大數據行銷研究
Big Data Marketing Research

Tamkang University
淡江大學

社群運算與大數據分析
(Social Computing and Big Data Analytics)

1051BDMR08
MIS EMBA (M2262) (8638)
Thu, 12,13,14 (19:20-22:10) (D409)

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系
http://mail. tku.edu.tw/myday/

2016-11-25

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2016/09/16	中秋節 (調整放假一天) (Mid-Autumn Festival Holiday)(Day off)
2	2016/09/23	大數據行銷研究課程介紹 (Course Orientation for Big Data Marketing Research)
3	2016/09/30	資料科學與大數據行銷 (Data Science and Big Data Marketing)
4	2016/10/07	大數據行銷分析與研究 (Big Data Marketing Analytics and Research)
5	2016/10/14	測量構念 (Measuring the Construct)
6	2016/10/21	測量與量表 (Measurement and Scaling)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
7	2016/10/28	大數據行銷個案分析 I (Case Study on Big Data Marketing I)
8	2016/11/04	探索性因素分析 (Exploratory Factor Analysis)
9	2016/11/11	確認性因素分析 (Confirmatory Factor Analysis)
10	2016/11/18	期中報告 (Midterm Presentation)
11	2016/11/25	社群運算與大數據分析 (Social Computing and Big Data Analytics)
12	2016/12/02	社會網路分析 (Social Network Analysis)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
13	2016/12/09	大數據行銷個案分析 II (Case Study on Big Data Marketing II)
14	2016/12/16	社會網絡分析量測與實務 (Measurements and Practices of Social Network Analysis)
15	2016/12/23	大數據情感分析 (Big Data Sentiment Analysis)
16	2016/12/30	金融科技行銷研究 (FinTech Marketing Research)
17	2017/01/06	期末報告 I (Term Project Presentation I)
18	2017/01/13	期末報告 II (Term Project Presentation II)

Outline

- Social Computing
- Big Data Analysis

教育部資通人才培育計畫
社群運算與巨量資料
課程四大模組

(1)「社群媒體」(Social Media)
(政治大學)

(2)「資料科學」(Data Science)
(政治大學)

(3)「分析技術」(Analytics Technology)
(高雄大學) (淡江大學)

(4)「領域應用」(Domain Application)
(淡江大學) (政治大學)

1.「社群媒體」(Social Media)(政治大學)

- 探討 社群媒體和資料分析的概念，以個案方式教學

2.「資料科學」(Data Science) (政治大學)

- 探討 Data Thinking 和 EDA 等，與DSP或痞客邦合作

3.「分析技術」(Analytics Technology) (高雄大學) (淡江大學)

- 列舉重要的分析方法，包括社會網絡分析，文字探勘分析技術簡介。
  - * 社會網絡分析 (高雄大學)
  - * 社會網絡量測 (高雄大學)
  - * 社會網絡分析工具 (高雄大學)
  - * 文字探勘分析技術簡介 (淡江大學)

4.「領域應用」(Domain Application) (淡江大學) (政治大學)

- 區分 Domain Knowledge ，聚焦探討各種商業行銷和輿情分析等
  - * 社群媒體行銷分析 (淡江大學)
  - * 社群媒體情感分析 (淡江大學)

Social Computing

Social Network Analysis

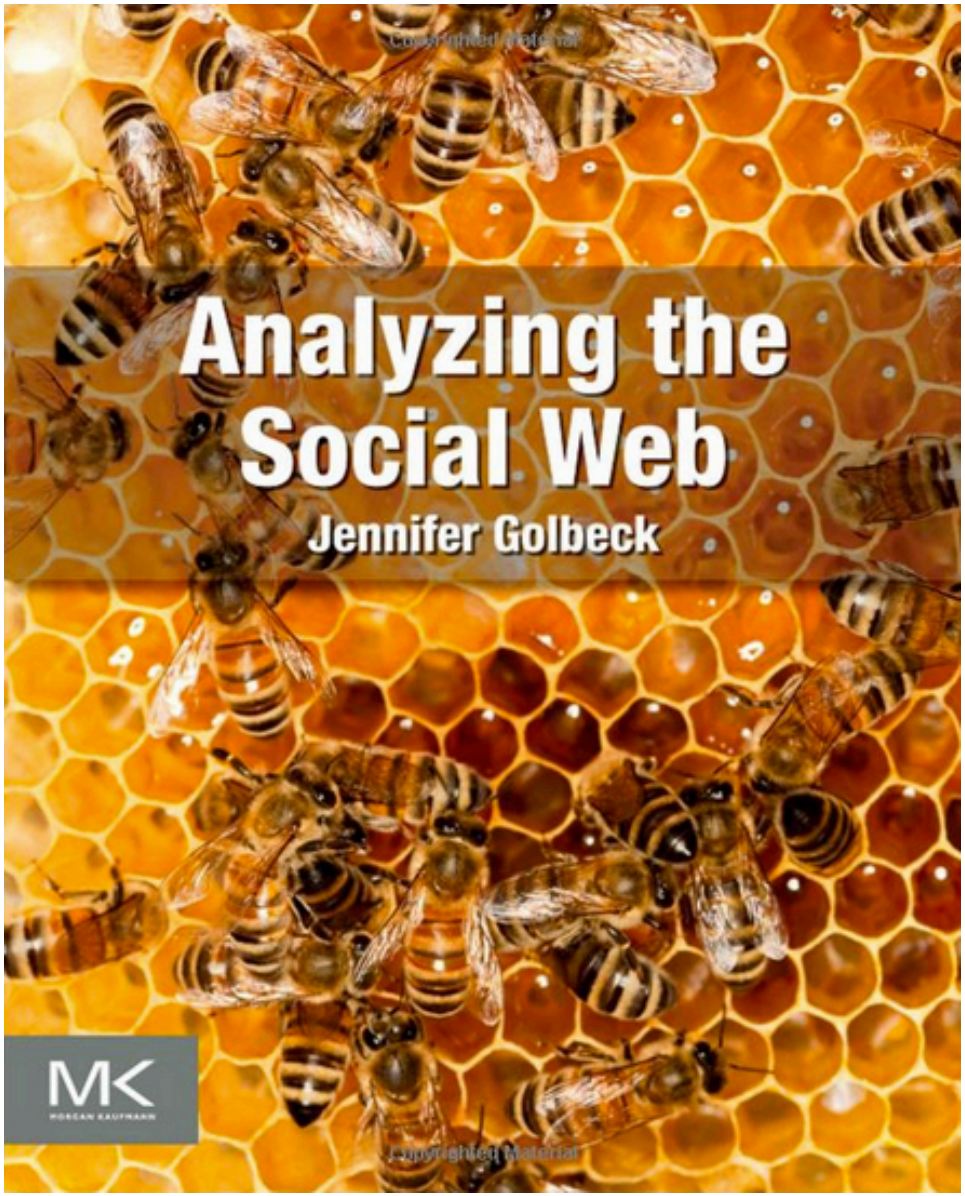
Social Computing

- Social Network Analysis
- Link mining
- Community Detection
- Social Recommendation

Business Insights with Social Analytics

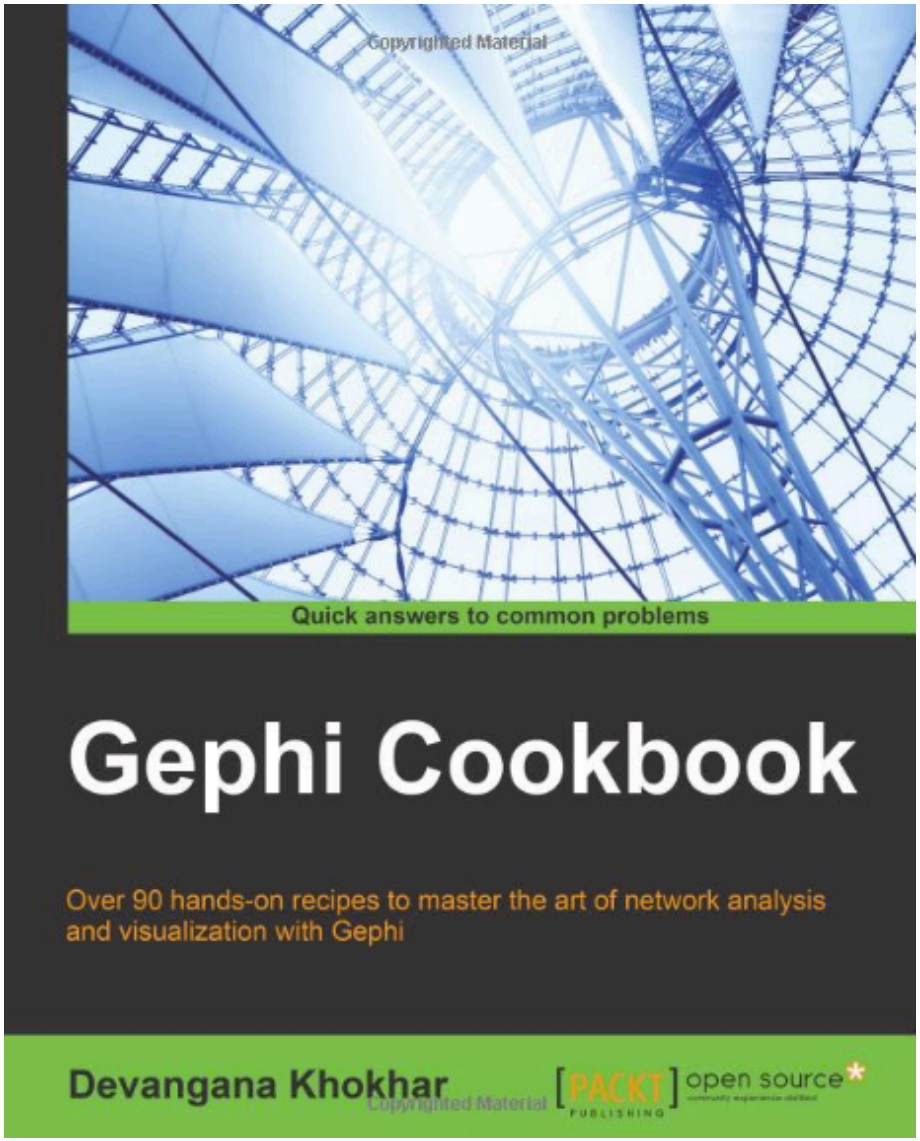
Analyzing the Social Web:
Social Network Analysis

Jennifer Golbeck (2013), **Analyzing the Social Web**, Morgan Kaufmann



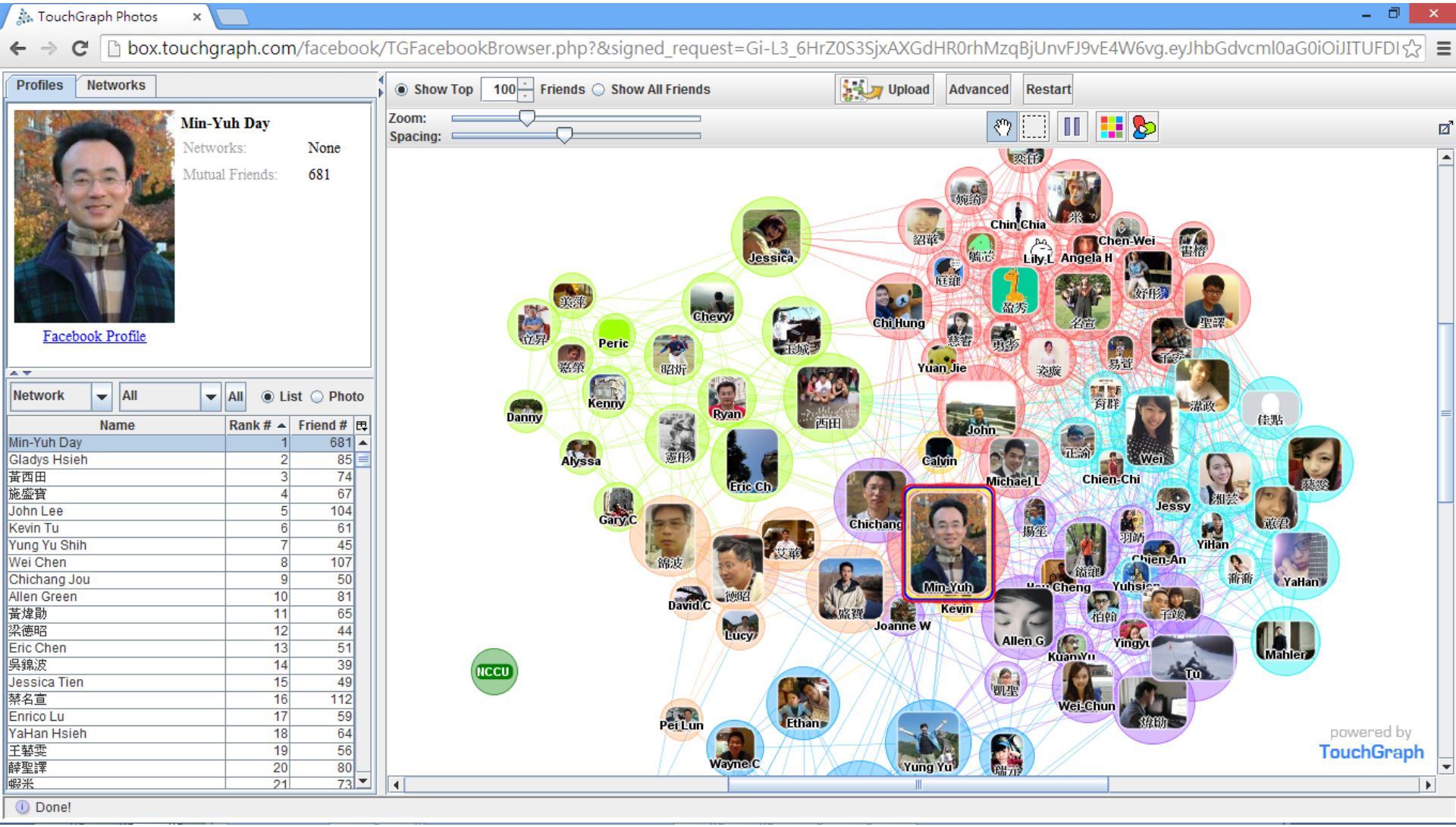
Source: http://www.amazon.com/Analyzing-Social-Web-Jennifer-Golbeck/dp/0124055311

Devangana Khokhar (2015), Gephi Cookbook, Packt Publishing



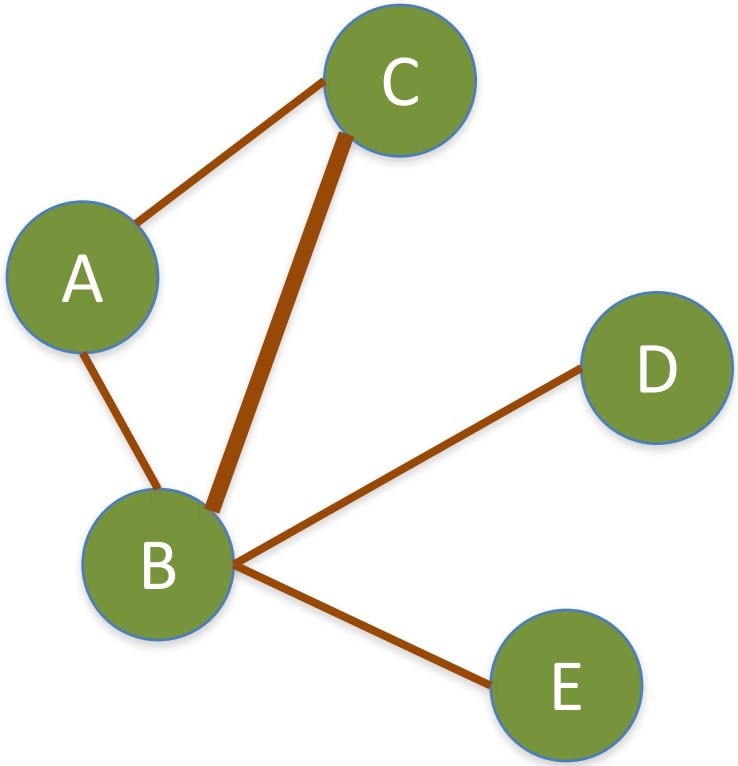
Source: http://www.amazon.com/Gephi-Cookbook-Devangana-Khokhar/dp/1783987405

Social Network Analysis (SNA)
Facebook TouchGraph



Graph Theory

Graph



Graph

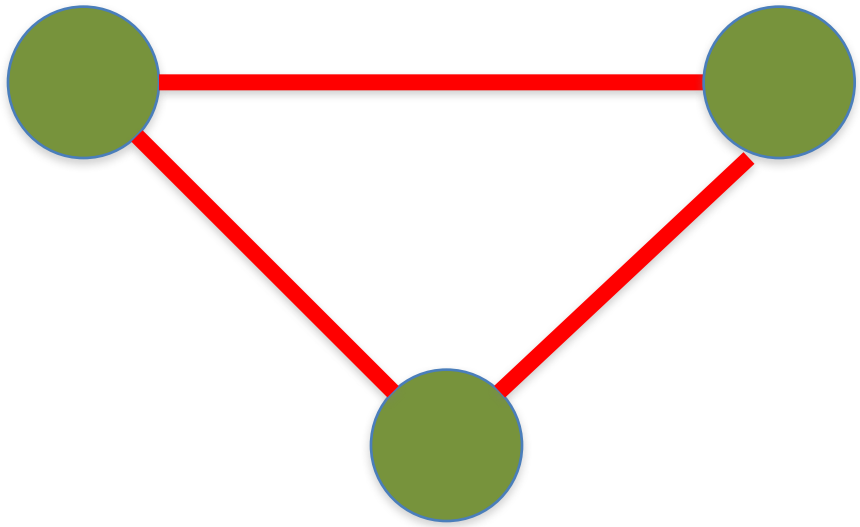
g = (V, E)

Vertex (Node)

Vertices (Nodes)

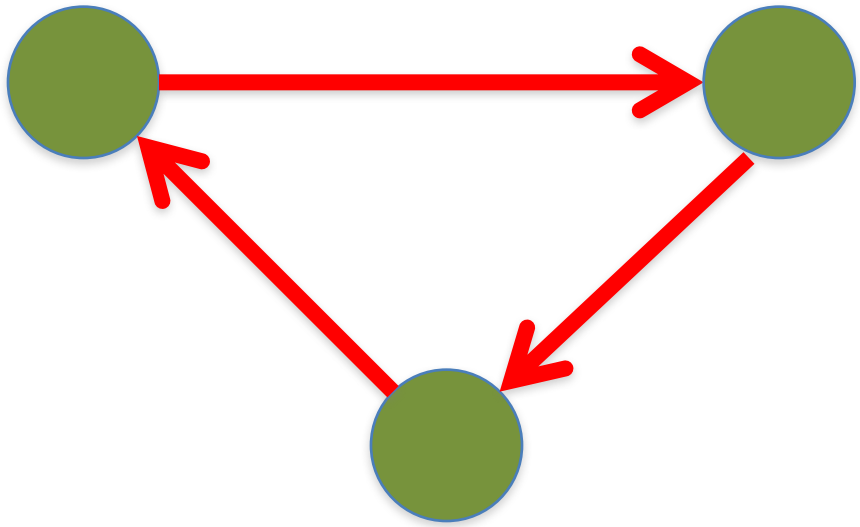
Edge

Edges

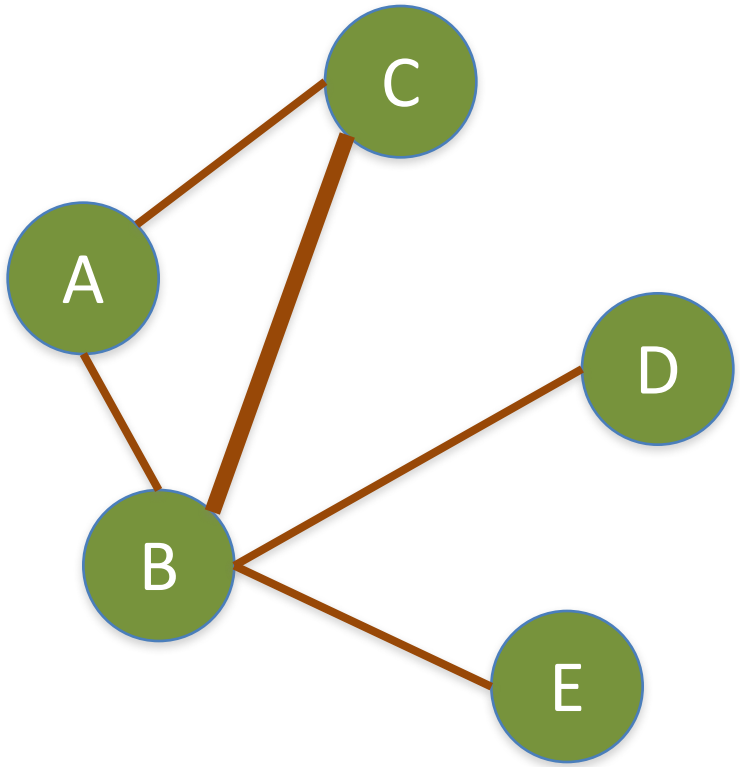


Arc

Arcs

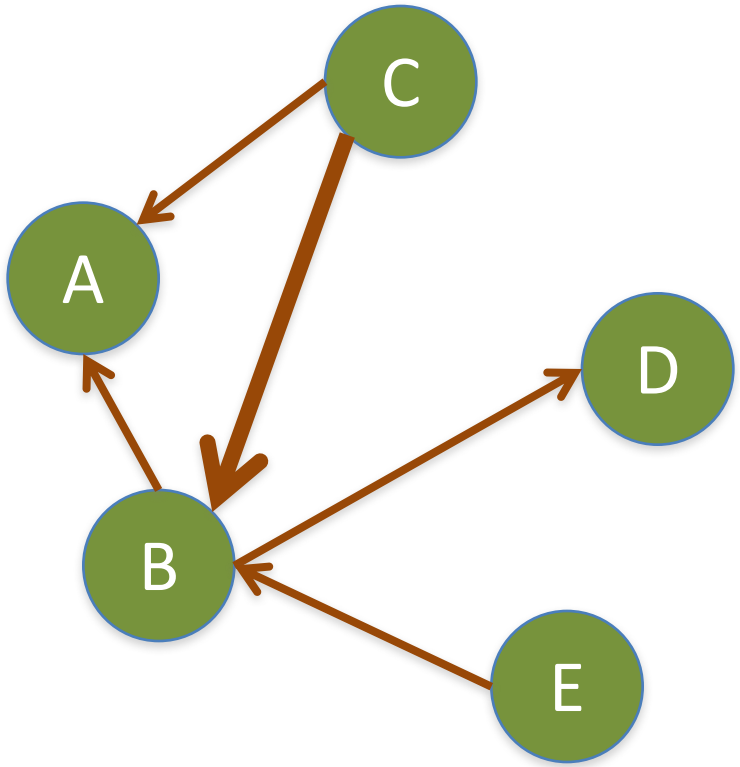


Undirected Graph



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

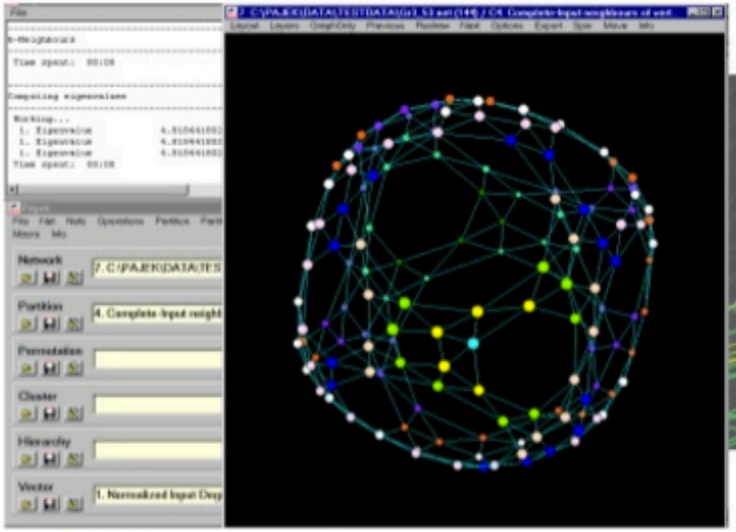
Directed Graph

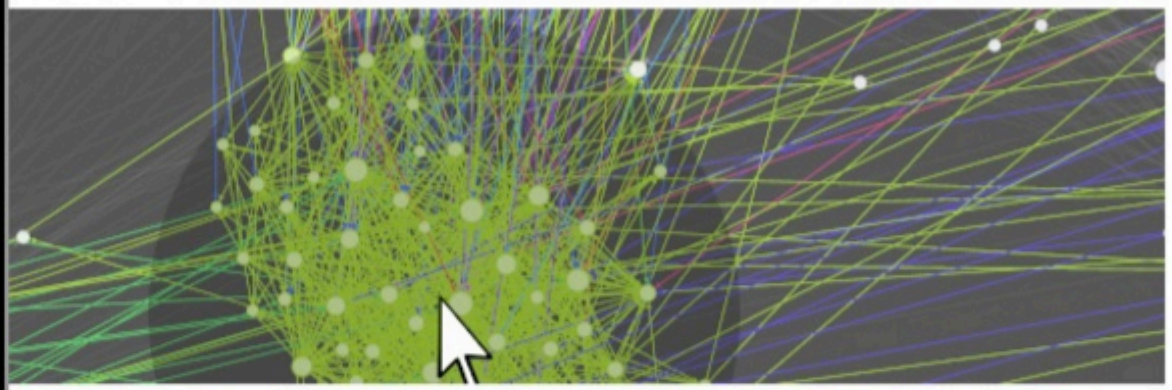


Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Measurements of Social Network Analysis

Exploratory Network Analysis





1 see the network

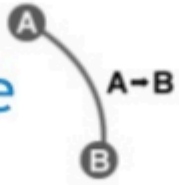
1st graph viz tool: Pajek (1996)
Vladimir Batagelj, Andrej Mrvar

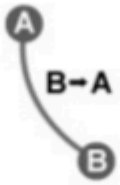
2 interact in real time

Gephi prototype (2008)
group, filter, compute metrics...

3 build a visual language

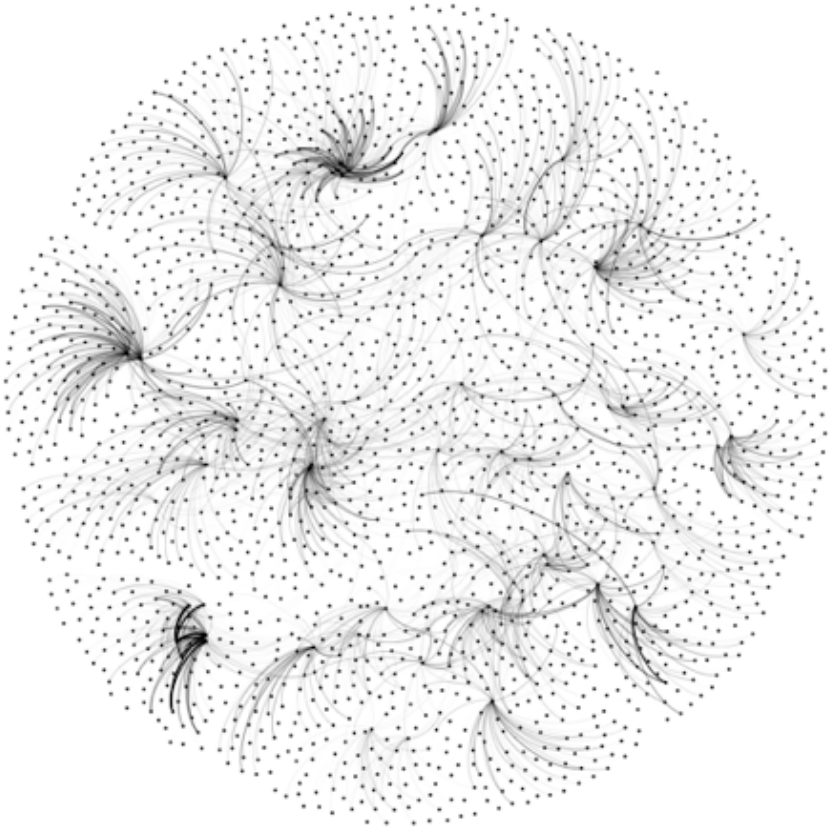
size by rank, color by partition,
label, curved edges, thickness...

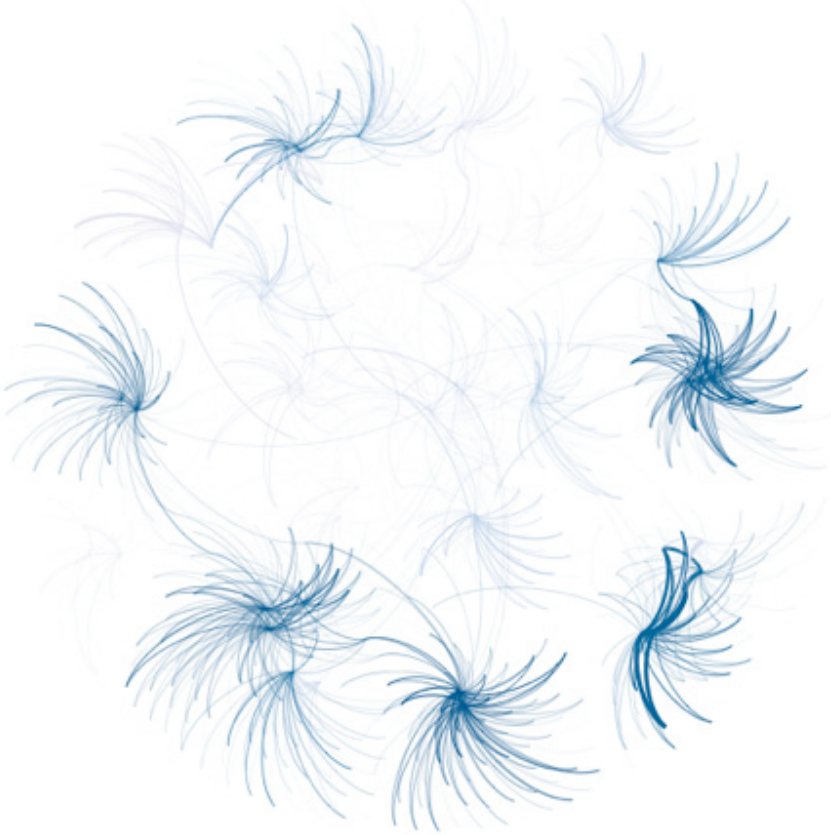




Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

Looking for a “Simple Small Truth”? What Data Visualization Should Do?





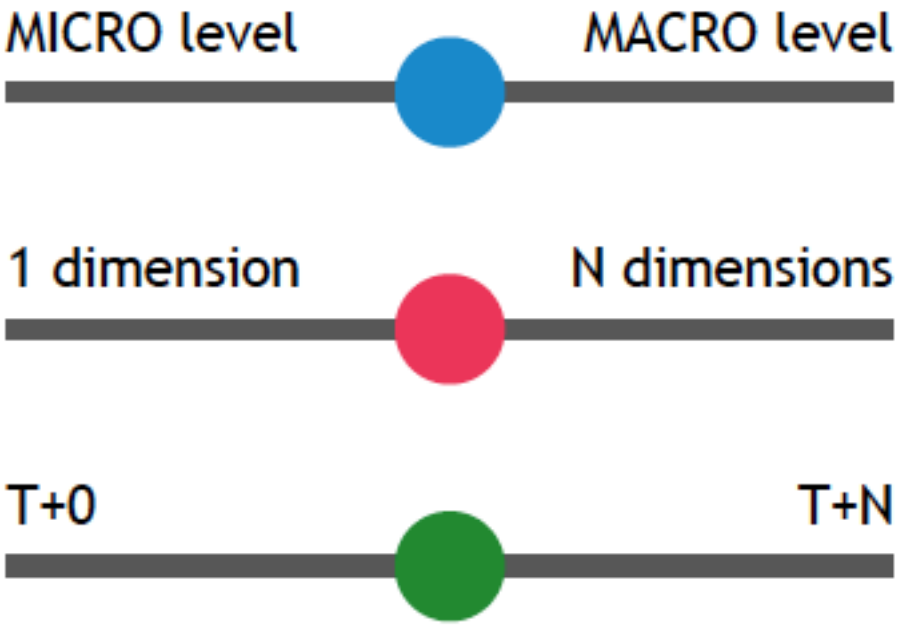
1. Make complex things **simple**
2. Extract **small** information from large data
3. Present **truth**, do not deceive

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

Measurements

Looking for Orderness in Data

Make varying 3 cursors simultaneously to extract meaningful patterns



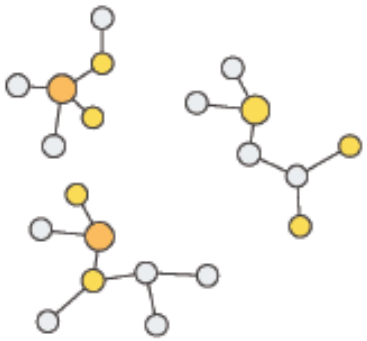
at different levels

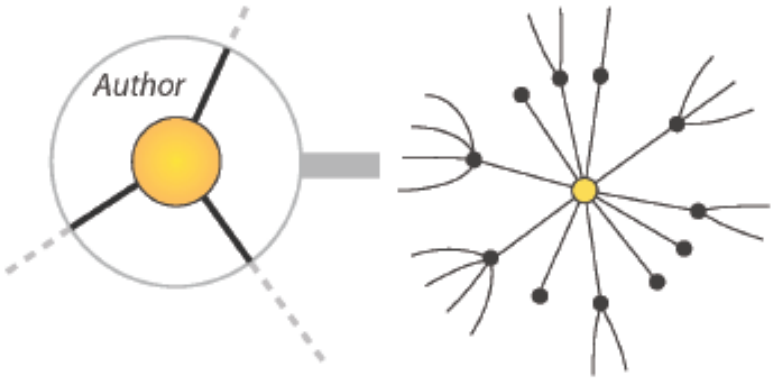
on multiple dimensions

at time scale

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

“Zoom” cursor on Quantitative Data





Global
- connectivity
- density
- centralization

Local
- communities
- bridges between communities
- local centers vs periphery

Individual
- centrality
- distances
- neighborhood
- location
- local authority vs hub

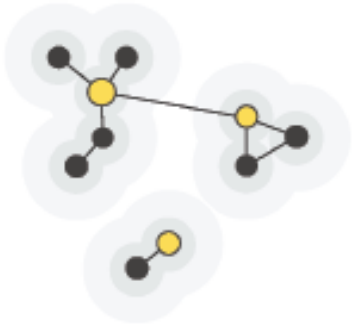
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

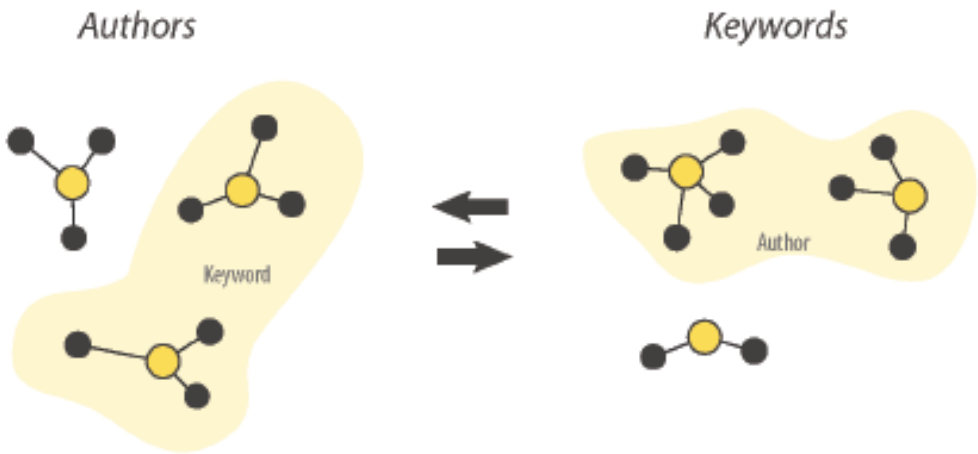
“Crossing” cursor on Quantitative Data

Social
- who with whom
- communities
- brokerage
- influence and power
- homophily

Semantic
- topics
- thematic clusters

Geographic
- spatial phenomena





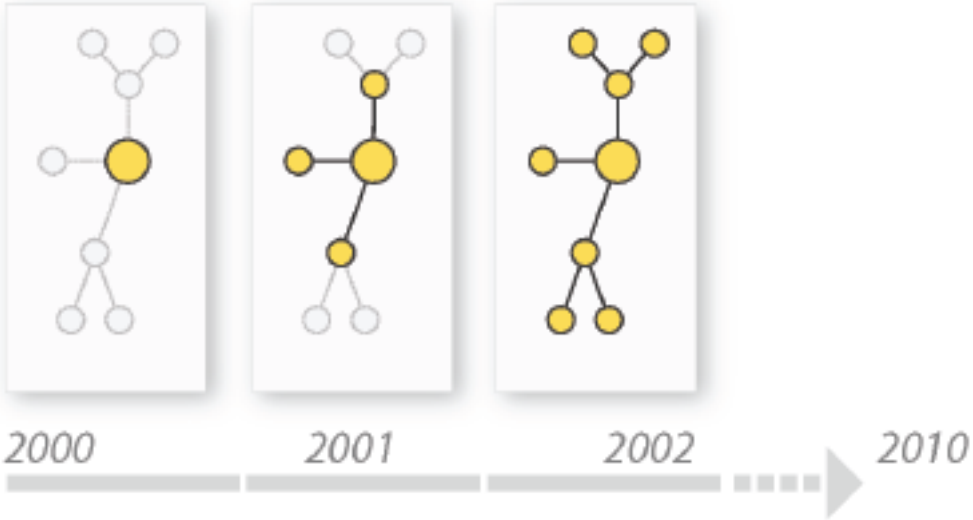
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

“Timeline” cursor on Temporal Data

Evolution of social ties

Evolution of communities

Evolution of topics



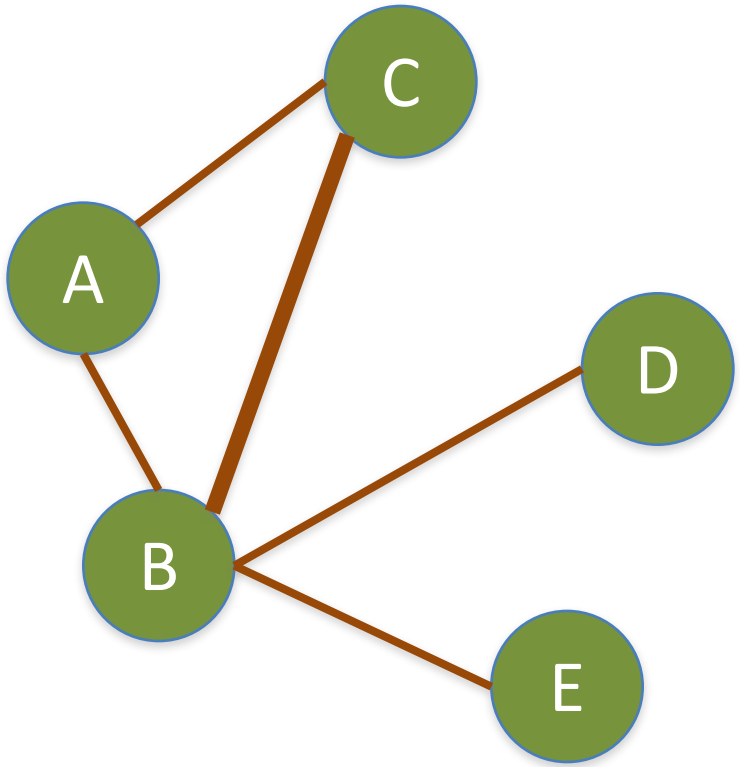
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

SNA Guideline

# nodes	
1 - 100	lists + edges in bonus, focus on qualitative data
100 - 1,000	How attributes explain the structure? • easy to read, "obvious" patterns • focus on entities (in context) • metrics are tools to describe the graph (centrality, bridging...) • links help to build and interpret categories of entities challenge: mix attribute crossing and connectivity
1,000 - 50,000	How the structure explains attributes? • hard to read, problem of "**hidden signals**": track patterns with various layouts and filtering • focus on structures • metrics are tools to build the graph (cosine similarity...) • categories help to understand the structure challenge: pattern recognition
> 50,000	require high computational power

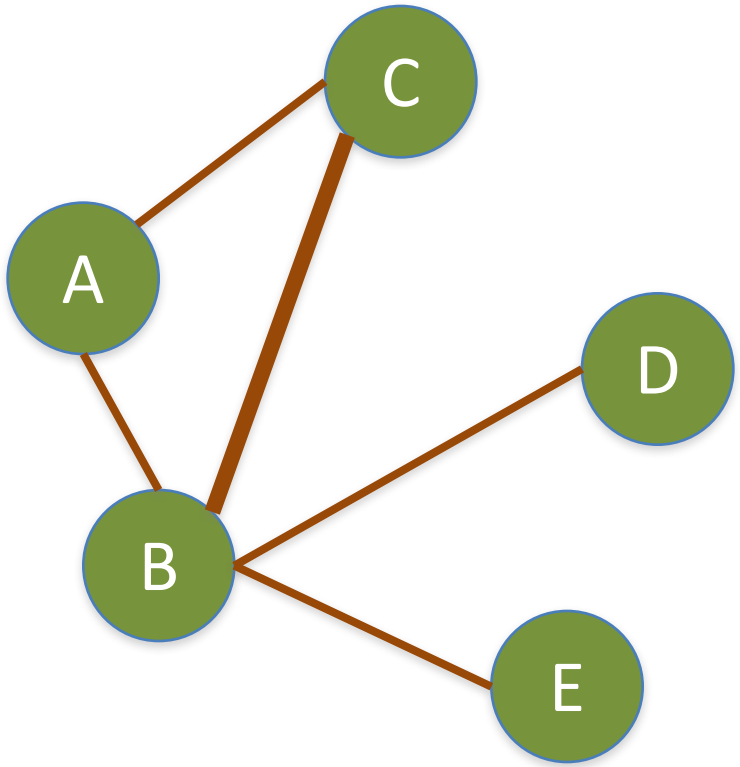
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

Degree



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Degree



A: 2

B: 4

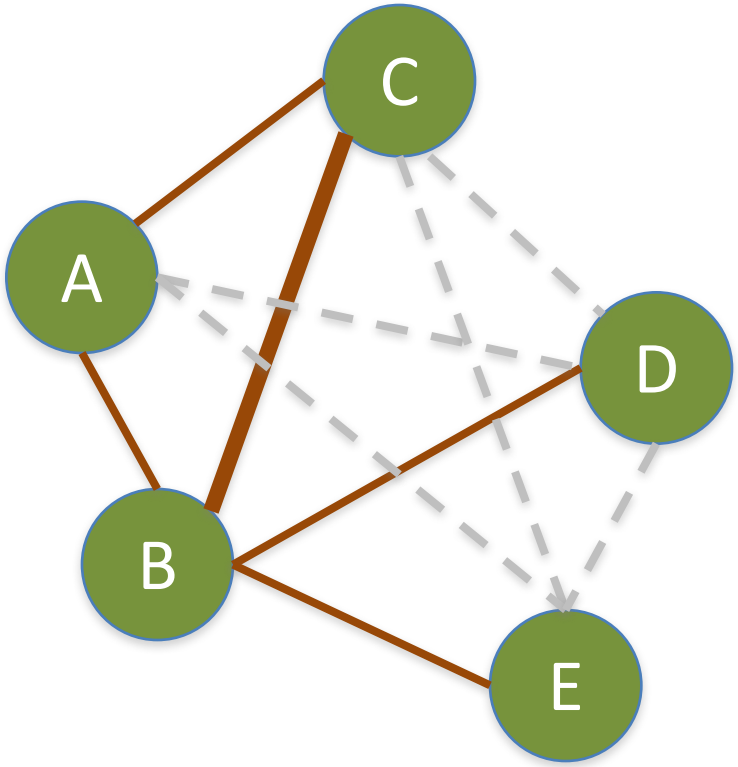
C: 2

D:1

E: 1

Source: https://www.youtube.com/watch?v=89mxOdwPfxA

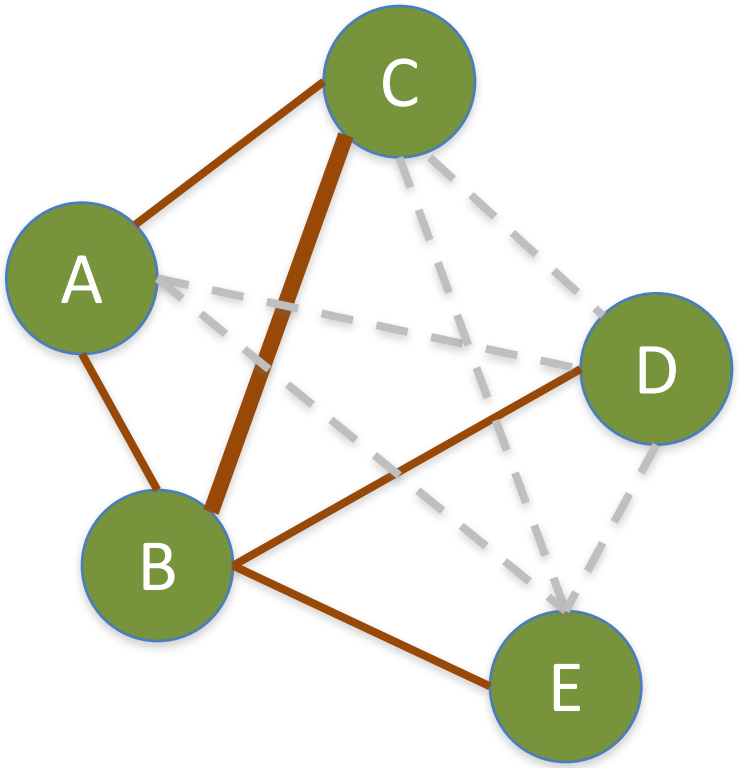
Density



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

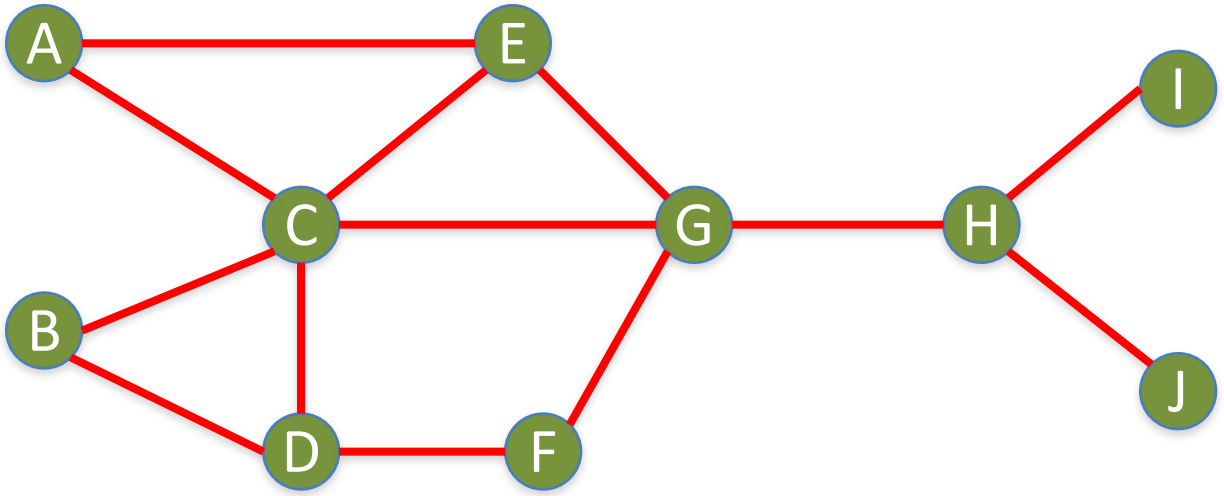
Density

Edges (Links): 5
Total Possible Edges: 10
Density: 5/10 = 0.5



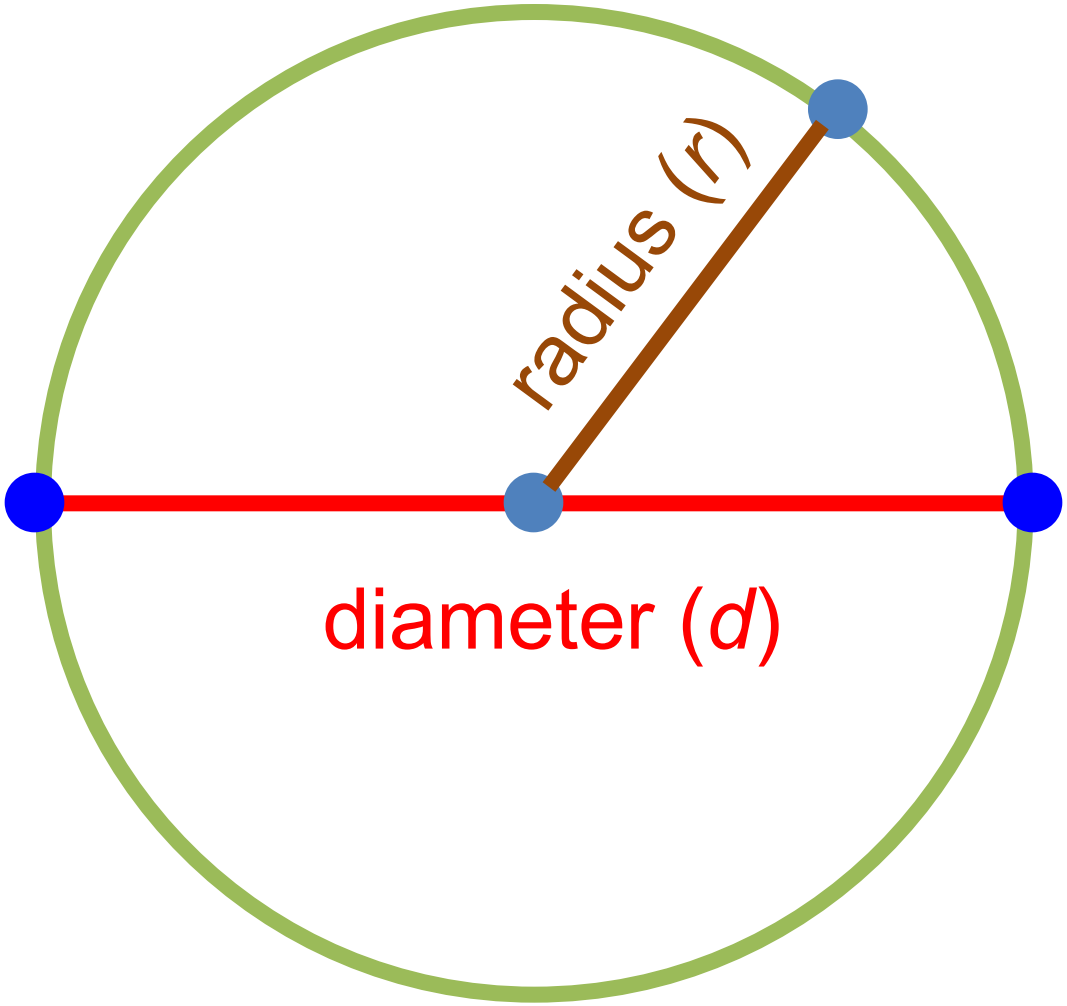
Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Density

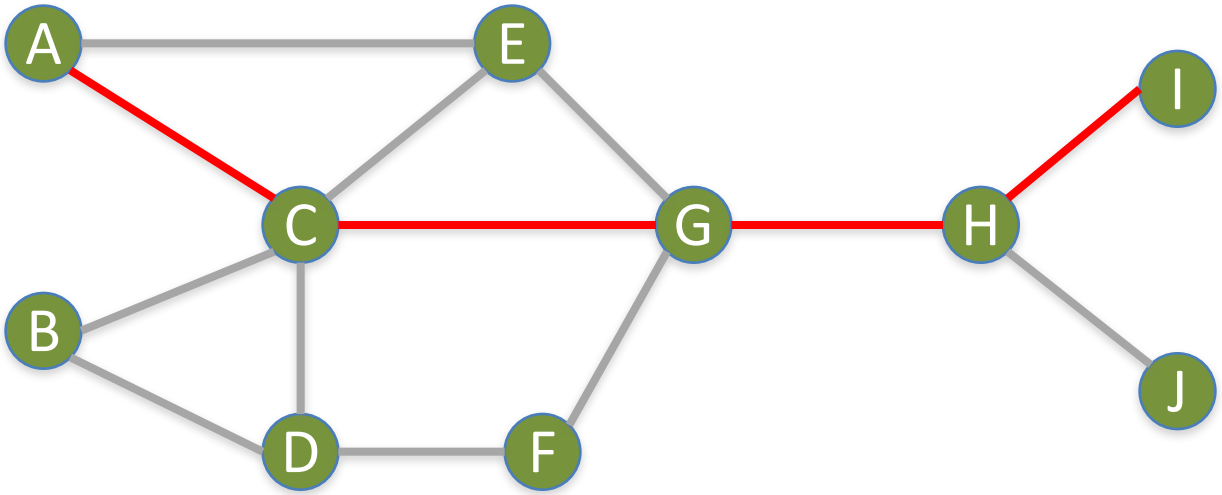


Nodes (n): 10
Edges (Links): 13
Total Possible Edges: (n * (n-1)) / 2 = (10 * 9) / 2 = 45
Density: 13/45 = 0.29

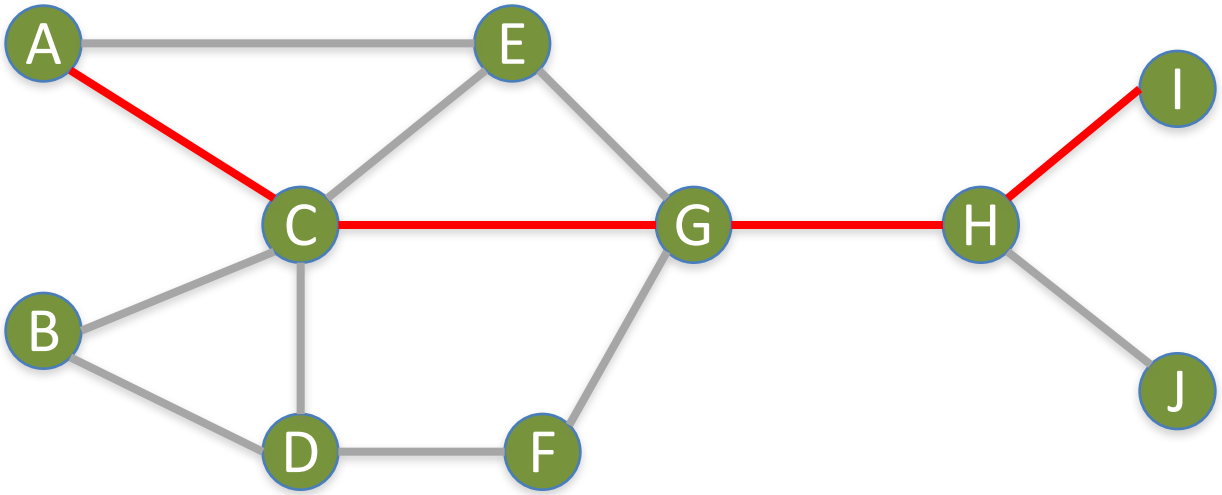
Diameter



Diameter

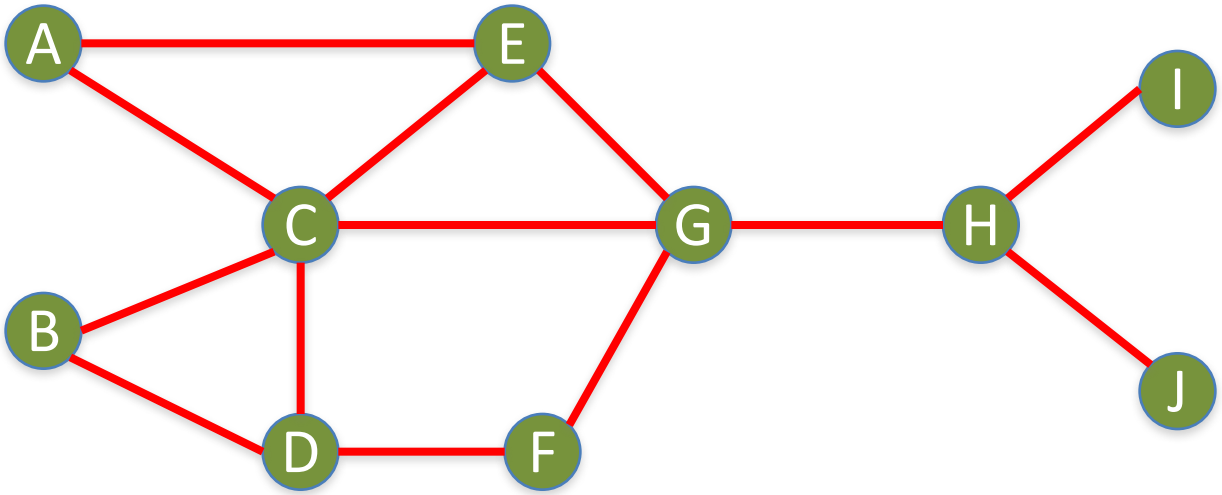


Diameter
Geodesic Path (Shortest Path)



A→ I : Diameter = 4

Which Node is Most Important?



Centrality

- Important or prominent actors are those that are linked or involved with other actors extensively.
- A person with extensive contacts (links) or communications with many other people in the organization is considered more important than a person with relatively fewer contacts.
- The links can also be called **ties**.
  A central actor is one involved in many ties.

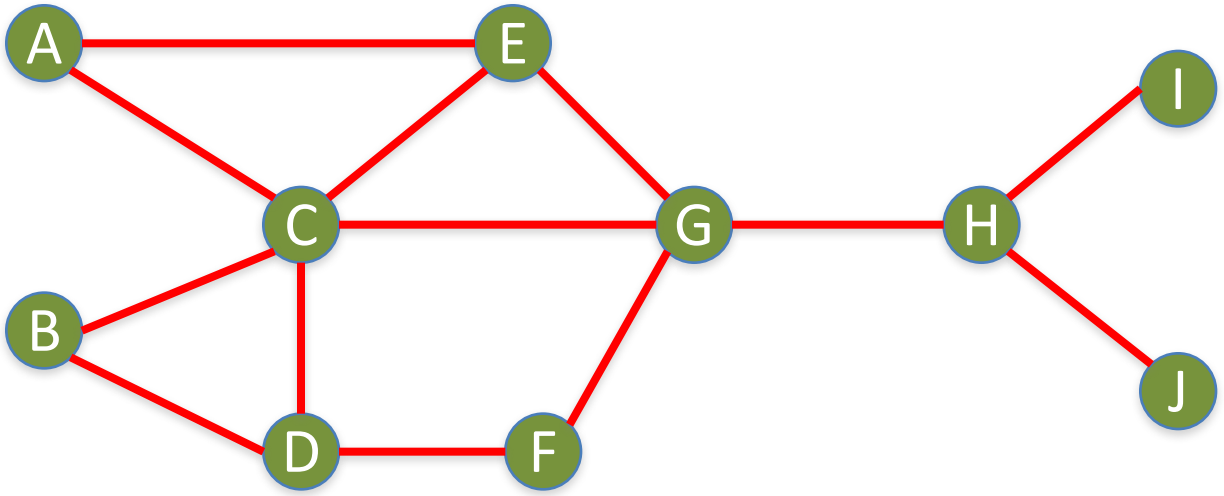
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data"

Social Network Analysis (SNA)

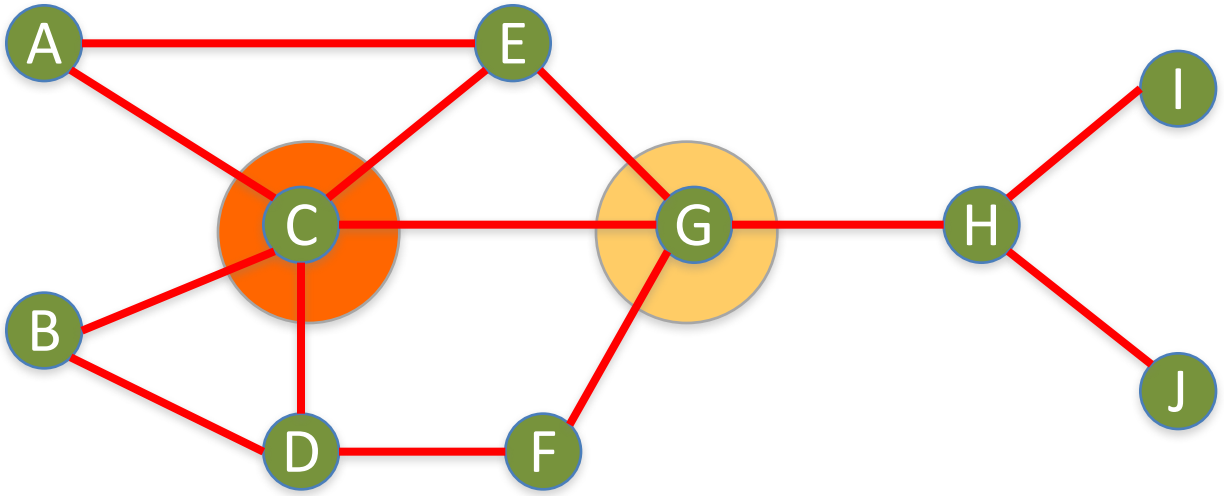
- Degree Centrality
- Betweenness Centrality
- Closeness Centrality

Degree Centrality

Social Network Analysis: Degree Centrality



Social Network Analysis: Degree Centrality



Node	Score	Standardized Score
A	2	2/10 = 0.2
B	2	2/10 = 0.2
C	**5**	**5/10 = 0.5**
D	3	3/10 = 0.3
E	3	3/10 = 0.3
F	2	2/10 = 0.2
G	**4**	**4/10 = 0.4**
H	3	3/10 = 0.3
I	1	1/10 = 0.1
J	1	1/10 = 0.1

Betweenness Centrality

Betweenness centrality:

Connectivity

Number of shortest paths going through the actor

Betweenness Centrality

$$C_B(i) = \sum_{j<k} g_{ik}(i) / g_{jk}$$

Where g_{jk} = the number of shortest paths connecting jk

$g_{jk}(i)$ = the number that actor i is on.

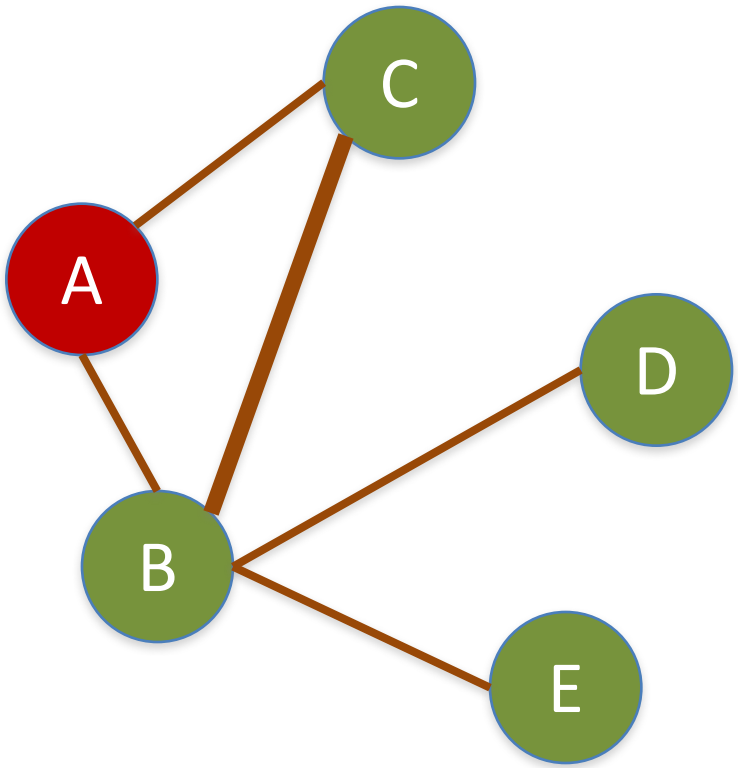
Normalized Betweenness Centrality

$$C'_B(i) = C_B(i) / [(n-1)(n-2)/2]$$

Number of pairs of vertices excluding the vertex itself

Source: https://www.youtube.com/watch?v=RXohUeNCJiU

Betweenness Centrality

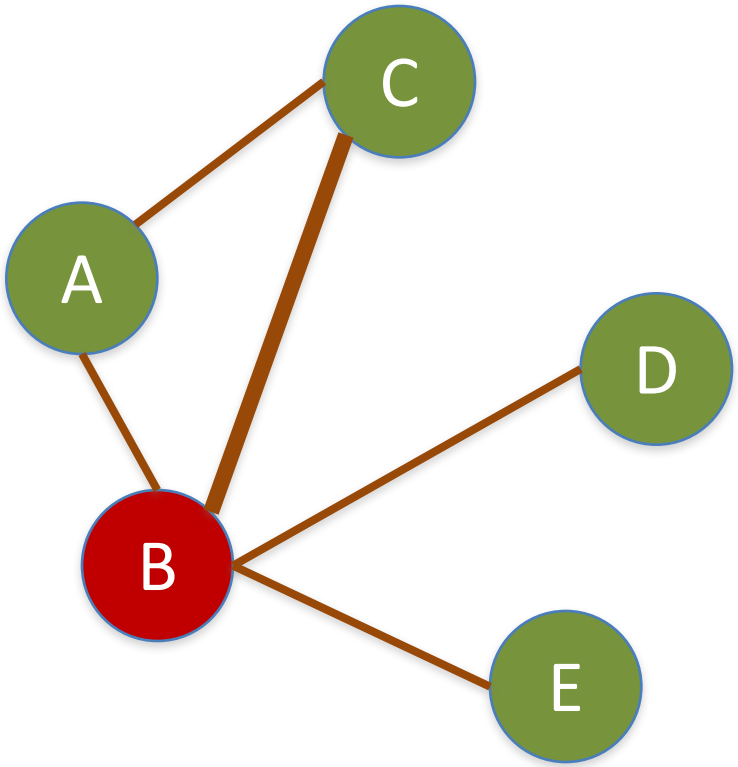


A:
B→C: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
C→D: 0/1 = 0
C→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

A: Betweenness Centrality = 0

Betweenness Centrality

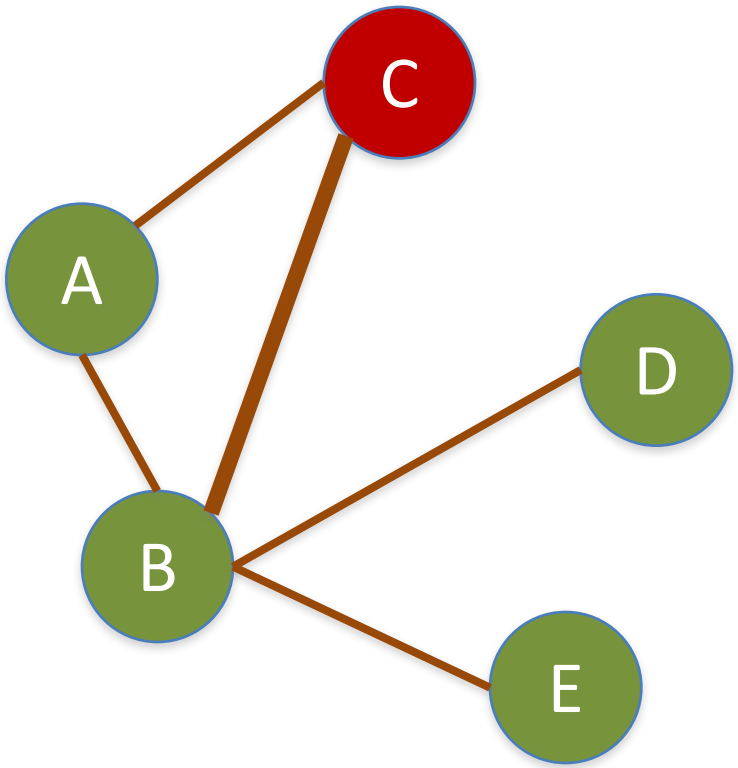


B:
A→C: 0/1 = 0
A→D: 1/1 = 1
A→E: 1/1 = 1
C→D: 1/1 = 1
C→E: 1/1 = 1
D→E: 1/1 = 1

Total: 5

B: Betweenness Centrality = 5

Betweenness Centrality

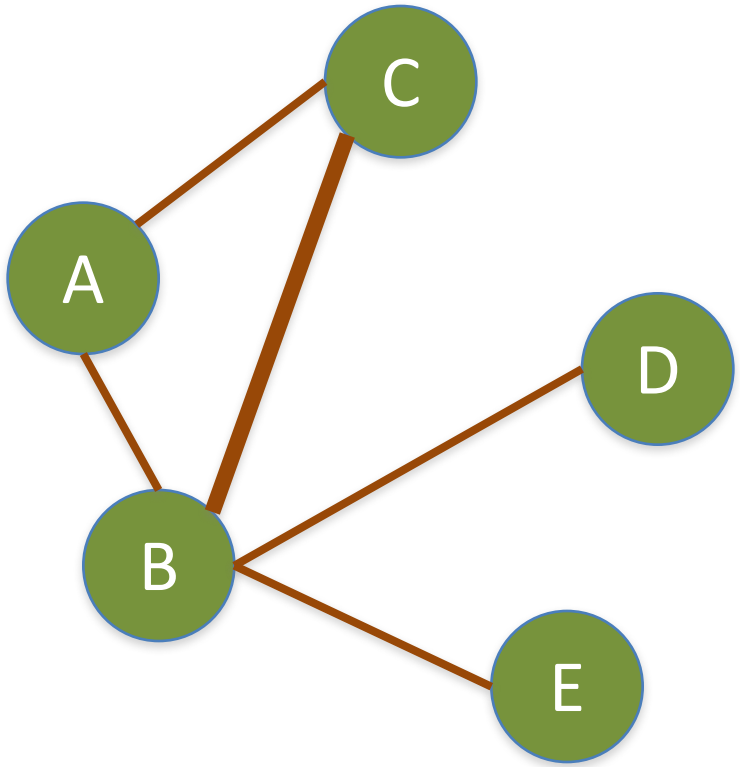


C:
A→B: 0/1 = 0
A→D: 0/1 = 0
A→E: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

C: Betweenness Centrality = 0

Betweenness Centrality



A: 0

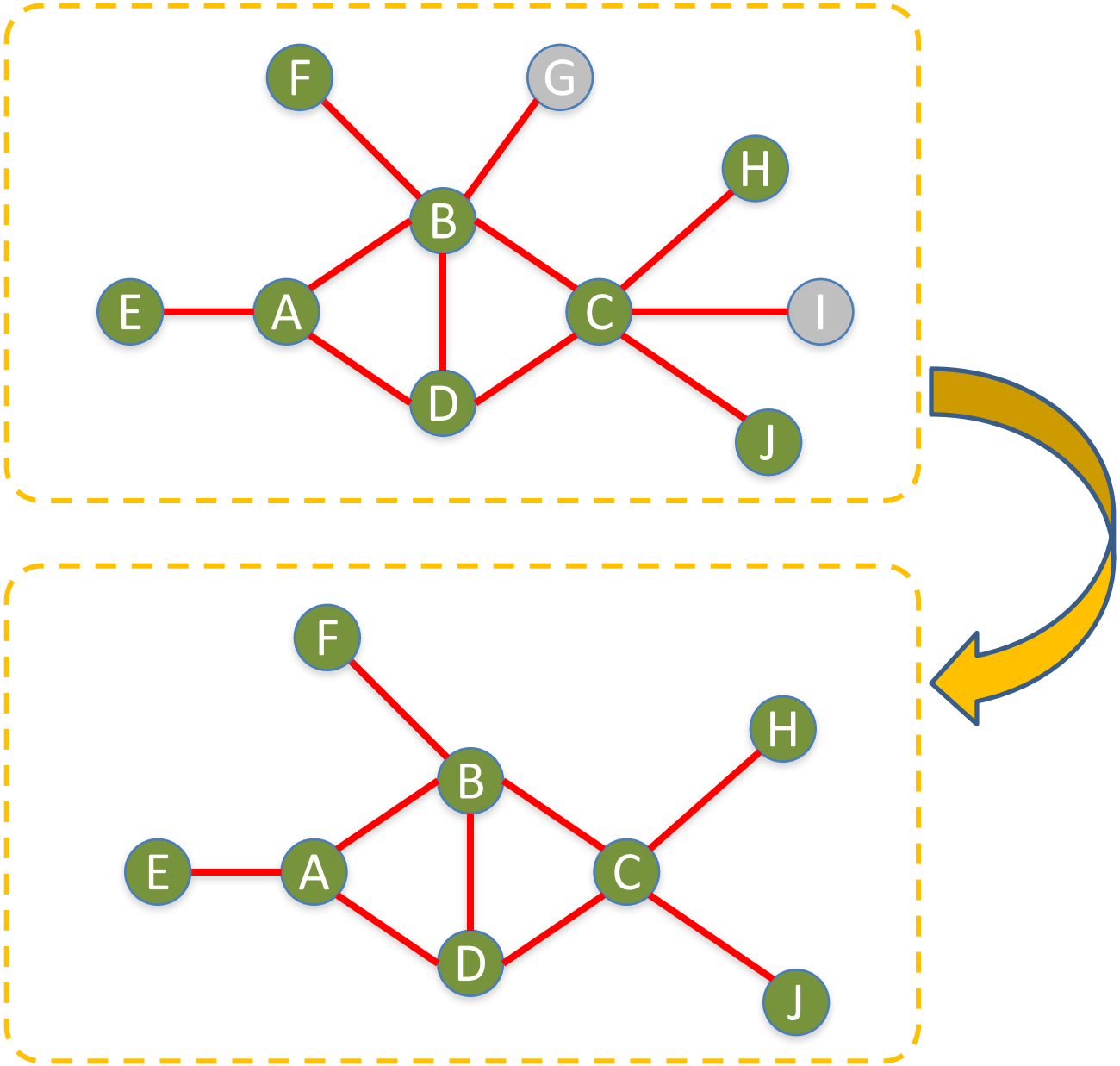
B: 5

C: 0

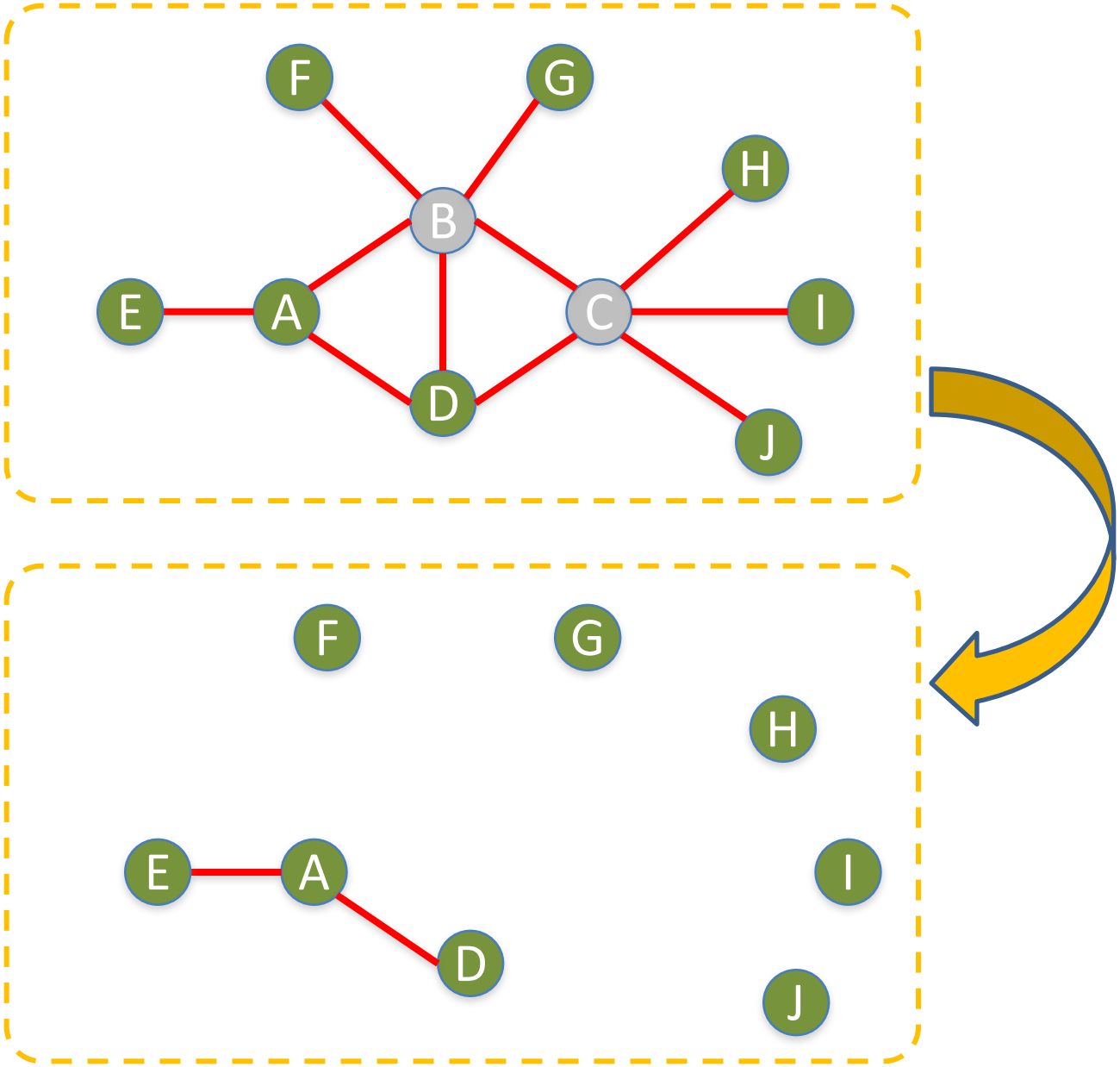
D: 0

E: 0

Which Node is Most Important?

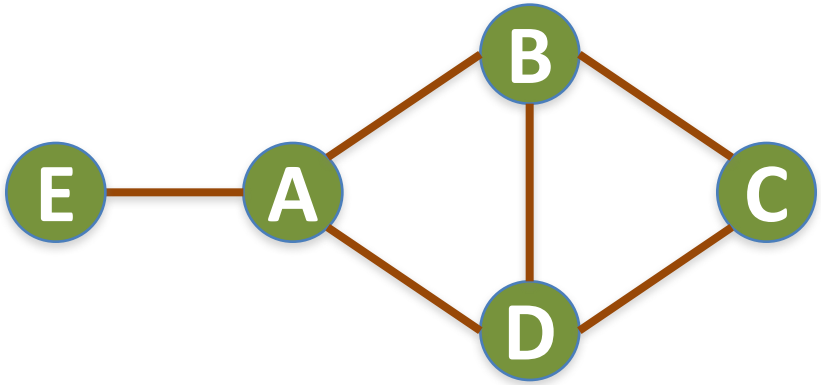


Which Node is Most Important?

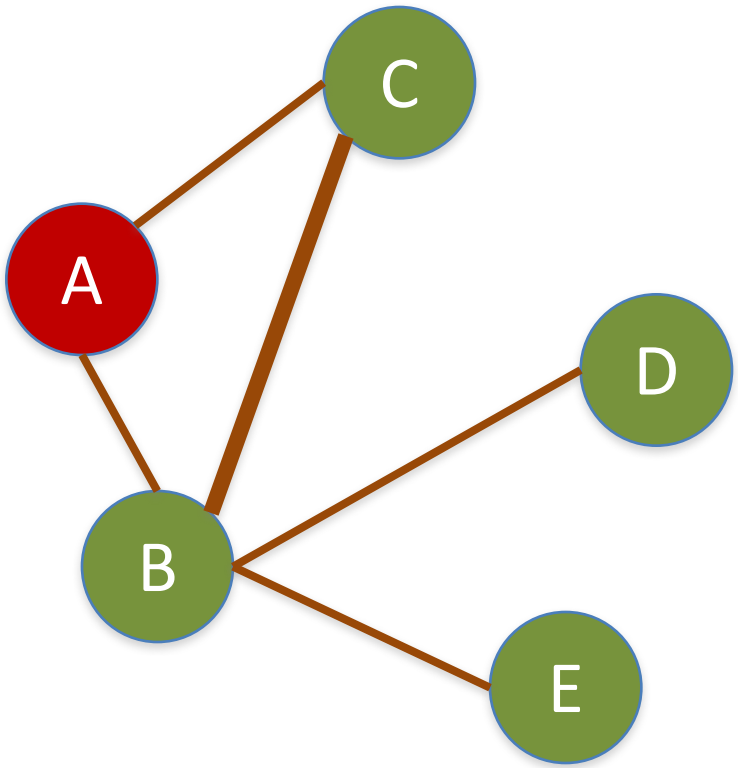


Betweenness Centrality

$$C_B(i) = \sum_{j<k} g_{ik}(i) / g_{jk}$$



Betweenness Centrality



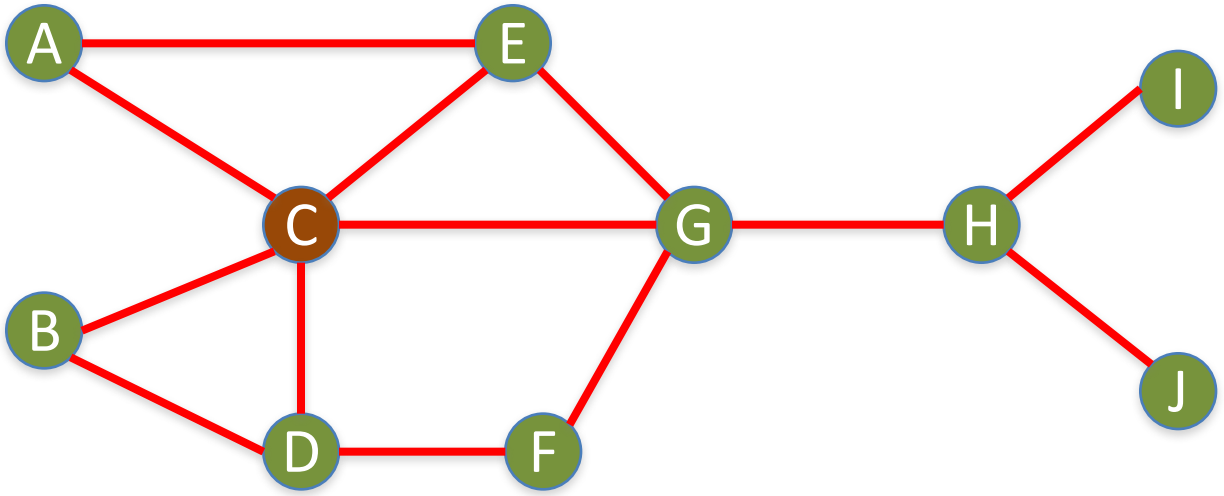
A:
B→C: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
C→D: 0/1 = 0
C→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

A: Betweenness Centrality = 0

Closeness Centrality

Social Network Analysis: Closeness Centrality

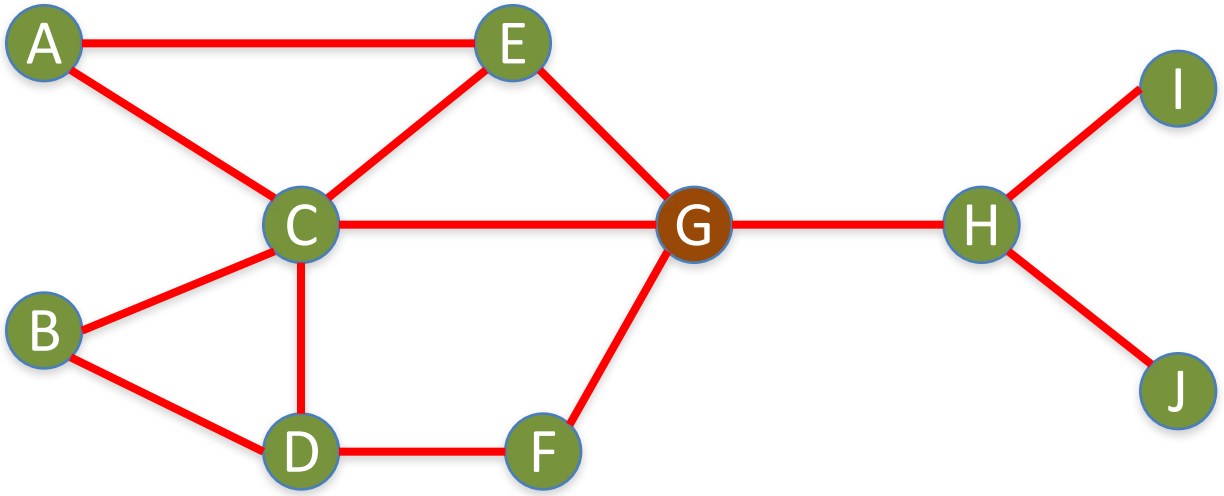


C→A: 1
C→B: 1
C→D: 1
C→E: 1
C→F: 2
C→G: 1
C→H: 2
C→I: 3
C→J: 3

Total=15

C: Closeness Centrality = 15/9 = 1.67

Social Network Analysis: Closeness Centrality

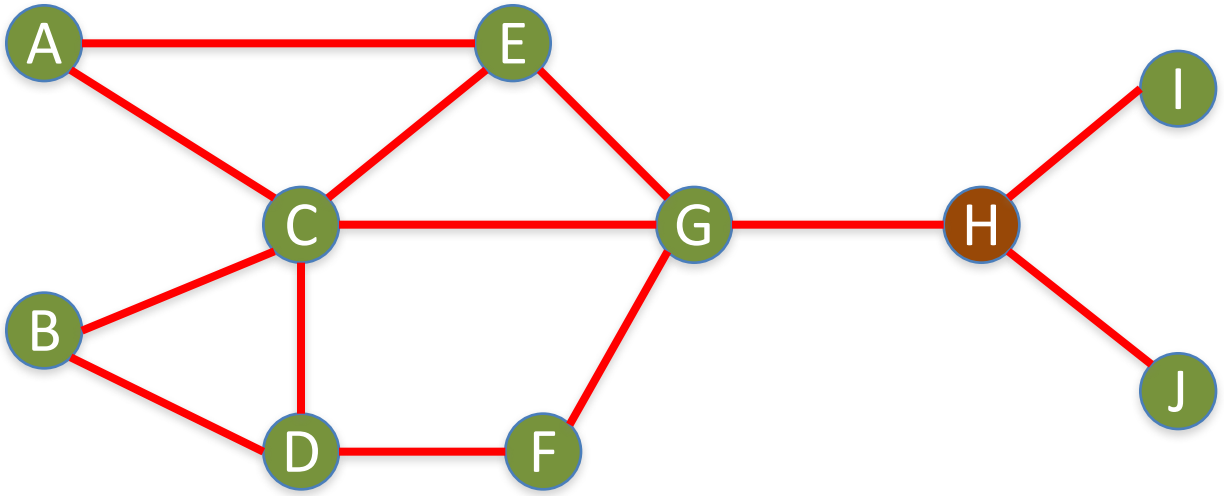


G→A: 2
G→B: 2
G→C: 1
G→D: 2
G→E: 1
G→F: 1
G→H: 1
G→I: 2
G→J: 2

Total=14

G: Closeness Centrality = 14/9 = 1.56

Social Network Analysis: Closeness Centrality

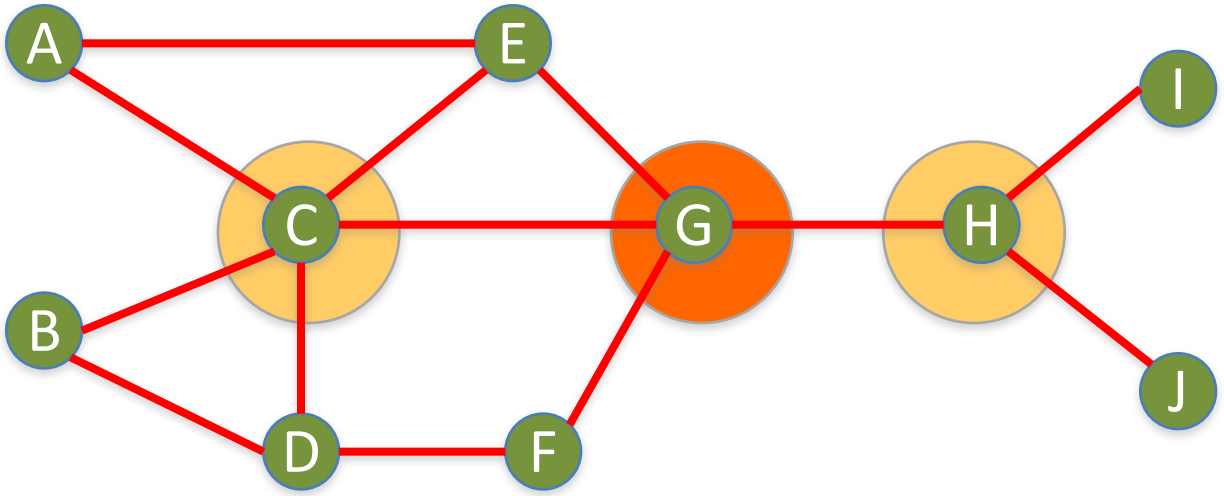


H→A: 3
H→B: 3
H→C: 2
H→D: 2
H→E: 2
H→F: 2
H→G: 1
H→I: 1
H→J: 1

Total=17

H: Closeness Centrality = 17/9 = 1.89

Social Network Analysis: Closeness Centrality



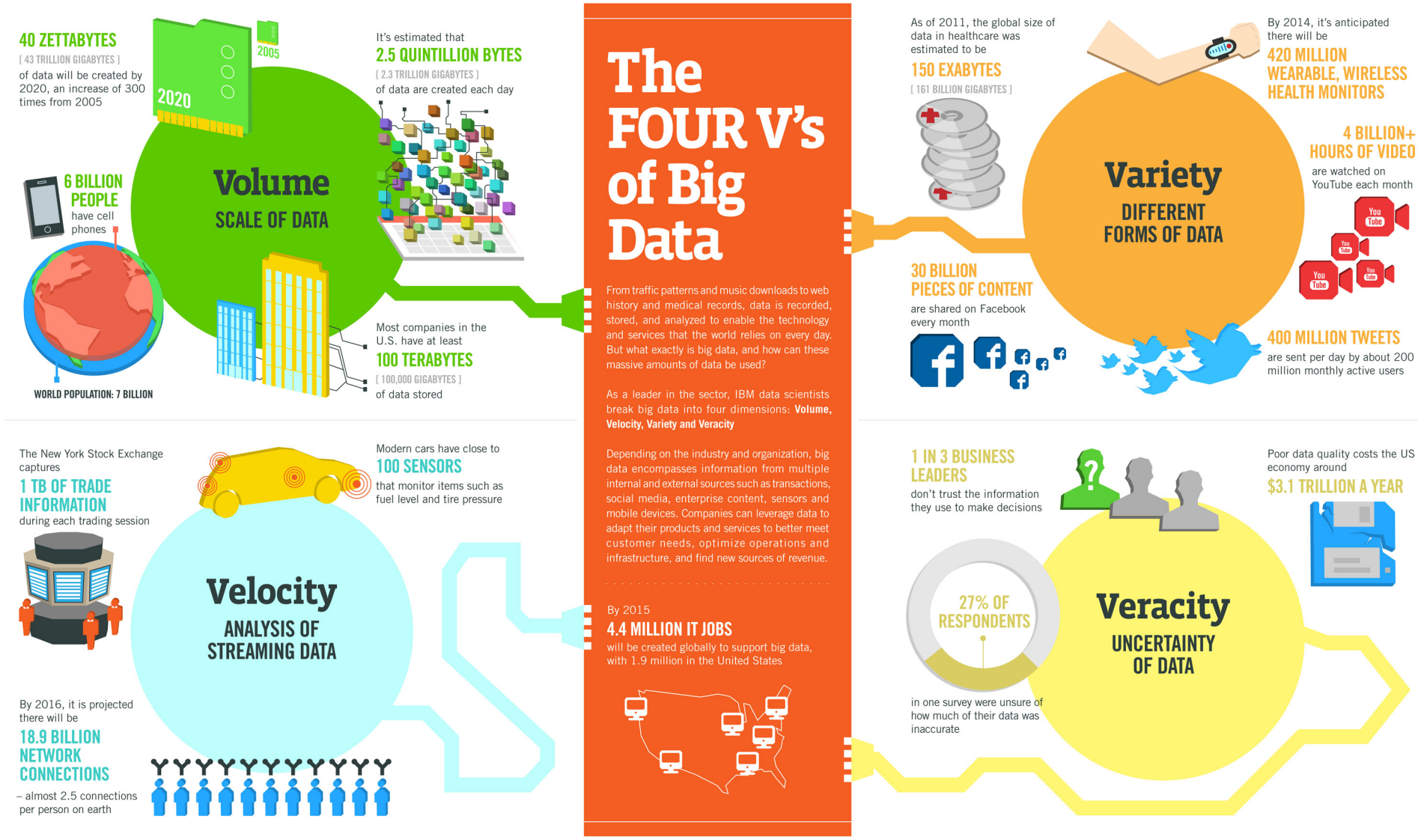
G: Closeness Centrality = 14/9 = 1.56 1

C: Closeness Centrality = 15/9 = 1.67 2

H: Closeness Centrality = 17/9 = 1.89 3

Big Data Analytics

Big Data 4 V



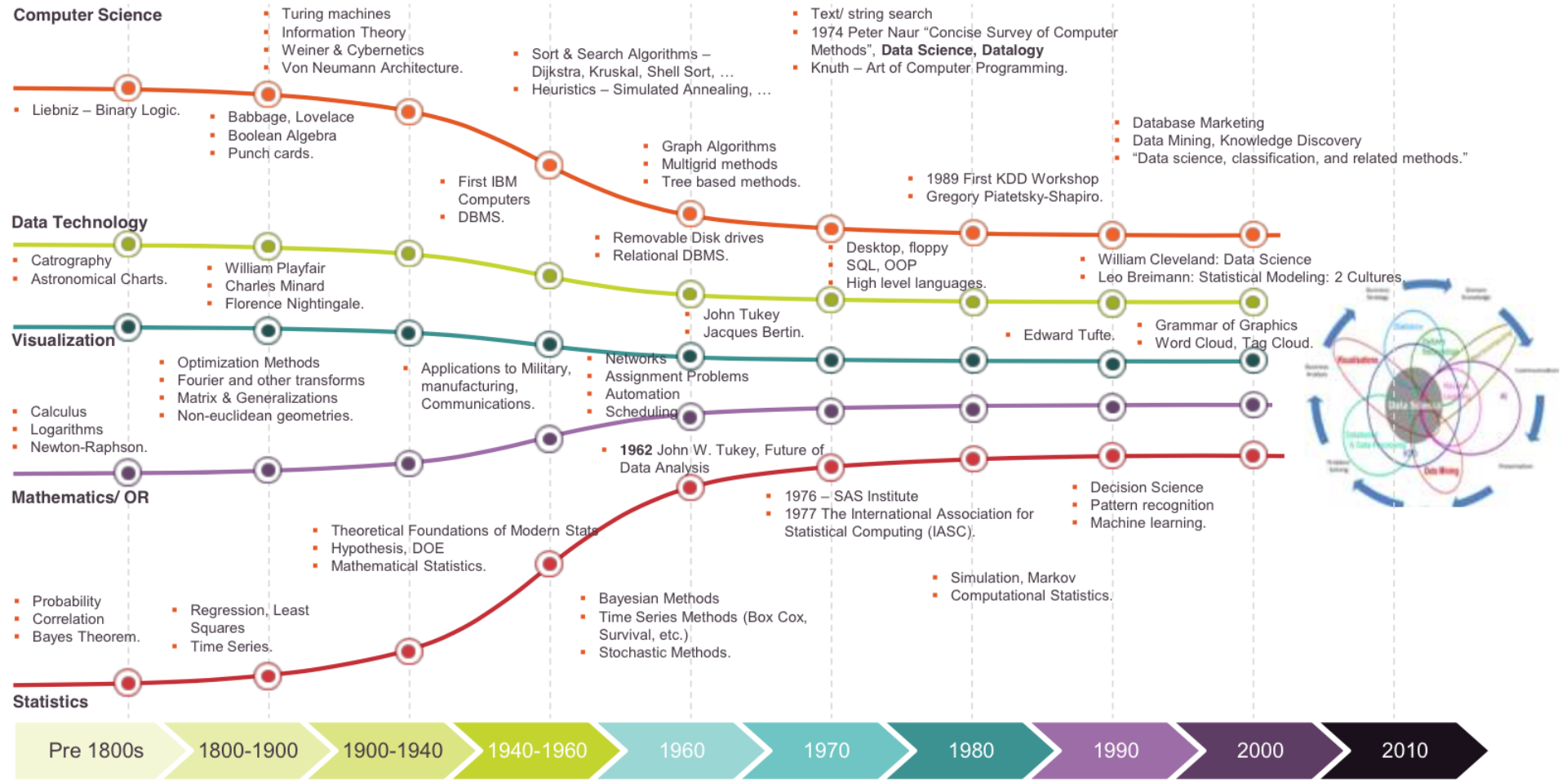
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

IBM

Source: https://www-01.ibm.com/software/data/bigdata/

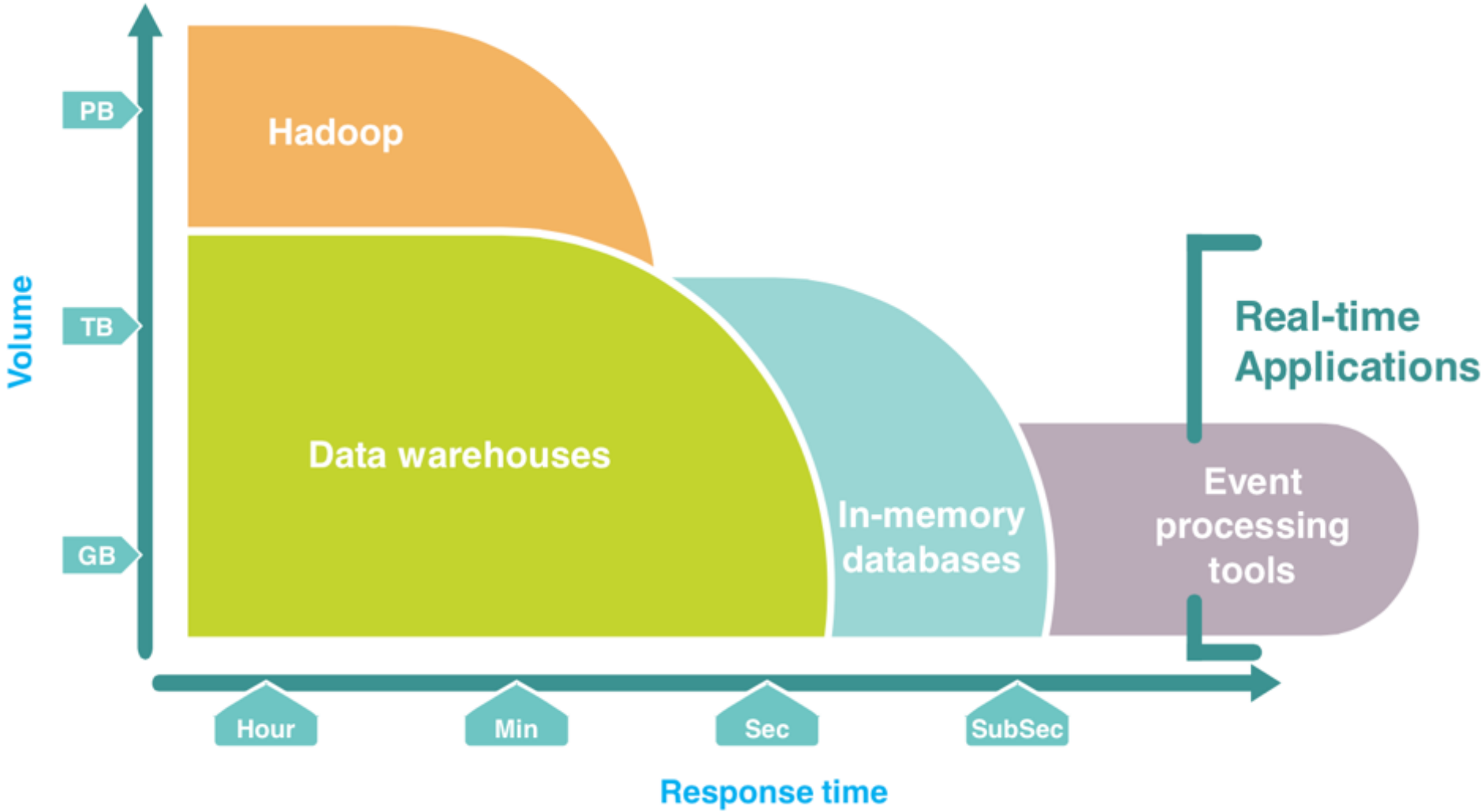
Value

History of Data Science



Source: http://www.kdnuggets.com/2015/02/history-data-science-infographic.html

Big Data Technologies are Enabling a New Approach

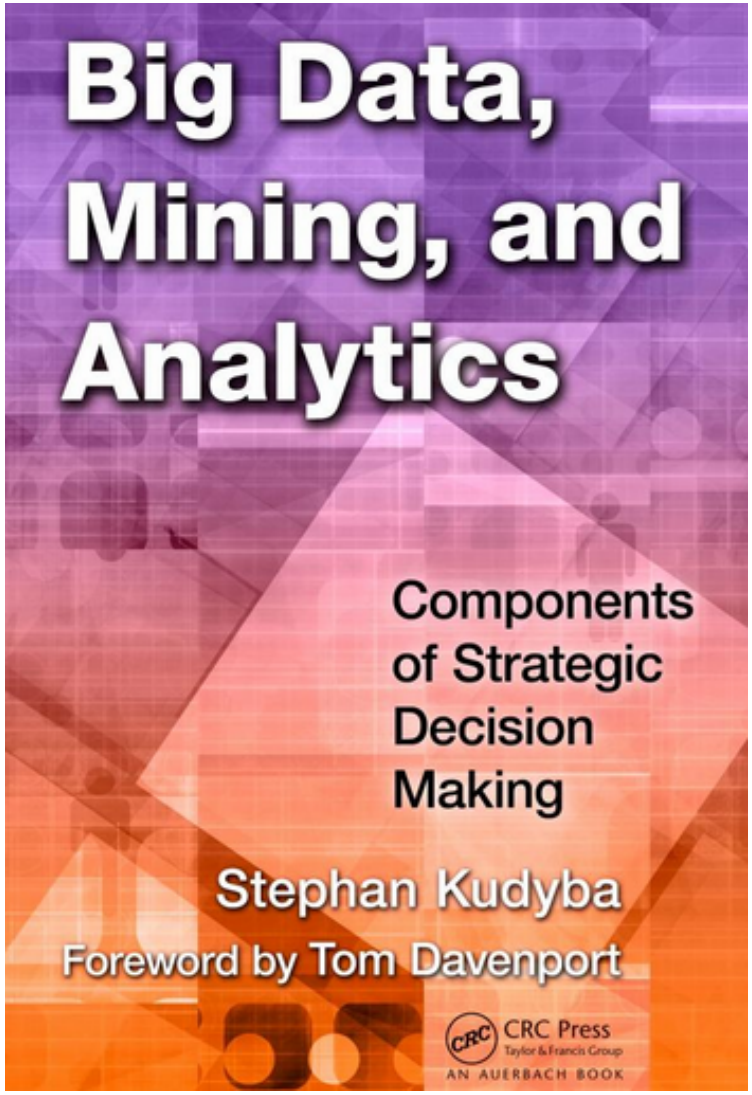


Source: http://www.doclens.com/119898/think-1-13-big-datas-impact-on-analytics/

Big Data Analytics and Data Mining

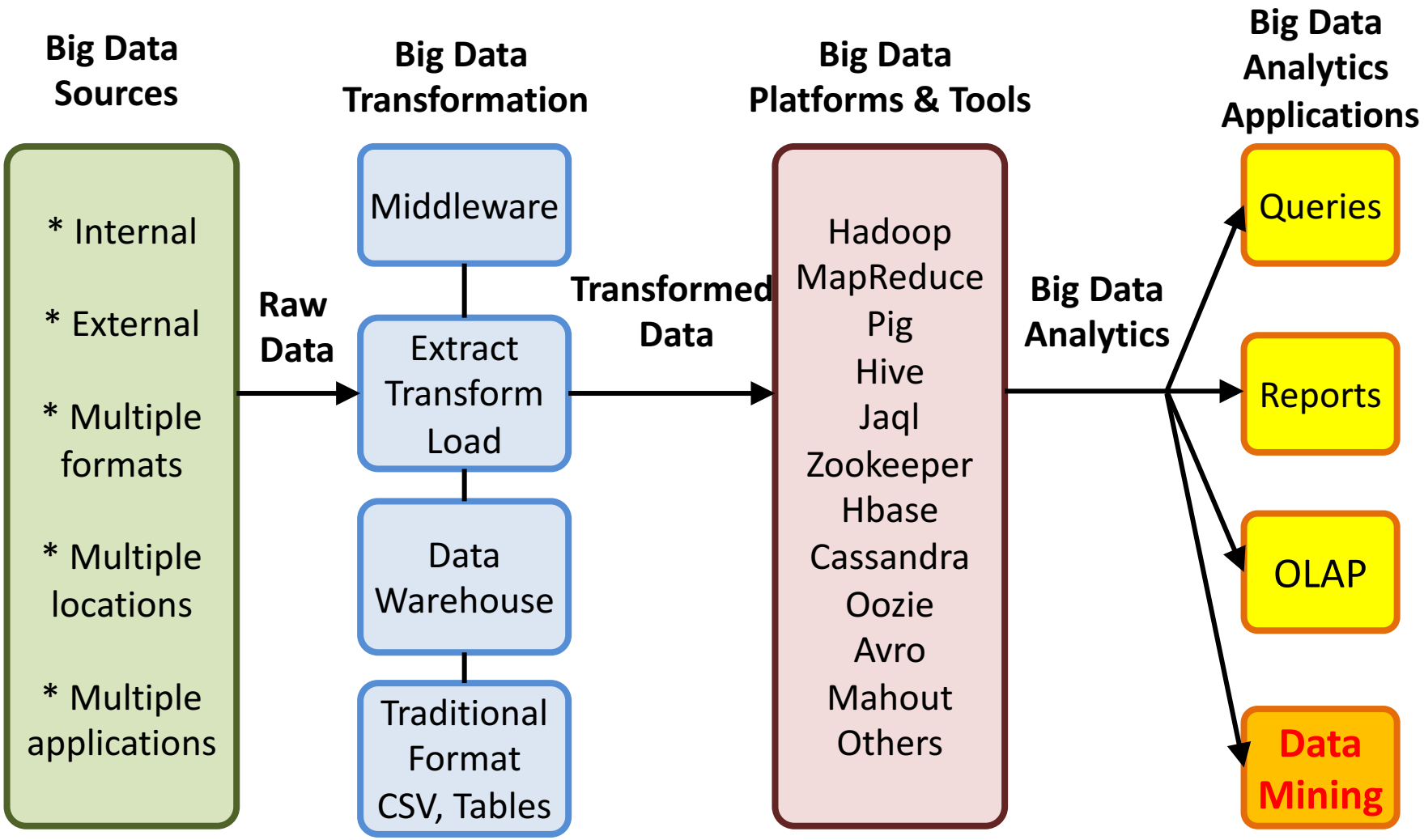
Stephan Kudyba (2014),

Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications



Source: http://www.amazon.com/gp/product/1466568704

Architecture of Big Data Analytics



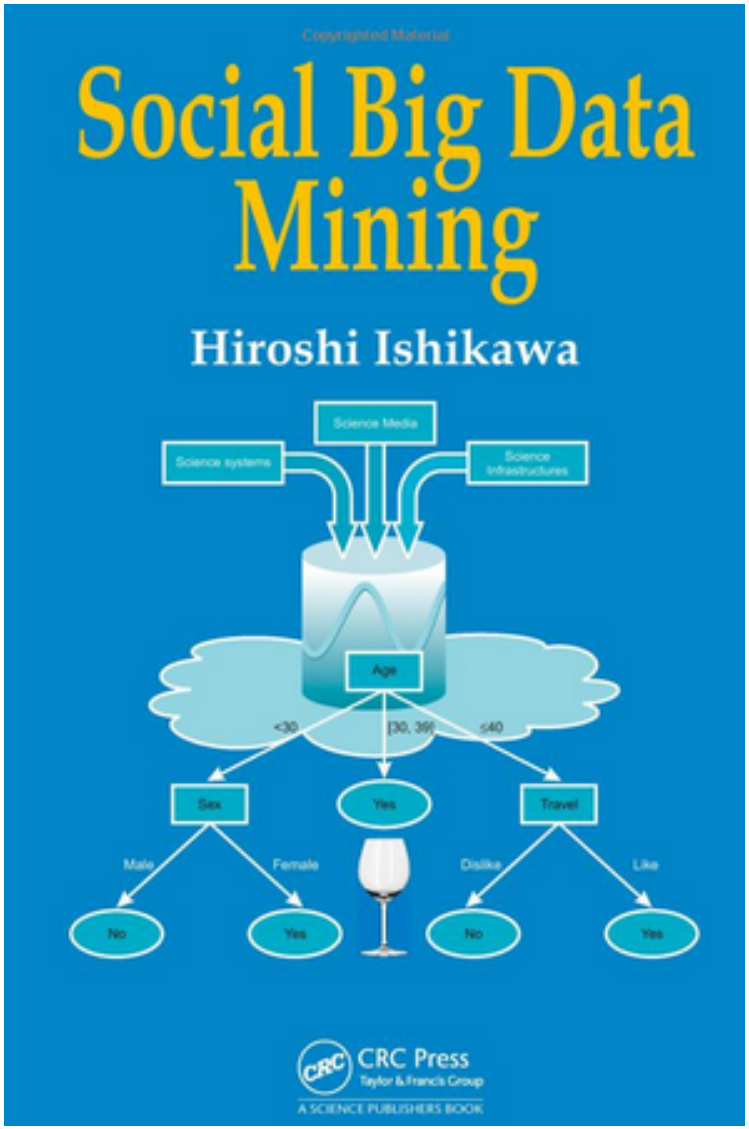
Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Architecture of Big Data Analytics

Source: Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications

Social Big Data Mining

(Hiroshi Ishikawa, 2015)



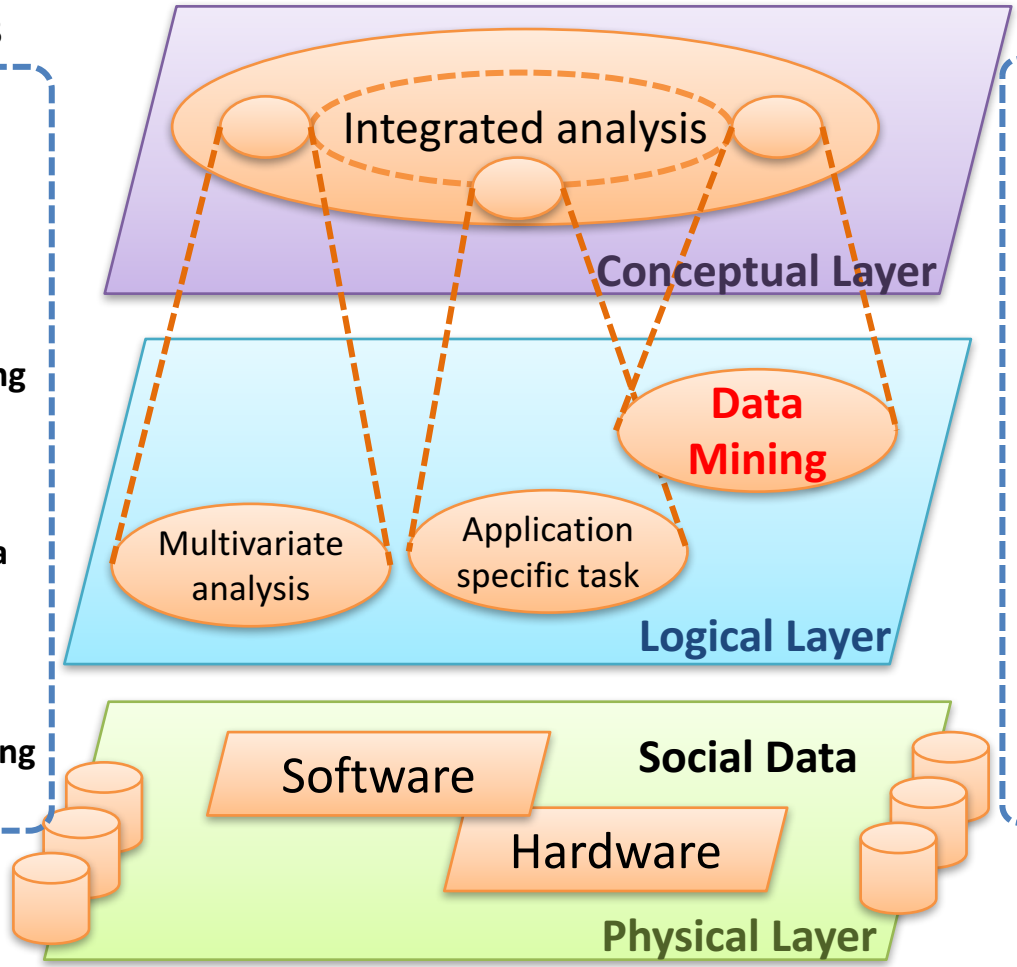
Source: http://www.amazon.com/Social-Data-Mining-Hiroshi-Ishikawa/dp/149871093X

Architecture for Social Big Data Mining

(Hiroshi Ishikawa, 2015)

Enabling Technologies

- Integrated analysis model
- Natural Language Processing
- Information Extraction
- Anomaly Detection
- Discovery of relationships among heterogeneous data
- Large-scale visualization
- Parallel distrusted processing

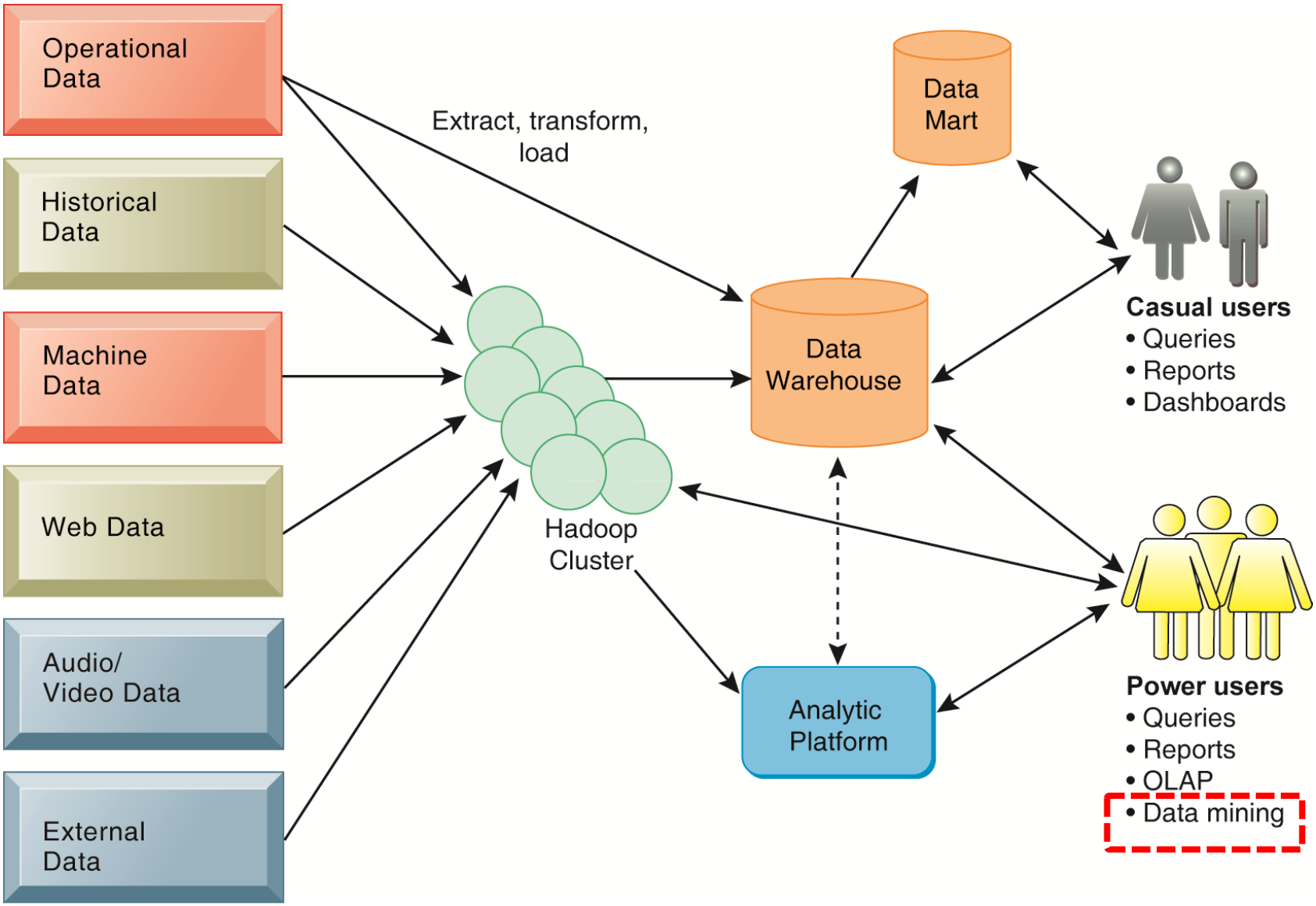


Analysts

- Model Construction
- Explanation by Model
- Construction and confirmation of individual hypothesis
- Description and execution of application-specific task

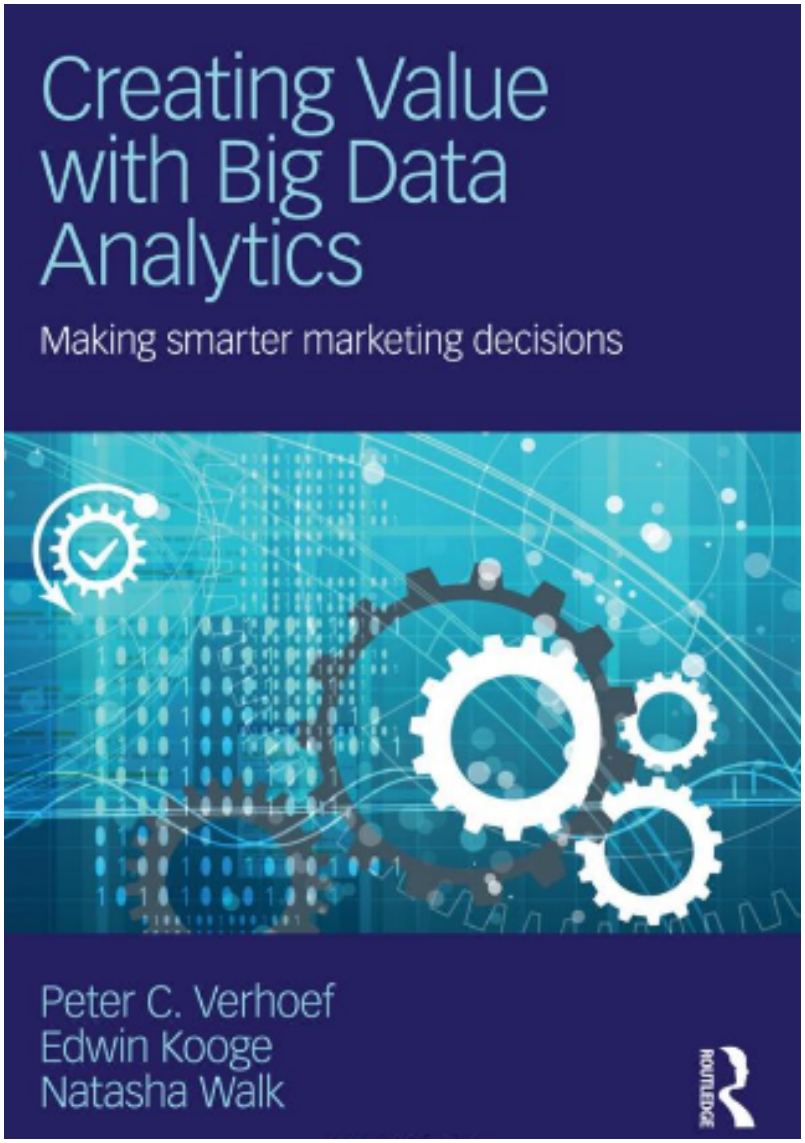
Source: Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press

Business Intelligence (BI) Infrastructure



Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.

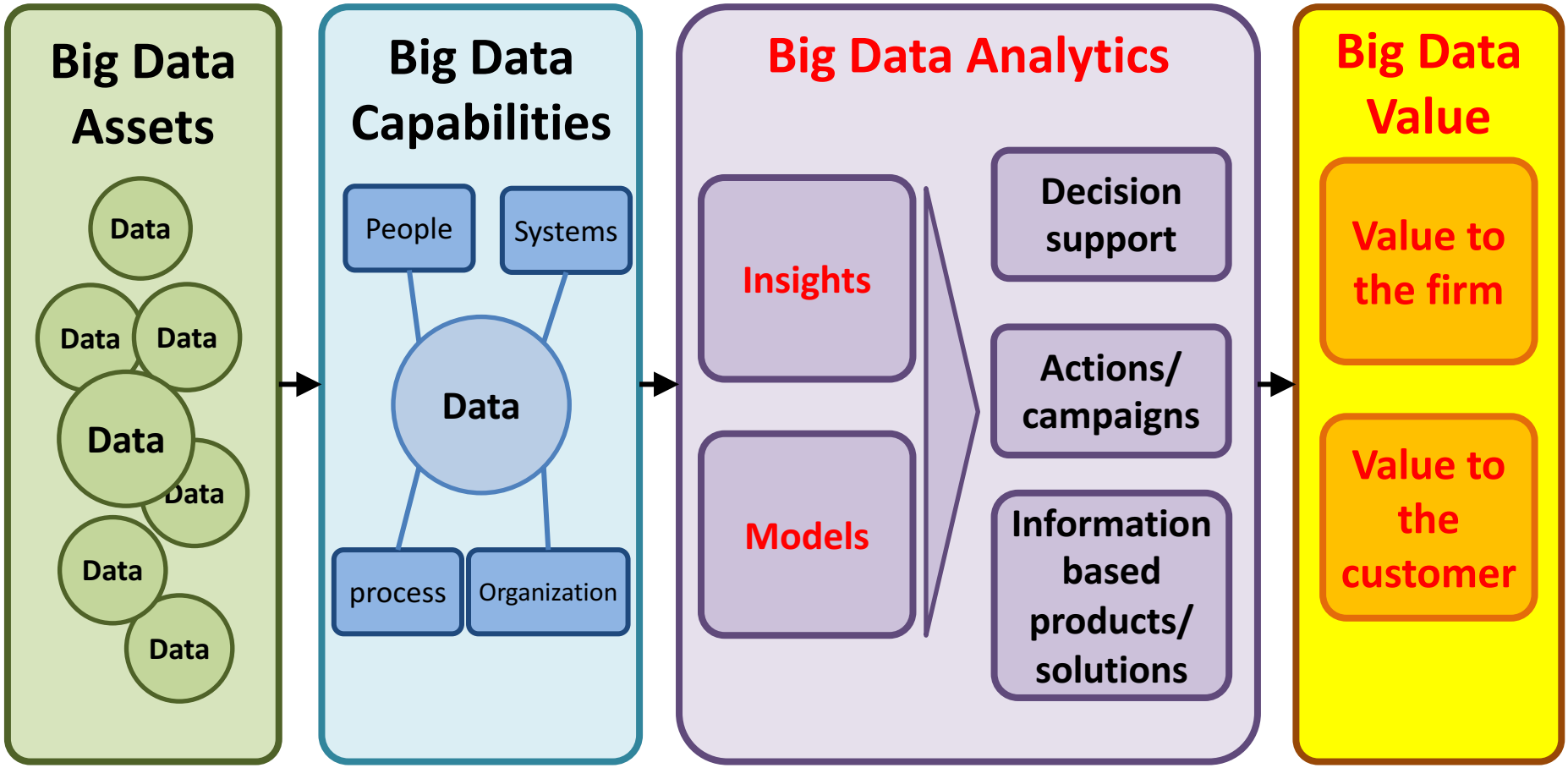
Creating Value with Big Data Analytics: Making Smarter Marketing Decisions, Peter C. Verhoef and Edwin Kooge, Routledge, 2016



Source: https://www.amazon.com/Creating-Value-Big-Data-Analytics/dp/1138837970

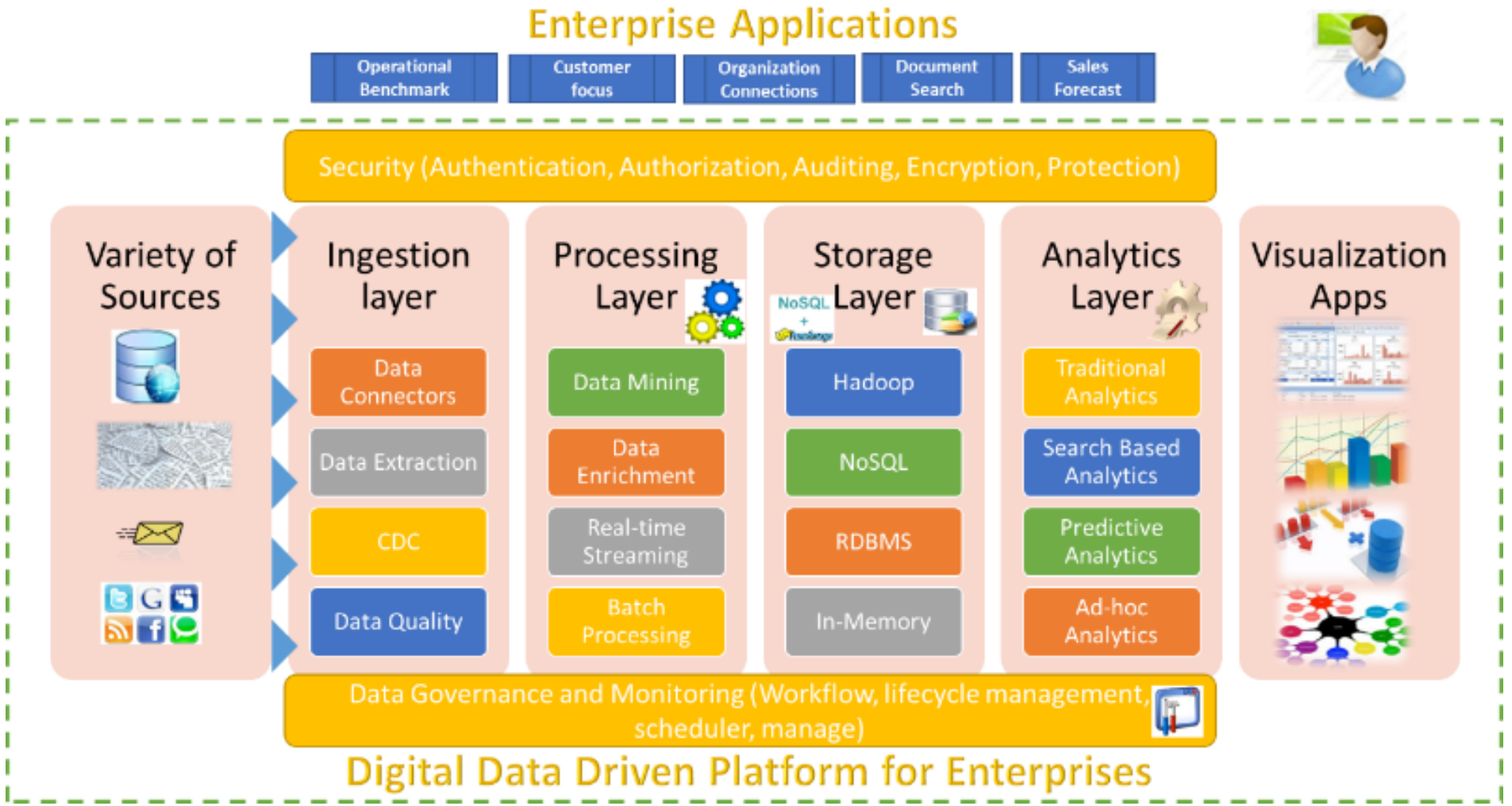
Big Data Value Creation Model

Creating Value with Big Data Analytics: Making Smarter Marketing Decisions



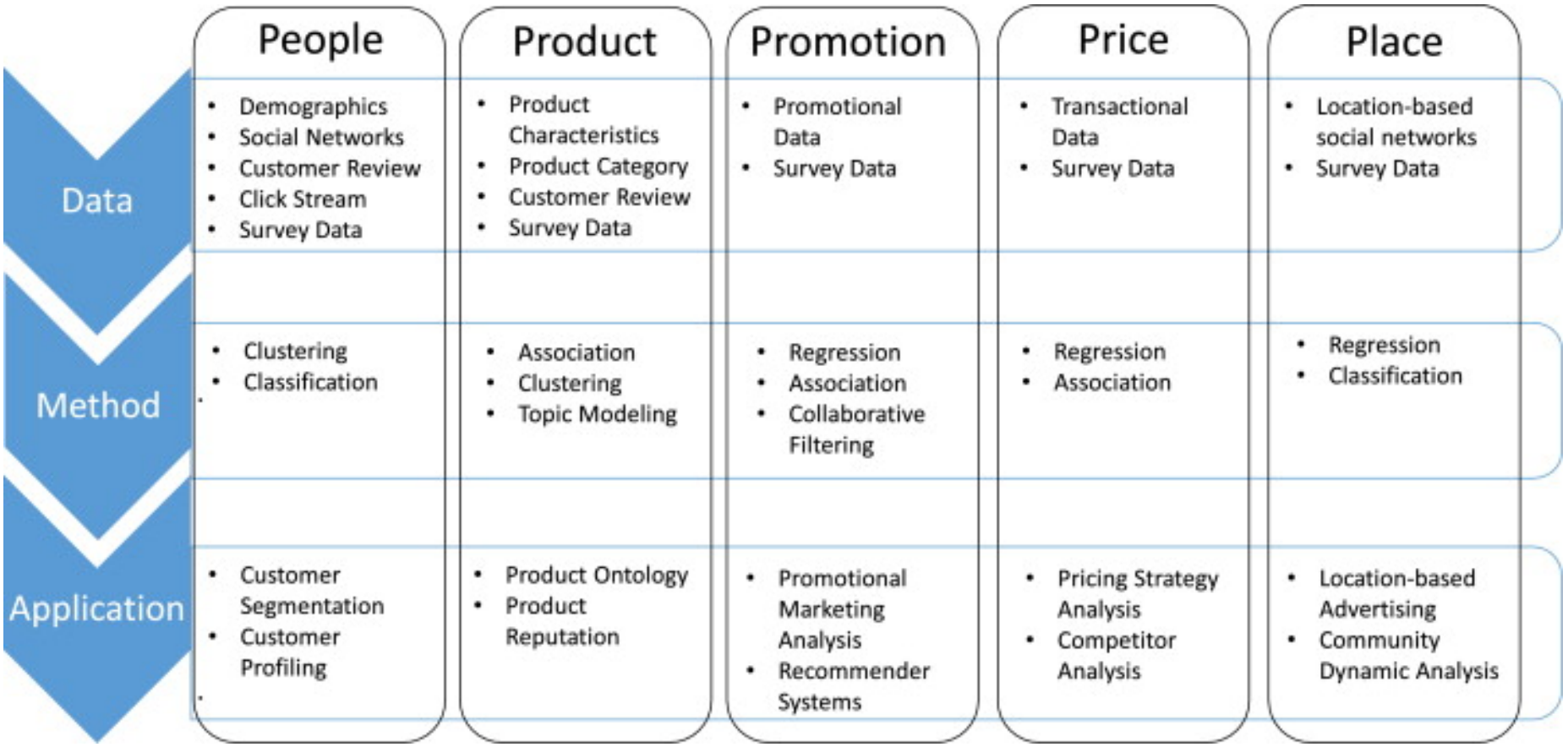
Source: Peter C. Verhoef and Edwin Kooge (2016), Creating Value with Big Data Analytics: Making Smarter Marketing Decisions, Routledge

Digital Data Platform for Enterprises
Big Data Analytics



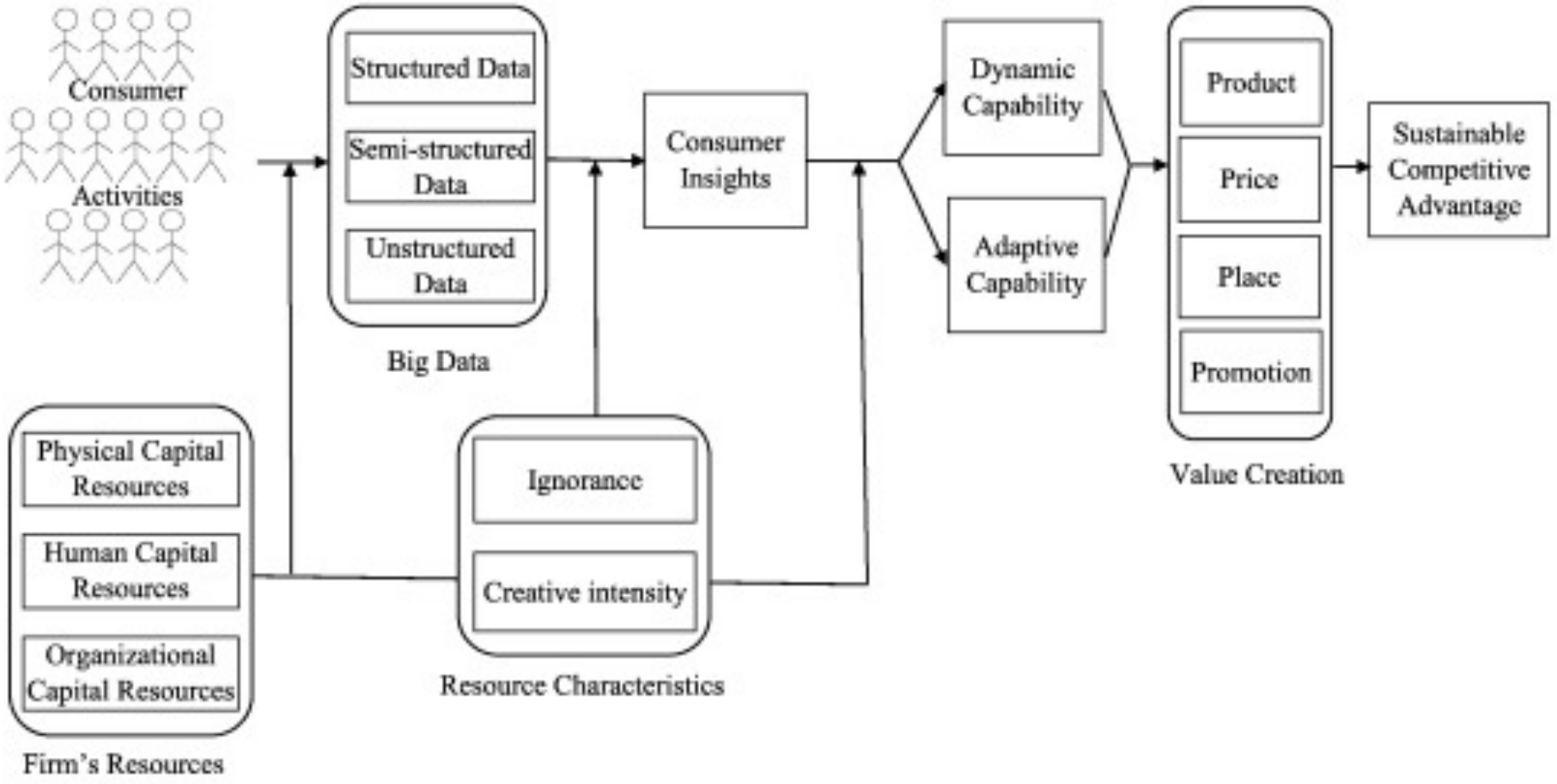
Source: https://blog.persistent.com/index.php/2015/05/05/is-your-enterprise-data-platform-ready-for-the-dive-into-digital-transformation/

A Marketing Mix Framework for Big Data Management



Source: Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research*,*2*(1), 28-32.

A resource-based view of the impact of Big Data on competitive advantage



Source: Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.

LeCun, Yann,
Yoshua Bengio,
and Geoffrey Hinton.

"Deep learning."

Nature 521, no. 7553 (2015): 436-444.

REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input.

Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification. Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations. An image, for example, comes in the form of an array of pixel values, and the learned features in the first layer of representation typically represent the presence or absence of edges at particular orientations and locations in the image. The second layer typically detects motifs by spotting particular arrangements of edges, regardless of small variations in the edge positions. The third layer may assemble motifs into larger combinations that correspond to parts of familiar objects, and subsequent layers would detect objects as combinations of these parts. The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.

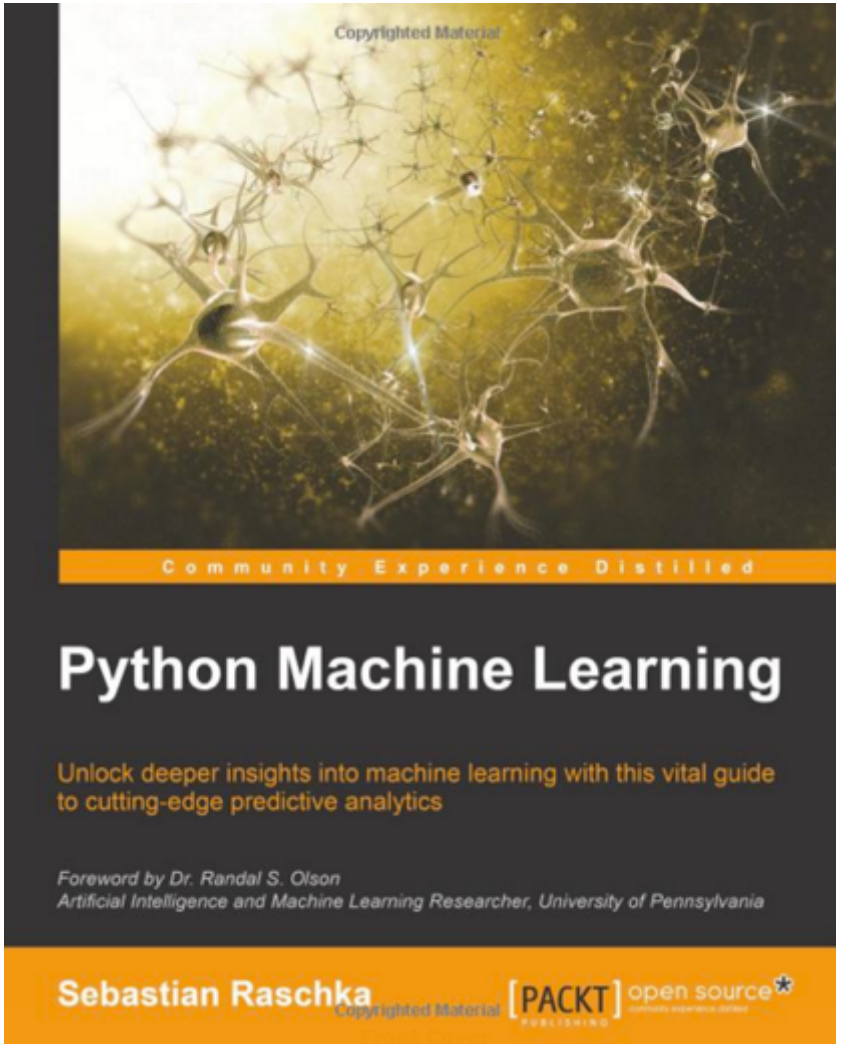
Deep learning is making major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years. It has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition to beating records in image recognition[1–4] and speech recognition[5–7], it has beaten other machine-learning techniques at predicting the activity of potential drug molecules[8], analysing particle accelerator data[9,10], reconstructing brain circuits[11], and predicting the effects of mutations in non-coding DNA on gene expression and disease[12,13]. Perhaps more surprisingly, deep learning has produced extremely promising results for various tasks in natural language understanding[14], particularly topic classification, sentiment analysis, question answering[15] and language translation[16,17].

Source: LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

Sebastian Raschka (2015),

Python Machine Learning,

Packt Publishing

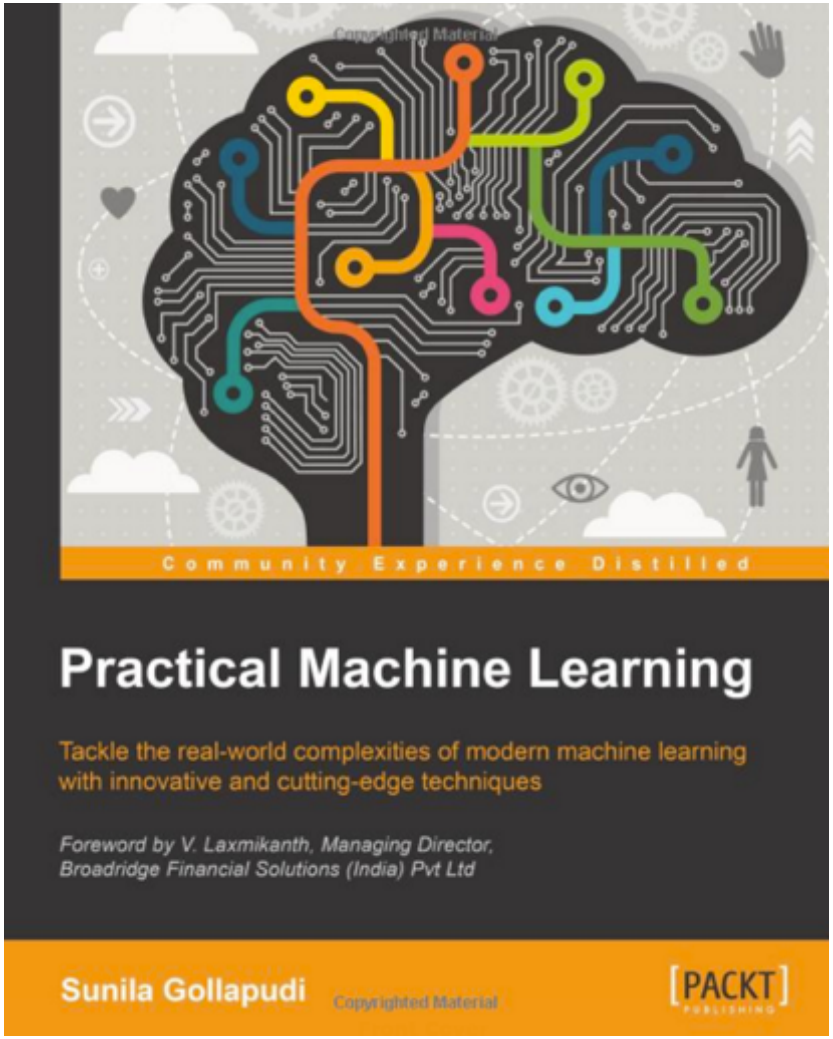


Source: http://www.amazon.com/Python-Machine-Learning-Sebastian-Raschka/dp/1783555130

Sunila Gollapudi (2016),

Practical Machine Learning,

Packt Publishing



Source: http://www.amazon.com/Practical-Machine-Learning-Sunila-Gollapudi/dp/178439968X

Machine Learning Models

- Deep Learning
- Association rules
- Decision tree
- Clustering
- Bayesian
- Kernel
- Ensemble
- Dimensionality reduction
- Regression Analysis
- Instance based

Source: Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing

Data Scientist
資料科學家

Source: http://www.ibmbigdatahub.com/infographic/what-makes-data-scientist

Deep Learning

Intelligence from Big Data

Source: https://www.vlab.org/events/deep-learning/

Data Scientist:

The Sexiest Job of the 21st Century

(Davenport & Patil, 2012)(HBR)

Source: Davenport, T. H., & Patil, D. J. (2012). Data Scientist. *Harvard business review*

Data Scientist:

The Sexiest Job of the 21st Century

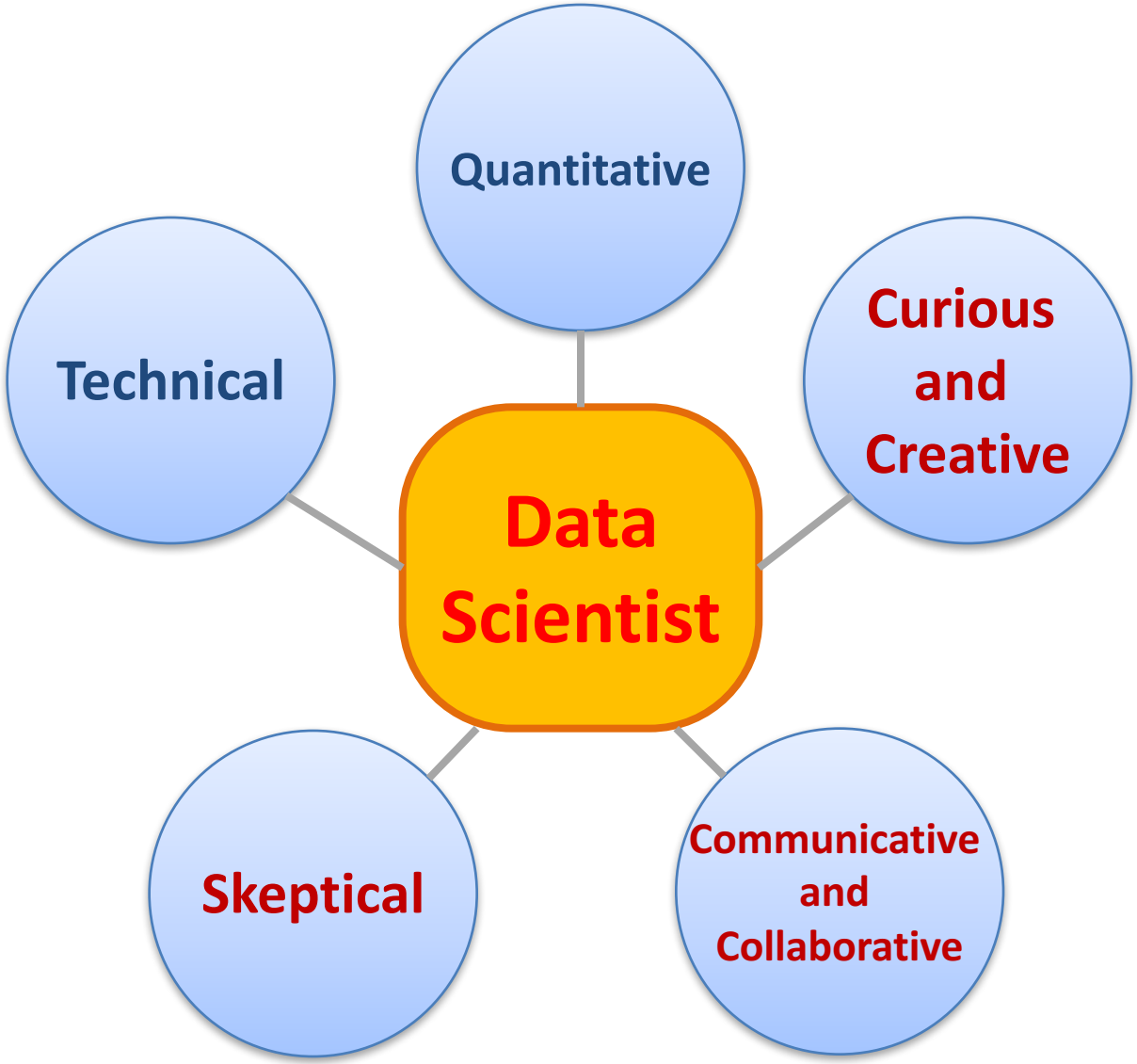
Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

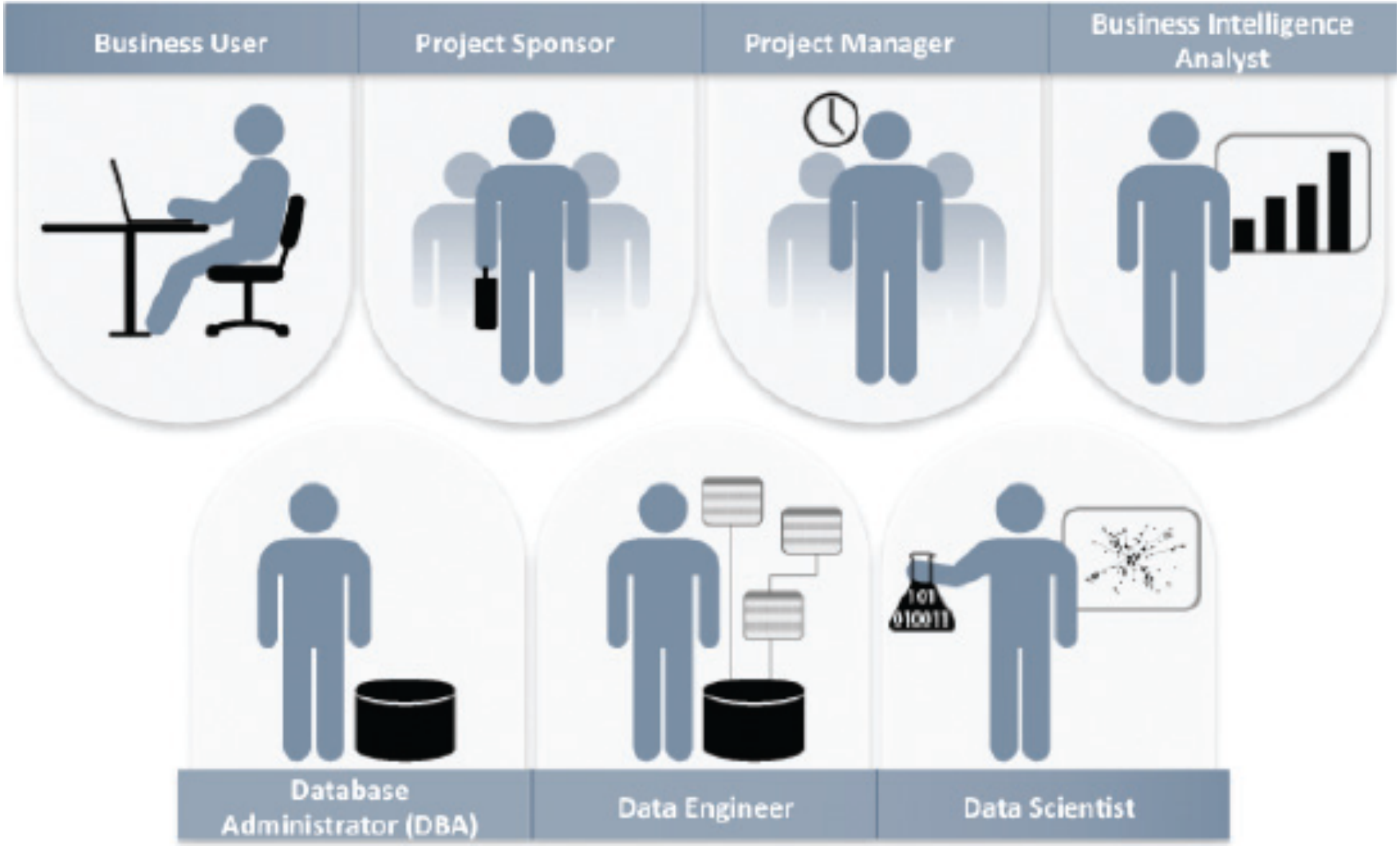
Source: Davenport, T. H., & Patil, D. J. (2012). Data Scientist. *Harvard business review*

Data Scientist Profile



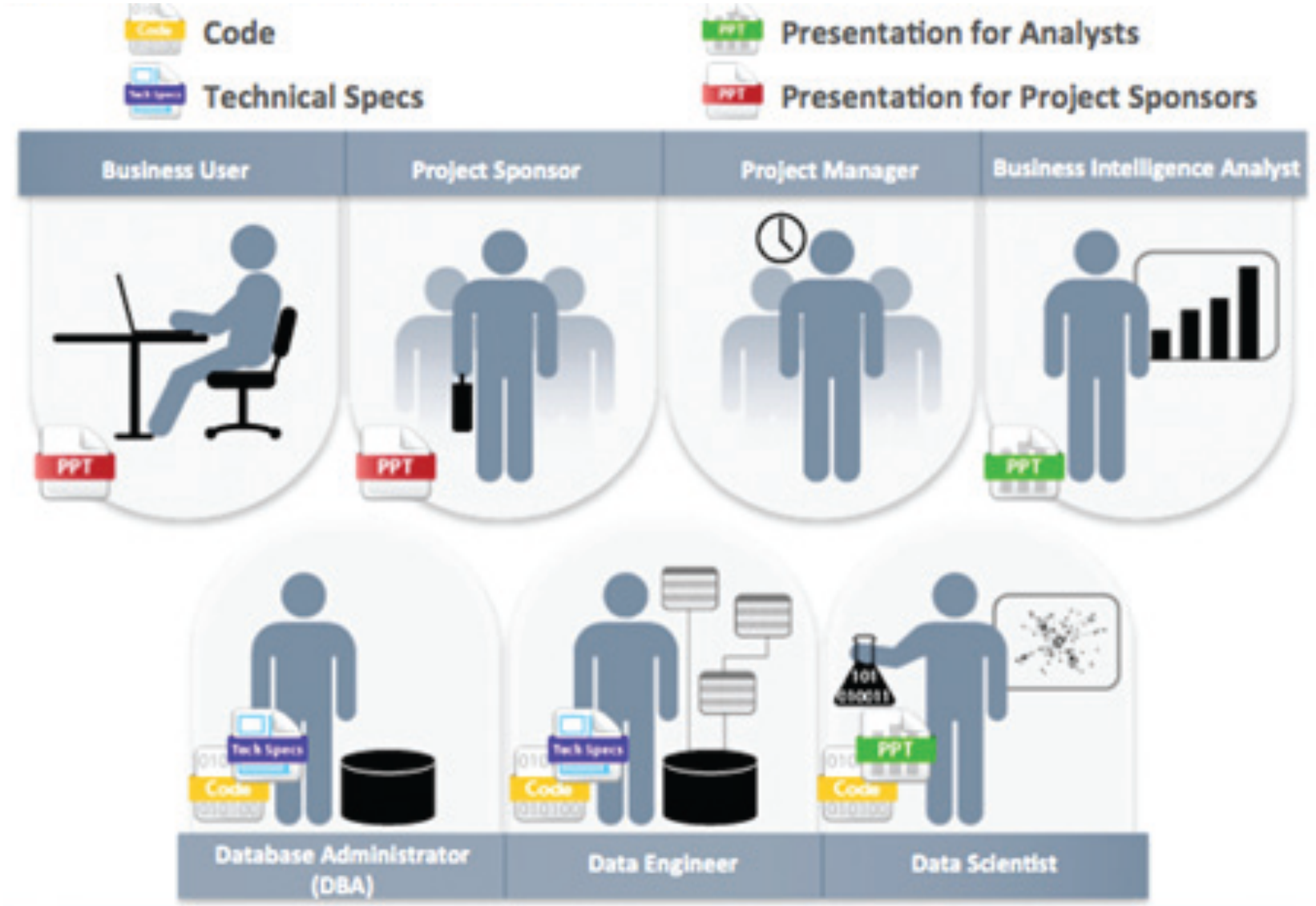
Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Key Roles for a Successful Analytics Project



Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Key Outputs from a Successful Analytics Project



Source: EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015

Data Science vs.
Big Data vs. Data Analytics

Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Data Science vs. Big Data vs. Data Analytics

WHAT ARE THEY?

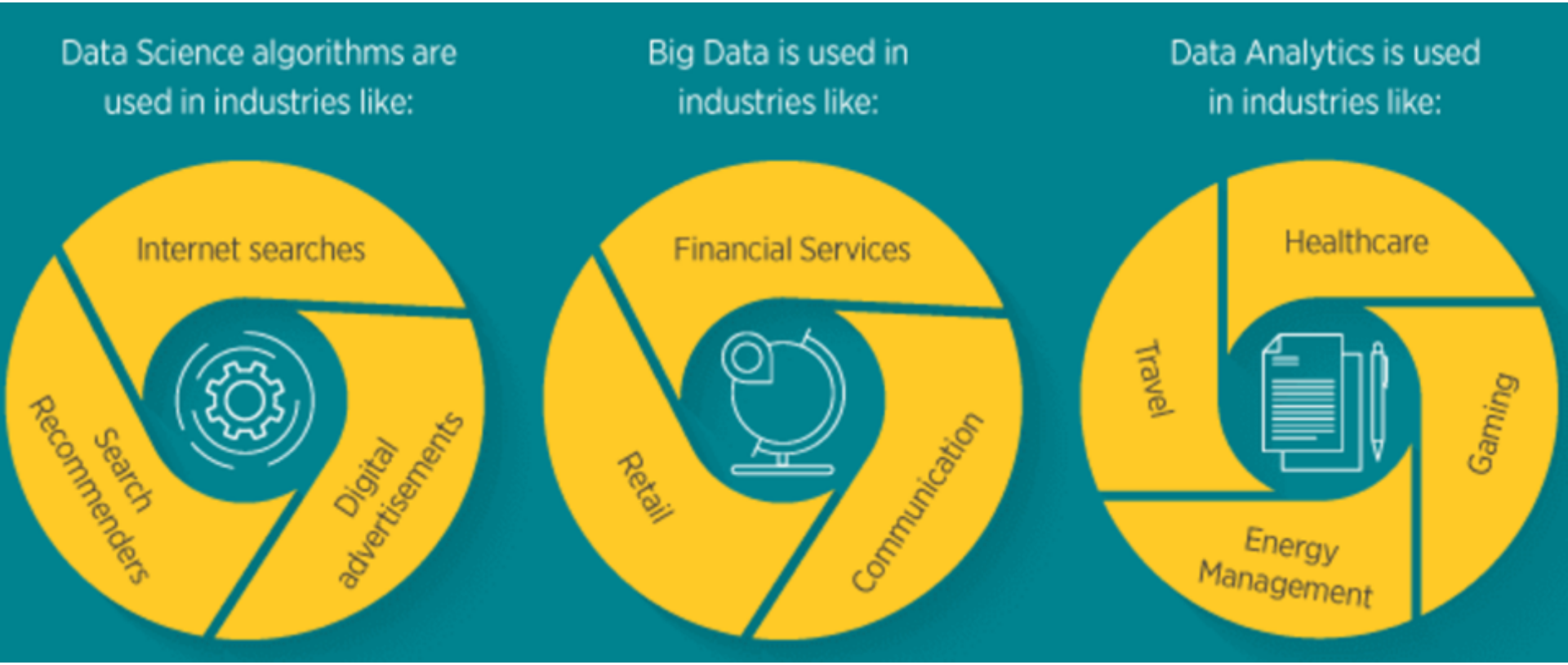
Data Science is a field that comprises of everything that related to data cleansing, preparation, and analysis.

Big Data is something that can be used to analyze insights which can lead to better decision and strategic business moves.

Data Analytics Involves automating insights into a certain dataset as well as supposes the usage of queries and data aggregation procedures.

Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

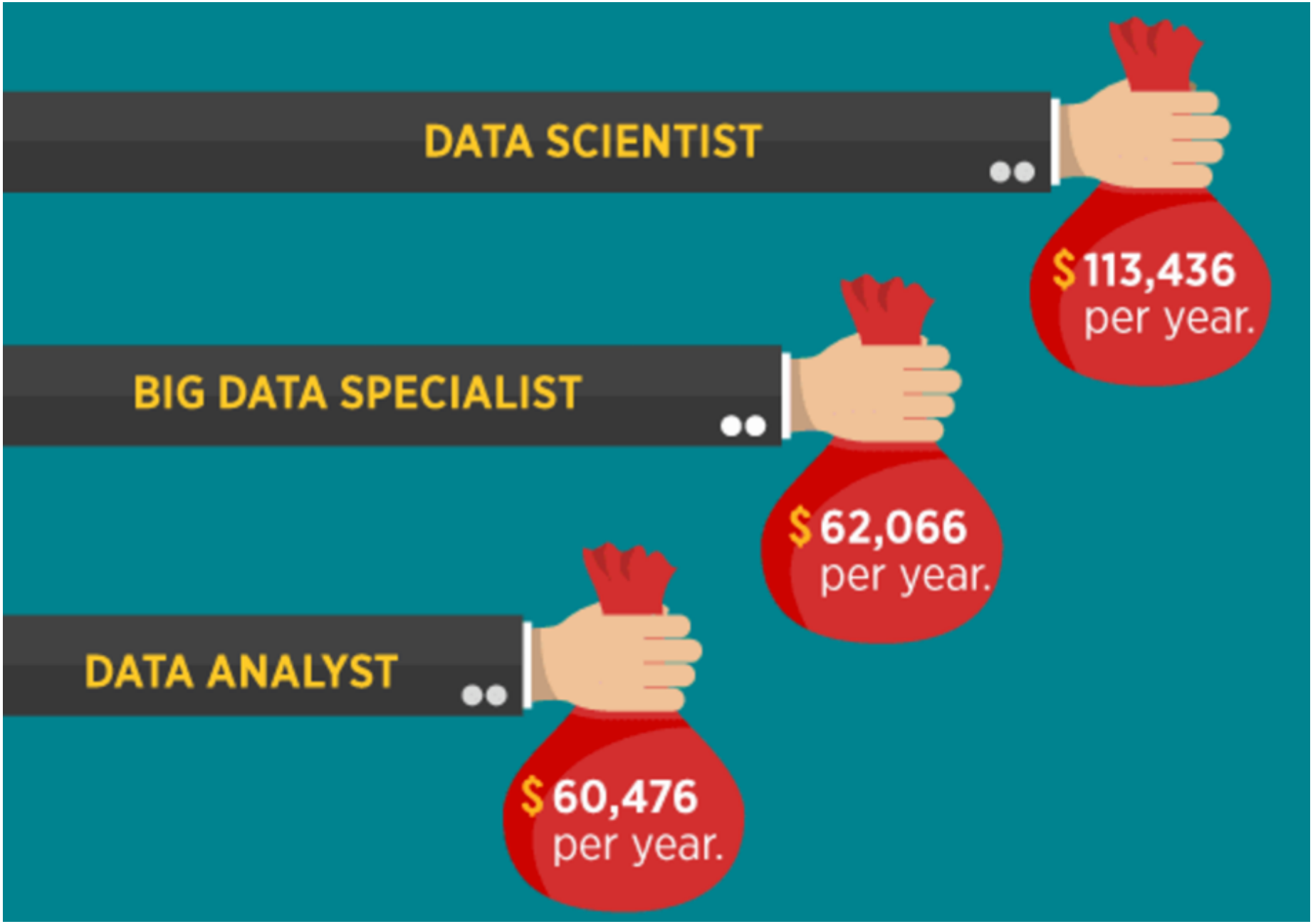
What are they used?



Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

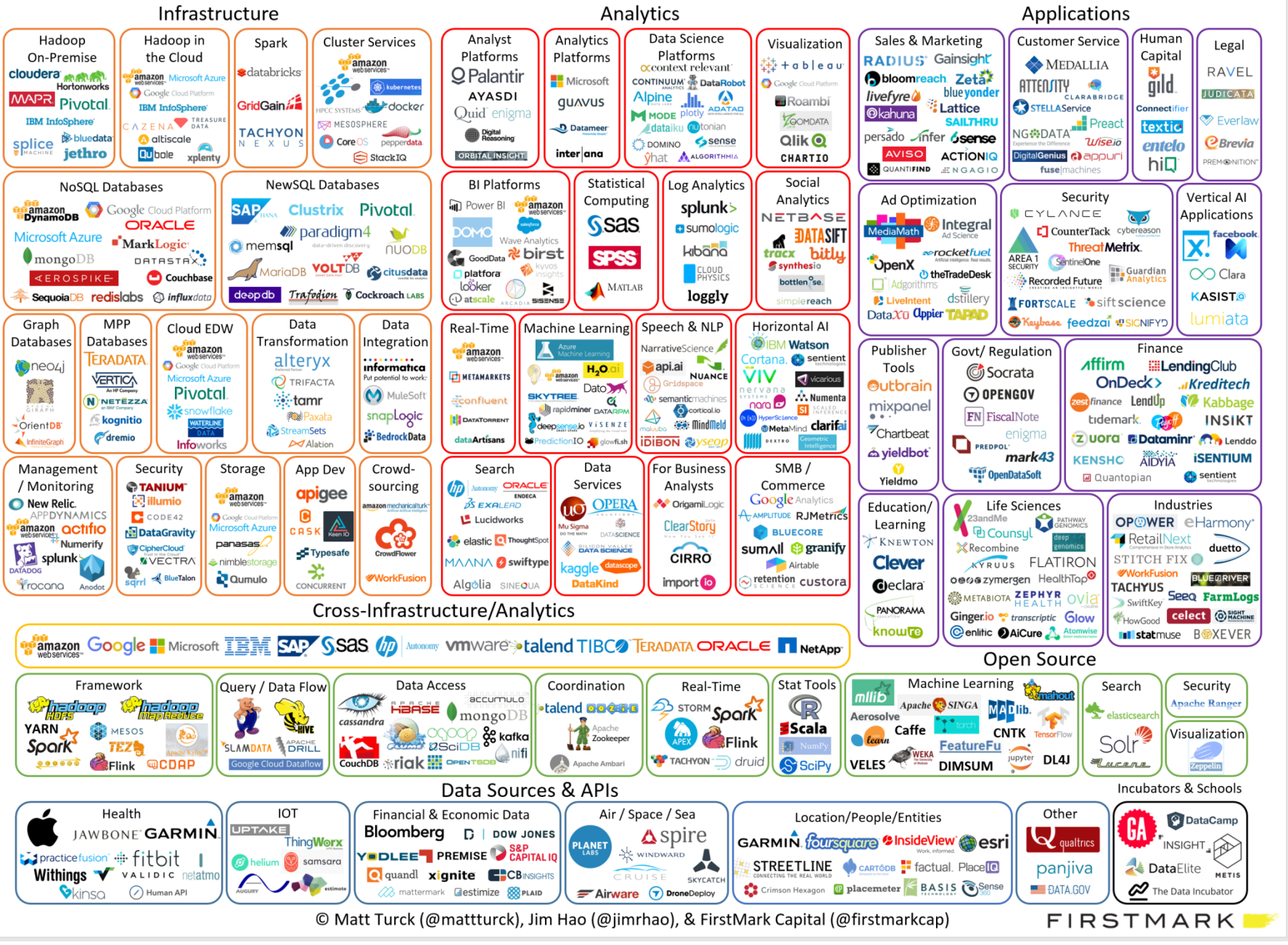
Data Science
What are the Skills Required?

Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article



Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Big Data Landscape 2016



Source: http://mattturck.com/wp-content/uploads/2016/01/matt_turck_big_data_landscape_full.png

Summary

- Social Computing
- Big Data Analysis

References

- Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press
- Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann
- Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing
- Devangana Khokhar (2015), Gephi Cookbook, Packt Publishing
- EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015
- Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications