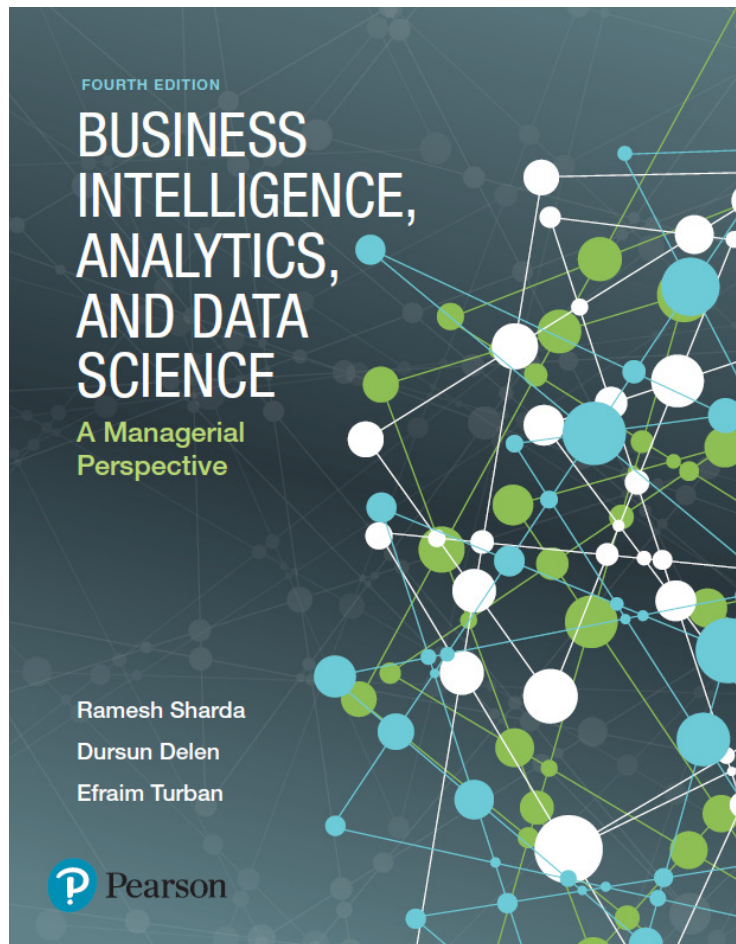


# Business Intelligence, Analytics, and Data Science: A Managerial Perspective

Fourth Edition



## Chapter 5 – Part B

Predictive Analytics II: Text, Web, and Social Media Analytics ...

# Sentiment Analysis

- Sentiment → belief, view, opinion, and conviction
- Sentiment analysis is trying to **answer** the question “What do people feel about a certain topic?”
- By analyzing data related to opinions of many using a variety of automated tools
- Used in variety of domains, but its applications in CRM are especially noteworthy (which related to customers/consumers’ opinions)

# Sentiment Analysis Applications

- Voice of the customer (VOC)
- Voice of the Market (VOM)
- Voice of the Employee (VOE)
- Brand Management
- Financial Markets
- Politics
- Government Intelligence
- ... others



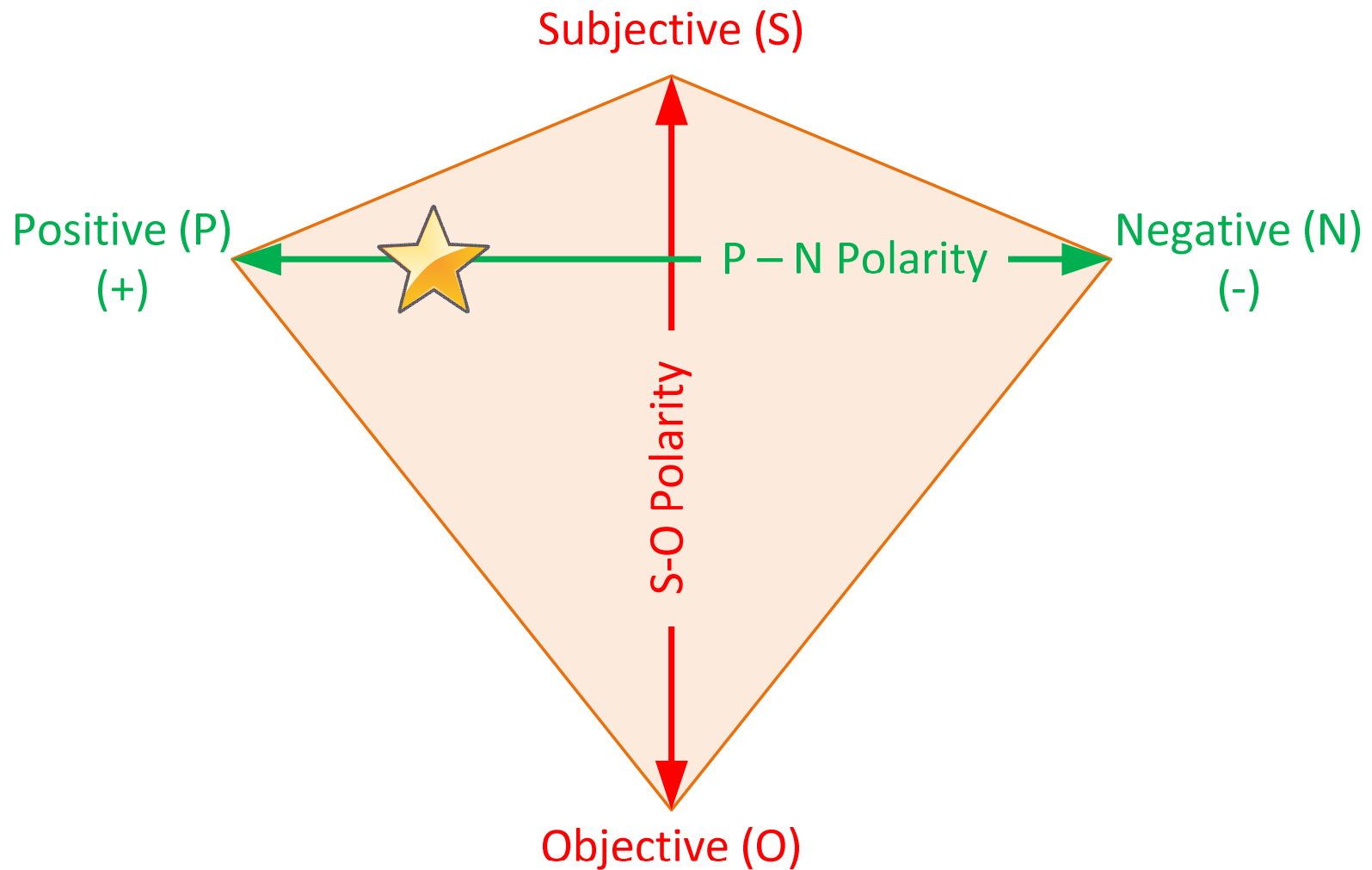
# Sentiment Analysis Process

- **Step 1** – Sentiment Detection
  - Comes right after the retrieval and preparation of the text documents
  - It is also called detection of objectivity
    - **Fact** [= objectivity] versus **Opinion** [= subjectivity]
- **Step 2** – N-P Polarity Classification
  - Given an opinionated piece of text, the goal is to classify the opinion as falling under one of two opposing sentiment polarities
    - **N** [= negative] versus **P** [= positive]

# Sentiment Analysis Process

- **Step 3 – Target Identification**
  - The goal of this step is to accurately identify the target of the expressed sentiment (e.g., a person, a product, an event, etc.)
    - Level of difficulty → the application domain
- **Step 4 – Collection and Aggregation**
  - Once the sentiments of all text data points in the document are identified and calculated, they are to be aggregated
    - Word → Statement → Paragraph → Document

# P-N Polarity and S-O Polarity



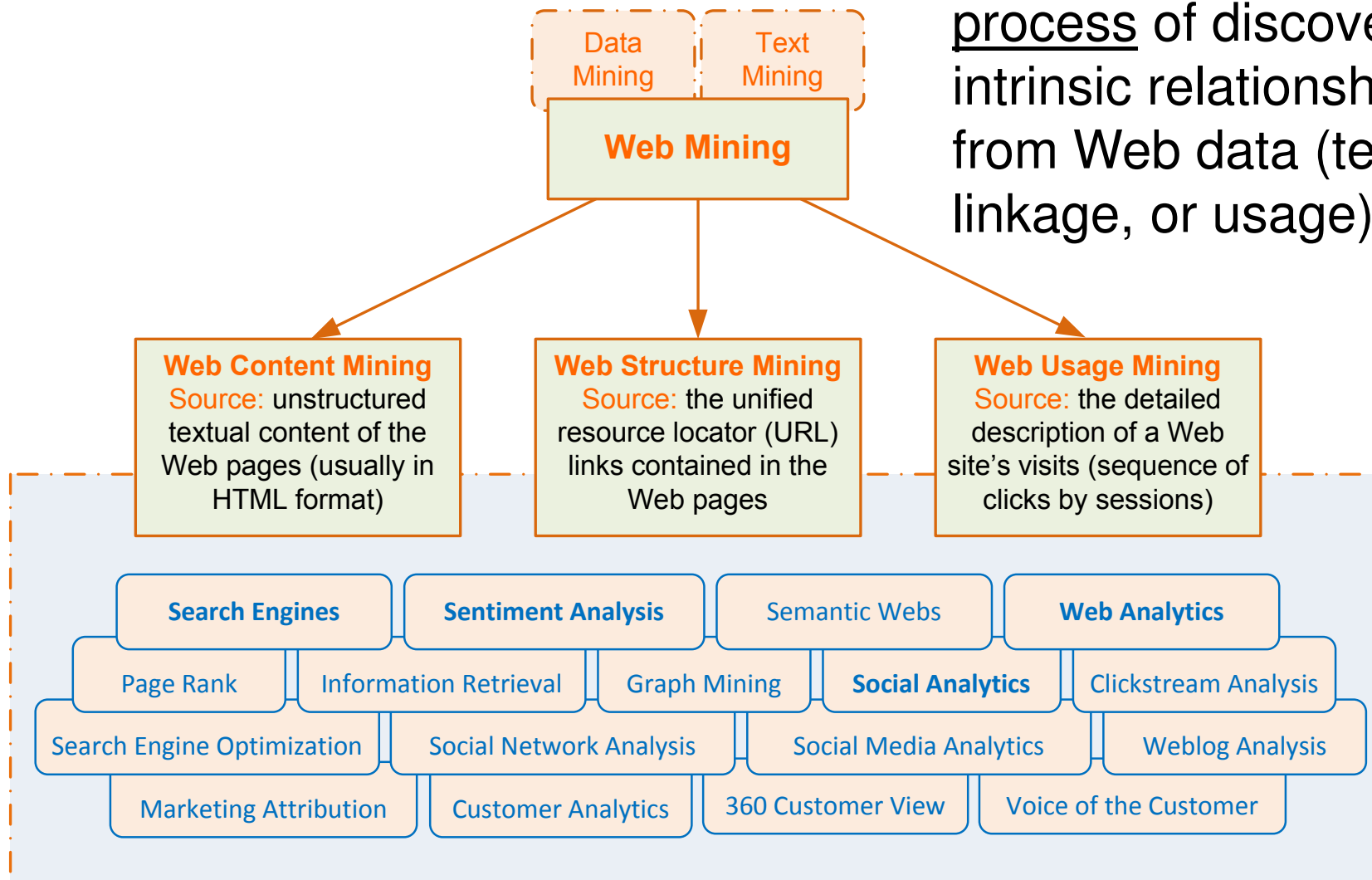
# Web Mining Overview

- Web is the largest repository of data
- Data is in HTML, XML, text format
- Challenges (of processing Web data)
  - The Web is too big for effective data mining
  - The Web is too complex
  - The Web is too dynamic
  - The Web is not specific to a domain
  - The Web has everything
- Opportunities and challenges are great!



# Web Mining

**Web mining** (or Web data mining) is the process of discovering intrinsic relationships from Web data (textual, linkage, or usage)



# Web Content/Structure Mining

- Mining the textual content on the Web
- Data collection via Web crawlers
- Web pages include hyperlinks
  - Authoritative pages
  - Hubs
  - Hyperlink-induced topic search (HITS) alg.

# Web Usage Mining

- Extraction of information from data generated through Web page visits and transactions...
  - data stored in server access logs, referrer logs, agent logs, and client-side cookies
  - user characteristics and usage profiles
  - metadata, such as page attributes, content attributes, and usage data
- Clickstream data
- Clickstream analysis

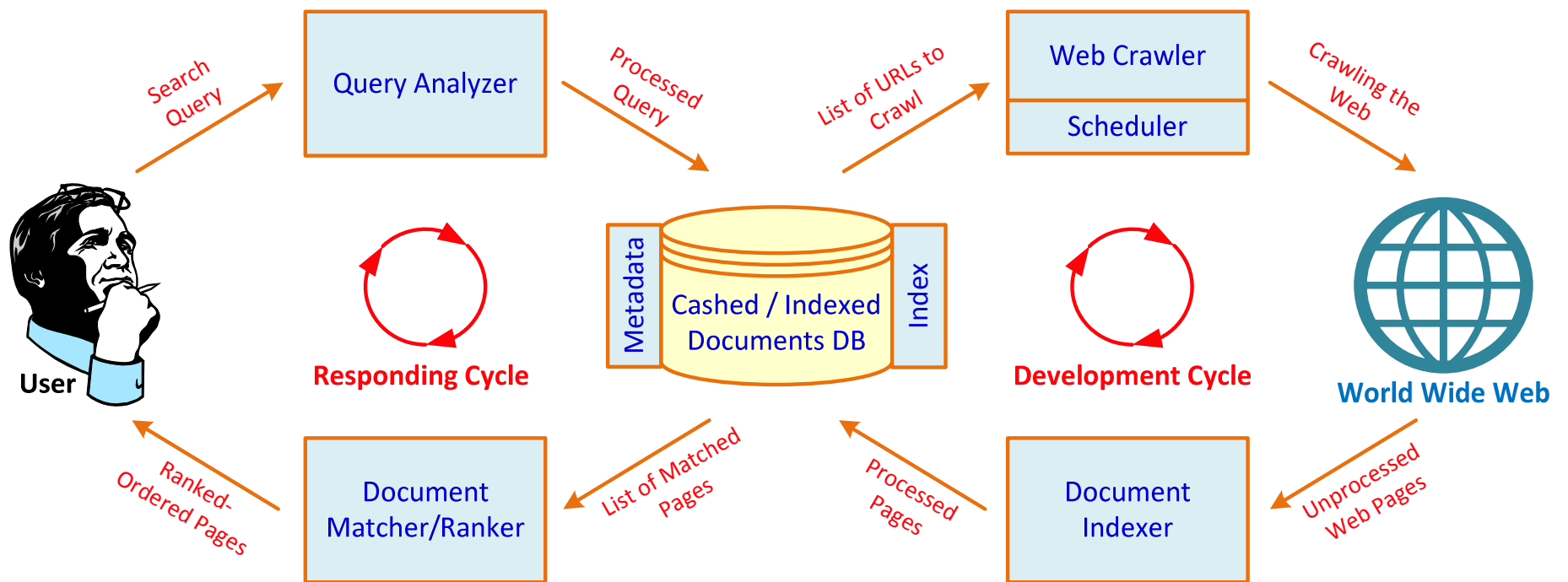
# Web Usage Mining

- Web usage mining applications
  - Determine the lifetime value of clients
  - Design cross-marketing strategies across products.
  - Evaluate promotional campaigns
  - Target electronic ads and coupons at user groups based on user access patterns
  - Predict user behavior based on previously learned rules and users' profiles
  - Present dynamic information to users based on their interests and profiles
  - ...

# Search Engines

- Google, Bing, Yahoo, ...
- For what reason do you use search engines?
- **Search engine** is a software program that searches for documents (Internet sites or files) based on the keywords (individual words, multi-word terms, or a complete sentence) that users have provided that have to do with the subject of their inquiry
- They are the workhorses of the Internet

# Structure of a Typical Internet Search Engine



# Anatomy of a Search Engine

## 1. Development Cycle

- Web Crawler
- Document Indexer

## 2. Response Cycle

- Query Analyzer
- Document Matcher/Ranker

# Search Engine Optimization

- It is the intentional activity of affecting the visibility of an e-commerce site or a Web site in a search engine's natural (unpaid or organic) search results
- Part of an Internet marketing strategy
- Based on knowing how a Search Engine works
  - Content, HTML, keywords, external links, ...
- Indexing based on ...
  - Webmaster submission of URL
  - Proactively and continuously crawling the Web

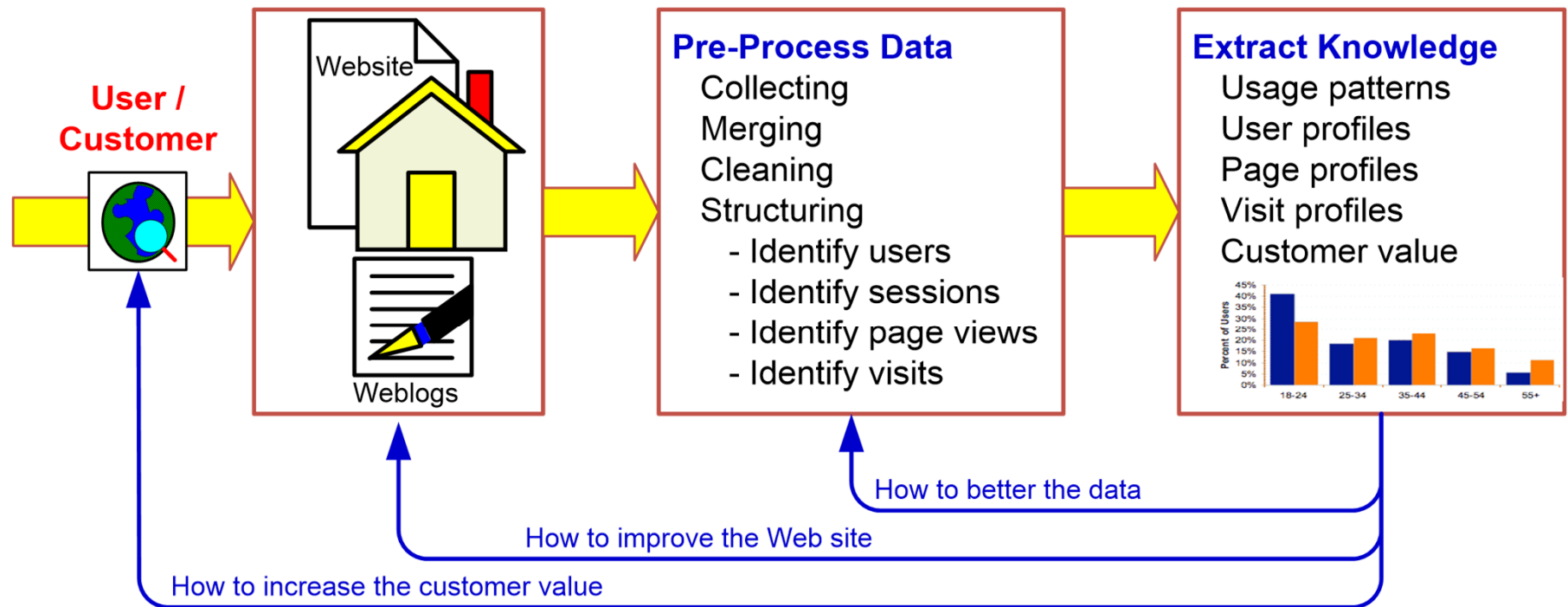


# Top 15 Most Popular Search Engines

(by eBizMBA, August 2016)

Rank	Name	Estimated Unique Monthly Visitors
1	Google	1,600,000,000
2	Bing	400,000,000
3	Yahoo! Search	300,000,000
4	Ask	245,000,000
5	AOL Search	125,000,000
6	Wow	100,000,000
7	WebCrawler	65,000,000
8	MyWebSearch	60,000,000
9	Infospace	24,000,000
10	Info	13,500,000
11	DuckDuckGo	11,000,000
12	Contentko	10,500,000
13	Dogpile	7,500,000
14	Alhea	4,000,000
15	ixQuick	1,000,000

# Web Usage Mining (Clickstream Analysis)



# Web Analytics Metrics

- Web site usability
  - How were the visitors using my Web site?
- Traffic sources
  - Where did they come from?
- Visitor profiles
  - What do my visitors look like?
- Conversion statistics
  - What does it all mean for the business?

# Web Analytics Metrics

## Web Site Usability

- Page views
- Time on site
- Downloads
- Click map
- Click paths

## Traffic Source

- Referral Web sites
- Search engines
- Direct
- Offline campaigns
- Online campaigns

# Web Analytics Metrics

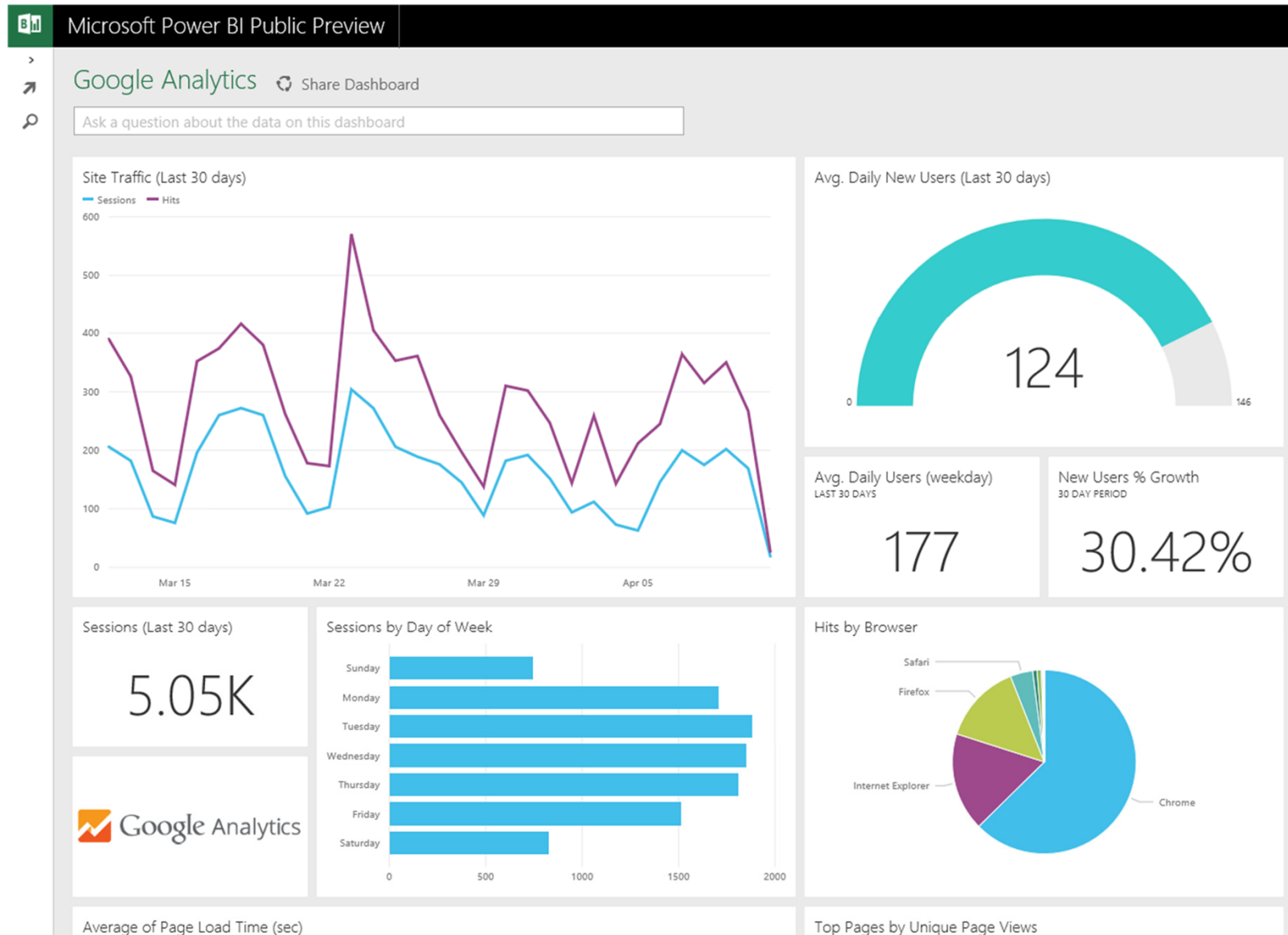
## Visitor Profiles

- Keywords
- Content groupings
- Geography
- Time of day
- Landing page profiles

## Conversion Statistics

- New visitors
- Returning visitors
- Leads
- Sales/conversions
- Abandonment/exit rate

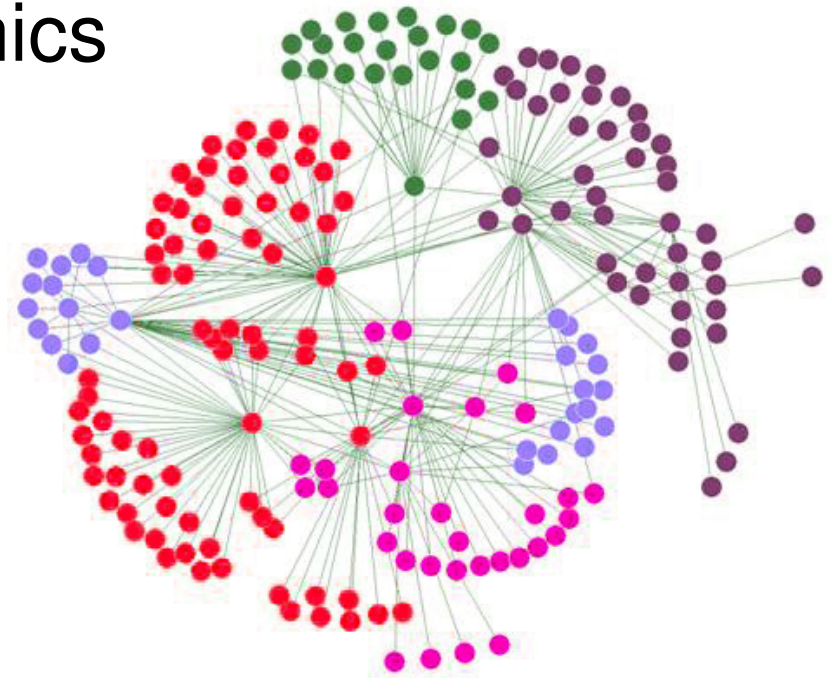
# A Sample Web Analytics Dashboard



# Social Analytics

## Social Network Analysis

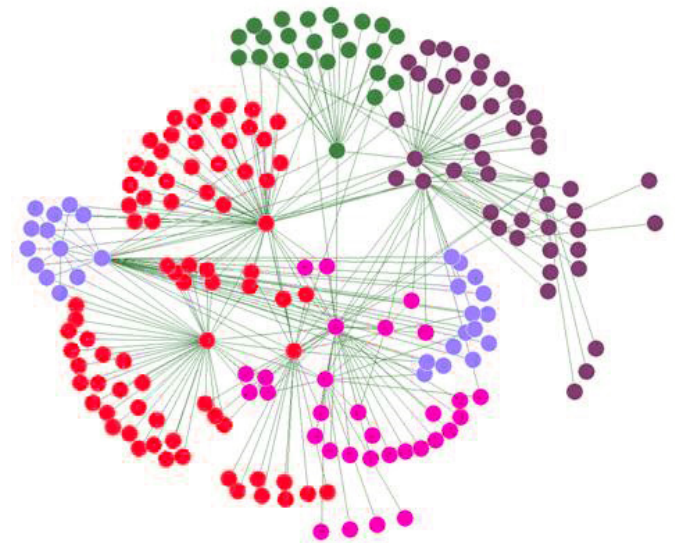
- **Social Network** - social structure composed of individuals linking to each other
- Analysis of social dynamics
- Interdisciplinary field
  - Social psychology
  - Sociology
  - Statistics
  - Graph theory



# Social Analytics

## Social Network Analysis

- **Social Networks** help study relationships between individuals, groups, organizations, societies
  - Self organizing
  - Emergent
  - Complex
- Typical social network types
  - Communication networks, community networks, criminal networks, innovation networks, ...





# Application Case 5.8

## Tito's Vodka Establishes Brand Loyalty with an Authentic Social Strategy

### Discussion Questions

1. How can social media analytics be used in the consumer products industry?
2. What do you think are the key challenges, potential solutions, and probable results in applying social media analytics in consumer products and services firms?

# Social Analytics

## Social Network Analysis Metrics

- **Connections**
  - Homophily
  - Multiplexity
  - Mutuality/reciprocity
  - Network closure
  - Propinquity
- **Distribution**
  - Bridge
  - Centrality
  - Density
  - Distance
  - Structural holes
- **Segmentation**
  - Cliques and social circles
  - Clustering coefficient
  - Cohesion

# Social Media

## Definitions and Concepts

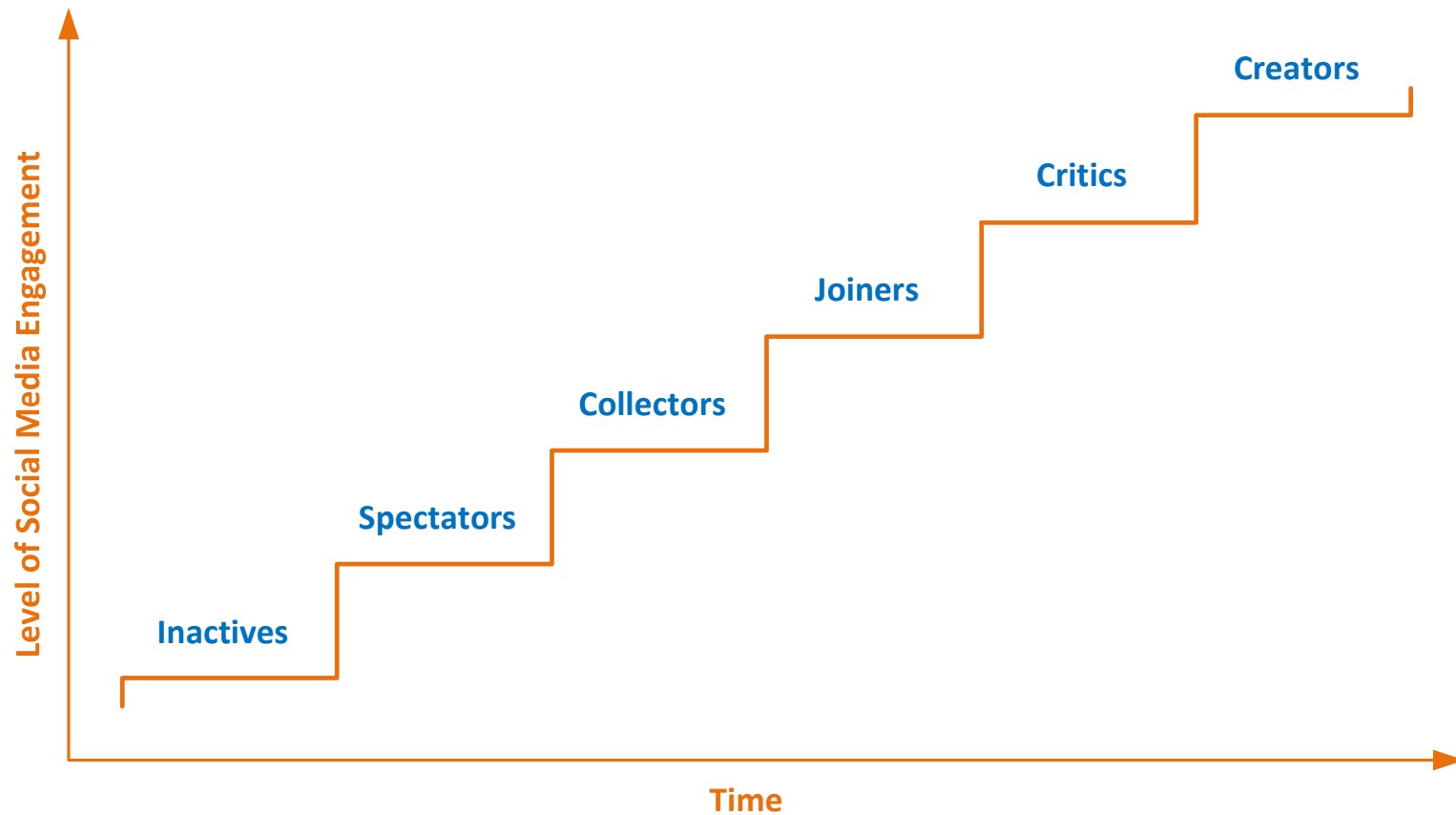
- Enabling technologies of social interactions among people
- Relies on enabling technologies of Web 2.0
- Takes on many different forms
  - Internet forums, Web logs, social blogs, microblogging, wikis, social networks, podcasts, pictures, video, and product reviews
- Different types of social media
  - Based on **media research** and **social process**

# Social versus Industrial Media

- Web-based social media are different from traditional/industrial media, such as newspapers, television, and film
- Differentiating characteristics
  - Quality
  - Reach
  - Frequency
  - Accessibility
  - Usability
  - Immediacy
  - Updatability

# How Do People Use Social Media?

- Different engagement levels



# Social Media Analytics

- It is the systematic and scientific ways to consume the vast amount of content created by Web-based social media outlets, tools, and techniques for the betterment of an organization's competitiveness
- Tools to measure social media impact:
  - Descriptive analytics
  - Social network analysis
  - Advanced analytics

# Best Practices in Social Media Analytics

- Think of measurement as a guidance system, not a rating system
- Track the elusive sentiment
- Continuously improve the accuracy of text analysis
- Look at the ripple effect
- Look beyond the brand
- Identify your most powerful influencers
- Look closely at the accuracy of your analytic tool
- Incorporate social media intelligence into planning