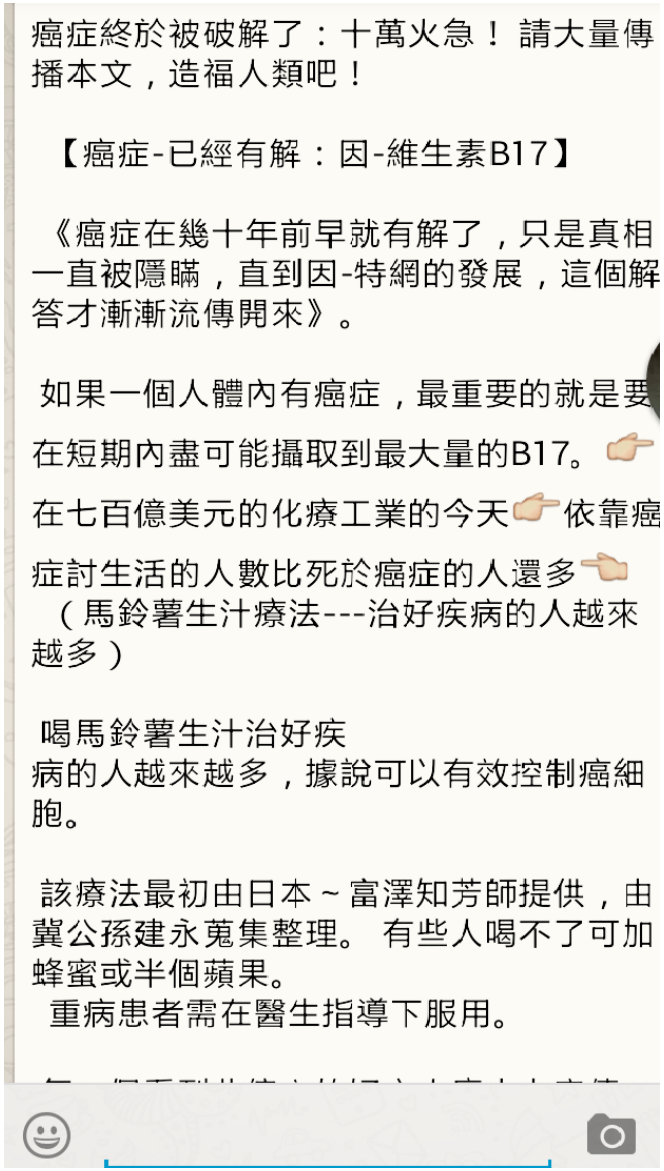
Search and ads

- An Internet search can return information that you may not want
- e.g., Google post advertisement beside search results
  - https://www.google.com/competition/howgoogleadswork.html#section2



Credibility

- Have you ever received false messages from friends?
- If we believe in rumors, false decisions can be made



Spread of False Contents

- News, blogs, and twitter messages posted by celebrities often contain false, irresponsible or even hurtful contents
- Causes emotional disturbance to the authors of the original messages!
- They also impact traditional social norms, and spread bad influence to teenagers

A circulated photograph after Boston bombing event: The girl is not dead and the man is not his boyfriend [27]

Minority Report, 2002

Right to be forgotten

- In May 2014, the European Court of Justice ruled against Google, on a case brought by a Spanish citizen who requested Google to remove a link to an article dated 1998 related to his foreclosed home

Source: http://en.wikipedia.org/wiki/Right_to_be_forgotten

Right to be forgotten

- Google received **12,000 requests** on the first day of compliance to remove link to personal details from its search results performed based on a person's name
- In 2014, **190,000 requests** are received

Response from HK Privacy Commissioner for Personal Data

- No absolute "right to be forgotten"
- Guidelines articulating 13 criteria to be considered when dealing with de-list requests, including:
  - whether the person plays a role in public life
  - whether public access to that information will protect them against his public or professional improper conduct;

Source: http://www.pcpd.org.hk/english/news_events/commissioners_message/blog_30122014.html

Should the following request be approved?

- a victim of physical violence wanted references to the assault removed from web searches of his/her name;
- a victim of rape requested removal of a link to a newspaper article about the crime;
- a girl requested removal of a link to explicit photographs of her taken by her ex-boyfriend;

Should the following request be approved?

- a sex offender who wanted recent information about his conviction de-linked;
- a person made multiple requests to remove 20 links to recent articles about his arrest for financial crimes committed in a professional capacity; and
- a public official requested removal of a link to a student organisation's petition demanding his removal.

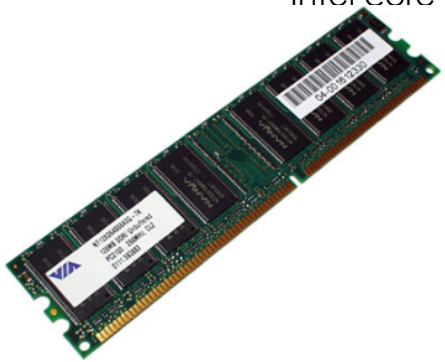
Organizing Big Data

Organizing and exploring Big Data

- How does Whatsapp organize data?
  - Users
  - Groups
  - Messages
  - Attachments
- Are cloud computing technologies adequate?

We are contributors of Big Data!

Intel core i7 CPU



A DDR3 RAM

Acknowledgement: Lily Chan, Director, Privacy, Data Protection & Cybersecurity, Microsoft

Data center

- Big data applications need storage of Petabytes of data (1 Petabyte = 1000 1TB hard disks)
- Data are kept in a **data center**, accessed through the Internet

Data Centers in Hong Kong

- Store Big Data, especially in Asian regions
- Need huge land space (170, 000 sq meters by 2015)
  - 17.8 hectares (h) in Tseng Kwan O
  - Hong Kong Science & Technology Parks

- Power consumption can be about 10-15 times that of a residential building
- Facilities: transformer rooms, UPS (uninterruptible power supply), backup power generators, air-conditioning systems, fire safety

Source: Mr. Charles Mok

Energy Consumption of Search

- A single Google Search was claimed to use the same amount of energy for boiling a kettle of water!
- Google's data center consumes 2 billion kWh per year
  - 0.1% of US total electricity consumption
  - Enough to power 200,000 homes a year

Green Data Centers

- Data center facilities account for 1.5% of world's total energy use in 2010.
- Promote renewable energy
  - Rather than relying on coal, nuclear, natural gas
- Energy saving measures in:
  - Construction and operation
  - Optimization of air flow and chiller systems
  - Use of central water cooling system

http://en.wikipedia.org/wiki/Green_computing

Google's water-based data center (rumor)

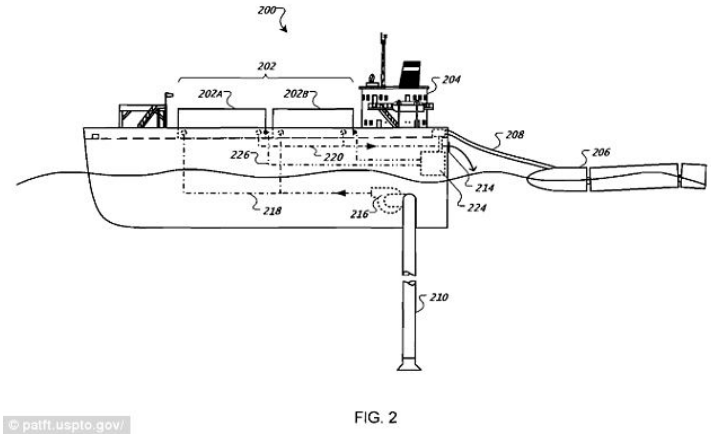
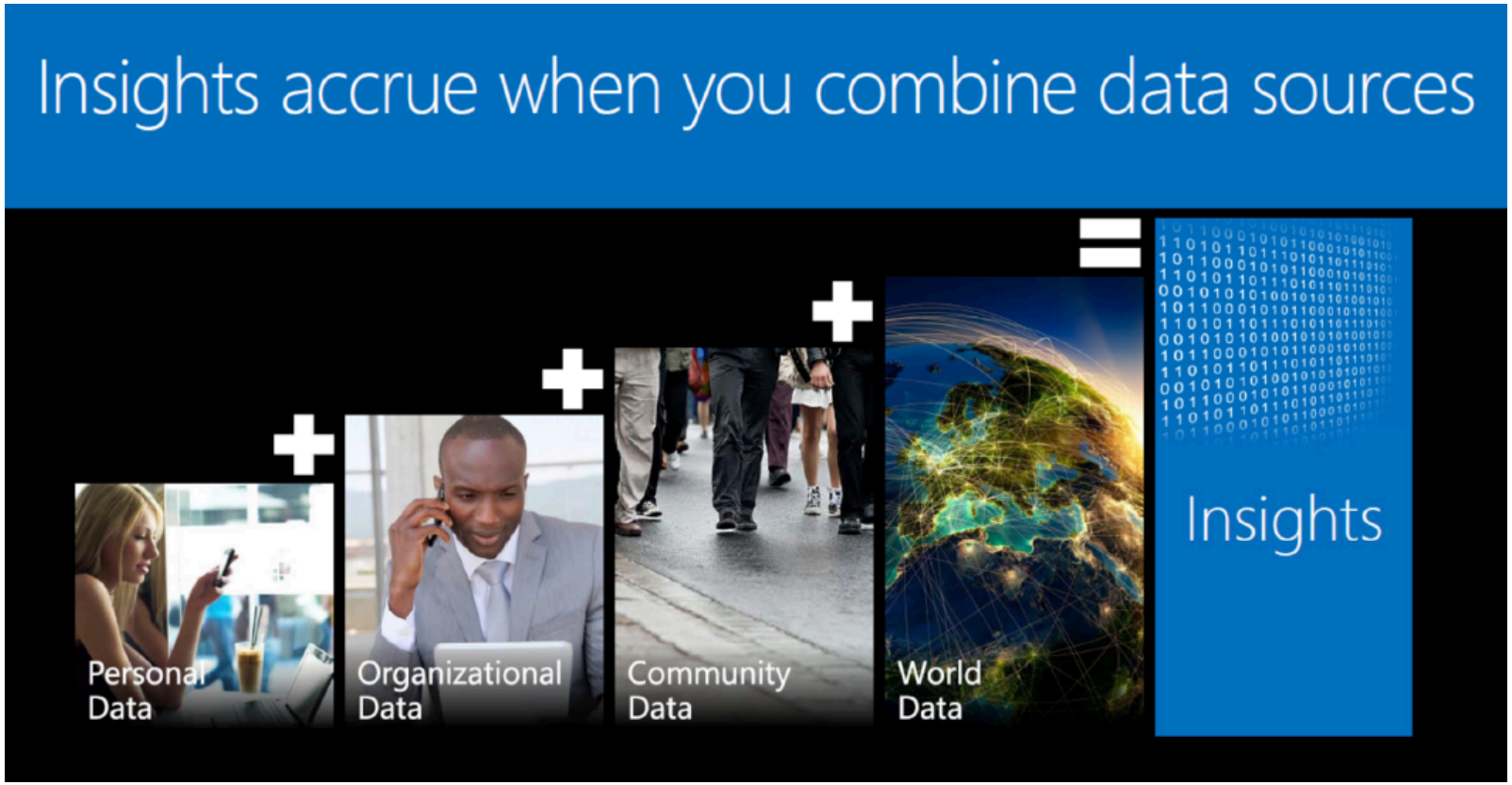


FIG. 2

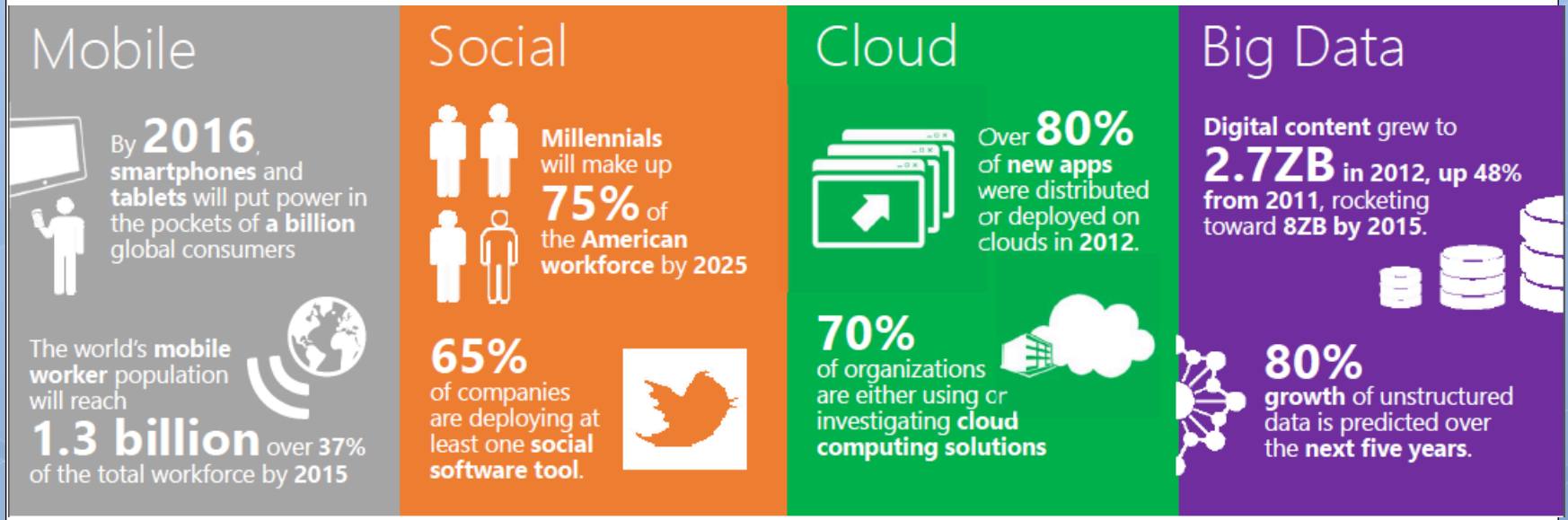
Why do they build the data center on water?

Source:
http://www.dailymail.co.uk/news/article-2477580/Googles-secret-water-based-data-center-Technology-giant-builds-huge-mystery-vessel-San-Francisco-Bay.html

Conclusions



Acknowledgement: Lily Chan, Director, Privacy, Data Protection & Cybersecurity, Microsoft

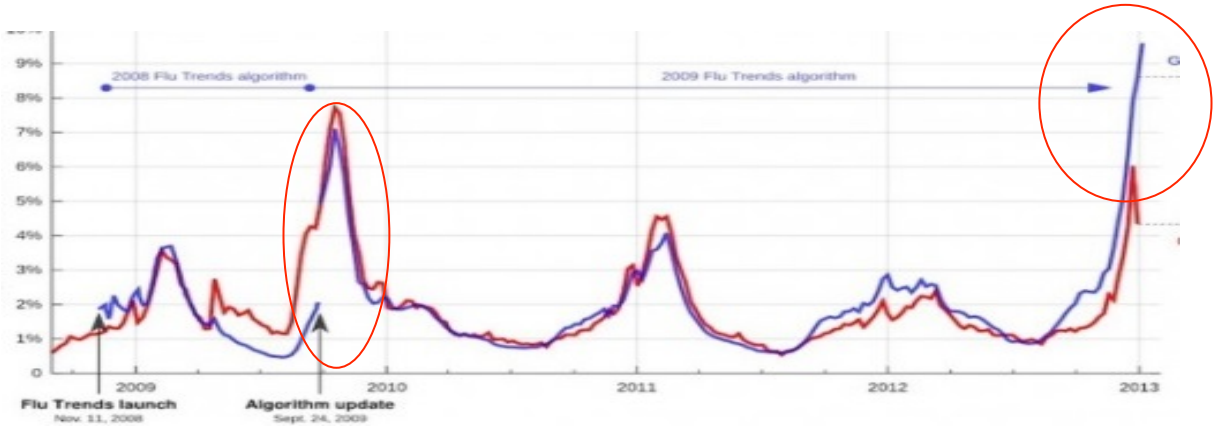


Acknowledgement: Lily Chan, Director, Privacy, Data Protection & Cybersecurity, Microsoft

Failures of big data analytics – Google flu prediction

It does not always work…

- Underestimated by half in 2009 when comparing with CDC data
- Overestimated by half in 2012 when comparing with CDC data
- Predictor of flu or predictor of winter?
- A black-box approach makes it hard for people to judge (EPIC president calls for the need of "algorithm transparency")

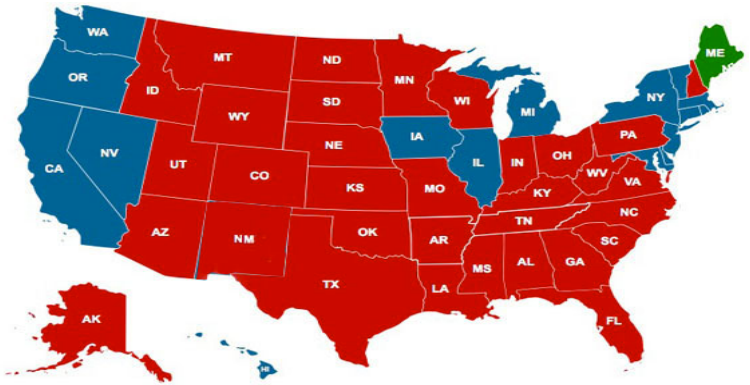


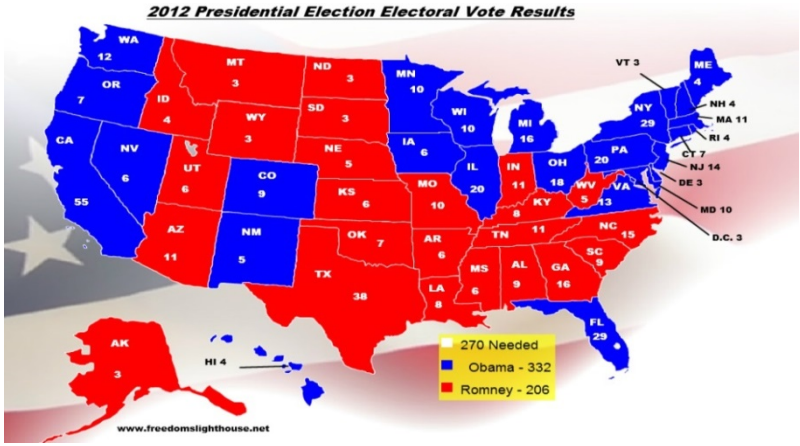
Acknowledgement: Henry Chang, Law and Tech Centre, HKU

Failures of big data analytics – US presidential election

"Past performance does not guarantee future results…"

- In 2012, Colorado professors built a data model that correctly "backward predicted" the eight US presidential election results since 1980
- But it failed to forward predict the 2012 election…



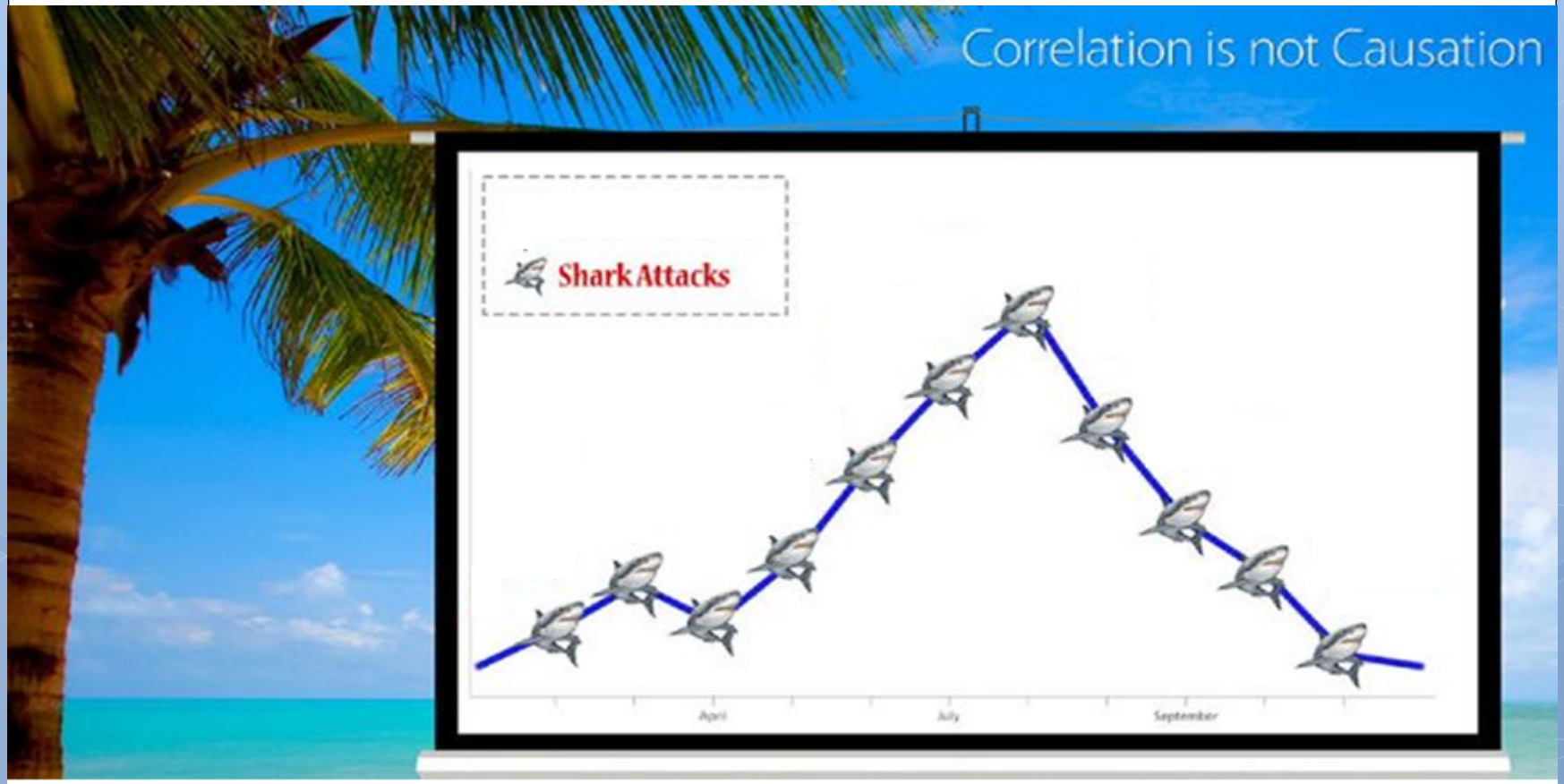


Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The reality of big data analytics

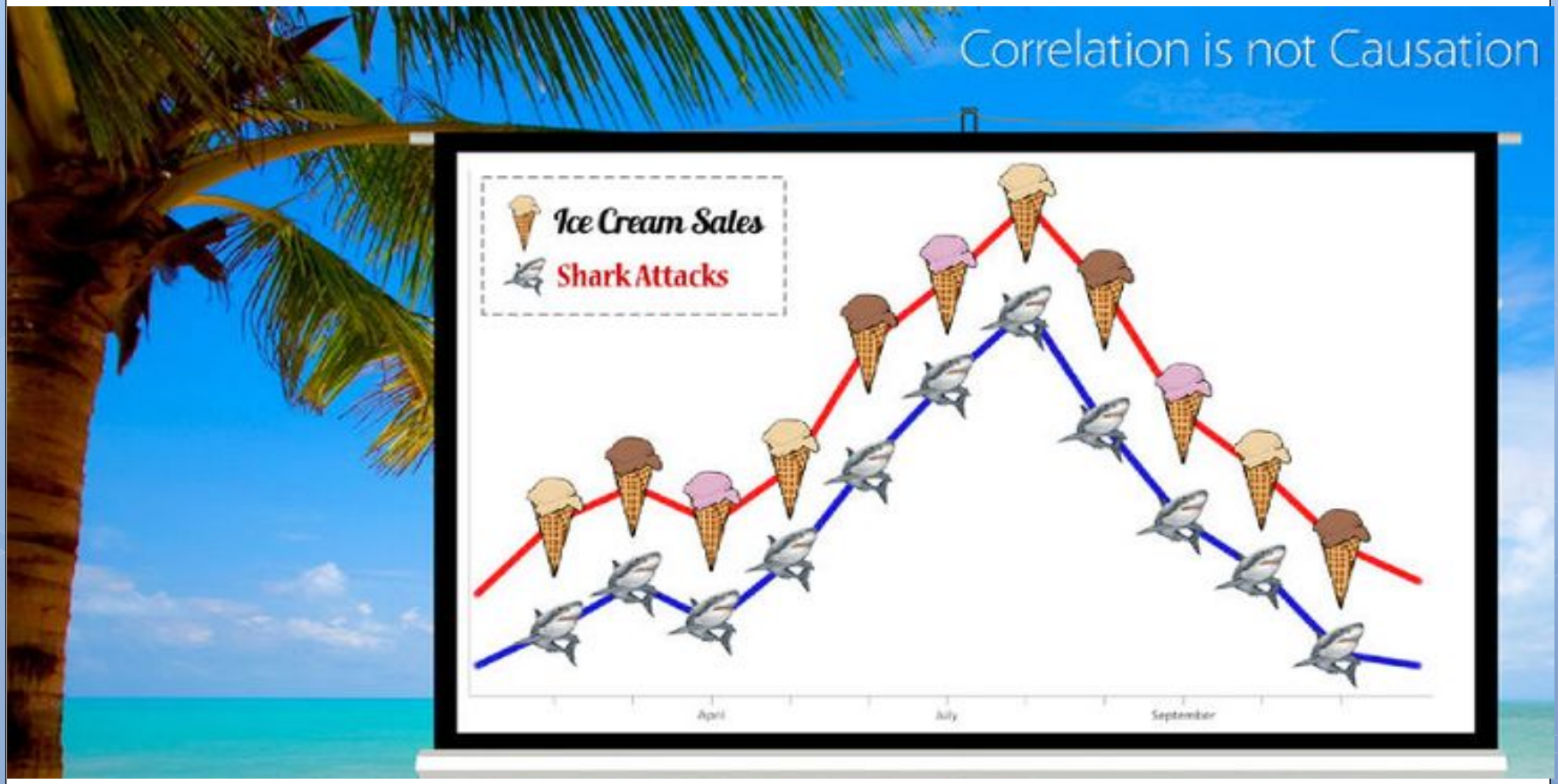
Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The reality of big data analytics



Acknowledgement: Henry Chang, Law and Tech Centre, HKU

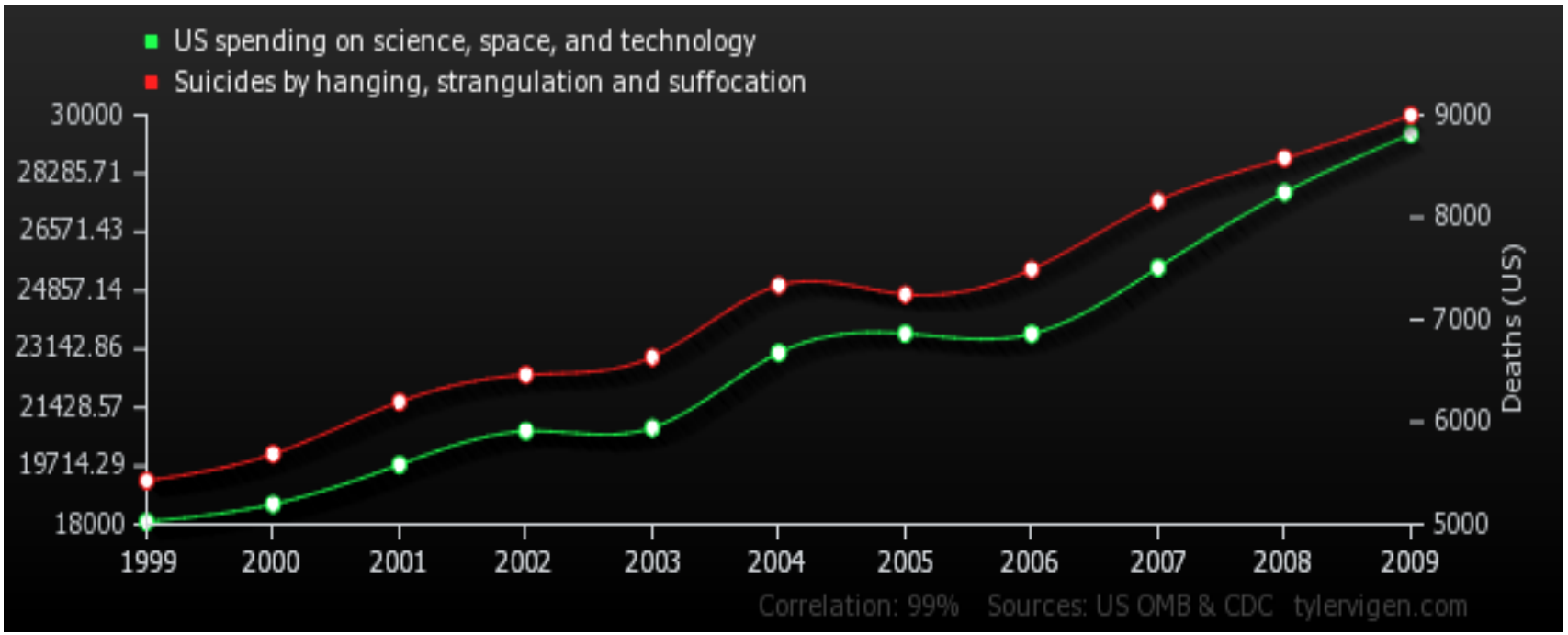
The reality of big data analytics



Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The (academic) reality of big data analytics

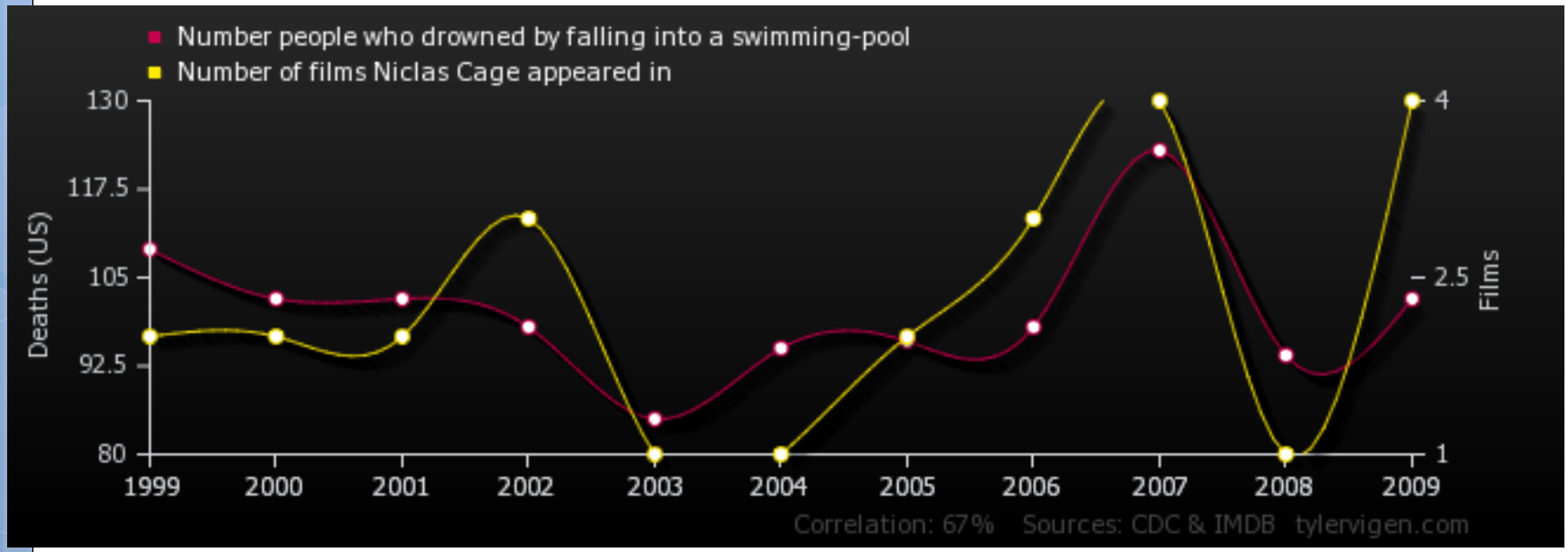
US spending on science, space, and technology reveals ***Suicides by hanging, strangulation and suffocation***?



Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The (academic) reality of big data analytics

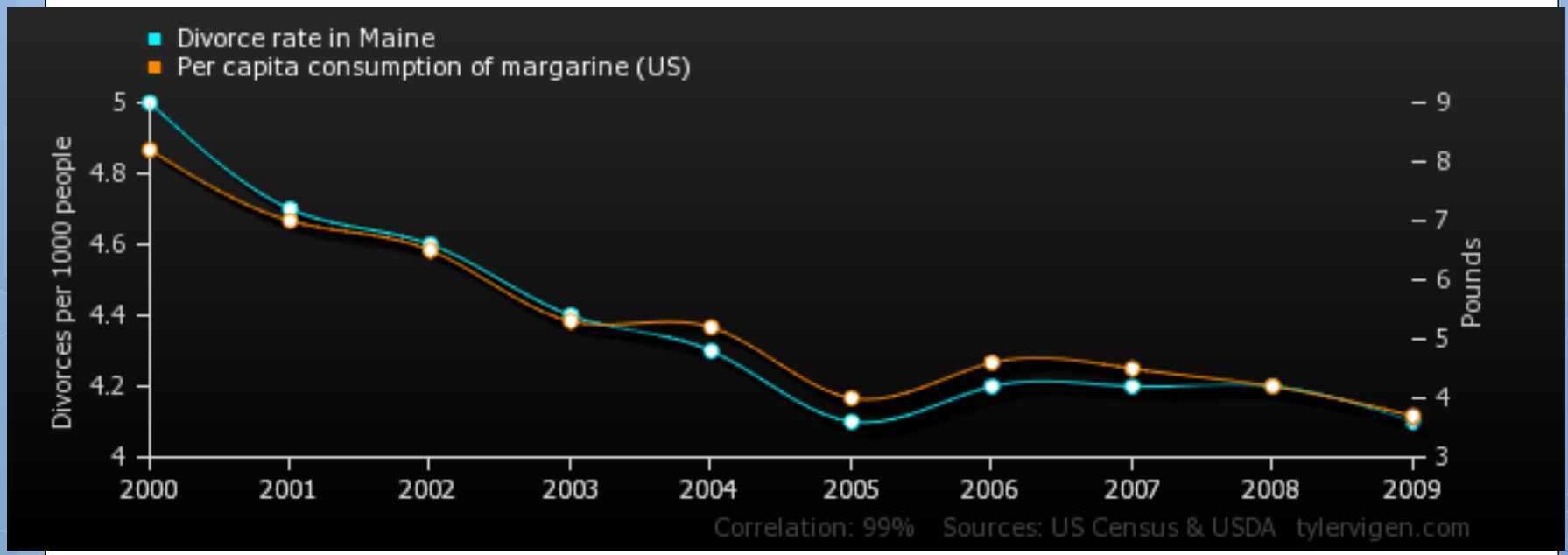
Number of Nicolas Cage films reveals ***Swimming-pool drowning*?**



Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The (academic) reality of big data analytics

Divorce rate in Maine reveals ***Per capita consumption of margarine*** ?



Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The reality of big data analytics

Big data analytics:

- ✓ Correlation
- ✗ Causation

Acknowledgement: Henry Chang, Law and Tech Centre, HKU

The reality of big data analytics

- You are a meat-lover.
- Your car insurance fee is increased because the increase company found that meat-lover tends to issue more insurance claim.
- Are you happy?

Big Data Projects

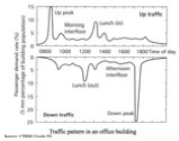
- Propose a new application of Big Data useful to Hong Kong
- Perform an analysis on
  - the existing technology or software implemented in other places in the world, their drawbacks
  - How your solution can improve the situation
  - How your solution should be implemented in Hong Kong

Big Data Projects

- Discuss, in various aspects about or not your application of Big Data:
  1. Is it making our life better?
  2. What is its impact on moral and social values?

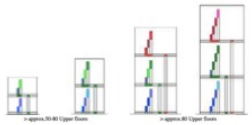
Wait or Walk,
No Longer a Question

Big Data Collection & Theoretical Background



Elevator user behavior pattern could be observed and "predicted"

Collect Data from the human behaviour and elevator system calculation



Building a Big data Repository with simulation.

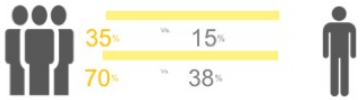
Visualising the Building

Application

Calculate the possible waiting times and give the suggestion

Update with the Portable App and Calculate the Calorie and Electronic Saving

Benefit



Energy Saving & Build your body shape

Share your smart decision with your friends!

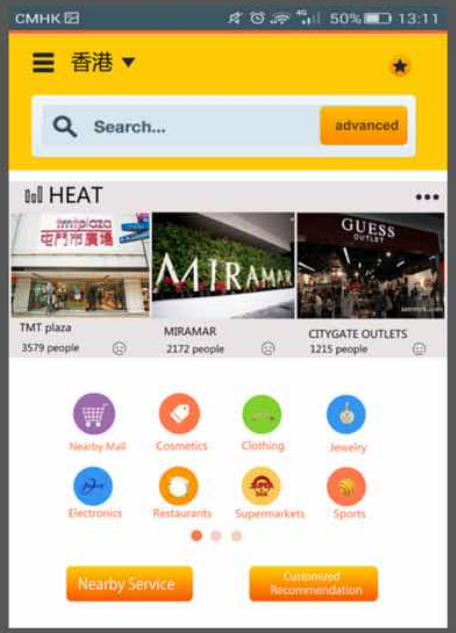
HK Comfy Shopping

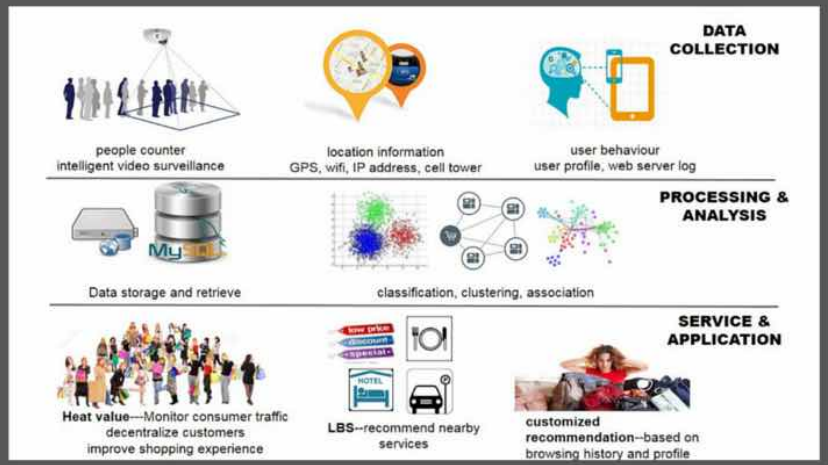
"not crowded, not hassled!"

Monitoring real-time population traffic (HEAT)

Pushing location-based information

Recommending customized products





香 港 大 學 計 算 機 科 學 系

THE UNIVERSITY OF HONG KONG

DEPARTMENT OF

COMPUTER SCIENCE

About us

HKUCS ranks No. 12 in 2015 QS World University Subject Ranking

Our Goals

- Develop students' careers
- Create high research impact
- Publish top-quality papers

Our graduates – recent faculties

- Oxford University, UK
- U of Southampton, UK
- University of Liverpool, UK
- King's college, U of London, UK
- University of Leicester, UK
- Arizona State U, USA
- Wayne State U, USA
- National University of Singapore
- Tsinghua, Taiwan
- University of Hong Kong, HK
- Hong Kong Polytechnic University, HK
- Hong Kong City University, HK
- Aalborg University, Denmark
- Bilkent U, Turkey
- Nanjing U, China

Our graduates - recent postdocs

- U Pittsburgh, USA
- U of Arizona, USA
- U of Cleveland, USA
- Yale, USA
- Purdue U, USA
- U of Southampton, UK
- U of Liverpool, UK
- Max Planck Institute, Germany
- U of Zurich, Switzerland
- Aalborg University, Denmark

Our research groups

- Algorithms and Bioinformatics – 5 faculties
- Data and Software Engineering – 4 faculties
- Systems and Networking – 4 faculties
- Graphics, Vision, HCI – 5 faculties
- Information Security and Forensics – 4 faculties
- Programming Language -1 faculty

Thank you!

Reynold Cheng
Computer Science, HKU
http://www.cs.hku.hk/~ckcheng/

ckcheng@cs.hku.hk

References

[1] Big data. Wikipedia. URL: http://en.wikipedia.org/wiki/Big_data

[2] How Much Information? University of California, San Diego. URL: http://hmi.ucsd.edu/howmuchinfo.php

[3] IBM DB2. URL: http://www.ibm.com/software/data

[4] Big Data Now: 2012 Edition from O'Rielly Media, Inc.

[5] IBM. URL: http://www-01.ibm.com/software/data/bigdata/

[6] Big data Analytics: Turning Big Data into Big Money. Frank Ohlhorst. John Wiley & Sons, Inc. 2013.

[7] Moore's law. Wikipedia. URL: http://en.wikipedia.org/wiki/Moore's_law

[8] Hadoop. URL: http://hadoop.apache.org/