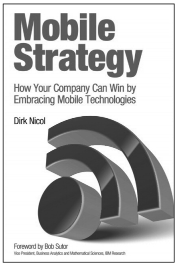
Social Media Analytics

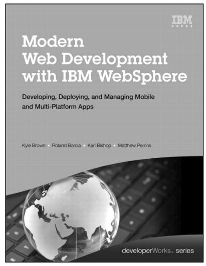
Techniques and Insights for Extracting Business Value Out of Social Media

Matthew Ganis • Avinash Kohirkar

Foreword by Ed Brill
IBM Vice President, Social Business Cloud: Deployment and Adoption

Related Books of Interest





Mobile Strategy

How Your Company Can Win by Embracing Mobile Technologies

By Dirk Nicol

ISBN-10: 0-13-309491-X

ISBN-13: 978-0-13-309491-6

Mobile Strategy gives IT leaders the ability to transform their business by offering all the guidance they need to navigate this complex landscape, leverage its opportunities, and protect their investments along the way. IBM's Dirk Nicol clearly explains key trends and issues across the entire mobile project lifecycle. He offers insights critical to evaluating mobile technologies, supporting BYOD, and integrating mobile, cloud, social, and big data. Throughout, you'll find proven best practices based on real-world case studies from his extensive experience with IBM's enterprise customers.

Modern Web Development with IBM WebSphere

Developing, Deploying, and Managing Mobile and Multi-Platform Apps

By Kyle Brown, Roland Barcia, Karl Bishop, Matthew Perrins

ISBN-10: 0-13-306703-3

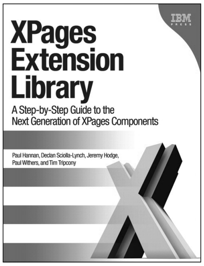
ISBN-13: 978-0-13-306703-3

This guide presents a coherent strategy for building modern mobile/web applications that are fast, responsive, interactive, reusable, maintainable, extensible, and a pleasure to use.

Using well-crafted examples, the authors introduce best practices for MobileFirst development, helping you create apps that work superbly on mobile devices and add features on conventional browsers. Throughout, you'll learn better ways to deliver Web 2.0 apps with HTML/JavaScript front ends, RESTful Web Services, and persistent data. Proven by IBM and its customers, the approach covered in this book leads to more successful mobile/web applications—and more effective development teams.

Sign up for the monthly IBM Press newsletter at ibmpressbooks.com/newsletters

Related Books of Interest



XPages Extension Library

A Step-by-Step Guide to the Next Generation of XPages Components

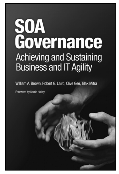
By Paul Hannan, Declan Sciolla-Lynch, Jeremy Hodge, Paul Withers, Tim Tripcony

ISBN-10: 0-13-290181-1

ISBN-13: 978-0-13-290181-9

XPages Extension Library is the first and only complete guide to Domino development with this library; it's the best manifestation yet of the underlying XPages Extensibility Framework. Complementing the popular *Mastering XPages*, it gives XPages developers complete information for taking full advantage of the new components from IBM.

Combining reference material and practical use cases, the authors offer step-by-step guidance for installing and configuring the XPages Extension Library and using its state-of-the-art applications infrastructure to quickly create rich web applications with outstanding user experiences.

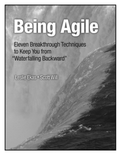


SOA Governance

Achieving and Sustaining Business and IT Agility

Brown, Laird, Gee, Mitra

ISBN: 978-0-13-714746-5

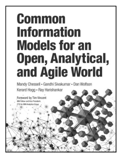


Being Agile

Eleven Breakthrough Techniques to Keep You from "Waterfalling Backward"

Ekas, Will

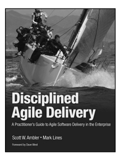
ISBN: 978-0-13-337562-6



Common Information Models for an Open, Analytical, and Agile World

Chessell, Sivakumar, Wolfson, Hogg, Harishankar

ISBN: 978-0-13-336615-0

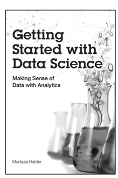


Disciplined Agile Delivery

A Practitioner's Guide to Agile Software Delivery in the Enterprise

Ambler, Lines

ISBN: 978-0-13-281013-5

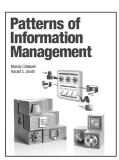


Getting Started with Data Science

Making Sense of Data with Analytics

Haider

ISBN: 978-0-13-399102-4



Patterns of Information Management

Chessell, Smith

ISBN: 978-0-13-315550-1

Sign up for the monthly IBM Press newsletter at ibmpressbooks.com/newsletters

This page intentionally left blank

Social Media Analytics

Social Media Analytics

Techniques and Insights for Extracting Business Value Out of Social Media

Matthew Ganis
Avinash Kohirkar

IBM Press
Pearson plc

New York • Boston • Indianapolis • San Francisco
Toronto • Montreal • London • Munich • Paris • Madrid
Cape Town • Sydney • Tokyo • Singapore • Mexico City

ibmpressbooks.com

IBM Press Program Managers: Steven M. Stansel, Ellice Uffer

Cover design: IBM Corporation

Associate Publisher: Dave Dusthimer

Marketing Manager: Dan Powell

Executive Editor: Mary Beth Ray

Editorial Assistant: Vanessa Evans

Development Editor: Box Twelve Communications

Technical Editors: Deborah DeLosa, Ajay Raina

Managing Editor: Kristy Hart

Designer: Alan Clements

Project Editor: Andy Beaster

Copy Editor: Chuck Hutchinson

Indexer: Ken Johnson

Senior Compositor: Gloria Schurick

Proofreader: Sarah Kearns

Manufacturing Buyer: Dan Uhrig

Published by Pearson plc

Publishing as IBM Press

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Library of Congress Control Number: 2015949068

ISBN-13: 978-0-13-389256-7

ISBN-10: 0-13-389256-5

Text printed in the United States on recycled paper at R.R. Donnelley in
Crawfordsville, Indiana.
First printing: December 2015

To the Ganis Gang—Always and Forever

—Matt Ganis

I dedicate this book to my mother, a person who has given me so much and who at age 80 is still one of the most inquisitive persons I know!

—Avinash Kohirkar

Contents

Appendices

Foreword

In the decade since social networking was born, we have seen the power of platforms that unite humanity. Across our professional and personal lives, social platforms have truly changed the world. Social media has been the tool to ignite revolutions and elections, deliver real-time news, connect people and interests, and of course, drive commerce. In 2005, industry analysts were skeptical about how blogging and its successors could ever be used in business; today every single social channel has both B2C and B2B offerings sprinkled generously throughout the content.

As businesses figured out that they could use social networks to interact directly with their customers and prospects, questions were immediately generated about efficacy and ROI. Was it just hype and noise, or were new audiences being reached and new opportunities created? For the first several years, the only way to answer these questions was anecdotally. Many brands and businesses viewed social media warily, feeling that nothing good could come from engaging in online discussions directly.

Things changed as the technology matured to offer tools for social listening. Whether for business, politics, or news, organizations learned they could identify trends and patterns in all the flotsam and jetsam of online content. Another leap forward occurred as analytics engines were applied to the vast stream of unstructured data, when suddenly big-picture profiles and behaviors could be identified.

Today, organizations of all sizes and missions are looking for ways to make sense of the information available on the social web. Analyzing social media, the right way at least, is now just as important as a brand presence or advertising strategy. When done correctly, the insights available can shape decisions, make organizations more responsive, and quell negative press before it takes off.

In *Social Media Analytics,* Matt Ganis and Avinash Kohirkar have set out a thorough approach to gaining business insights from social media. Matt and Avinash understand this challenge. Each has built his career on data analysis and insights, and they have specifically looked at social content for the last several years. They have examined key vectors of social participation, including reach, eminence, engagement, and activation. They understand how to filter out noise and focus on relevant insight, building the right tools and conducting the right studies to demonstrate trends, correlations, and results.

Social Media Analytics provides much-needed understanding of both what can be accomplished by examining social streams and why such insights matter. In the first part, the book looks at data identification, sources, determining relevancy, and time horizons. In Part II, several chapters explain ways to find data—what tools, how to understand output, and getting deep into the insights themselves. Part III goes further into interpreting data, looking at potential shortcomings of social analysis and useful ways of sharing insight through visualization.

Social media has evolved quickly from the initial hype, through the naysayers, and to a point where it is no longer viewed as optional. Today, however, there are so many social channels, devising a strategy for sharing and leveraging the online conversation can make the difference between success and failure.

I invite you to think back nostalgically to the days of focus groups, printed surveys, and controlled messages. As those tools of the past have faded out, they've been replaced with a veritable deluge of information. *Social Media Analytics* will help you devise the right strategy to make data-driven decisions rather than reacting to that one nasty tweet, looking at the overall story your customers and prospects are sharing online.

Ed Brill
Vice President, Social Business
IBM Corporation
Chicago, September 2015

Preface: Mining for Gold (or Digging in the Mud)

In *The Adventure of the Six Napoleons* by Arthur Conan Doyle, the famous sleuth Sherlock Holmes remarks to his sidekick, Watson:

> "The Press, Watson, is a most valuable institution, if you only know how to use it."

That statement, when applied to the wealth of data in social media channels today (loosely, "the press"), has never been more true. Companies are always looking for an "edge" in an attempt to find ways to remain relevant to their ever more vocal set of constituents. They are struggling to position themselves as trusted advisors or suppliers in a cut-throat environment of competitors, where consumers use public opinion (both good and bad) to share information and experiences at the speed of light (literally). When looking to explore this deluge of social media data, we must think and act like detectives. Careful investigations can, at times, lead to many revealing insights. This can be both time consuming and complex; it is work that requires a careful, methodical effort and not only requires patience and perseverance, but at times also requires a creative streak or spark of insight.

This book, aimed at executives (or analysts) responsible for understanding public opinion, brand management, and public perceptions, attempts to look at the processes and insights needed when attempting to answer questions within this massive amount of unstructured data we call *social media*.

Just What Do We Mean When We Say Social Media?

A social media website doesn't just give you information, but rather it is built around a way to interact with you while allowing access to the information. This interaction could be collecting comments or suggestions on a blog or voting on a specific topic—allowing users to have a voice in a conversation as opposed to simply reading others' opinions—this is why we call it a *social media conversation*.

Think of print media or a static web page or website as a one-way street, much like reading a newspaper or listening to a report on television; you have very limited ability to give your thoughts on the matter. (Radio talk shows at least allow users to call in to express their opinions—although ultimately they have the ability to limit the conversation by cutting off the call at any point.) Social media can be considered a two-way street that enables communication between end users. Social media gives users on the Internet the ability to express their opinions and interact with each other at speeds unheard of in the past with traditional media. This popularity of social media continues to grow at an exponential rate.

Why Look at This Data?

Consider one of the most famous cases of using Twitter to watch for customer satisfaction issues: @ComcastCares. As *BusinessWeek*'s Rebecca Reisner [1] said, Frank Eliason is probably the best known and most successful customer care representative in the world (or at least the United States). In April 2008, Eliason's team started monitoring Twitter traffic for mentions of his company, Comcast, made by disgruntled customers. (Comcast is one the largest providers of entertainment, information, and communications services and products in the United States, providing cable television, broadband Internet, and telephone services.) His tactic was to watch Twitter and immediately reach out to these customers who expressed dissatisfaction with Comcast's customer service. The idea was to quiet the spread of any negative sentiment amongst Comcast customers, while providing a sense of personal touch to these frustrated clients.

According to a 2011 report (Eliason has since left Comcast for greener pastures), the new Comcast customer care division processed about 6,000 blog posts and 2,000 Twitter messages per day, which resulted in faster customer response times that directly translate into improved customer satisfaction indexes. While Comcast is not analyzing social media per se, it is watching issues related to perceived poor quality so that it can quickly address issues and interact with these customers.

How Does This Translate into Business Value?

According to Eliason, Comcast was able to understand issues on Twitter far in advance of their call centers (that is, when customers would call in to tell of a problem) [2]. For example, during the NHL playoffs, a sports network carried by Comcast went off the air. People used Twitter to complain about Comcast, claiming the problem was poor service. However, in reality, all of the other networks were offline as well due to a lightning strike. The Comcast call center was able to find out the reason within a few minutes of it happening and was able to put up an automated message telling people what happened. In this case, Comcast estimated that it was able to save $1.2 million by putting up a message about the outage. Customers were able to listen to the message and hang up rather than call in to complain, thus using valuable call center resources.

As another example, consider a new product launch. The marketing team spent hundreds of hours determining the best way to disseminate the message of your new offering, and the company has spent millions on advertising, yet there appears to be lackluster acceptance.

Why?

One way to listen to the man on the street is to scan various social media outlets such as discussion forums, blogs, or chatter on sites like Twitter or LinkedIn. Perhaps you can pick up on messages or customer perceptions of your product or brand. Perhaps when you look at the discussion around your product, you'll see something similar to the situation shown in Figure I.1.

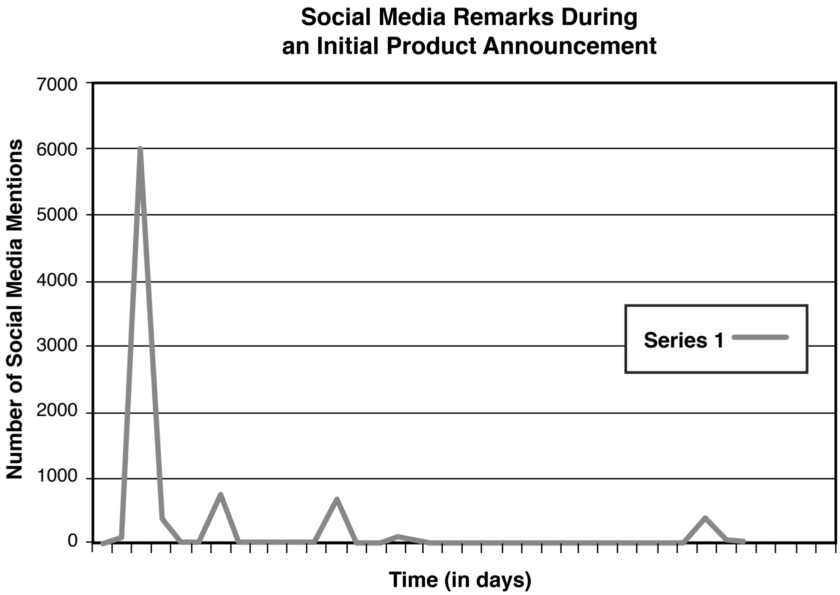


Figure I.1 Social media remarks during an initial product announcement.

This graph was produced for one of the projects we worked with during its launch. Note the steep rise in conversation at the initial launch. Social media conversations went from 0 to more than 6,000 mentions over the course of a few days. This is great! But look at what happens next. The level of conversations fell off rapidly, with just a few isolated spikes in conversation (which were later revealed to be additional announcements). So in this case, it wasn't so much that potential customers didn't like what they saw in the marketplace (of course, that may be the reason for the lack of conversation), but it appears more likely that the marketing campaign wasn't resonating with the public to pick up and carry on the conversation. We look at this particular case in a bit more detail later, but the message here is that a simple analysis within social media can quickly reveal where your business plan might have gone awry.

The Book's Approach

"I keep six honest serving men; they taught me all I know; their names are What and Why and When and How and Where and Who."

—Rudyard Kipling [3]

The process of social media analysis involves essentially three steps: *data identification*, *data analysis,* and finally *information interpretation*. In explaining each of these steps, we provide important insights and techniques that can be used to maximize the value derived at every point during the process. The approach we take is to first define a question to be answered (such as "What is the public's perception of our company in the light of a natural disaster?"). In attempting to analyze these questions, we suggest that analysts think like detectives, always asking the important questions "Who? What? Where? When? Why? and How?" These questions help in determining the proper data sources to evaluate, which can greatly affect the type of analysis that can be performed.

Data Identification

Any social media investigation is only as good as the data in which you are searching. The first part of this book explores proper *data identification*—or where to look in this vast social media space. In searching for answers, keep in mind that we will be searching through massive amounts of unstructured data, all in an attempt to make sense out of what we find in the process. Once we uncover some interesting artifacts, we will be transforming them into (hopefully useful) information. In the long run, the ultimate business objective is to derive real business insight from this *data*, turning the *information* we've gleaned from these sources into actionable *knowledge*.

In the first part of this book, we explore the source of the data that will be under analysis. To ensure that what we are collecting is the proper data or it explores the correct conversations, we look into questions such as these:

- Whose opinions or thoughts are we interested in?
- Where are the conversations about the topic in question happening?
- Do we need to look at the question back in time or just current discussions?

Data Analysis

In Part II of the book, we explore the data analysis techniques that can be utilized in answering questions within the data collection. Again, putting on our detective hats, we return to our "honest serving men" as described previously by Rudyard Kipling and explore a variety of topics.

How we want to look at this newly uncovered information is important. A data model is used to represent the unstructured data we collect and is an important (and complex) part of answering our questions. These data models are living and breathing entities that need to change over time or when newly discovered insights need to be incorporated into the model. These relatively long-running models tend to be complex and difficult to finalize, and as a result, many people may want to take a less-detailed view of the information. Many choose a real-time view of the data, where watching metrics or trends in real time (or near real time) provides a valuable, yet low-cost, set of insights. As an alternative between long-running analysis and a real-time view lies a structured search model that allows for the searching of common words or phrases within a dataset in an attempt to reveal some insightful information. Each type of analysis has its pros and cons, many of which are explored within this section.

In an attempt to understand *what* people are saying, we begin to explore some of the interpretations of the data, looking at simple metrics such as:

- In a collection that contains Twitter data related to a new product or service, what is the top hashtag?
- Are those hashtags positive or negative in their sentiment?
- What is the volume of conversation about the product or service? (Are people talking about it?)

Other techniques used to discern what people are talking about include the use of word clouds or the collection of top word groups or phrases. These visualizations can help analysts understand the types of conversations that are being held about the company or service in question. More advanced analysis may include the use of a relationship matrix that attempts to understand the interrelationship between concepts or terms (for example, how is the public's view of customer service correlated with perceived cleanliness of a store?).

Marketing teams will be sponsoring advertising campaigns or coming out with press releases at strategic points during a new product release or during a particular point in time—all in an effort to attract new customers while exploiting the loyalty of their existing customer base. But is their message reaching the intended audience? The question of *where* people are talking becomes important in evaluating the outlets that people use when discussing a topic. If the company is advertising mainly in trade journals but there is a large amount of conversation happening in Twitter, would the message be better spread via microblogging? (Or perhaps the use of microblogging can augment the marketing message?) Along those same lines, if we stand on a box in the center of a square and preach our message, do we want to do it in the middle of the night when the square is empty, or at lunchtime when the square is bustling with traffic. The same is true in the social media space: *when* we choose to disseminate information may be just as important as *where*.

Information Interpretation

Once we have all of this data reduced into information nuggets, making sense of the information becomes paramount. In Part III, we demonstrate that the insights derived can be as varied as the original question that was posed at the beginning of the analysis. In some cases, the goal is not only to identify *who* is doing the talking in our analysis but, more importantly, who is *influencing* the conversation or who is influential in their thoughts and opinions. It's important to remember what SunTzu once said: "Keep your friends close, and your enemies closer." The identification of the "movers and shakers" can be important in social media; these are the individuals we want to follow or attempt to get closer to in order to have them use their influence for us as opposed to against us. In other cases, *what* people are saying about a particular issue or topic is the object of the research. For example:

- Are people excited about the newly designed web experience that our company just released, or are they talking about the difficulty in finding information within our website?
- How critical are the outsourcing decisions that we just made to the brand perception of our company or product?

- What were the key issues or topics that people cited when they were expressing negative sentiment?

In our experience, we have also encountered cases in which the *where* is the most important finding. For a newly launched marketing campaign, is the conversation happening more in company-sponsored venues, or is it also happening in neutral venues? Analysis and insights around *when* are also important. For example, is the sentiment for your company becoming negative around the same time as the sentiment for a key competitor (perhaps indicating a downturn in your market)? More importantly, has sentiment for your company or brand gone negative while the competition has gone positive?

Why You Should Read This Book

According to *Merriam-Webster*, *Definition of SOCIAL MEDIA: forms of electronic communication (as Web sites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content (as videos)* [4]. Use of social media has grown exponentially over the past eight years (see Figure I.2) [2]. Thus, social media is a major contributor to the explosive growth of big data in our world.

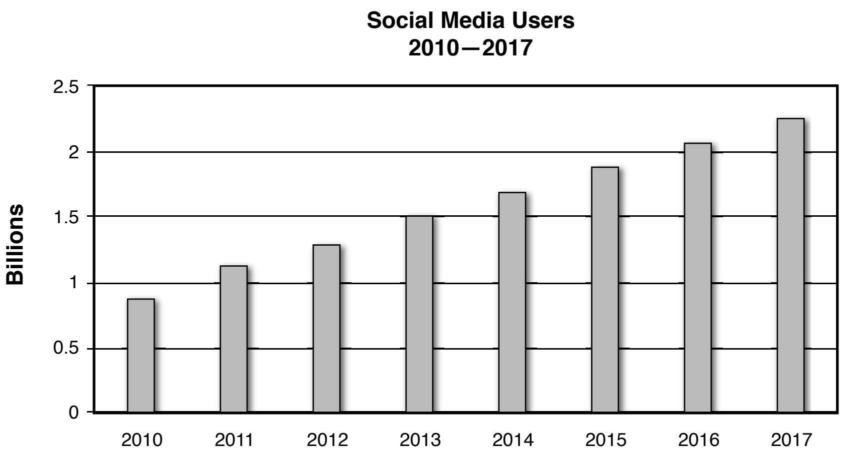


Figure I.2 Growth of social media users 2010–2017

Research has shown that the growth of social media use is far from over. According to Internet Statistics and Market Research Company eMarketer, in a report published in June 2013 [5], the current prediction is that one in four people across the globe will participate in social media by 2014. That's an incredible number. Consider also:

- Asia-Pacific will have largest social network population worldwide through 2017.
- The Middle-East and Africa will have the second largest audience starting next year, because their population penetration rates are among the lowest.
- Asia-Pacific has the largest user base with 777 million people, where 44.8% of social network users are expected at the end of the year.
- The higher penetration of Internet users in India, Japan, Australia, South Korea, Brazil, Mexico, Russia, Middle-East, and Africa has helped to revise the estimated number of social media users in 2013 by 100 million.
- In 2014, the Middle East and Africa (MEA) region emerged as the second largest social media hub, with more than 248 million users surpassing Latin America in regions in the next year.
- By the beginning of 2015, India was expected to surpass the United States as the second largest social media country after China.

IBM CEO Ginni Rometty has called big data the next great natural resource [6]. Getting in on the "ground floor" of anything can be challenging, but if you want to turn this natural resource into business value "gold," you should read this book.

This book will serve the needs of a number of business users. Those users who are new to the subject will get a good overall understanding of the domain by reading the entire book. Those users who have some familiarity with either one or more of the sections of this book will be able to get additional techniques and methodologies to add to their repertoire.

To enable you to apply the content from this book to your unique situation, we have included a number of case studies. The techniques and findings we present here are primarily based on over three years' worth of hands-on experience in executing a variety of social media analytics projects for IBM and IBM's clients. To protect proprietary information, we've edited the cases for illustrative purposes.

For example, we analyzed Twitter content for about a month before the 2014 Grammy Awards were announced and identified a list of potential winners. When the actual results came in, every single one of them was in the top three choices that we had predicted.

These are just some of the examples of value that people are finding by mining this new natural resource. We cover a variety of these use cases throughout the book. People have even used this new capability to fine-tune multimillion dollar marketing campaigns. And, in some cases, people have used analysis of Twitter data during the first two days of a conference and created talking points for an executive presentation on the third day.

By reading this book, you will get a broad understanding of the following topics:

- What are the various types of social media analysis that can be done?
- How do we collect the right kind of data for a project?
- How do we analyze the data using a variety of tools and techniques to get the value from it?
- How do we interpret the results and apply them for real business value?

What This Book Does and Does Not Focus On

A lot of good books out there are targeted at social business marketing managers and focused on how to effectively utilize social media channels to market their brand, their goods, and their services. We do not focus on that approach in this book.

This book is also not directed at technologists, architects, and programmers looking to implement the most effective technology solutions for social media analytics. We provide some information that might be helpful for this type of an audience, but this book is not primarily directed toward them.

This book also does not focus on a single technology platform or a single tool and therefore does not serve as a user manual for one of these products. The intention is to provide enough information to business users so that you can either build your technology solutions or buy solutions to serve your needs for extracting business value out of social media and textual content.

Even though this book is primarily targeted to business users, we cover several technical aspects at length to equip business users with enough knowledge to extract value from this book. Subsequent chapters cover enough detail, but what follows is a list of some of these key technology concepts with a high-level description.

- **Big Data**—Big data is usually characterized by a large volume of data, a large variety of data, and data that is moving at a large velocity (speed). For example, this includes the content flowing through the cables of your local cable TV provider during prime time or content being streamed by Netflix during the screening of an episode of *House of Cards*!
- **Natural Language Processing (NLP)**—NLP involves analysis of words used in our language. A simple application of NLP is a word cloud. A more complex example of NLP includes analyzing streams of conversations and identifying dominant themes.
- **Sentiment Analysis**—This is a special case of natural language processing. In this case, the content is analyzed by software and interpreted to identify if positive, negative, or a neutral sentiment is being expressed. For example, the sentence "I am very happy with the latest release of Product XYZ" is treated as expressing a positive sentiment, whereas the sentence "The installation process for Product XYZ is very difficult" is treated as negative. An example of neutral sentiment is "Product XYZ is supports platform A and system B."

Endnotes

[1] Reisner, Rebecca. Comcast's Twitter Man, *Business Week*, January 2009.

[2] Bernoff, Josh, and Ted Schadler. *Empowered: Unleash Your Employees, Energize Your Customers, Transform Your Business*. Harvard Business Press, 2010.

[3] Kipling, Rudyard. *The Elephant's Child: From the Just So Stories*. ABDO, 2005.

[4] "Social Media." Merriam-Webster.com. Merriam-Webster, n.d. Web. Sept. 21 2015. http://www.merriam-webster.com/dictionary/social media.

[5] Social Networking Reaches Nearly One in Four Around the World. See more at: http://www.emarketer.com/Article/Social-Networking-Reaches-Nearly-One-Four-Around-World/1009976#sthash.B7RoQKQs.dpuf.

[6] Lenzner, Robert. IBM CEO Ginni Rometty Crowns Data As The Globe's Next Natural Resource, *Forbes*, March 2013.

Acknowledgments

Matt Ganis

To be able to work my whole career with the latest technology, in fields that blossom into multimillion dollar industries, is truly a dream come true for the geek inside me. However, to be able to work in that field with my good friend is truly a blessing. Thank you, Avinash. The last few years have been a very wild ride!

However, I could never have achieved the success I have (in both my professional and personal life) if it weren't for the love of my life, my wife, Karen. She's stood by me through thick and thin, and if there is one person in this whole world that I can always count on, it's her. No matter what the circumstances, whether good or bad, I know I can turn to her for advice, a smile, or a shoulder to lean on. I love you, Karen. You make each day special for not only me, but our family. If there was ever a glue that bound a family together, it's you.

I also want to thank my children, Matthew and Taylor. I know it embarrasses you when I cheer the loudest or ask the dumbest questions, but I hope you can forgive the proudest father in the world. Matthew, I may not have your brains, and Taylor, I may not have your athletic ability, but I can promise you I will always have an endless amount of love for you both.

It's said that every journey begins with a single step. My journey started many years ago—more than I'd like to remember. But for all the sacrifices they made, from putting me through college to slugging through the snow to look through my telescope at "little white dots," thank you, Mom and Dad. You've always been there for me. You never wavered in your encouragement and always showed a pride in everything I did. Thanks for starting me down this path.

Avinash Kohirkar

I remember the day in 2011 when I was discussing a new career opportunity with Steve Wright. He wanted me to co-lead an offering of Social Analytics within IBM with Matt Ganis. As I look back, that was a very key day in my IBM life. What ensued was a wonderful ride in the world of social media analytics. I have had the great privilege of working with Matt, a technologist at heart, with a unique drive and passion to extract business benefit out of technology. The idea of this book would not have been possible without Matt. Thank you, Matt, for asking me to collaborate on this project!

In 1988, I was working with System Software for Unisys in Camarillo, California. I remember telling my wife, "I enjoy what I am doing so much that I can't believe they pay me to do this stuff!" I have been extremely fortunate that in my entire career I have felt like that pretty much all the time. I thank my wife, Smita Kohirkar, for enabling me to have a career like this by providing me unwavering support in every facet of my life. There have been numerous times in my career when my work consumed me and my attention, and she took care of me and my family during these times with a smile on her face. I was always a techie and a geek, but I learned a lot about the rest of life from her. The last few months have been especially busy for me while working on this book, and I could not have done it without her understanding and her support. Thank you very much!

I also want to thank my children, Neeraj and Sneha. Despite being the best kids a dad can have, they have been an inspiration for me. Neeraj, who has an amazing ability to learn anything in a very short amount of time, inspires me with his unlimited enthusiasm and passion for life. Sneha, who was always wise beyond her years, never ceases to amaze me with her unique ideas and unique perspectives. I love you all so much!

Joint Acknowledgments

IBM, like many large companies, is full of all kinds of personalities and interesting individuals. To call it a unique place to work is doing it a huge injustice. We've had the distinct pleasure of working with (and for) a number of really special people.

Our career change into analytics and, in particular, social media analytics is due to one person: Stephen Wright. Steve was the visionary leader who saw the potential in analytics and was our chief supporter, cheerleader, and, when we needed it, critic. IBM needs more Stephen Wrights. IBM is lucky to have him, and we were fortunate enough to be able to work for him during this exciting venture into this world of big data, analytics, and cloud computing. Steve, if ever someone were owed a huge debit of thanks, it's you. It was a truly a career highlight being part of the Enterprise Web Strategy and Experience team under your leadership.

Of course, Steve wasn't alone in his desire to see analytics used within the enterprise. We owe a huge debt of thanks to our management chain, specifically John Rosato and Ajay Raina. These are two great leaders who were always willing to support and trust us as we developed our analytics offerings into a world-class operation. As our services grew in sophistication, we joined forces with Liam Cleaver and James Newswanger as they formed IBM's Social Insights Group. To both of you, thank you for your support and willingness to allow us to grow our customer base.

As authors, we are indebted to both Ajay Raina and Debbie Delosa for reviewing the entire manuscript and providing us with invaluable feedback and critiques.

Thanks as well to Santosh Borse, Mila Gessner, and Chris Gruber for their technical and analytics leadership in executing a variety of social analytics projects over the years. We have used examples in this book that were based on individual contributions of these highly skilled analysts and technical wizards. (Santosh, how you pull off some of your magic still amazes us to this day!)

About the Authors

Dr. Matthew Ganis, a member of IBM's Academy of Technology, is currently an IBM Senior Technical Staff Member located in Somers, New York. His current areas of interest include social media analytics, the Internet of Things, and Agile software methodologies. He is an adjunct professor of computer science and astronomy at Pace University in Pleasantville, New York, where he teaches at both the undergraduate and graduate level.

Dr. Ganis holds a BS degree in computer science and an MBA in information systems from Pace University, an MS degree in astronomy from the University of Western Sydney, Australia, and a doctorate of professional studies in computing from Pace University. He has authored or co-authored over 40 papers in both of his fields of interest, ranging from programming techniques, computer system administration, computer networking, and topics on stellar evolution and radio astronomy. He is also the proud co-author of *A Practical Guide to Distributed Scrum* published by IBM Press.

In his 30-year career at IBM, he has been responsible for a number of technological advances such as the creation of the first enterprise firewalls for IBM; the creation of highly available World Wide Web platforms to support the Atlanta, Sydney, and Nagano Olympics (which secured Dr. Ganis and his team a spot in the *Guinness Book of World Records* for the highest sustained rate of Internet web traffic); and the proliferation of advanced software development techniques across IBM's worldwide development laboratories.

He can be found on LinkedIn (https://www.linkedin.com/in/mattganis), on Twitter as @mattganis, or on his blog at http://mganis.blogspot.com.

Avinash Kohirkar is currently Manager of Social Business Adoption in IBM. His current areas of interest include deployment and adoption of social technologies within an enterprise, social engagement dashboards, and social media analytics. Avinash Kohirkar holds a BS degree in electronics and communications engineering from Osmania University (India), an MS degree in industrial engineering from NITIE (India), and an MBA in finance from California State University. He has contributed to IBM white papers and has given numerous presentations on social analytics in IBM and outside IBM. He has authored a number of articles on this subject that have been published in the *Cutter IT Journal* and *Infosys Lab Briefings*.

In his 19-year career at IBM, he has leveraged technologies such as e-business, Web 2.0, social collaboration, social graph technologies, big data, and social media and text analytics for the business benefit of IBM and IBM's customers. He is recognized as a thought leader in the project management profession within IBM and is certified as Executive Project Manager at the highest level within IBM. He has held several technical, business, and management positions during his career: Architect, Development Manager, Project Manager, Project Executive, Associate Partner, Project Executive, and Business Manager.

He can be found on LinkedIn (https://www.linkedin.com/in/AvinashKohirkar) and on Twitter as @kohirkar.

This page intentionally left blank

3

Whose Comments Are We Interested In?

All opinions are not equal. Some are a very great deal more robust, sophisticated, and well supported in logic and argument than others.

—**Douglas Adams**, *The Salmon of Doubt* [1]

Up to this point, we have concerned ourselves with what data to analyze while ensuring that what we selected is germane to our topic. In this chapter, we explore how important it is to determine *whose* comments we are interested in. A few examples are as follows:

- If we are interested in getting *objective* feedback on a product from a specific company, we might want to make sure that we can identify or exclude this company's employees from the pool of content under analysis.
- Similarly, we need to ask: Are we interested in comments from the general public, or are we interested in the comments of C-level employees (that is, chief marketing officers or chief information officers)?
- Also, are we interested only in people who have a positive bias toward a company or those with a strong negative bias?

Looking for the Right Subset of People

At the beginning of a social analytics project, analysts spend a fair amount of time thinking about the ultimate goals of the project and the results that we expect to get at the conclusion of the project. This upfront analysis will go a long way in determining the appropriate target segment of the analysis.

During the definition of a typical social media analysis project, requesters will (or should) explicitly point out the "who" (whose opinion are they interested in?) or will give the researcher or the model builder sufficient hints or guidance. Various attributes can be used to segment or target the audience that we're interested in. Some of them are described in the following sections.

Employment

Do we want the opinions of employees or nonemployees?

For example, if a company launches a new product or service and wants to see how the marketplace is reacting to that product or service in social media, it might prefer to exclude the comments of its own employees. In other situations, we might exclusively focus on the employee population if the intent is to learn how they are responding to a new product, service, or strategy. In a project that we worked on, IBM was interested in learning about the marketplace reaction of a brand-new product type. The marketing team specifically asked us to exclude the comments and sentiments of IBMers to understand sentiment from "neutral" people so as not to bias the results.

Sentiment

Are we looking for comments from people with a positive bias or negative bias?

For example, if the object of social media analysis is to detect customer support issues, it makes sense to focus only on posts with a clear negative bias. You might argue that highlighting positive customer experiences is just as important and probably needs to be considered as well. Another common use case involves trying to compare the sentiment about a variety of products that a company is providing to the marketplace. In this situation, we may consider opinions from all ranges of demographics and keep score about the number of positive, negative, or neutral comments. Sometimes, the purpose of a project is merely to find how many people or comments mention the company's product versus a competitor's product. In this case, we may (initially) ignore sentiment and consider all comments without exclusions.

A few years ago, there was a civil movement called Occupy Wall Street in the United States. Numerous people congregated around specific commercial buildings to express their silent protests against what they believed to be unfair practices. During this time, as a validation of some of our analytics capabilities, we built an experimental social listening model to detect whether there was any impact to an IBM location where some key customer meetings were being conducted. In this case, we built a model that focused on snippets of information that may have negative sentiment about IBM and then specifically looked for any mentions of protests or civil actions.

In many cases, sentiment is a result of an analysis phase. However, in some instances, the scope and nature of the project determine whether we should include comments only from people who have either a favorable view or an unfavorable view of our topic. In cases like these, we are able to take this information into account in the very initial phase of the project and focus only on a specific subset of people.

Location or Geography

Do we want to focus on comments from people who live in a specific location?

One of the projects that we were involved in dealt with issues around water in South Africa. In this particular project, we were clearly interested in comments from people in South Africa about the variety of issues and questions around the current and future needs and use of clean and healthy water. Sometimes we may be interested in comments from all over the world, but valuable insights can emerge when we classify the analytics by region.

Language

Is the language of the content important to us?

Some projects require us to understand what is specifically being said about a company's product or service in a particular local language. For example, if a company wants to do some market research around the market's appetite for a machine translation tool in Spanish-speaking countries, it will be interested in content contributed by individuals in the Spanish language.

Age

Is the age of content author important to the project at hand?

There is a lot of discussion in popular media about the work habits of Generation Xers. Those in Generation X (or Gen X) were born after the Western Post–World War II baby boom. As a point of reference, most consider those with birth dates ranging from the early 1960s to the early 1980s as being part of this demographic. If a company's Human Resources department wanted to study the experience of its newly hired Gen Xers, we would have to determine a way to segment the population based on age.

Gender

Are we specifically interested in comments of men or women?

Gender also becomes an important attribute upon which we may segment audience for a particular project. If an organization is creating training and educational materials to encourage more women to pursue higher studies in science- and mathematics-related disciplines, it may choose to focus exclusively on comments and feedback from women. Similarly, if a health-care company is undertaking research about male-pattern baldness, it would be served well by segmenting its audience to include only men.

In one case, we were asked to evaluate the comments that were made in social media during the introduction of a new movie trailer. Our client was interested not only in the reaction to the trailer, and by association the movie itself, but also if certain themes resonated with either males, females, or both. Again, the goal was to determine not only likeability of the movie, but also keys in how to market it.

Profession/Expertise

Do we need opinions from anybody in general, or do we need opinions from people who are working in a specific profession (such as the IT profession) in a specific industry (such as automotive)?

For example, if IBM is interested in learning about the reaction to the cognitive computing capabilities of IBM Watson in the area of health care, it is probably interested in the opinions of corporate users as opposed to home users.

Eminence or Popularity

Are we interested in opinions only from people of certain standing in the domain of the topic area?

A major aspect of a social media campaign for companies involves identifying who might be an "influencer" in a particular topic area or industry. For performing this type of analysis, we tend to spend a lot of time in developing rules to ensure we are able to narrow the solution space to identify a small subset of individuals that a company should target its marketing messages to.

Role

When dealing with social media analysis within a company's intranet, are we interested in segmenting based on a specific job role?

For example, we are working on a project that computes a social scorecard for employees based on their participation in social media. There are some roles in which the job demands a lot of collaboration in social media, and then there are some people who might be working on highly specialized or highly sensitive projects in which they may not be allowed to share information in social media. Here, the type of role is very important in interpreting scores.

Specific People or Groups

Are we really interested in narrowing down our analysis to comments about or comments from a specific individual or a specific set of individuals?

A couple of years ago, we were asked to build an application to capture and display sentiment in near real time about tennis players participating in the US Open. In this case, we used names of players, their nicknames, and a variety of other aliases to ensure we were targeting the right segment. In another example, we were asked to identify how people in social media were reacting to a Lance Armstrong interview with Oprah Winfrey.

Do We Really Want ALL the Comments?

In Chapter 1, we discussed the concept of bias—or the skewing of a dataset based on a potentially inappropriate set of authors. Perhaps *inappropriate* is too strong of a word, but in some cases you might want to exclude the comments of your company's employees. At IBM, we tend to look at ourselves as one of the best customers of our products and services, but sometimes IBMers are also among our most vocal critics. If we are looking to understand the true concerns or thoughts of our external customers and clients, we may want to exclude the subset of IBMers from the conversation. This is an example of the employment attribute that we discussed previously. Again, the purpose isn't to exclude because these comments aren't valuable, but in the spirit of openness and true sentiment or feelings, it may be useful to separate the comments.

In one example, we were asked to look at the social media activity around a new product launch. The client's concern was that while there was a tremendous amount of money and time being invested in the various marketing campaigns, the sales hadn't picked up as much as had been anticipated. A quick analysis of the discussion around the topic showed the level of activity over a four-week period (see Figure 3.1).

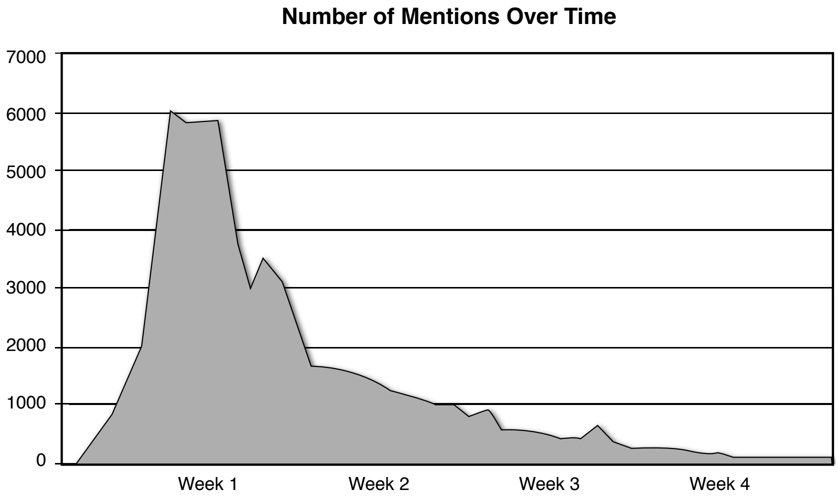


Figure 3.1 Social media remarks during an initial product announcement.

This graph shows the number of mentions of the particular product over time. It's rather clear from this simple graphic that in the beginning, there was quite a bit of hype or discussion around this product launch, but over a short period of time, the discussion continued to decline almost to zero mentions.

What was even more disturbing about this analysis was who was having the conversations. We quickly looked at the top contributors to this thread of conversation and turned up the list shown in Figure 3.2.

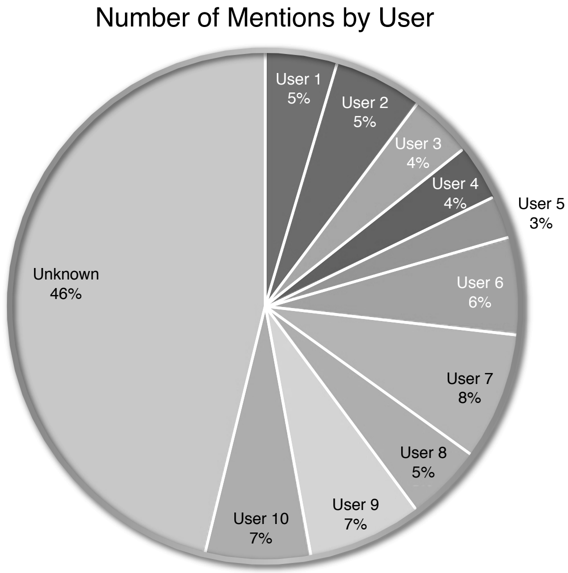


Figure 3.2 Top contributors to social media remarks during an initial product announcement.

A manual lookup of the top 10 users in this conversation revealed that at least 9 of them were employees of the company and represented nearly half the conversation (47%).

The conclusion we drew was that in the various social media and news venues, the employees were chatting about the new release, but given the slope of the curve in Figure 3.1, that conversation didn't sustain itself. After the employees stopped talking, there was virtually no conversation. Clearly, a new marketing plan was needed since what was being said wasn't being repeated, commented on, or perhaps even resonating with the public.

Are They Happy or Unhappy?

I'll never forget the time I [Matt] was traveling to Las Vegas to speak at a trade show. It was a long flight, but when we landed and the plane was taxiing to the gate, I simply tweeted "Viva Las Vegas" and was almost instantly greeted with a return tweet for a hotel/casino special. Someone was actually watching for conversation about the city, not just me, to send a special offer.

Watching or monitoring social media for customer issues is still a growing trend. It provides the ability to respond to issues in a timely fashion as well as gives opportunities for additional business opportunities.

Consumers are using Twitter to either ask questions about product- and service-related issues or to air complaints with increasing regularity. A study by Sprout Social found that social media messages eliciting a direct response from companies had risen by 178% from 2012 to 2013 [2]. To stay competitive, companies are choosing to watch for negative terms or concepts being used around a brand and head off a potential customer satisfaction problem later.

By listening to customer feedback in Twitter, companies like JetBlue have been able to build their reputation as responsive customer service organizations. Think about this from the consumers' perspective. Airline delays can be one of the most common causes of customer frustration. Not only do these delays happen often, but those being delayed or inconvenienced can be pretty vocal about their feelings, especially when there is nothing to do but sit in an airline terminal with their smart phones.

Acknowledging this fact, @JetBlue ensures the company is responsive to its customers because it understands the importance of continued customer loyalty. JetBlue not only engages with happy customers but also responds to and helps frustrated customers as quickly as possible.

According to an article in *AdWeek* [3], due to a downpouring of rain in the Northeast that grounded most of JetBlue's planes, the company was facing a public relations storm that seemed unlikely to go away anytime soon. On this particular occasion, passengers were trapped in their planes (on the tarmac) in New York City for hours—going nowhere and growing more annoyed by the minute. In many cases, passenger delays stretched into days while over 1,000 flights were ultimately canceled.

Needless to say, customer concerns and outcries ran rampant. However, through social media channels, then-CEO David Neeleman reached out to travelers of JetBlue to *personally* apologize for the issues and presented the company's plans to improve service. The use of social media outlets to enable an open atmosphere of communication coupled with the company's willing to admit (publically) its mistakes went a long way to turn a bad situation good.

The lesson?

Listening to the right content (in some cases, customer dissatisfaction) can provide an added vehicle to achieving customer loyalty and goodwill.

JetBlue leveraged YouTube (a popular video-sharing site) to explain the service failure and describe how it planned to improve its operations as a part of its effort to control the situation. Again, it did this by posting an apology by founder and then-CEO David Neeleman shortly after the trouble began. As a result, the company built a relationship with its customers.

This use of a social media source coupled with JetBlue's complete openness and willingness to take responsibility helped to push it over the media reports and resume its standing as a consumer favorite. What's important is that despite the negative news coverage and complaints by consumer advocacy groups, the airline was able to keep its place atop the J.D. Power North America Airline Satisfaction Study for low-cost carriers going on 11 years in a row [4]!

So when we think about who we want to listen to, the answer, of course, is everybody. But by segmenting the comments into those with positive sentiments and those with negative sentiments, we can quickly respond to those urgent customer issues.

Location and Language

There are times when understanding the mood or the thoughts of a particular region of the world is of main importance. For example, if we are interested in understanding the social opinions or concerns of youths in India, monitoring data from the United States isn't all that practical. Just to be complete in this thought, however, while we understand that there may be some spillover discussion in US-based traffic about conditions in India, the likelihood of finding any significant content is probably not worth the effort of having to discover it in a vast sea of other (unrelated) data. Obviously, this is a decision that needs to be made by each data scientist or organization; our intent is simply to point out where there may be value in looking only at a particular region in the world.

As an example, consider the diagram shown in Figure 3.3; it shows social media mentions for a particular bank we were working on an analysis for. The bank had recently made some announcements and was interested to see if there was an increase or decrease in social media traffic as (perhaps) a result of the media attention. Figure 3.3 shows a summary of the top 10 languages for all of the media mentions we were able to collect over the previous two days.

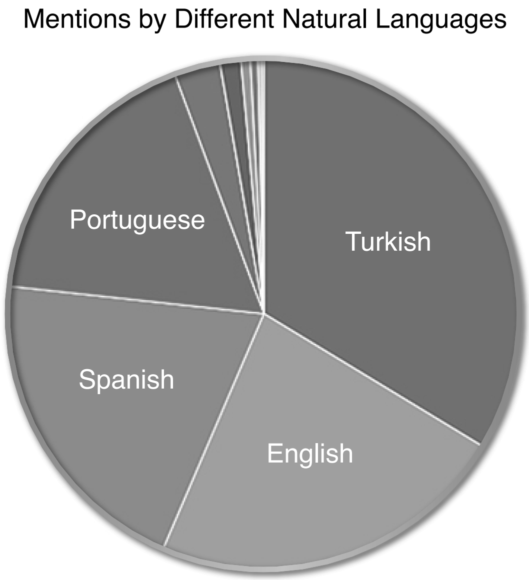


Figure 3.3 Top 10 languages used in mentions.

What we were able to see was a large amount of traffic coming not from English (US) speaking individuals, but from Turkish social media participants. Not only that, but it appeared that Portuguese and Spanish numbers were almost equally as high. What was more interesting was that the announcements were made in the United States.

One of the interesting facts to gather would obviously be the location of the individuals making the comments. In some cases, this information is easy to retrieve—for example, through the use of GPS technology on mobile devices. In the case of Twitter, the use of geolocation can allow someone to find tweets that have been sent from a specific location. This could be a country, a city, or multiple regions around the world. When a Twitter user opts in to allow location-based services on his or her Twitter account, Twitter uses geotagging to categorize each tweet by location and makes that information available to subscribers of the data. In theory, this would give users of that data the ability to track tweets sent from a specific city or country. Unfortunately, the statistics on the use of this feature aren't promising (yet), with only about 10% of the total population enabling the feature [5].

Lacking the exact geolocation, we could make the assumption that those posting in Turkish, for example, were originating their tweets from Turkey.

It may not be a perfect one-to-one match, but lacking any other information, it's the best we could do.

In this case, the bank in question had made an announcement (in the US press) about some branch closings in Europe. From the backlash we were able to mine from social media sources, it appears that those most widely affected customers were located in Spanish-speaking countries as well as Turkey. While we don't know exactly how the bank handled this situation (our job was simply to discover any potential issues), we do know it immediately focused customer relations on branches and banking in those regions in an effort to minimize any fallout from its announcements.

Age and Gender

Understanding the demographics of just who is using social media to communicate is an important step in being able to understand what is being said about a company or brand.

Some of the current data provided by the Pew Research Center [6] around social media can give us a better idea of who is generating all of the traffic (and who is listening). Let's not make a mistake here: according to this work, approximately 74% of Internet users are engaged in some form of social media (that's over 2.2 billion individuals). While we've tried to summarize some of the more simple statistics in Table 3.1[7], some numbers should stand out:

- In the 18–29-year-old bracket, there is 89% usage.
- The 30–49-year-old bracket sits at 82%.
- In the 50–64-year-old bracket, 65% are active on social media.
- In the 65-plus bracket, 49% are using social media.

Time spent online using social media shows [8]:

- The United States at 16 minutes of every hour
- The Australians at 14 minutes for every hour
- The United Kingdom users at 13 minutes

And while we're at it, remember that 71% of users' social media access comes from a mobile device [9], and women tend to dominate most of the social media platforms [10].

Table 3.1 Social Media Demographics of Prominent US Sites as of December 2014

Social Media Site	Percent of *Males* polled that participated	Percent of *Females* polled that participated	Ages 18–29	Ages 30–49	Ages 50–64	Ages 65 and older
Facebook	66%	76%	84%	79%	60%	45%
Twitter	17%	18%	31%	19%	9%	5%
Instagram	15%	20%	37%	18%	6%	1%
Pinterest	8%	33%	27%	24%	14%	9%
LinkedIn	24%	19%	15%	27%	24%	13%

Ultimately, we would like to include some of this demographics information in an analysis, but the knowledge of this information is just as useful. If, for example, we were wondering what the issues were surrounding health care (or other issues) post retirement in social media, we would be hard-pressed to find much discussion by that demographic in places such as Instagram or Twitter (since the number of participants in the 65 and older demographic seems to be quite low). That's not to say the chatter wouldn't be out there; there could be significant discussion by the children of those users in the 30–39-year-old demographic, but again, it may come with a different perspective. Similarly, based on this table, if we were interested in the content from females, Pinterest might be a good venue to consider.

Eminence, Prestige, or Popularity

What does it mean to be eminent? There are a number of online presentations and seminars on increasing your social media eminence, or "digital footprint." What are some attributes of eminent people? They tend to be in a position of superiority or distinction. Often they are high ranking or famous (either worldwide or within their social community or sphere of influence) and have a tremendous amount of influence over those who hear what they have to say.

For example, if the president of the United States (or any world leader) makes a comment on some social or economic issue, that comment is usually picked up by the press and is on everyone's lips by the time the evening news comes on (more so if it's a controversial topic). These leaders are highly influential and can literally change the minds or perspectives of millions of people in a relatively short time span. On the other hand, if coauthor Avinash Kohirkar makes a public statement about the same topic, the results are vastly different. He may influence family and friends, but the net effect of his comments pale in comparison to those that are viewed with a higher degree of eminence.

So what do these users do to lay claim to being popular, prestigious, or eminent?

People who are perceived to have a high degree of social media eminence publish high-quality articles or blog entries. Other users rush to see what they have to say (and often repeat it or are influenced by it). Highly eminent people are seen as those individuals who add value to online business discussions. Their eminence is further bolstered by others who have rated their contributions as valuable and have tagged them for reuse by others. In Chapter 11, we talk about how social analytics can be used to determine eminence!

It stands to reason that we would want to know what these people are saying. We also want to know if something was said in the social media concerning our brands or products. It does make a difference if a comment was made by a simple techie (such as Avinash) or a world leader.

One of the challenges in using eminence (or influence) as a metric is determining how to quantify it. There is a lot of discussion and debate in the industry about this topic, and there are lots of tools and approaches that people are using to measure influence [11]. To illustrate this point here, we are going to make some assumptions and come up with a simple formula.

In some of our work, we make the following assumptions:

- Influential people are those who often have their comments repeated.
- Influential people tend to have many people following them (that is, the interest in what they have to say is high).

Based on these assumptions, we defined a simple metric called "reach" that is a quantifiable way to determine how widespread someone's message could be. Reach, to us, is simply the number of things that a person has said multiplied by the number of people listening. Is this metric perfect? No. But it is something to watch for: a person with a large reach is saying a lot and is also reaching a wide audience. Granted, someone could be blabbering about some topic on social media and posting thousands of messages, all being received by a small handful of listeners. If that's a concern, simply look to modify the definition of influence to something like that shown in Figure 3.4.

Method 1:

$$\boxed{\text{Reach} = \text{Followers} \times \text{Messages}}$$

Method 2:

$$\boxed{\text{Reach} = (\text{Followers} \times \text{Messages}) \times \frac{\text{Followers}}{\text{Messages}}}$$

Figure 3.4 Simple formulas for calculating influence.

It is possible for a company to use the concept of influencers to effectively communicate a key marketing message broadly. Consider the effect a well-known industry analyst who is constantly talking about security in financial institutions such as banks could have on the perception of various institutions. In addition, if we follow this analyst, we will come to understand the social media venues that this analyst and others like him or her participate in. As an example, let's assume that IBM acquired a company that specializes in fraud detection for banks. Our marketing teams in IBM will be served well by posting about this event on the venues that this analyst is already quite active in. If the analyst is impressed by the acquisition and chooses to "like" it or "share" it, that message will be received by a large number of his or her followers.

How do we measure how influential someone is? Or how do we measure how effective a person's messages are? We can look to see if that person has talked about a specific product or service and then measure the sales of that product or service to see if there is an increase (or decrease). However, that would be a difficult measurement and, quite honestly, wouldn't represent the image or perception of the product or service, which could, at a later date, affect the sales.

Instead, we chose to look at someone's reach, or how far and wide this person's message *could be* spread. Figure 3.4 shows an example of how reach could be computed in a message system such as Twitter (although it's equally applicable to any systems where a post is made and others follow that posting).

In Figure 3.4, we show that an individual's reach can simply be calculated in one of two ways:

- **Method 1**—Multiply the number of messages sent by the number of people that could read that message. If someone sends 1,000 messages and 10 people are following that person, the combined message has a calculated score of 10,000 (see Table 3.2).
- **Method 2**—Multiply the number of messages sent by the number of people that could read the message and then multiply that result by the ratio of followers to messages.

Table 3.2 Example of Determining Someone's Reach in Social Media

Followers	Messages	Reach (Method 1) (Followers * messages)	Ratio	Reach (Method 2)
10	1,000	10,000	0.01	100
200	50	10,000	4	40,000

In method 2, we've add another factor to our equation: the ratio of the number of followers to the number of messages produced. Doing so effectively gives more weight to the person with a larger following. This produces perhaps a more meaningful score for our metric, where we might be more inclined to focus on the comments of the second user rather than those of the first.

Summary

As you can see, as we're moving forward in these chapters, we're trying to get more and more specific about the data that is under analysis. In this chapter, we discussed the concept of the individual in the conversation, or the who. It's a huge point that we need consider in any kind of analysis we're looking to perform. Remember, if you're looking to understand the societal issues in, say, India, does it make sense to include opinions or thoughts of those people in the United States? Perhaps. But at a minimum, we believe you should at least consider breaking out the views of Indians to better understand your question at hand.

If the public chatter about a new movie contains the words *childish*, *silly*, or *waste of time*, is it relevant? That depends. If the movie is geared for children, and those are the views of adults, perhaps not. Remember, sometimes it's not what is said, but who is saying it!

Endnotes

[1] Adams, Douglas, *The Salmon of Doubt,* 2002, Random House Publishing.

[2] McCauley, Andrew, "Nice to Tweet You: 3 Ways to Use Twitter for Customer Service," *Modern Marketing Blog,* April 22, 2014. Retrieved from http://www.responsys.com/blogs/nsm/social-media-marketing/nice-tweet-3-ways-use-twitter-customer-service/.

[3] Giantasio, David, "JetBlue Knows How to Communicate with Customers in Social, and When to Shut Up: Mastering the Transparency Game," *AdWeek*, September 9, 2013.

[4] Maxon, Terry, "J.D. Power Study Puts Alaska Airlines, JetBlue Airways at Top of Customer Satisfaction," *The Dallas Morning News*, May 13, 2015. Retrieved from http://aviationblog.dallasnews.com/2015/05/j-d-power-study-puts-alaska-airlines-jetblue-airways-at-top-of-customer-satisfaction.html/.

[5] "An Exhaustive Study of Twitter Users Across the World," Beevolve, Inc., October 10, 2012. Retrieved from http://www.beevolve.com/twitter-statistics/.

[6] "Social Networking Fact Sheet," *Pew Research Center*, 2014. Retrieved from http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/.

[7] See http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/.

[8] See http://www.computerworld.com/article/2496852/internet/americans-spend-16-minutes-of-every-hour-online-on-social-nets.html.

[9] See http://blogs.adobe.com/digitalmarketing/mobile/adobe-2013-mobile-consumer-survey-71-of-people-use-mobile-to-access-social-media/.

[10] Clifford, Catherine, "Women Dominate Every Social Media Network—Except One," *Entrepreneur*, March 4, 2014. Retrieved from http://www.entrepreneur.com/article/231970.

[11] Bullock, Lilach, "What Are the Best Tools to Measure Social Media Influence?," Yahoo Small Business. Retrieved from https://smallbusiness.yahoo.com/advisor/best-tools-measure-social-media-influence-153738338.html.

Index

Symbols

A

B

C

E

J

T

U

V

W

Y