Chapter 10

Cultural Anthropology through the Lens of *Wikipedia*

Peter A. Gloor, Joao Marcos, Patrick M. de Boer, Hauke Fuehres, Wei Lo, and Keiichi Nemoto

Contents

10.1 Introduction

Over the last 10 years, the web has become a mirror of the real world (Gloor et al. 2009). More recently, the web has also begun to influence the real world: societal events such as the Arab Spring and the Chilean student unrest have drawn a large part of their impetus from the Internet and online social networks. In the meantime, *Wikipedia* has become one of the top 10 websites,* occasionally beating daily newspapers in the actuality of most recent news. Be it the resignation of German national soccer team captain Philipp Lahm or the downing of Malaysian Airlines Flight 17 in Ukraine by a guided missile, the corresponding *Wikipedia* page is updated as soon as the actual event happens (Becker 2012, Futterer et al. 2013).

In an ongoing project at the MIT Center for Collective Intelligence started in 2008 (Iba et al. 2010), we are using *Wikipedia* as a "socioscope" to study different aspects of intercultural human evolution. We draw on different-language Wikipedias to provide a mirror of today's historical understanding of world history through different cultures to measure differences in gender equality, track emotionality and positive sentiment in different-language news, and assess whether war and conquest or science and art are considered more important in different cultures.

We are using Wikipedias in different languages as a window into the "soul" of different cultures, replacing anthropological fieldwork with statistical analysis of the treatment given by native speakers of a culture to different subjects in *Wikipedia*. One of the most popular categories in *Wikipedia* is the people pages, talking about the most important people of all ages. Wikipedians have put together "notability criteria" that clearly define if a person deserves inclusion into *Wikipedia* or not. In earlier work, we have used the people pages in *Wikipedia* to identify the most influential thought leaders, idea givers, and academics through social network analysis techniques (Frick et al. 2013) and have calculated maps of the most influential people in the different-language Wikipedias (Kleeb et al. 2012).

In this chapter, we introduce Wikihistory, a dynamic temporal map of the most influential people of all times in four different Wikipedias (English, German, Chinese, and Japanese) and then look at gender distribution in the English, Portuguese, Spanish, and German Wikipedias among people considered noteworthy, resulting in a comparison among cultures on the subject of gender equality. In addition, we also compare sentiment and emotionality of news articles in the English, Spanish, Portuguese, and German Wikinews, a hand-curated news page that is part of *Wikipedia*.

---

* http://www.alexa.com/topsites.

10.1.1 Bias of **Wikipedia**

For different parts of our analysis, we rely on the English, German, Chinese, Japanese, Portuguese, and Spanish Wikipedias. While the German, Japanese, and Portuguese Wikipedias are more or less representative of their language spaces, things are more complex for the English, Chinese, and Spanish Wikipedias. Owing to the global dominance of the English language, the English *Wikipedia* is by far the largest, with 1.4 million monthly page edits in the United States (with U.S. population of 308 million) and 486,000 page edits in the United Kingdom (with a UK population of 62 million). To a smaller extent, English Wikipedians also exhibit some political bias, as they have a reputation of having a leftist liberal bent. In a study, Greenstein and Zhu (2012) found that compared to the early days of *Wikipedia*, the liberal bias, while still there, has been reduced. They found that phrases like "civil rights" and "trade deficit" favored by Democrats are still more prominent than phrases like "economic growth" and "illegal immigration" favored by Republicans. We speculate that the other Wikipedias we analyzed might show AQ1 a similar bias introduced by most *Wikipedia* editors being intellectuals, journalists, and academics, which have been shown to in general having left-leaning sympathies (Brint 1984, Blodget 2009).

While the English *Wikipedia* could claim to reflect the dominant view of the world, it comes with a heavy bias toward the worldview of the United States and United Kingdom. This is in contrast to the Chinese *Wikipedia*, with 51,000 monthly edits in Hong Kong (with a population of 7 million), 46,000 monthly edits in Taiwan (with a population of 23 million), and 35,000 monthly edits in mainland China (with a population of 1.3 billion people). As *Wikipedia* is officially blocked in China, editing is done to a large part in Hong Kong and Taiwan; however, as we will see in the following, based on the importance of communist heroes in the Chinese *Wikipedia*, we suspect that Chinese censors are actively editing the Chinese *Wikipedia*.

The Spanish *Wikipedia* on the other hand draws on a widely distributed editor base in Spain, but also many contributors in the more populous Latin American countries such as Colombia and Mexico and some very active editor communities in Chile and Argentina.

10.2 Methodology for Network Creation

Our goal was to identify a social network of all notable people in the history of mankind. Given our subjective view of the past and the fact that our collective memory forgets facts over time, we had to approximate this social network using a proxy: we only consider people that made it into *Wikipedia*, for whom we assume that they fulfill Wikipedia's notability criteria. In this social graph,
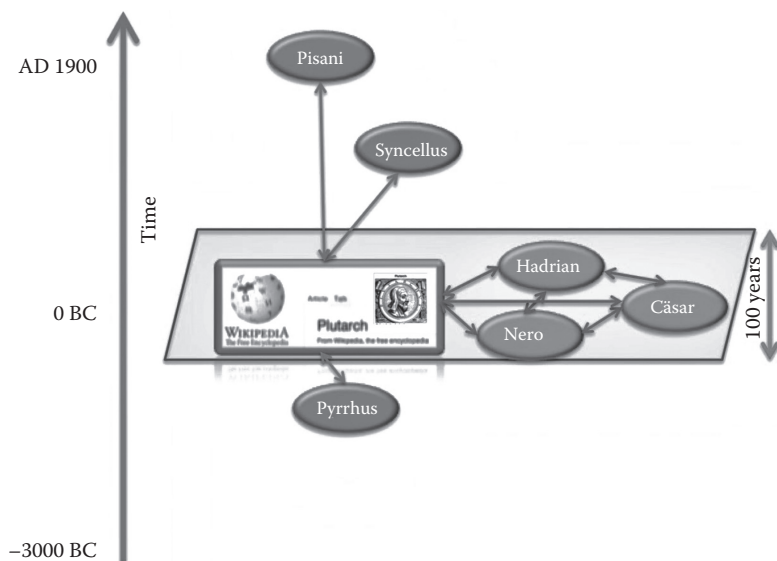
Figure 10.1   Construction of a social network of a historical personality.

people are considered vertices. For every underlying people page on *Wikipedia*, all references to other people are used to infer edges to other vertices. To put it in simple words, a link between two people is constructed if one is mentioned on the *Wikipedia* page of the other. The amount of mentions is used to calculate the weight of each directed edge. As a second requirement, a link between two people can only exist if both of them were living at the same time. This is done to keep the "social element," as a relationship embodies some sort of interaction between two people, which is hard to imagine if one interaction partner is dead. Of course, this connectivity is filtered through the eyes of twenty-first-century *Wikipedia* editors. For the English *Wikipedia*, we therefore start with the 900,000 pages tagged as "people pages." All 900,000 people pages are dated by extracting the dates of birth and death of each individual. Using this information, 4900 networks through history, from 3000 BC to 1950 CE, are calculated, as shown in Figure 10.1.

For instance, in the graph shown in Figure 10.1, from all the links to the page about Plutarch, only the links from and to Hadrian, Caesar, and Nero are kept, while the links to Pyrrhus, who died well before Plutarch was born, and the pages to medieval historian Syncellus and modern historian Pisani are ignored. Repeating this process leads to the 4900 unique networks mentioned earlier (less for the Chinese and Japanese *Wikipedia*, as their history does not go as far back). For each of these networks, the most central people in the graph are determined using the PageRank algorithm (Page et al. 1999). To get a second selection criterion

among all the influencers, their indegree, that is, other people pages pointing back to them, is taken. As has been shown in 30 years of social network analysis research, the structural properties of individual actors correspond to their real-world influence (Wassermann and Faust 1994).

10.3 The Top 50 of All Times

Who are the most important people of all times? The answer to this question is quite different in the United States, the United Kingdom, and German-speaking countries than it is in China and Japan. Looking at the top 10 and top 50 people lists also confirms that most English language *Wikipedia* editors come from the United States and the United Kingdom, while Chinese language editors come from Taiwan, Hong Kong, and mainland China.

10.3.1 *English* Wikipedia *Analysis*

The 50 most important people in the English *Wikipedia* from 3000 BC to AD 1900 are primarily politicians (26 kings and generals), followed by religious leaders (13), and poets and historians (11) (Table 10.1).

The disproportionally large role of historians clearly stands out as they shape our view on the past. It seems that it pays to be a historian, to write one's own place in history. This is clearly shown by Sidney Lee, a Victorian professor of English and history, who wrote 800 biographies and thereby secured his place in the annals.

Not only is a nineteenth-century biographer under the top 10 influencers of all times (which is also an artifact of our collection method), but also classical historians like Polybius, Tacitus, and Plutarch get very high ranks. Treating biographers and historians well so they write positively about world leaders is of course no new insight: Roman emperor Vespasian was paying historians Tacitus, Suetonius, Josephus, and Pliny the Elder, and in return they speak suspiciously well about him, shaping his positive image in history. Caesar and Winston Churchill took this concept one step further, writing their history themselves. As today's history is written in *Wikipedia*, the conclusion seems obvious: treat Wikipedians well! This means for example to give them more recognition or provide research funding for *Wikipedia*-related work.

The next three networking pictures illustrate the English *Wikipedia* Leadership networks through the ages, with snapshots taken at 600 BC (Figure 10.2), at AD 0 AQ2 (Figure 10.3), and at AD 600 (Figure 10.4). In correspondence with the availability of written records, in 600 BC the Greek philosophers, documented in classical texts by Roman librarians, and Babylonian kings, documented in writings on the walls of historical buildings, are in the center, but there is also a noticeable cluster of Chinese leaders, as well as historical figures around Buddha. AQ3

Table 10.1   Top 50 Most Important People of All Times in Four Wikipedias (Most Important at the Top)

	English	Chinese	Japanese	German
1.	*George W. Bush*	*Mao Zedong*	Ikuhiko Hata	*Adolf Hitler*
2.	William Shakespeare	*Yuan Shikai*	*Tokugawa Ieyasu*	Johann Wolfgang von Goethe
3.	Sidney Lee	Jay Chou	*Toyotomi Hideyoshi*	Aristotle
4.	**Jesus**	*Oda Nobunaga*	*Adolf Hitler*	Benedikt XVI.
5.	*Charles II of England*	*Tokugawa Ieyasu*	*Oda Nobunaga*	Platon
6.	Aristotle	*Emperor Gaozong of Tang*	*Hirohito*	Martin Luther
7.	*Napoleon*	*Cao Cao*	*Tokugawa Hidetada*	*Otto von Bismarck*
8.	Muhammad	*Kangxi Emperor*	*Tokugawa Iemitsu*	Johannes Paul II.
9.	*Charlemagne*	*Emperor Huizong of Song*	*Chiang Kai-shek*	Johann Heinrich Zedler
10.	Plutarch	*Yongle Emperor*	*Tokugawa Ienari*	Johann Sebastian Bach
11.	*Julius Caesar*	*Kangxi Emperor*	*Emperor Meiji*	*Wilhelm II. (Deutsches Reich)*
12.	*William III of England*	*Hongwu Emperor*	*Tokugawa Tsunayoshi*	*Karl V. (HRR)*
13.	Homer	*Jiajing Emperor*	*Tokugawa Yoshimune*	Wolfgang Amadeus Mozart
14.	**Bede**	*Koxinga*	*Emperor Go-Daigo*	*Napoleon Bonaparte*
15.	**Athanasius of Alexandria**	*Wang Shichong*	*Minamoto no Yoritomo*	Richard Wagner

(*Continued*)

Table 10.1 (*Continued*)   Top 50 Most Important People of All Times in Four Wikipedias (Most Important at the Top)

	English	Chinese	Japanese	German
16.	Dante Alighieri	*Emperor Daizong of Tang*	*Emperor Go-Shirakawa*	Georg Wilhelm Friedrich Hegel
17.	**Gautama Buddha**	*Emperor Xuanzong of Tang*	*Ashikaga Takauji*	*Marcus Tullius Cicero*
18.	*Tiberius*	*Zhengtong Emperor*	*Ashikaga Yoshimitsu*	Pius XII.
19.	**Cyril of Alexandria**	*Emperor Xizong of Tang*	*Emperor Go-Toba*	Ludwig van Beethoven
20.	**Bernard of Clairvaux**	*Qianlong Emperor*	*Emperor Saga*	Karl Marx
21.	**Moses**	*Dong Zhuo*	*Ashikaga Yoshimasa*	Friedrich Schiller
22.	*Tacitus*	*Yuwen Tai*	*Ashikaga Yoshinori*	*Joseph Goebbels*
23.	*Edward III of England*	*Qianlong Emperor*	*Emperor Toba*	*Abraham Lincoln*
24.	*Justinian I*	*Huan Xuan*	*Emperor Kanmu*	*Heinrich Himmler*
25.	*David*	*Emperor Huizong of Song*	*Emperor NinmyÅ*	Denis Diderot
26.	*Ashoka*	*Emperor Xuanzong of Tang*	*Emperor Daigo*	*Benito Mussolini*
27.	Origen	*Emperor Renzong of Song*	*Emperor Junna*	*Thomas Jefferson*
28.	*Septimius Severus*	*Temür Khan, Emperor Chengzong of Yuan*	*Emperor Seiwa*	Immanuel Kant
29.	Polybius	*Emperor Taizong of Song*	*Emperor Wu of Liu Song*	*Friedrich II. (HRR)*
30.	Confucius	*Fu Jian*	*Emperor Montoku*	Heinrich Heine

AQ4

(*Continued*)

Table 10.1 (*Continued*) Top 50 Most Important People of All Times in Four Wikipedias (Most Important at the Top)

	English	*Chinese*	*Japanese*	*German*
31.	*Alexander Severus*	*Sima Ying*	*Emperor Uda*	Bertolt Brecht
32.	**Patriarch Eutychius of Alexandria**	*Emperor Taizong of Song*	*Fujiwara no Michinaga*	*Andrew Jackson*
33.	*Tutankhamun*	*Wang Mang*	*Emperor Wen of Liu Song*	Gerhart Hauptmann
34.	*Akhenaten*	*Emperor Shunzong of Tang*	*Cao Cao*	Franziskus (Papst)
35.	*Ramesses II*	*Shi Jingtang*	*Emperor Xiaowu of Liu Song*	*Ludwig XIV.*
36.	**Pope Benjamin I of Alexandria**	Wang Yangming	*Emperor Xiaowen of Northern Wei*	Franz Liszt
37.	*Teti*	*Hong Xiuquan*	*Emperor Ming of Liu Song*	*Konrad Adenauer*
38.	*Amenemhat II*	*Napoleon III*	*Emperor Wu of Southern Qi*	Plutarch
39.	*Pepi II Neferkare*	*Emperor Xiaowu of Liu Song*	*Emperor Wu of Liang*	*George Washington*
40.	*Merneith*	*Sima Shi*	*Sun Quan*	Pablo Picasso
41.	Terence	*Möngke Khan*	*Augustus*	Friedrich Nietzsche
42.	*Cato the Elder*	*Chang Yuchun*	*Pompey*	Albert Einstein
43.	*Charles Martel*	*Emperor Ming of Han*	*Tiberius*	*Barack Obama*
44.	*Gilgamesh*	*Tuoba Huang*	*Mark Antony*	*Augustus*
45.	**Deborah**	*Yuwen Huaji*	*Alexander the Great*	Alexander von Humboldt

Table 10.1 (*Continued*)   Top 50 Most Important People of All Times in Four Wikipedias (Most Important at the Top)

	English	*Chinese*	*Japanese*	*German*
46.	*Lugalbanda*	*Mu Ying*	Plato	*Josef Stalin*
47.	*Kubaba*	*Emperor Wen of Liu Song*	Cicero	*Heinrich IV. (HRR)*
48.	*Fu Xi*	**Zhu Xi**	*Emperor Wu of Han*	*Benjamin Franklin*
49.	*Henry I of England*	**Confucius**	*Ptolemy I Soter*	*Angela Merkel*
50.	Petrarch	*Emperor Ming of Southern Qi*	*Octavia the Younger*	Friedrich Engels

Italics denote politicians, bold religious leaders, and normal font scientists and artists.
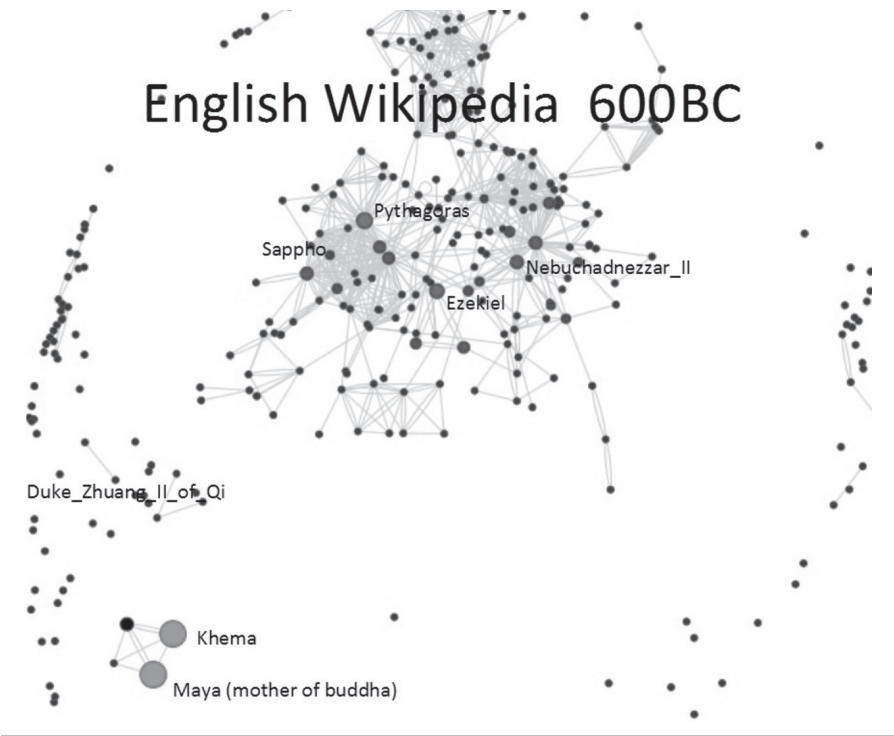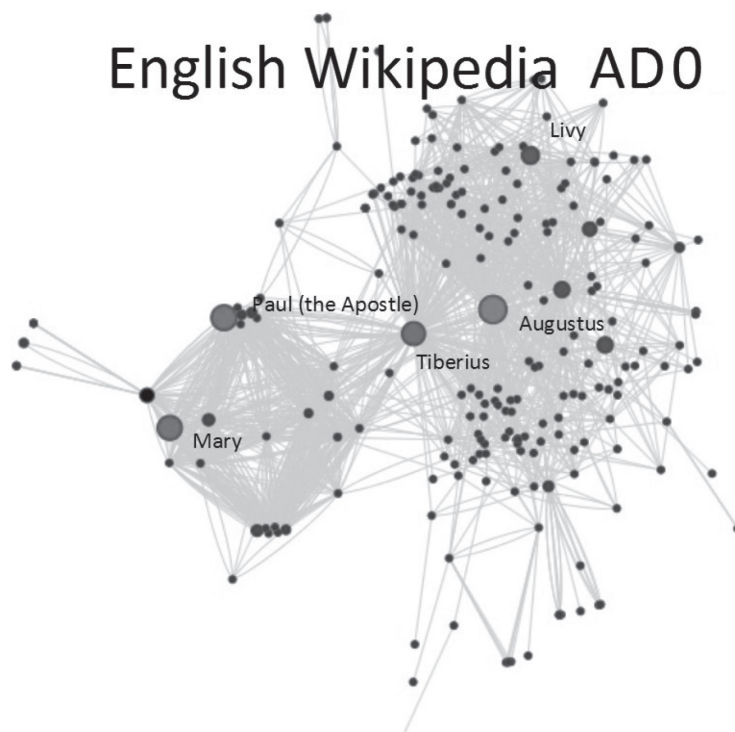
AQ5

Figure 10.2   Network of the most important people in the English *Wikipedia* in 600 BC.

AQ6

AQ7 **Figure 10.3   Network of the most important people in the English *Wikipedia* in AD 0.**

Around AD 0, Roman emperors, consuls, and writers are in the center, as well as biblical figures. Jesus has not been born yet, but his mother Mary has a prominent position in the biblical cluster.

Around AD 600, the Catholic Church with popes, monks, and saints stands out. Greek Orthodox church fathers around Babai the Great form a separate cluster. The Chinese emperors and generals form another cluster of almost similar size, representative for the subsequent blossoming of the Tang dynasty starting in 618, commonly seen as the high point in Chinese civilization.

10.3.2 *Chinese* Wikipedia *Analysis*

Eminent leaders in the Chinese *Wikipedia* show a markedly different picture. Other than in the English *Wikipedia*, the top 50 leaders of all times are mostly emperors and generals. While in the English *Wikipedia* only 26 out of 50 people fell into that category; in the Chinese *Wikipedia* there are only 4 people who are not political and military leaders, among them are 3 Confucian scholars including
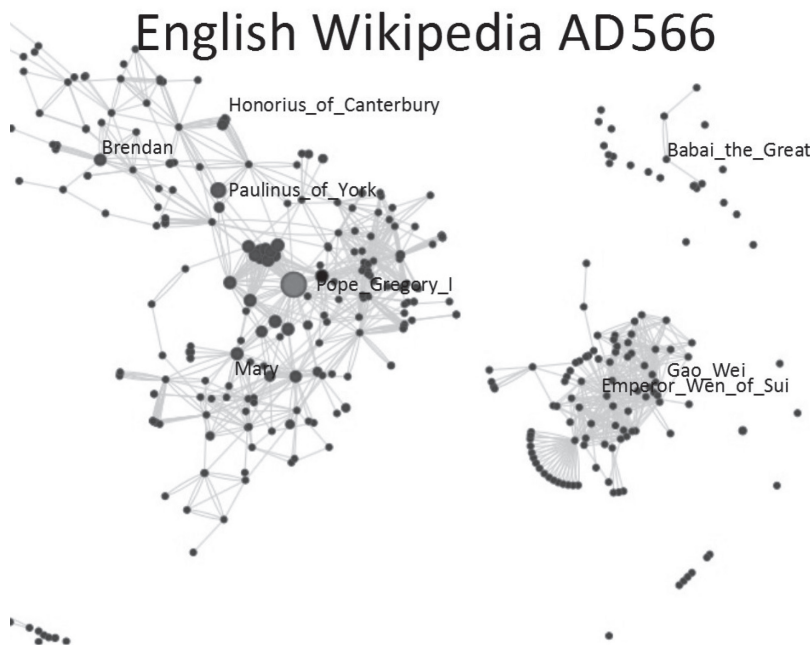
Figure 10.4   Network of the most important people in the English *Wikipedia* in AD 0.

Confucius himself, plus one of the most popular Chinese pop musicians (who is mostly unknown outside Chinese-speaking countries).

As Tables 10.1 and 10.2 illustrate in China and Japan, only famous warriors and politicians had a chance to make it into the top 10 and top 50—the East seems far less religious than the West in their view on the past—while the Western Wikipedias are more balanced with half of the top 10 and the top 50 of all times being religious leaders, artists, or scientists. Historians play a special role. Both Sidney Lee and

Table 10.2   Distribution of Different People Categories in Four Wikipedias among the Top 50 People of All Times

	English	*Chinese*	*Japanese*	*German*
Politicians	26	46	47	23
Religious leaders	11	1	0	5
Artists/scientists	13	3	3	22
Cultural in-group	10	48	31	31

Ikuhiko Hata, a nineteenth-century Japanese military biographer, owe their prominent position to their prolific biography writing, as they get many backlinks from the references on the pages of contemporary politicians they wrote about.

Note that due to the smaller size of the Chinese *Wikipedia*, the indegrees of the top 50 Chinese leaders are much lower than for the English *Wikipedia* leaders. This means that in the lower ranks of the top 50 a few more incoming links can have a large effect on the ranking of a person. Furthermore, Chinese leaders do not go back as far in history as the Egyptian pharaohs and the ancient kings of the Middle East, where a combination of year-ring dating and listing of dynasties allows to date ancient leaders back to 3000 BC. The oldest entries in both Wikipedias which are Chinese are from about the time 800 BC. Mythical earlier emperors like the Yellow Emperor who might have reigned around 3000 BC are not included in the dead people category, as historians are not certain if he ever really existed as a single real person.

The Chinese *Wikipedia* leaders have far fewer links to people outside China than leaders in the non-Asian Wikipedias we have looked at. One could therefore see them as more inward-looking, that is, China-focused. Napoleon III (not Napoleon I) and Tokugawa Ieyasu are the only non-Chinese/Japanese among the top 50 leaders. The English *Wikipedia* is far more diverse, including Confucius, Buddha, and Fu Xi besides military leaders from ancient Egypt and Sumeria together with the leaders of the Western world.

A big restriction of the Chinese *Wikipedia* is the censorship executed by the Chinese government, which might influence the scant inclusion of Western politicians, scientists, and religious leaders.

10.3.3 *Japanese* Wikipedia *Analysis*

Other than the Chinese *Wikipedia*, the Japanese *Wikipedia* is not subject to any censorship and can thus rightfully claim to represent the opinion and collective intelligence of the Japanese population. The top person article by indegree, just like the article about the historian Sidney Lee in the English *Wikipedia*, is about a military historian who wrote about all the politicians, mostly generals, referenced in the Japanese *Wikipedia*. Amazingly, the fourth most central article in the Japanese *Wikipedia* is about Adolf Hitler.

10.3.4 Cultural Chauvinism

There are striking differences in the number of out-group leaders, that is, leaders not part of the language sphere of a particular *Wikipedia*, included among the top 50. While the English *Wikipedia* includes 80% non-English leaders among the top 50, only two non-Chinese made it into the top 50 of the Chinese *Wikipedia*: Napoleon III and Tokugawa Ieyasu. The Japanese *Wikipedia* is slightly more balanced, with almost 40% non-Japanese leaders, half of them Chinese Emperors and people like Adolf Hitler, Plato, Cicero, and Augustus.

A part of that effect might also come from the different sizes and ethnic profiles of the editors of the different Wikipedias, as the English *Wikipedia* probably includes articles written about Chinese leaders written by Chinese and Japanese living abroad, while the articles in the Chinese and Japanese *Wikipedia* are most likely written by ethnic Chinese and Japanese.

10.4 Differences in Gender Equality

In a subsequent analysis, we studied the percentage of female important people in different cultures. Although, in theory, *Wikipedia* people pages are tagged by gender, many of them do not include this mapping. We therefore created a heuristic algorithm that collects keywords in different languages and analyzes the frequency of those words on each page. It counts, for example, "she," "he," and "herself," for the English *Wikipedia*. In the Portuguese *Wikipedia*, we used words such as "ela" and "ele." The same approach was used for the German and Spanish Wikipedias. To verify the accuracy of this method, we compared the results with a subset of people pages that include the gender tag. We found that this approach got the gender right in more than 90% of the people pages in those four languages.

10.4.1 Top 50 Gender Analysis

While many societies have made great progress toward gender equality over the last 100 years, we wanted to measure the pace of this progress. We tracked the evolution of the share of women among the people represented in the different Wikipedias. We found that the 50 most important women in the Wikipedias from 2000 BC to AD 2000 are almost exclusively in two categories, namely, politicians (mostly queens) and artists (e.g., writers, actresses, and singers) (Table 10.3).

Table 10.4 lists all the names in the four Wikipedias. The first group consists of politicians and queens. In this group, the most prominent member is Queen Elizabeth II, showing that people's fascination with the British Royals is unbroken.

Table 10.3　Split between Politicians and Artists among the Top 50 Most Influential Women of All Times in Different Wikipedias

	German	*English*	*Spanish*	*Portuguese*
Politicians	10	14	12	11
Artists	38	36	35	34
Others	2	0	3	5
% Politicians	20.83	28.00	25.53	24.44

Table 10.4   Top Most Important Female Leaders of All Times in Four Wikipedias

	German	*English*	*Spanish*	*Portuguese*
1.	**Elizabeth II.**	**Elizabeth II**	**Michelle Bachelet**	**Maria II de Portugal**
2.	**Angela Merkel**	**Queen Victoria**	Marilyn Monroe	Madonna
3.	**Maria Theresia**	Joan Baez	**Isabel I de Inglaterra**	**Isabel II do Reino Unido**
4.	Katie Fforde	Simone Signoret	**María Teresa I de Austria**	**Dilma Rousseff**
5.	Marlene Dietrich	**Benazir Bhutto**	Mary Pickford	**Maria I de Portugal**
6.	Rosamunde Pilcher	Eudora Welty	Katharine Hepburn	**Vitória do Reino Unido**
7.	**Margaret Thatcher**	Gertrude B. Elion	Elizabeth Taylor	Judy Garland
8.	Hannah Arendt	Betty White	**Hillary Clinton**	Cher
9.	Greta Garbo	Annette Bening	**Margaret Thatcher**	Fernanda Montenegro
10.	Elfriede Jelinek	Eunice Kennedy Shriver	Greta Garbo	Barbra Streisand
11.	Sigrid Roth	**Ruth Bader Ginsburg**	**María Estela Martínez de Perón**	Meryl Streep
12.	Maria (Mutter Jesu)	Susan Hayward	**Vivian Malone Jones**	Marilyn Monroe
13.	Thea Leitner	Mary J. Blige	**Eva Perón**	Gal Costa
14.	Ella Fitzgerald	**Margaret Beckett**	Madonna	Sophia Loren
15.	**Maria Stuart**	**Barbara Jordan**	Gabriela Mistral	Britney Spears
16.	Marilyn Monroe	Zora Neale Hurston	Barbra Streisand	**Marta Suplicy**

(*Continued*)

Table 10.4 (*Continued*)   Top Most Important Female Leaders of All Times in Four Wikipedias

	German	*English*	*Spanish*	*Portuguese*
17.	Jane Fonda	Rosalynn Carter	Britney Spears	Christina Aguilera
18.	Bette Davis	Sarah Vaughan	Jane Fonda	Roberta Flack
19.	Elizabeth Taylor	Heidi Klum	Marlene Dietrich	Maggie Smith
20.	Billie Holiday	**Helen Clark**	**Juana de Arco**	Rihanna
21.	Nelly Sachs	Donna de Varona	Whitney Houston	Helen Mirren
22.	Liza Minnelli	Julie Walters	Virginia Woolf	Elis Regina
23.	Diana Ross	Patti Smith	Christina Aguilera	Kate Winslet
24.	Katharine Hepburn	**Harriet Tubman**	Mariah Carey	**Maria Teresa da Áustria**
25.	**Beatrix (Niederlande)**	Michelle Pfeiffer	George Sand	Angelina Jolie
26.	Joan Crawford	Allison Janney	Meryl Streep	Lauren Bacall
27.	Katarina Witt	Carole King	Aretha Franklin	Whitney Houston
28.	Shirley MacLaine	Alice Munro	Ella Fitzgerald	Judi Dench
29.	**Elizabeth I.**	Faye Dunaway	Shakira	Marília Pêra
30.	Romy Schneider	Willa Cather	**Josefina de Beauharnais**	Janet Jackson
31.	Claudette Colbert	**Margaret Sanger**	Ingrid Bergman	**Isabel I de Castela**
32.	Toni Morrison	Kirstie Alley	Judy Garland	Nigar Jamal
33.	Ingrid Bergman	**Joanna of Castile**	Liza Minnelli	Jennifer Lawrence
34.	Barbra Streisand	Sheryl Crow	Simone de Beauvoir	Cate Blanchett

(*Continued*)

Table 10.4 (*Continued*)   Top Most Important Female Leaders of All Times in Four Wikipedias

	German	English	Spanish	Portuguese
35.	Whoopi Goldberg	Sally Ride	**María Josefa de Austria**	Mary Pickford
36.	Sigrid Undset	Jennifer Hudson	Ava Gardner	**Maria Antonieta**
37.	George Sand	Serena Williams	Billie Holiday	Penélope Cruz
38.	Nadine Gordimer	**Donna Shalala**	Samuel Johnson	Sissy Spacek
39.	Gabriela Mistral	**Shirley Chisholm**	Lady Gaga	Whoopi Goldberg
40.	Doris Lessing	Olivia Newton-John	**Angela Merkel**	**Hillary Clinton**
41.	Pearl S. Buck	Kristin Scott Thomas	Nicole Kidman	Jennifer Lopez
42.	**Marie Antoinette**	**Margaret, Maid of Norway**	Maria Callas	Ingrid Bergman
43.	Helen Hayes	Candice Bergen	Vanessa Redgrave	Barbara Stanwyck
44.	Meryl Streep	Wilma Rudolph	Joan Crawford	**Isabel I de Inglaterra**
45.	**Hillary Clinton**	**Martha Griffiths**	Mercedes Sosa	Jodie Foster
46.	Judy Garland	Antonina Houbraken	Janet Jackson	Shirley MacLaine
47.	Alice Munro	Claire Danes	Rita Hayworth	**Francisca de Bragança**
48.	Britney Spears	Marion Cotillard	Julia Roberts	Sandra Bullock
49.	Goldie Hawn	Sappho	Tina Turner	**Januária Maria de Bragança**
50.	**Isabella I. (Kastilien)**	Alice Hamilton	Rihanna	Lady Gaga

Politicians in bold.

In our metrics, she consistently leads the female leadership ranking. Moreover, in the German and Portuguese Wikipedias, we find German chancellor Angela Merkel and Brazilian president Dilma Rousseff among the top five positions in their corresponding countries, illustrating that in the last decades women have been accumulating more influence in global politics. Another interesting result is the position of American politician Hillary Clinton. She is present among the top 50 most influential women in all Wikipedias but the English. We can speculate that maybe she has more influence overseas than inside her own country due to her position as a foreign minister. We also have to note that this analysis was completed 2 years before Hillary announced her aspirations for the U.S. 2016 presidential race.

In the art categories, we find a more diversified distribution between writers, singers, and actresses. It also seems that the United States exerts a dominating influence in this area in the rest of the world. Many actresses who worked or work in the U.S. show business, such as Marilyn Monroe, Judy Garland, or Barbra Streisand, have a high PageRank in the different Wikipedias.

10.4.2 Longitudinal Gender Analysis

By analyzing the percentage of women over the last century, we can see how the fraction of women among the world's leaders is increasing over time (Figure 10.2). Our data show the Portuguese *Wikipedia* having the highest percentage of women among the Wikipedias we analyzed, followed by the Spanish and English.

One fact that may have influenced this result is the number of total people pages in the Portuguese *Wikipedia* (93,000), which is significantly lower than that in the English *Wikipedia*, which has 10 times more people pages, translating to a lower sample size. The graph shows a linear increase in the percentage of women until 1995 (Figure 10.5), when the percentage really explodes to parity between women and men in the English, Spanish, and Portuguese Wikipedias. This explosion, however, might also be an artifact of the smaller number of people pages for the last 20 years, as there are very few people, other than Justin Bieber or Malala Yousafzai (the youngest ever Nobel Peace Prize winner), that fulfill the notability criteria of *Wikipedia* at such a young age. Nevertheless, it seems that in this smaller sample of young high achievers we attain gender equality.

10.5 Comparing Different Cultures through Wikinews

We complement our historical analysis of *Wikipedia* with an analysis of the treatment of late breaking news by Wikipedians in Wikinews, a hand-curated news page of whatever Wikipedians consider most newsworthy on any given day. In particular, we use Wikinews to analyze sentiment and emotion in different cultures. Wikinews is part of *Wikipedia*. It is a free and open news website where users create the content based on digesting news from sources such as Reuters, Bloomberg, and CNN.
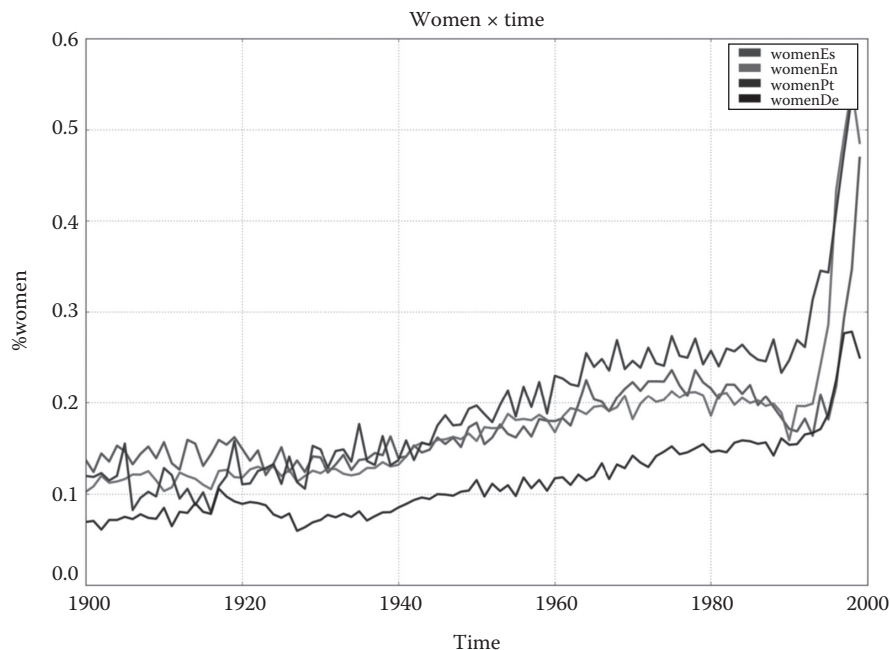
266 ■ *Social Network Analysis*

Figure 10.5 Percentage of women among Wikihistory important people from 1900 to 2000. Each line shows the ratio of women to men for the particular language.

The main focus in this project is to analyze what kinds of news are more relevant in each culture and the sentiment related to those news. For example, we want to investigate if some cultures focus more on bad or good news or whether some topics are considered more newsworthy in some cultures than in others.

10.5.1 Data Collection

We collected the news using the Wikinews pages in English, German, Spanish, and Portuguese. In those pages, we can find the most relevant events of each year through the eyes of the Wikipedians. In particular, whenever they consider a particular topic newsworthy, they will add a short sentence on the Wikinews page as an anchor and create a new *Wikipedia* page for the entire event. For each language, we collected all the links to the pages in *Wikipedia* and connected those links with each other. This approach creates four distinct networks, one for each language (see Figures 10.6 through 10.9). As we can see, some Wikipedias have a more complex and connected news network. This is not surprising because, for example, the English *Wikipedia* has more content and is more updated than the Portuguese. The total number of articles is also quite different between the different Wikipedias.
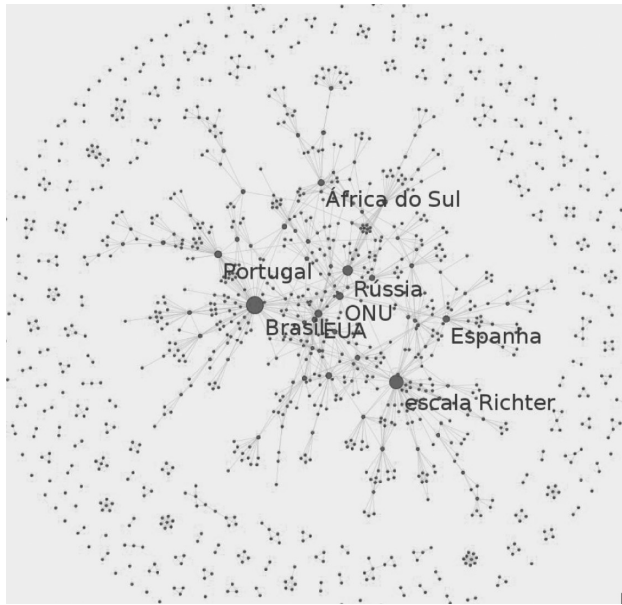
K26496_C010.indd   266   10/8/2016   10:25:31 AM

Figure 10.6   Portuguese Wikinews network.
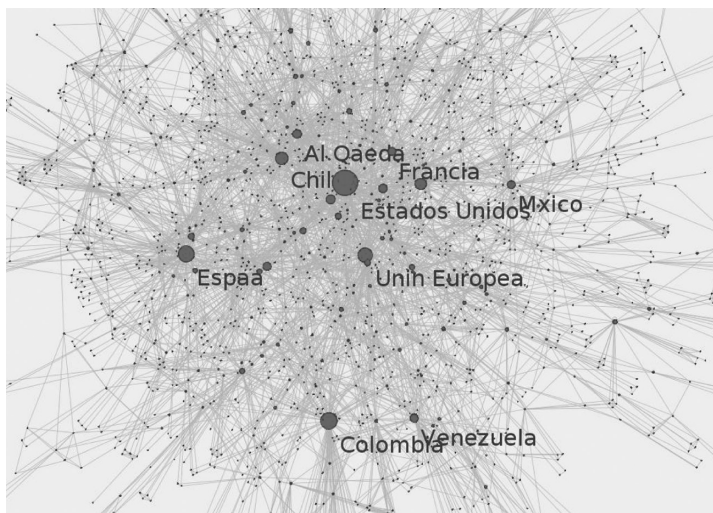

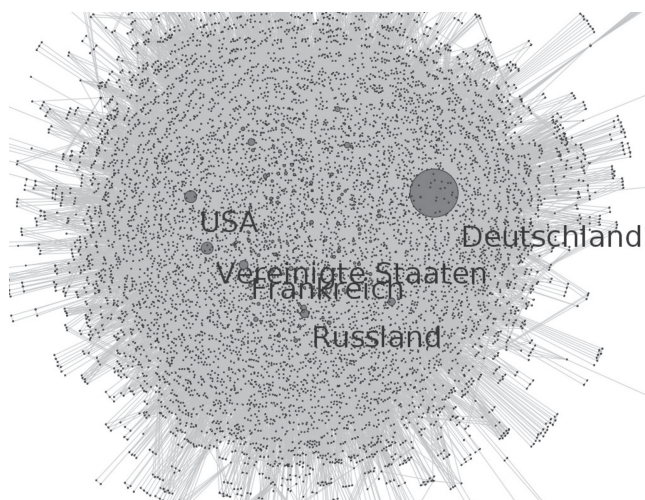**Figure 10.7   Spanish Wikinews network.**

Figure 10.8   German Wikinews network.

As we can see in the network pictures in Figures 10.6 through 10.9, the most central nodes are usually countries. News are linked to location pages related to where an event happened as opposed to one main subject like war, company names, or people. Unsurprisingly, we observe a heavy location bias, in which intracultural news get a much more detailed treatment. This means that *Wikipedia* editors of one
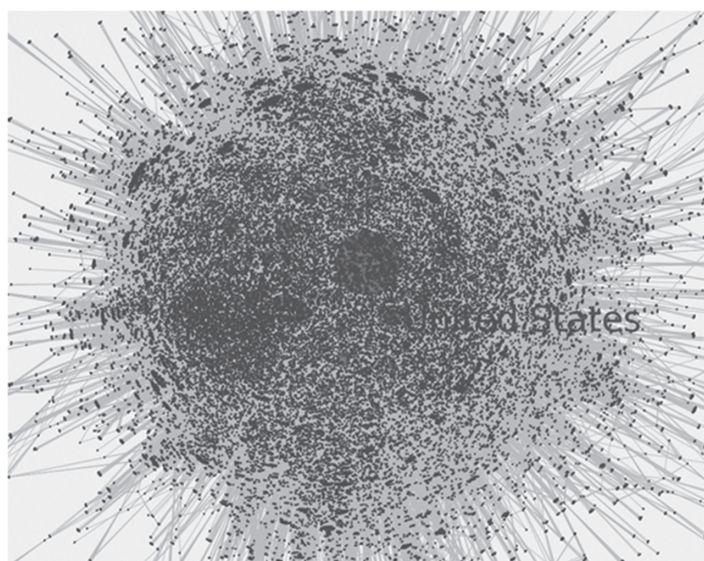

Figure 10.9   English Wikinews network.

language upload news that happened in the country with the same language more frequently than outside news (Figure 10.10). For the Spanish *Wikipedia*, due to the many countries sharing the same language, we observe a more equal distribution of topics by betweenness centrality. As many different countries are appearing on the Spanish Wikinews with similar centrality, this illustrates that people from many of those countries seem to be contributing. The "popularity" of one country in other cultures and languages can show how one country has the attention of others. For
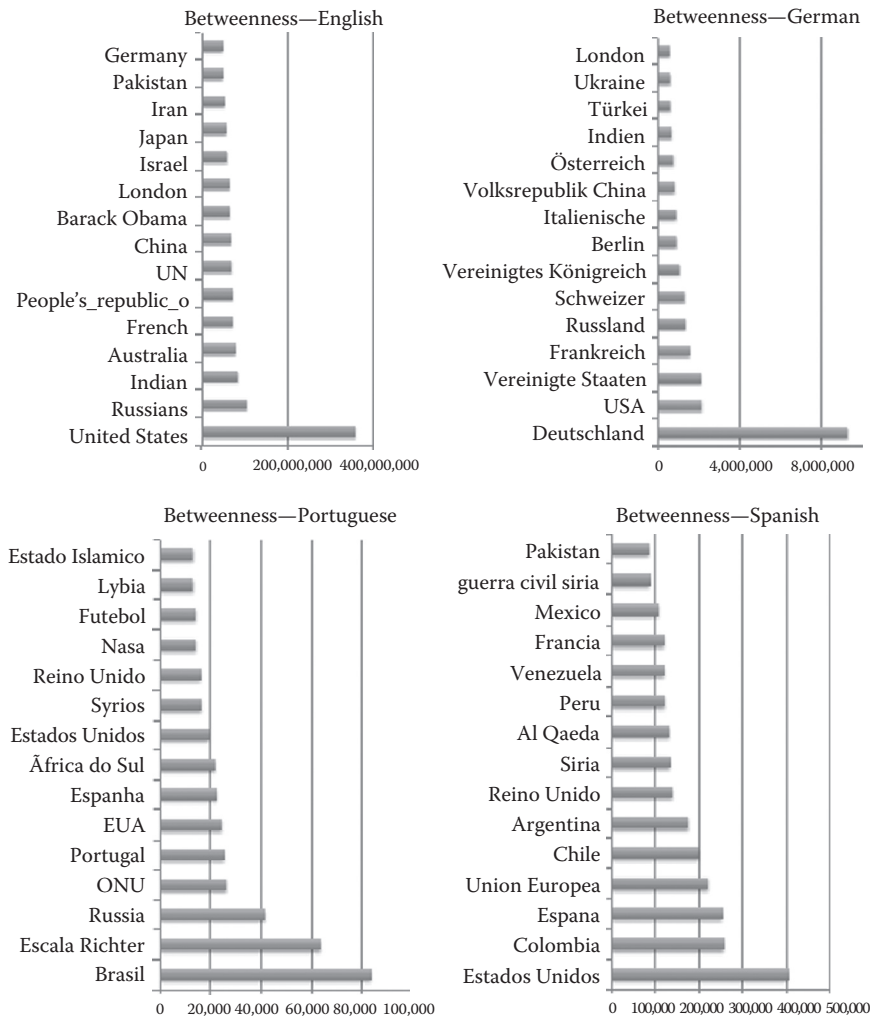

Figure 10.10 The most important topics by betweenness in the different-language Wikinews.
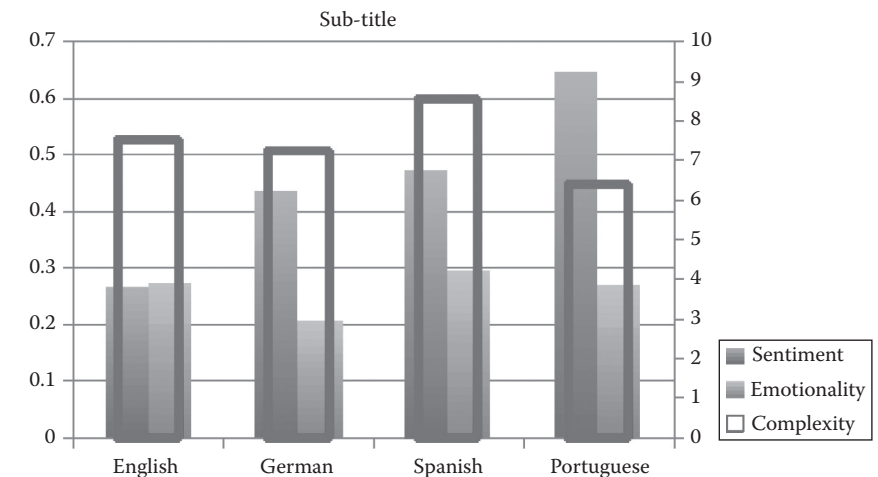
Figure 10.11   Sentiment, emotionality, and complexity of the pages linked to Wikinews in the different languages.

example, we can see how the United States is a big player on all Wikinews, illustrating U.S. dominance in global politics across all languages and cultures.

We not only looked at the key topics in the different-language versions but also compared the sentiment in the news, using automatic sentiment analysis (Brönnimann 2014). Using a machine learning approach, the computer measured the density of positive words like "great," "wonderful," and negative words like "horrible" or "bad." Emotionality is defined as the variation in positive and negative sentiment. Complexity measures the use of rare words compared to a simple standardized vocabulary. We used the sentiment, emotionality, and complexity functions of the tool Condor (Brönnimann 2014) to calculate these metrics for all four Wikinews data sets. The overall negativity of sentiment stands out, below the neutrality threshold of 0.5. While the text has been written by Wikipedians, they base their selection on general sources such as Bloomberg, Reuters, and CNN. This confirms the popular belief that people are much more interested in negative than in positive news. However, we also notice that some Wikinews data sets are more positive than others (Figure 10.11). For instance, the English language *Wikipedia* is the most negative while Portuguese and Spanish are the most positive. It also seems that Spanish Wikinews uses more complex language than the others, which might be due to the many countries that share the same language. Spanish speakers in Spain might be saying similar things in other words than speakers in Mexico, Chile, or Argentina. The same could be claimed for the English Wikinews, due to the U.S., UK, Australia, Canada, and New Zealand editing. However, in this case, the population of the United States is considerably larger than that of the others combined. German and Portuguese Wikinews basically just cover one

country, where words might be used more consistently, leading to lower complexity measures as calculated by our software. On the other hand, the Germans seem to be less emotional than the Spanish, Portuguese, or English, confirming a national stereotype of dry and matter-of-fact behavior.

10.6 Related Work

Charles Murray's list of "Human Accomplishments" (2003) shows the most innovative people in history. The list is led by Isaac Newton and Galileo Galilei. It is based on a compilation of history textbooks, where "accomplished" people have to be listed multiple times in different languages and sources. Murray's list has been combined in the Pantheon with *Wikipedia* (http://pantheon.media.mit.edu/methods). Pantheon is based on a list created from 11,340 *Wikipedia* people pages, where a person has to be listed in at least 25 different-language Wikipedias. The list has been hand-curated to remove obviously nonsignificant characters. Our analysis approach is purely based on data and not hand-curated, making a claim to higher objectivity. Second, to be included into the Pantheon, a person also needs to be listed in the Charles Murray's book whose inclusion criteria are to be listed in multiple school textbooks. Consistent with traditional western education, Pantheon seems to be biased toward Greek philosophers, as the top five entries are Aristotle, Plato, Jesus Christ, Socrates, and Alexander the Great.

10.7 Conclusions

The Internet enables researchers to more easily compile rankings of the most important people of all times (Murray 2003, Hidalgo 2014). Our work is unique in that we extract language-specific rankings that allow us to compare the worldview for dozens of different cultures. Probing the historical perspective of many different-language-specific Wikipedias enables an x-ray view into cultural and national characteristics, giving some confirmation to national stereotypes of Germans as "people of the book" or Japanese placing heavy emphasis on their warlike past.

We think that our preliminary research has just scratched the surface of digital humanities, illustrating the potential of employing the different-language versions of *Wikipedia* as a socioscope into the different facets of human nature and culture.

References   AQ9

Becker, M. (2012) Die Aktualität von Online-Enzyklopädien—Eine empirische Analyse am Beispiel Wikipedia. Diploma thesis, University of Cologne, Köln, Germany.
Blodget, H. (2009) Who the hell writes Wikipedia, anyway? *Business Insider*, January 3, 2009.   AQ10
    http://www.businessinsider.com/2009/1/who-the-hell-writes-wikipedia-anyway?

Brint, S. (1984) "New-Class" and cumulative trend explanations of the liberal political attitudes of professionals. *American Journal of Sociology*, 90, 30–71.

Brönnimann, L. (2014) Analyse der Verbreitung von Innovationen in sozialen Netzwerken. MSc thesis, University of Applied Sciences Northwestern Switzerland, Basel, Switzerland. http://swarmcreativity.blogspot.com/2014/03/how-to-measure-influ-ence-in-social.html.

Frick, K., Guertler, D., Gloor, P. (2013) Coolhunting for the world's thought leaders. *Proceedings Fourth International Conference on Collaborative Innovation Networks* (*COINs 2013*), Santiago de Chile, August 11–13, 2013.

Futterer, T., Gloor, P., Malhotra, T., Mfula, H., Packmohr, K., Schultheiss, S. (2013) WikiPulse—A news-portal based on Wikipedia. *Proceedings Fourth International Conference on Collaborative Innovation Networks* (*COINs 2013*), Santiago de Chile, August 11–13, 2013.

Gloor, P., De Boer, P., Lo, W., Wagner, S., Nemoto, K., Fuehres, H. (2015) Cultural anthropology through the lens of Wikipedia—A comparison of historical leadership networks in the English, Chinese, and Japanese Wikipedia. *Proceedings of the Fifth International Conference on Collaborative Innovation Networks* (*COINs15*), Tokyo, Japan, March 12–14, 2015.

Gloor, P., Krauss, J., Nann, S., Fischbach, K., Schoder, D. (2009) Web Science 2.0: Identifying trends through semantic social network analysis. *IEEE Conference on Social Computing* (*SocialCom-09*), Vancouver, British Columbia, Canada, August 29–31, 2009.

Greenstein, S., Zhu, F. (2012) Collective intelligence and neutral point of view: The case of Wikipedia. NBER Working Paper No. 18167.

Hidalgo, C. (2014) Pantheon: Mapping historical cultural production. http://pantheon.media.mit.edu.

Iba, T., Nemoto, K., Peters, B., Gloor, P. A. (2010) Analyzing the creative editing behavior of Wikipedia editors: Through dynamic social network analysis. *Procedia-Social and Behavioral Sciences*, 2(4), 6441–6456.

Kleeb, R., Gloor, P., Nemoto, K., Henninger, M. (2012) Wikimaps: Dynamic maps of knowledge. *International Journal of Organisational Design and Engineering*, 2(2), 204–224.

Lee, J. A., Soutar, G. N., Daly, T. M., Louviere, J. J. (2011) Schwartz values clusters in the United States and China. *Journal of Cross-Cultural Psychology*, 42(2), 234–252.

Murray, C. (2003). *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950*. HarperCollins.

Page, L., Brin, S., Motwani, R., Winograd, T. (1999) The PageRank citation ranking: Bringing order to the Web.

Wasserman, S., Faust, K. (1994) *Social Network Analysis: Methods and Applications*, Vol. 8. Cambridge University Press. http://www.infoplease.com/spot/womenstimeline1.html.

AQ11

AQ12

AQ13

AQ14

AUTHOR QUERIES

[AQ1]   Please check the sentence starting "We speculate that the other Wikipedias…" for clarity.

[AQ2]   Please check if the inserted citations for Figures 10.3 through 10.5 and Table 10.3 are okay.

[AQ3]   Please suggest if "around" in phrases "around Buddha" and "around Babai the Great" can be changed to "in the era of".

[AQ4]   Please check if the text "*Emperor NinmyÅ*" is okay.

[AQ5]   Please check whether the text "English Wikipedia in 600 BC "can be retained in Figure 10.2.

[AQ6]   Please check whether the text "English Wikipedia in AD 0 "can be retained in Figure 10.3.

[AQ7]   Captions of Figures 10.3 and 10.4 are exactly the same. Please check.

[AQ8]   Please check whether the text "English Wikipedia in AD 566 "can be retained in Figure 10.4.

[AQ9]   Please cite Gloor et al. (2015), Lee et al. (2011) in text.

[AQ10] Please provide page range for Blodget (2009).

[AQ11] Please provide publisher details for Greenstein and Zhu (2012).

[AQ12] Please provide date of access for Hidalgo (2014).

[AQ13] Please provide publisher location for Murray (2003), Wasserman and Faust (1994).

[AQ14] Please provide complete details for Page et al. (1999).