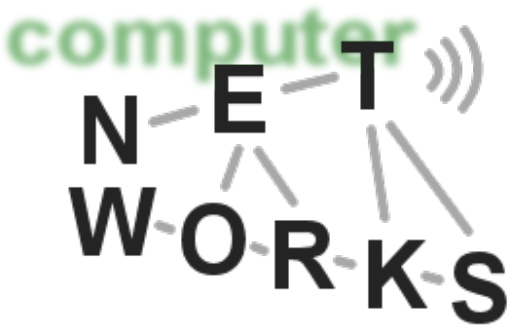
Introduction

Advanced Topics in Mobile Communications (AToMIC): Social Network in Mobile Big Data

Summer Semester 2016



computer NETWORKS

Organizational Information

- Course is held in English
- 5 ECTS credits
- AI: M.Inf.223.Mp, M.Inf.224.Mp, M.Inf.1223
- ITIS: 3.10
- Course Link: https://wiki.net.informatik.uni-goettingen.de/wiki/Advanced_Topics_in_Mobile_Communications_(AToMIC):_Social_Network_in_Mobile_Big_Data_(Summer_2016)

Organizational Information

- No exam, but 1 exercise, 1 presentation and 1 report
  - Exercise (20%)
  - Presentation (40%): first two Fridays of July
  - Report (40%): end of September

- Check the FlexNow registration deadlines

Course Overview

- This course covers:
  - An overview of the lecture
    - (1 lecture, today)
  - Different big data methods
    - classification and regression methodologies: unsupervised, supervised, and statistical data mining methods.
    - behavior detection algorithms: community detection solutions based on graph theory and factor analysis approaches
  - A couple of social network theories:
    - Small groups
    - Social structure theory
    - Social exchange theory
    - Social balance theory
    - Tie strength
    - Small world theory
    - Social capital, Guanxi

Course Overview

Date	Topic
15.04.2016	Introduction, mobile big data; literatures
22.04.2016	Big data methods (machine learning, data mining, semantic and structural analysis)
29.04.2016	Big data methods (cont.); data samples
06.05.2016	Social network theory
13.05.2016	Interdisciplinary methods and case study
20.05.2016	cancelled due to business trips
27.05.2016	cancelled due to business trips
03.06.2016	Mining opinion leaders and structure hole spanners
10.06.2016	cancelled due to business trips
17.06.2016	cancelled due to business trips
25.06.2016	Practical session
01.07.2016	Final presentations (1)
08.07.2016	Final presentations (2)
15.07.2016	cancelled due to business trip

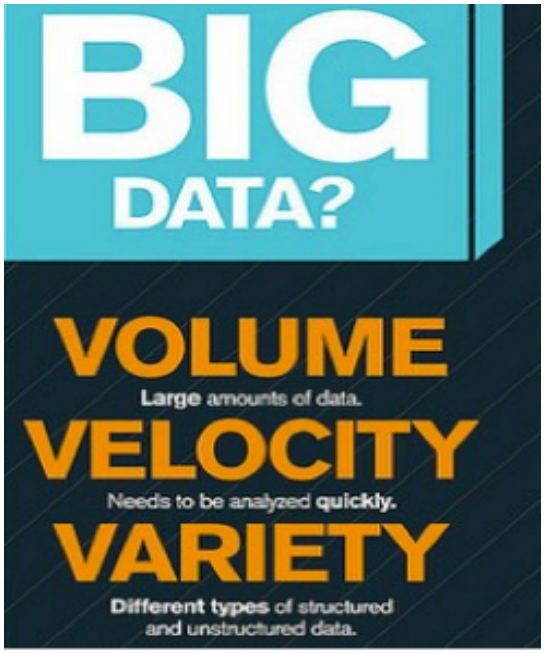
What is Big Data?

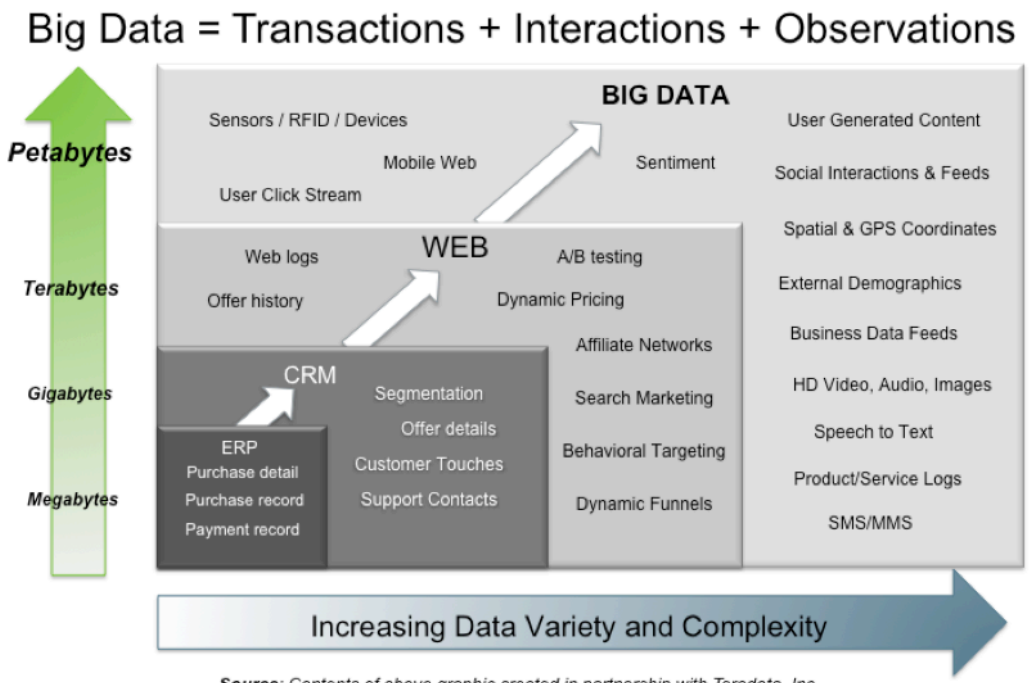
The concept of Big Data from Wikipedia:

- a term for data sets that are so large or complex that traditional data processing applications are inadequate
- the concept gained momentum in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of big data as the three Vs

What is Big Data?

- Three characteristics of Big Data V3s
  - Volume
  - Velocity
  - Variety

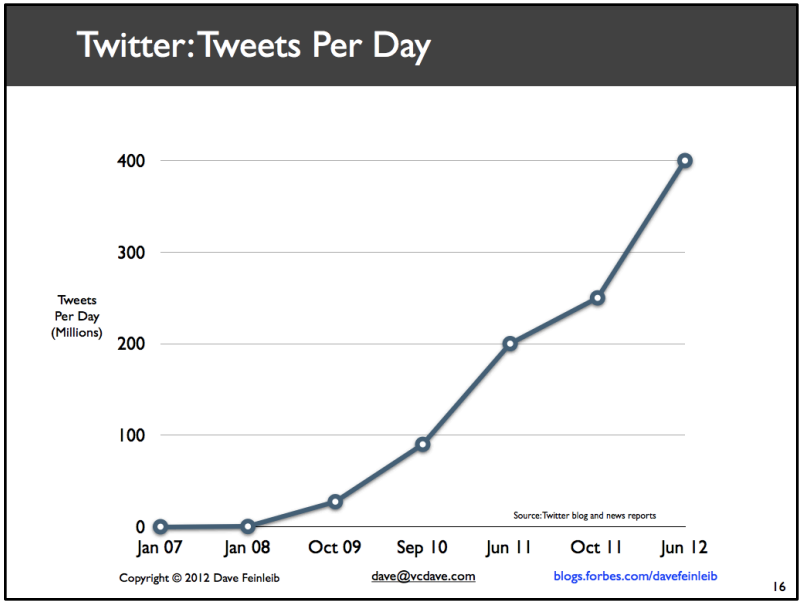


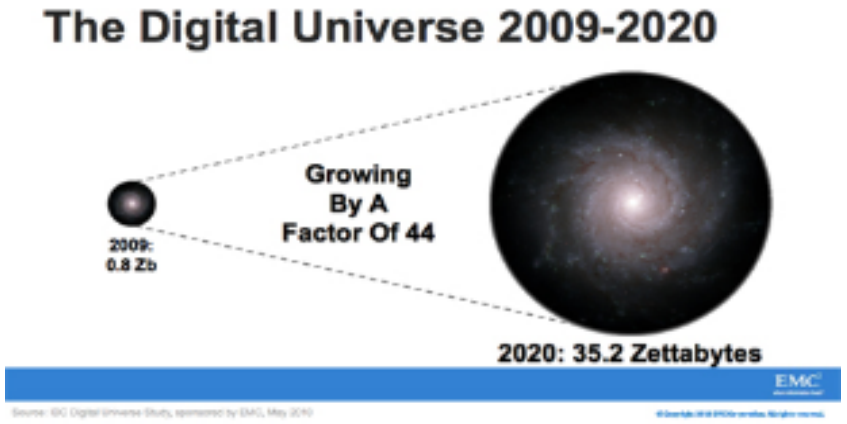


Source: *Contents of above graphic created in partnership with Teradata, Inc.*

What is Big Data?

- Volume
  - Data volume is growing exponentially
    - 500 million tweets are tweeted on Twitter everyday
  - Predict
    - 44x growth from 2009 to 2020
    - the volume of data in the world will grow to 35 Zettabytes in 2020
      - 1 ZB = 10^21 bytes
      - Kilobyte – Megabyte – Gigabyte – Terabyte – Petabyte – Exabyte – Zettabyte – Yottabyte





12+ TBs
of tweet data every day

2 TBs of data every day

25+ TBs of log data every day

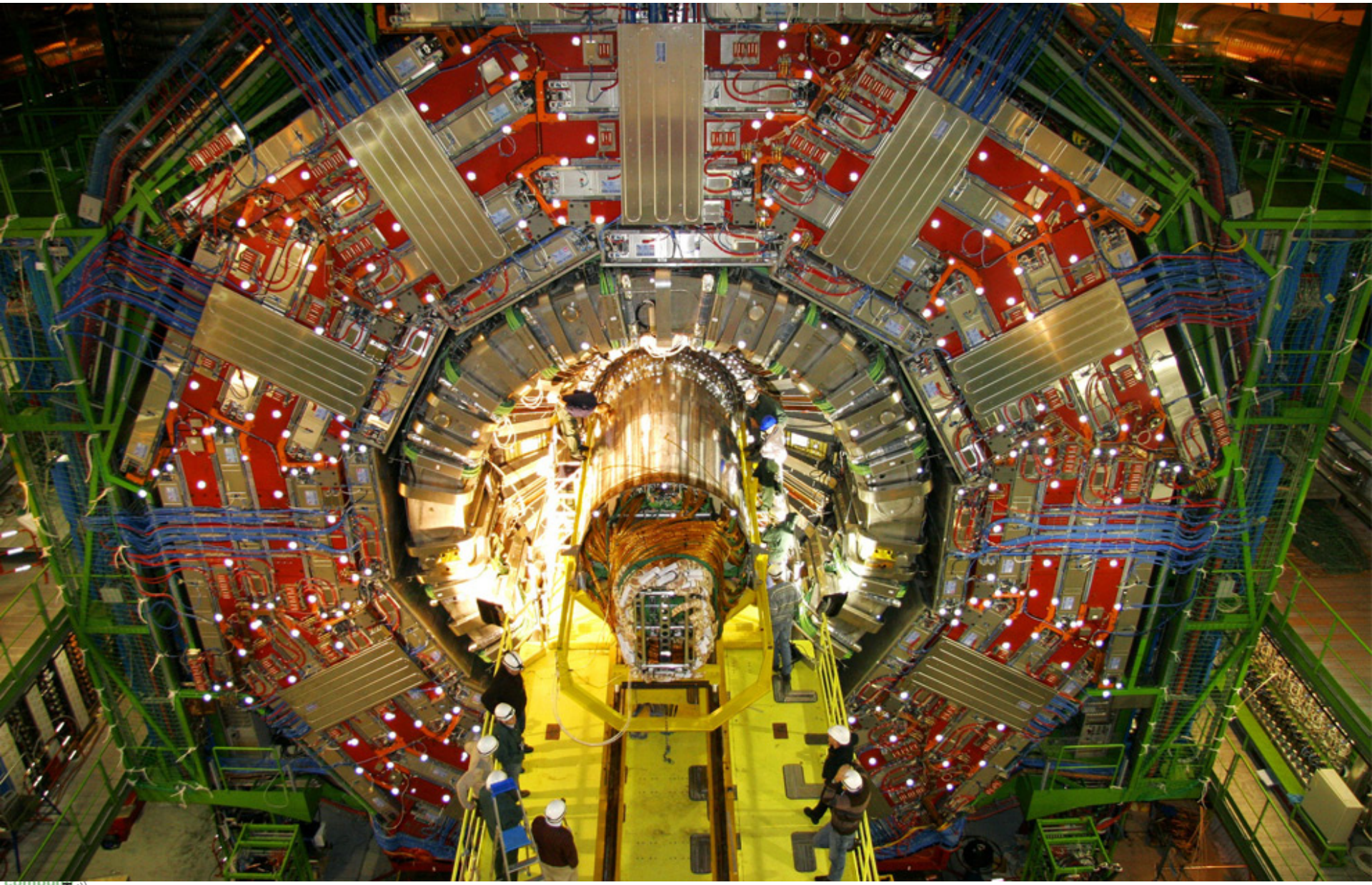
30 billion RFID tags today (1.3B in 2005)

4.6 billion camera phones world wide

100s of millions of GPS enabled devices sold annually

2+ billion people on the Web by end 2011

76 million smart meters in 2009… 200M by 2014



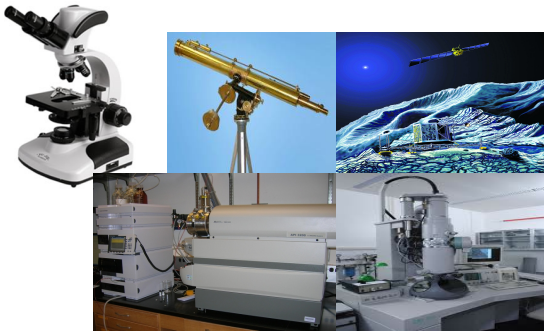
CERN's Large Hydron Collider (LHC) generates 15 PB a year

What is Big Data?

- Velocity
  - Data is generated fast and must be processed fast
  - RFID tags, sensors and smart metering are driving the need to process torrents of data in near-real time
    - Healthcare monitoring

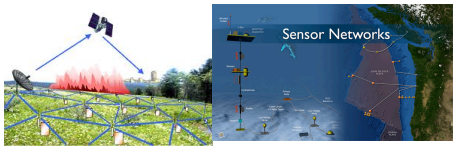
Real-time/Fast Data

Social media and networks
(all of us are generating data)



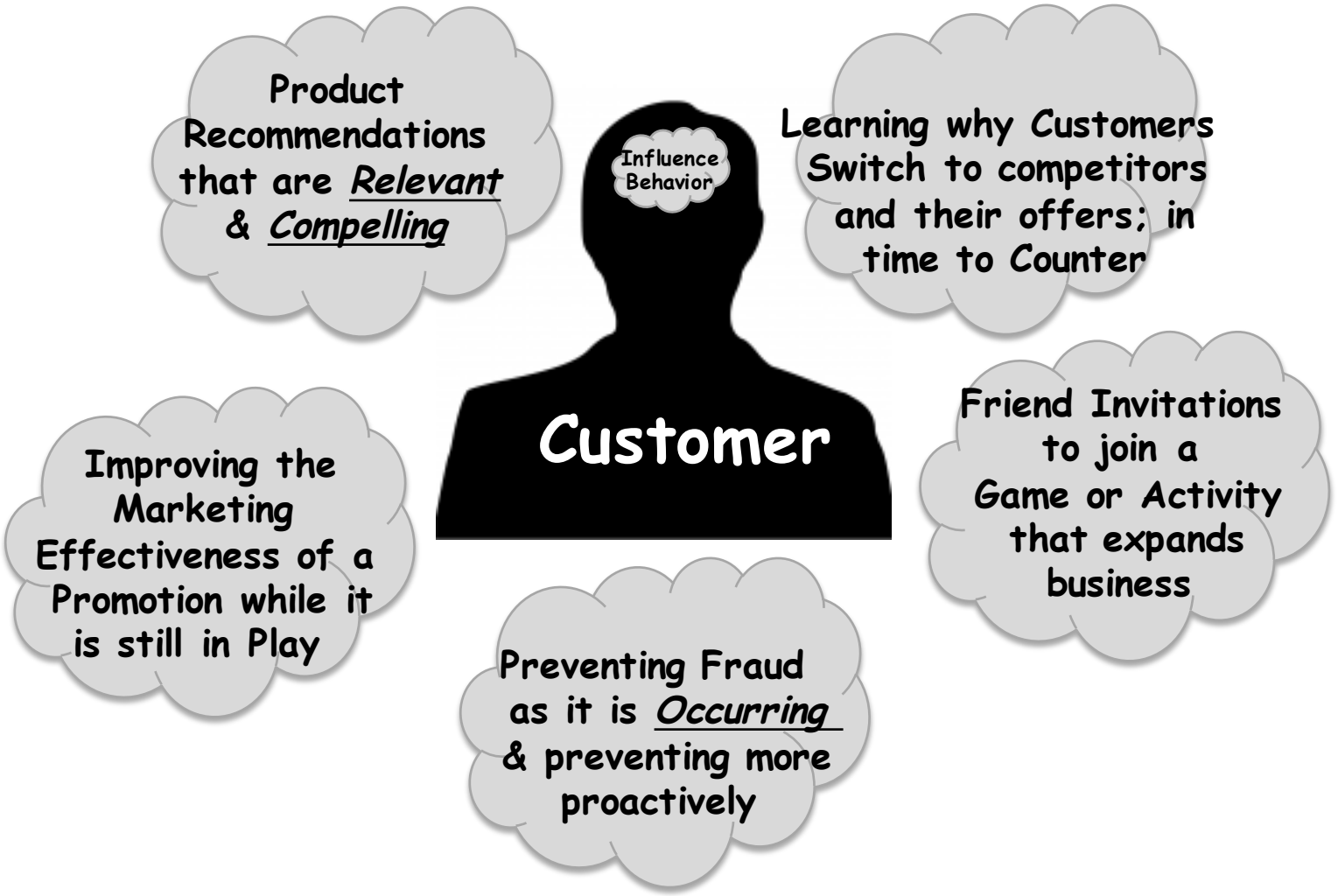
Scientific instruments
(collecting all sorts of data)

Mobile devices
(tracking all objects all the time)



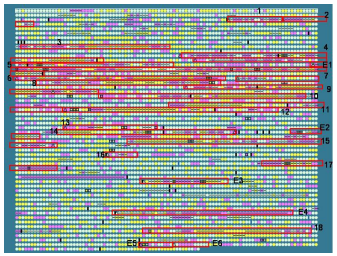
Sensor technology and networks
(measuring all kinds of data)

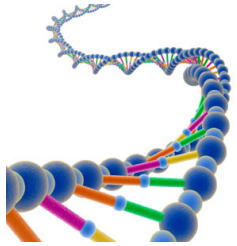
Real-Time Analytics/Decision Requirement

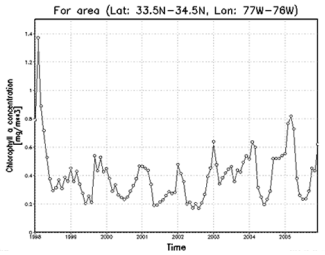


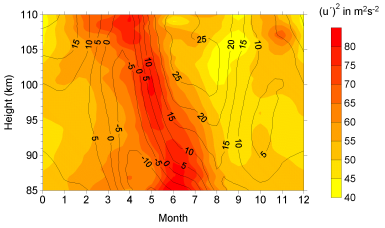
What is Big Data?

- Variety(Complexity)
  - Data comes in all types of formats
    - structured, numeric data in traditional database
    - unstructured text documents, email, video, audio, stock ticker data and financial transactions
  - A single application can be generating/collecting many types of data

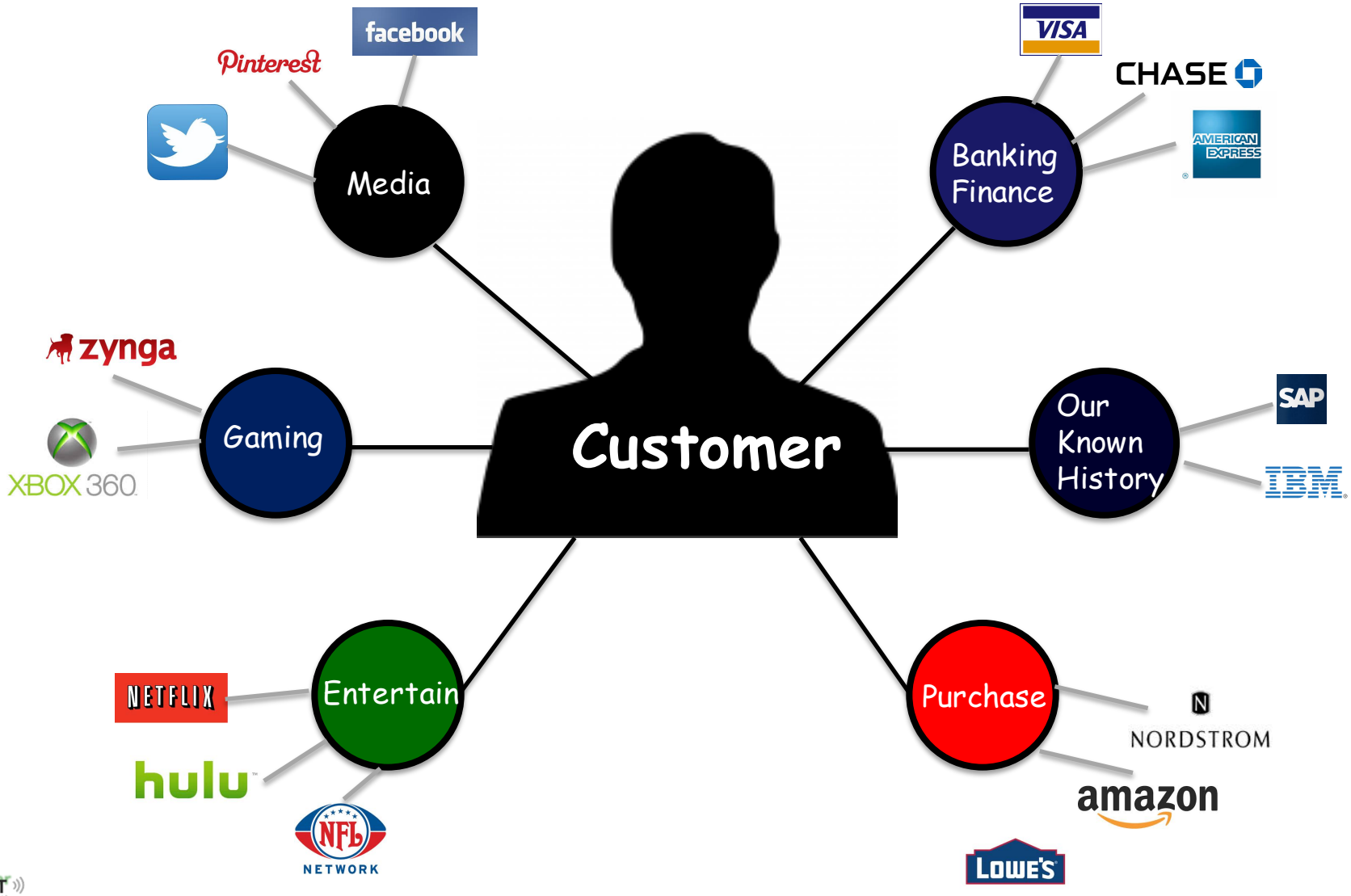






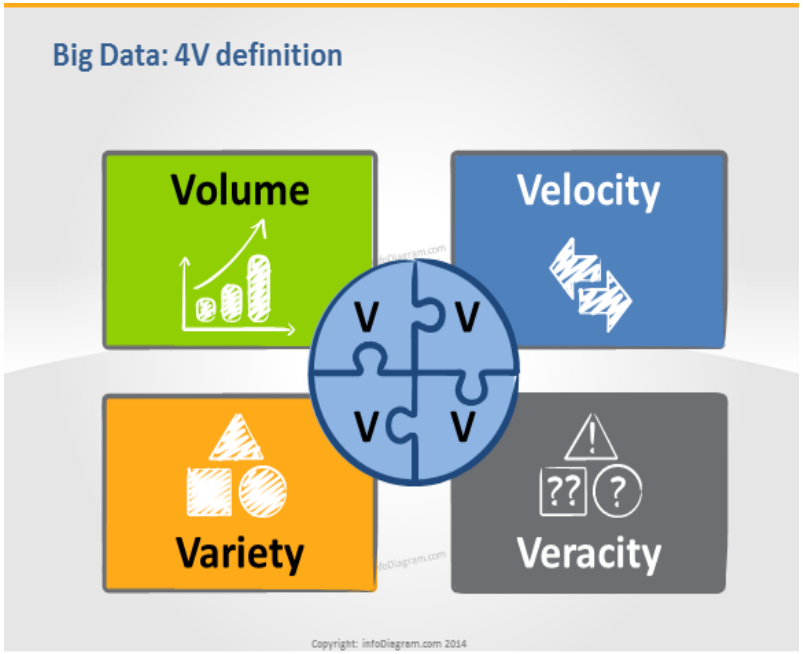


A single view to the customer



What is Big Data?

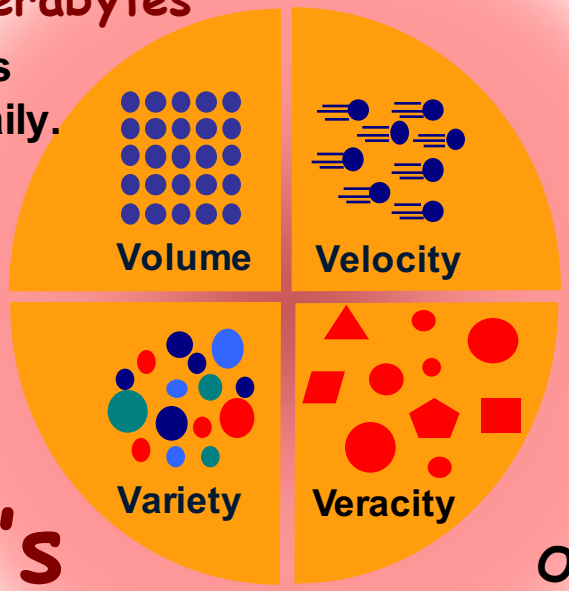
- Some give 4V definition
  - Veracity
    - It refers to the trustworthiness of the data
    - Can the manager rely on the fact that the data is representative?
    - Every good manager knows that there are inherent discrepancies in all the data collected



With Big Data, We've Moved into a New Era of Analytics

12+ terabytes
of Tweets
create daily.

5+ million
trade events
per second.



100's
of different
types of data.

Only 1 in 3
decision makers trust
their information.

The 5 Key Big Data Use Cases

Big Data Exploration
Find, visualize, understand all big data to improve decision making

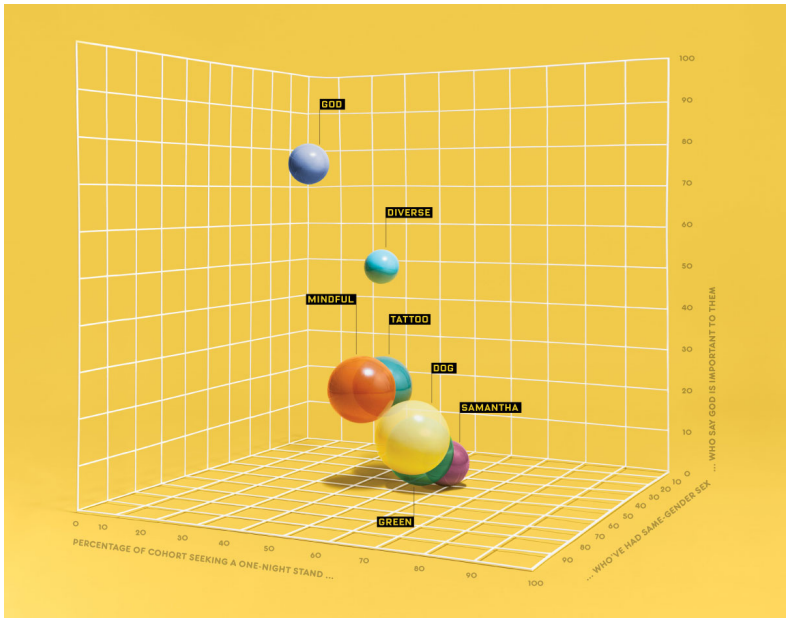
Enhanced 360° View of the Customer
Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources

Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time

Operations Analysis
Analyze a variety of machine data for improved business results

Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

How a Math Genius Hacked OkCupid to Find True Love



“Beer and diapers"

Enabler of Big Data – Social network research

- What is social network
  - a set of actors and the ties (resource flows) or relations (stable states) among them
    - close colleagues (*relation*) among teachers (*actors*)
    - one teacher (*actor*) provides help (*tie*) to another
    - communication (*tie*) between people (*actors*) in an organization
  - Social networking sites are main research objects
    - Offline: classmates network…
    - Online: Facebook, Twitter...

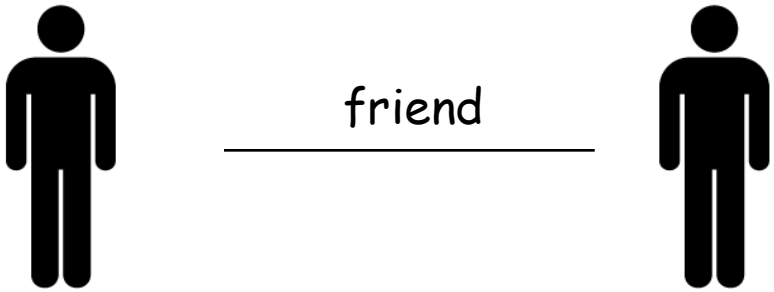
Online social networks

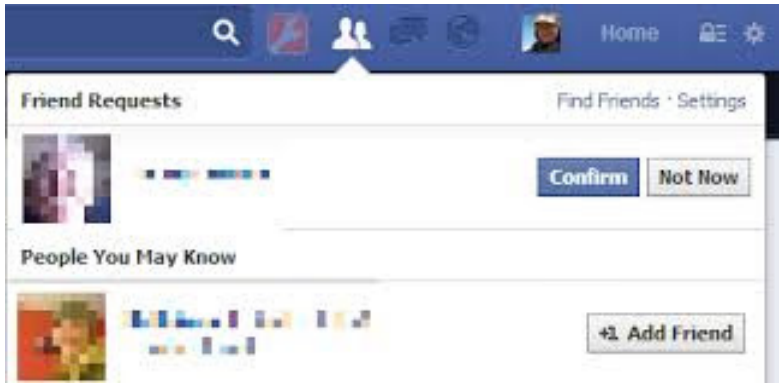
- Role of OSN
  - OSNs have reached 82% of the world's Internet-using population (1.2billion) (2011)
  - Social Networking accounts for 19%of all time spent online (2011)
  - Social Networking is the most popular online activity worldwide

- Variety of OSN
  - Various functions, social relationships, social networks

Facebook

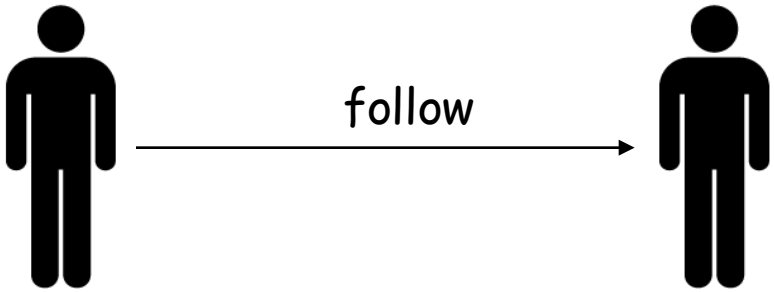
- Relationship
  - User A sends a friend request to user B and user B confirms the request from user A, then they are friends
  - Undirected social network





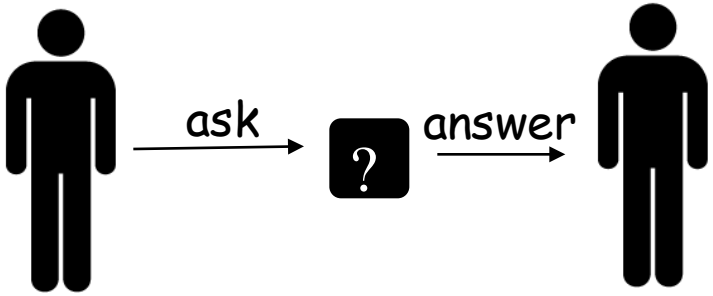
Twitter

- Relationship
  - User A follows user B
    - User A is user B's fan
    - User B is user A's friend
  - Directed social network

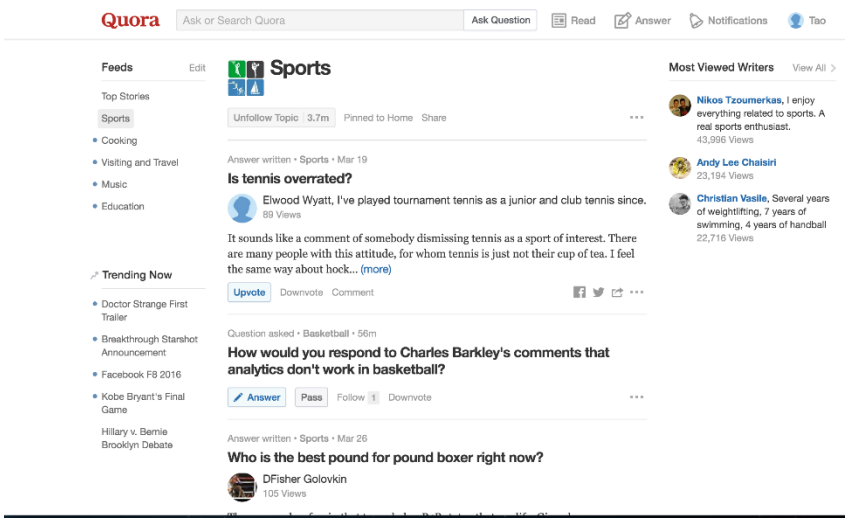


Quora

- Relationship
  - Question-answering
    - User A asks a question and user B answers the question
  - Directed social network



Quora



The popularity of mobile devices bring in mobile big data

- Background
  - more than one billion smart phones are in use that are producing vast amount of data
  - with the rapid development of the Internet-of-Things (IoT), much more data is automatically generated by millions of machine nodes with growing mobility
    - e.g., sensors carried by moving objects or vehicles

What is Mobile Big Data?

- Concept
  - making the results of big data analysis available on mobile devices
- Trend
  - the volume, velocity, and/or variety of this data is increasing extremely fast
  - Mobile Big Data (MBD) has been already in our lives and being enriched rapidly

Benefits of mobility

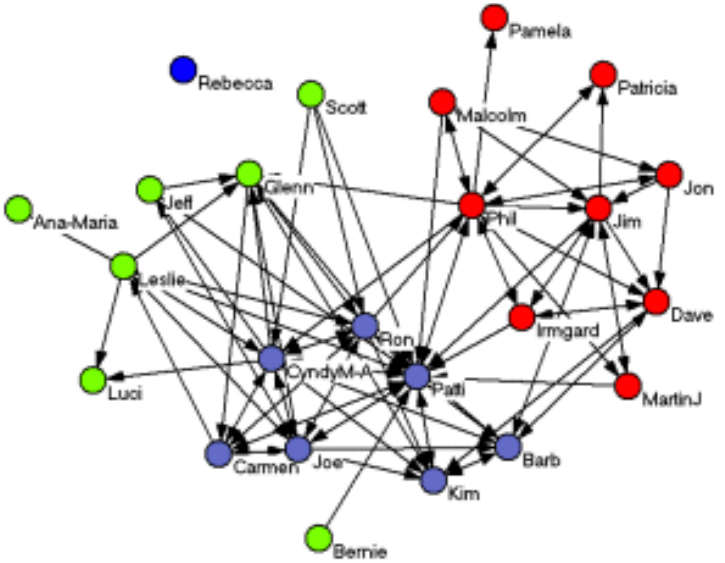
- Mobile devices bring
  - the availability of a large amount of sensors which gives a lot of opportunities
    - location
    - temperature
  - access to all your data at any time and anywhere

Opportunities from Mobile Big Data

- Opportunities from Mobile Big Data
  - Use social network analysis methods to understand behaviours and requirements of mobile users
    - Data mining and data crowd sourcing
    - Knowledge discovery
    - …

Social network analysis

- Social network analysis (SNA)
  - the mapping and measuring of relationships and flows between people, groups, organizations, computers or other information/knowledge processing entities.
  - the nodes in the network are the people and groups while the links show relationships or flows between the nodes



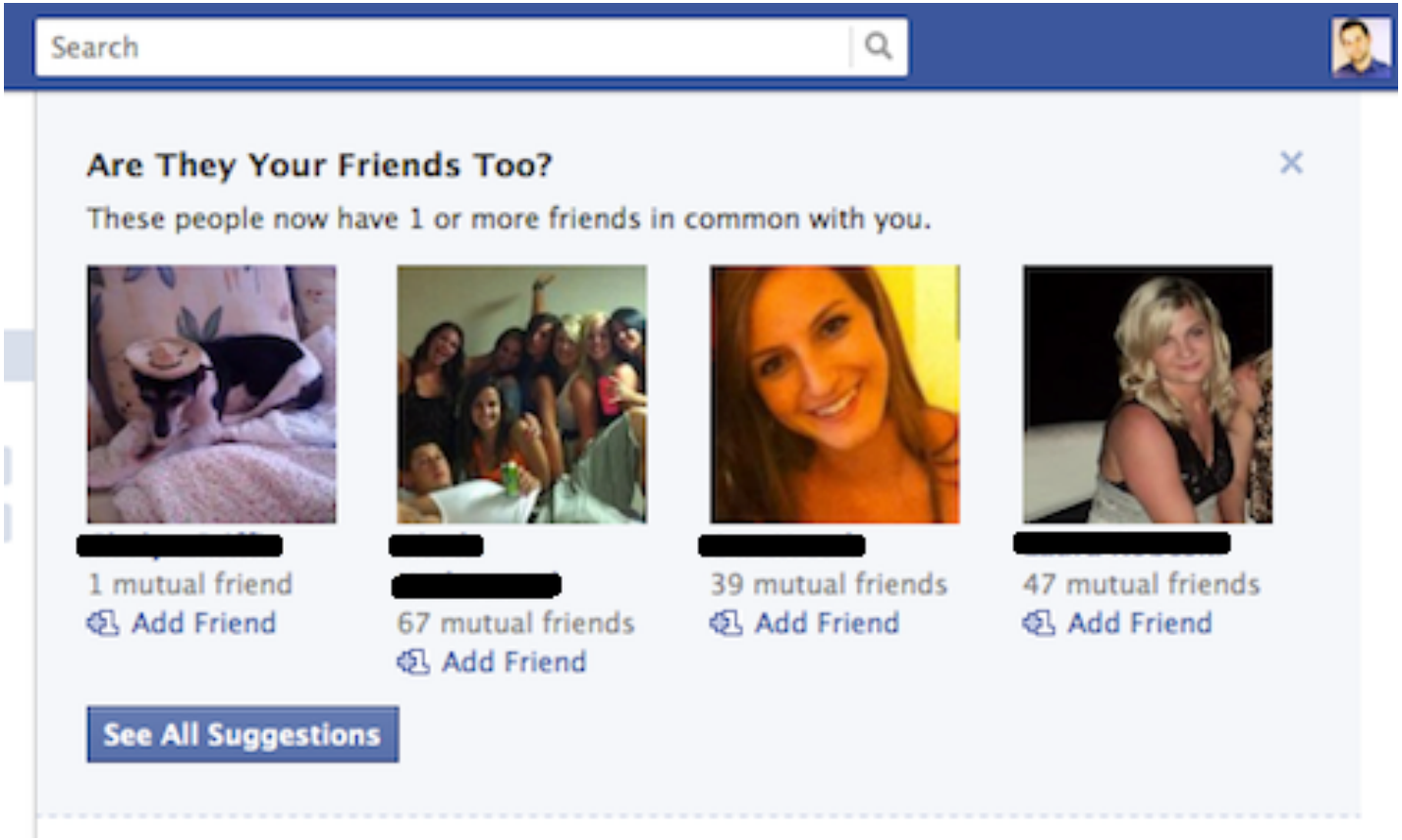
What can social network analysis do in Big Data?

- Applications of social network analysis in big data
  - Community identification
  - Influential user identification
  - Link prediction
  - Point of interest recommendation
  - Disease prediction
  - Crime prediction
  - Event monitoring
  - …

Community identification

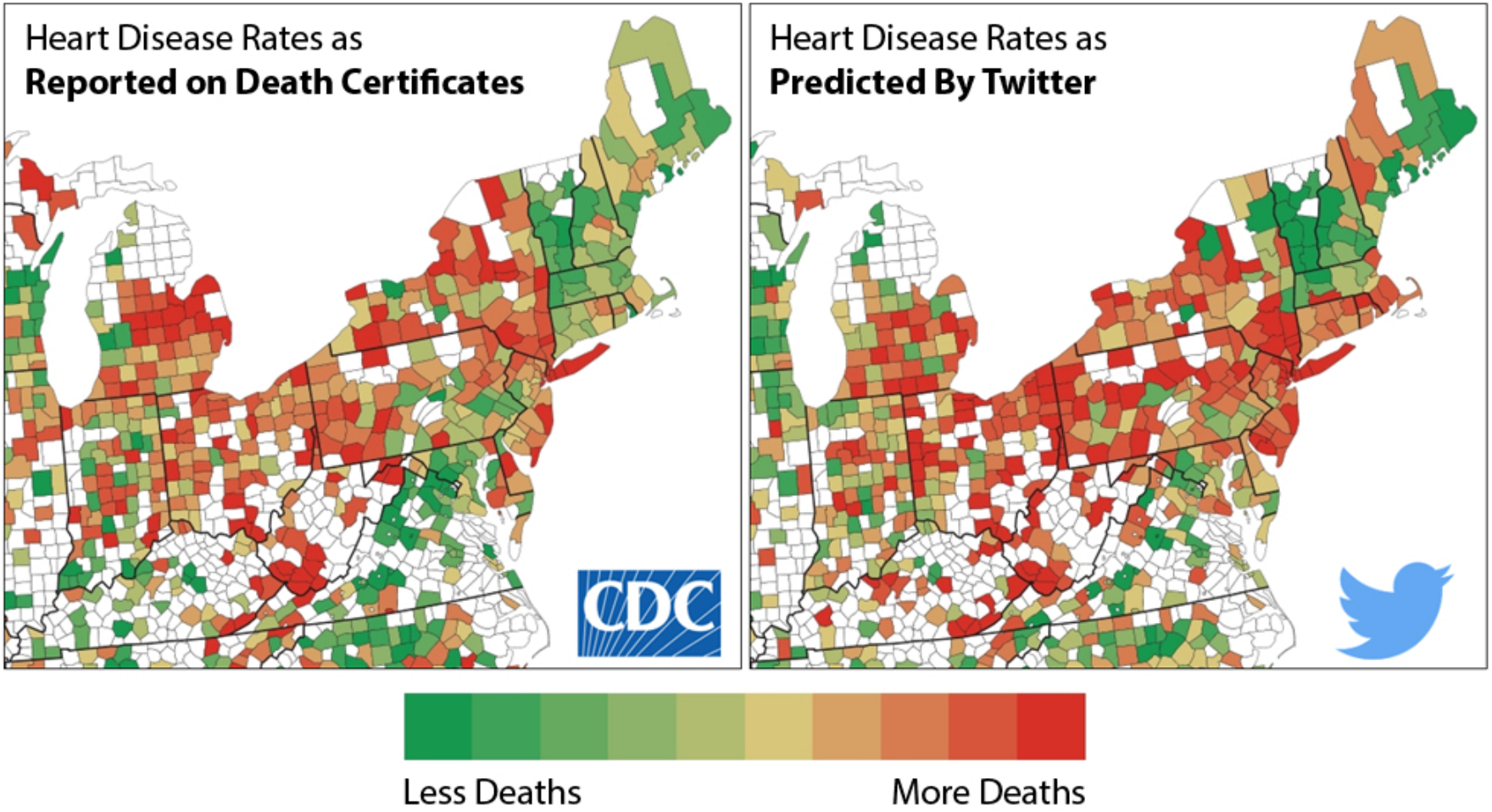
Influential user identification

Link prediction

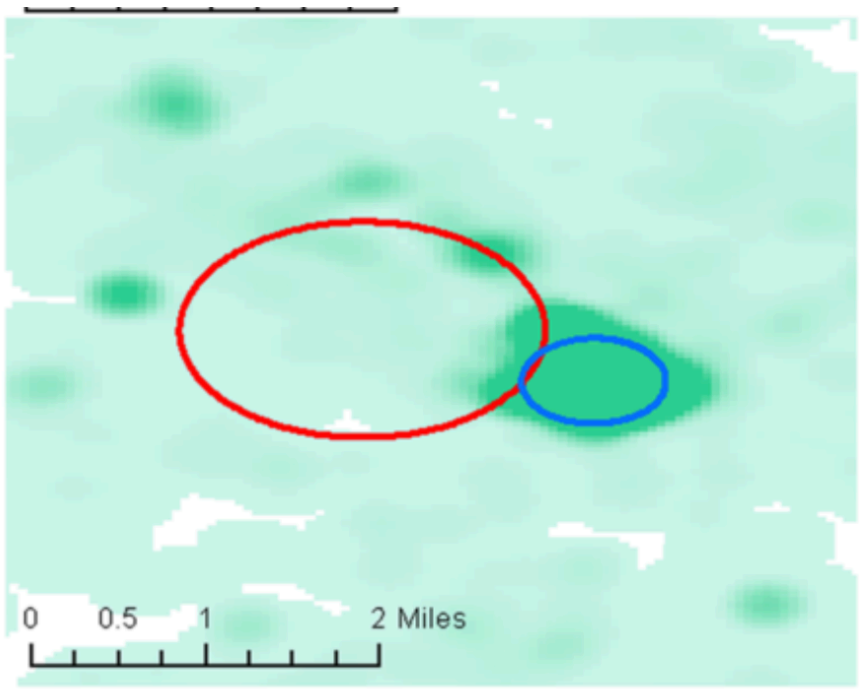


Point of interest recommendation

Disease prediction



Crime prediction



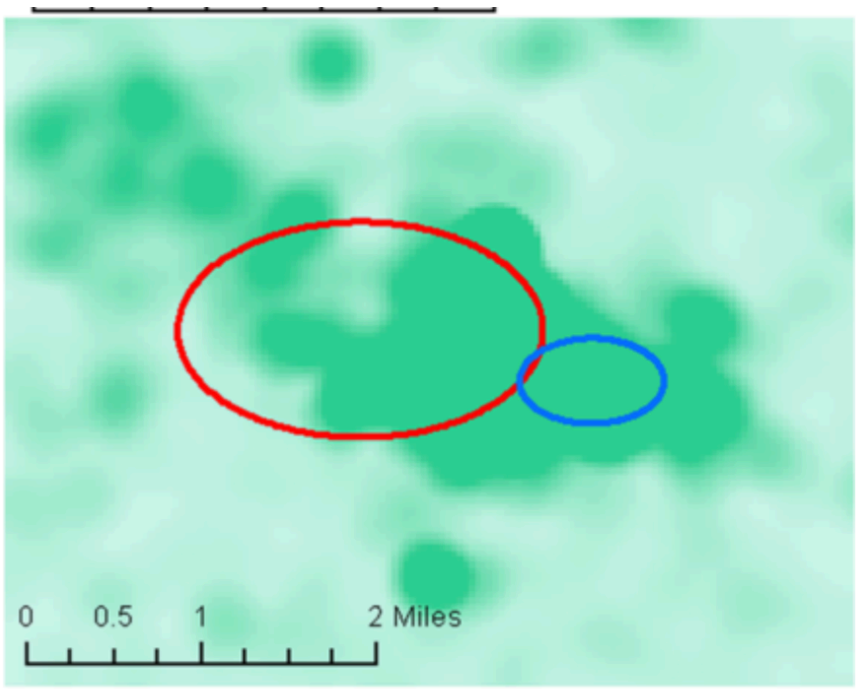
Crime Density

High : 55.2116

Low : 0

Cluster A

Clusrer B



Twitter Message Density

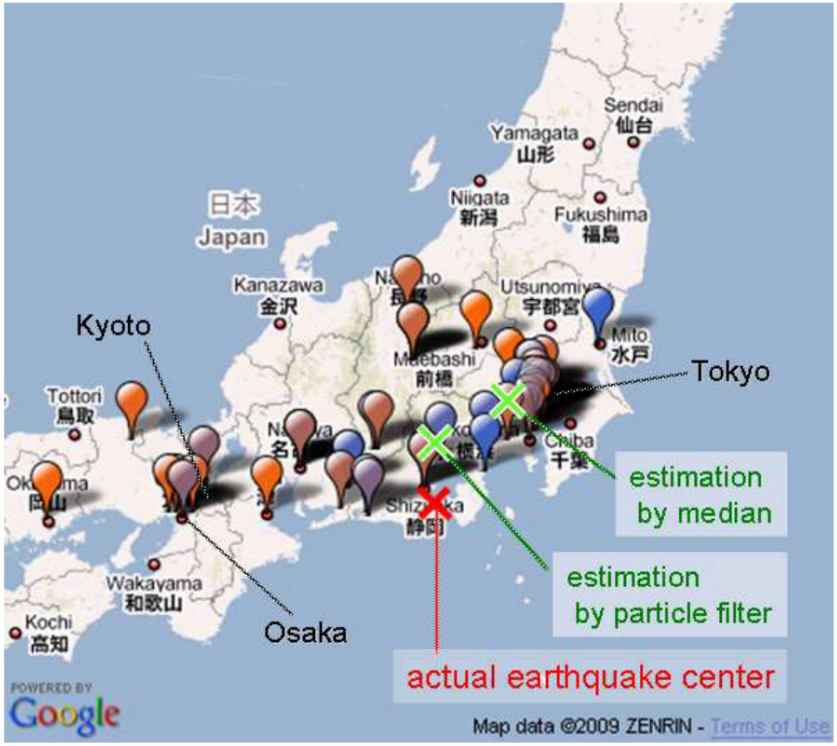
High : 118.196

Low : 0

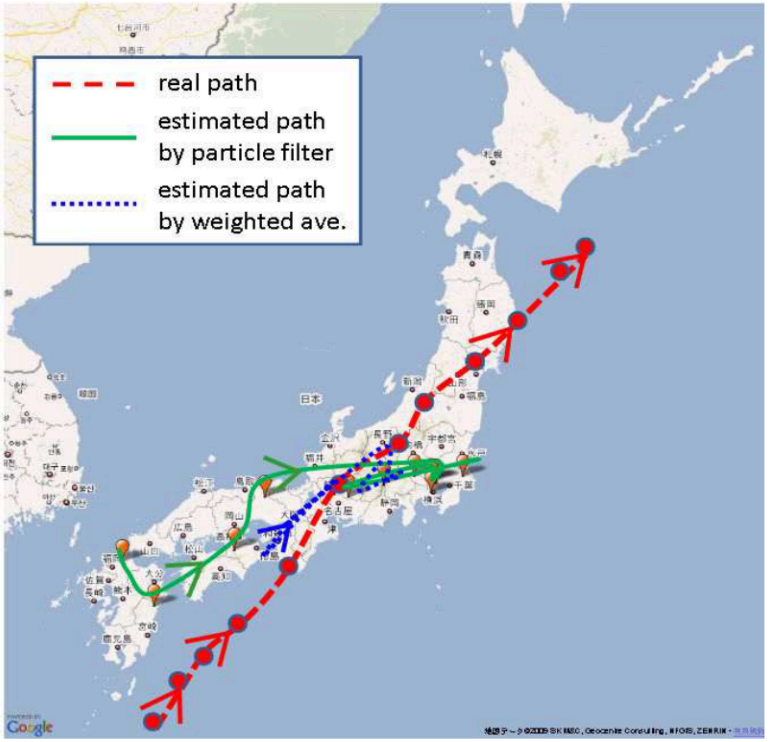
Cluster A

Clusrer B

Event monitoring



Earthquake location estimation based on tweets.



Typhoon trajectory estimation based on tweets.

Software Tools

- SNAP
  - C++ library for working with massive network datasets
- Pajek
  - Program for large network analysis
- NetworkX
  - Python package for the study of the structure of complex networks
- Gephi
- …

Literature for improvement

- E. Cho, S. A. Myers, J. Leskovec. Friendship and Mobility: User Movement in Location-Based Social Networks ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011.
- Wang H, Terrovitis M, Mamoulis N. Location recommendation in location-based social networks using user check-in data[C]//Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2013: 374-383.
- Scellato S, Noulas A, Lambiotte R, et al. Socio-Spatial Properties of Online Location-Based Social Networks[J]. ICWSM, 2011, 11: 329-336.
- Ugander J, Karrer B, Backstrom L, et al. The anatomy of the facebook social graph[J]. arXiv preprint arXiv:1111.4503, 2011.
- Benevenuto F, Rodrigues T, Cha M, et al. Characterizing user behavior in online social networks[C]//Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference. ACM, 2009: 49-62.
- N. Eagle, A. Pentland, and D. Lazer (2009), "Inferring Social Network Structure using Mobile Phone Data", Proceedings of the National Academy of Sciences (PNAS) Vol 106(36), pp. 15274-15278.

Outlook

- In the next lecture:
  - Introduction to big data methods
    - machine learning, data mining, semantic and structural analysis

References

- "Introduction to Big Data". http://www.cs.kent.edu/~jin/BigData/Lecture1.pptx
- "Social Networks". https://www.ischool.utexas.edu/~i385q/archive/sharma_social_networks.ppt
- "A New Era of Analytic". http://www.cs.bilkent.edu.tr/~guvenir/courses/CS418/OmerSever-BD_Presentation_1.1.ppt

Thank you!