HYPERLINK NETWORK SYSTEM AND IMAGE OF GLOBAL CITIES: WEBPAGES AND THEIR CONTENTS

by

Jae Soen Son

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography and Urban Regional Analysis

Charlotte

2014

Approved by:

_____
Dr. Jean-Claude Thill

_____
Dr. Qingfang Wang

_____
Dr. Eric Delmelle

_____
Dr. Min Jiang

ABSTRACT

JAE SOEN SON. Hyperlink network system and image of global cities: webpages and their contents. (Under the direction of Dr. JEAN-CLAUDE THILL)

A distinctive trend of globalization research is a conceptual expansion that mirrors the penetration of globalization in various aspects of life. The World Wide Web has become the ultimate platform to create and disseminate information in this era of globalization. Although the importance of web-based information is widely acknowledged, the use of this information in global city research is not significant yet. Therefore, the purpose of this research is to extend the concept of globalization to the efficiency of information networks and the thematic dimensionality of the conveyed images from webpages.

To this end, 264 global and globalizing cities are selected. The city hyperlink networks are constructed from the web crawling results of each city, and hyperlink network analysis measures the effectiveness of these hyperlink networks. The textual contents are also extracted from the crawled webpages, and the thematic dimensionality of the textual contents is measured by quantified content analysis and multidimensional scaling.

The efficiency of the hyperlink network in information flow is confirmed to be a new consideration that shapes the globality of cities. The cities with high efficiency of connections have faster and easier access, which means better structure for city image formation. Specifically, social networking websites are the center of this information flow. This means that social interactions on the Web play a crucial role to form the images of cities. Apart from the positivity and the negativity of the city image, the

dimensionality of cities on the thematic space denotes how they are expressed, discussed, and shared on the Web. The image status based on dimensions of globalization is an important starting point to city branding. It is concluded that a research framework handling information networks and images simultaneously deepens the understanding of how the structure and the contents on the Web affect the formation and maintenance of global city networks. Overall, this research demonstrates the usefulness of information networks and images of cities on the Web to overcome data inconsistency and scarcity in global city research.

DEDICATION

This dissertation is dedicated to my brilliant and supportive wife, Kyeong Jin Lee, for all of her constant devotion and encouragement.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Knight Distinguished Professor Jean-Claude Thill for the continuous support of my PhD study and research, for his profundity, enthusiasm, motivation, mentorship, and solicitude. His endless guidance helped me in all of the stages of research and writing this dissertation. I could not have imagined having a better advisor and mentor for my PhD study.

Besides my advisor, I would like to thank the rest of my dissertation committee for their encouragement and insightful comments. Prof. Qingfang Wang stimulated me to draw a big picture of my research and to have confidence in my research. Prof. Eric Delmelle advised me to elaborate important methodological parts of my research. Prof. Min Jiang suggested invaluable comments for the publication of this research in the near future.

During my PhD program I received financial support from the UNC Charlotte graduate school, the department of geography, and the graduate and professional student government (GPSG). Graduate Assistant Support Plan (GASP) of the graduate school and my department provided tuition waivers and assistantships. My department and GPSG also provided me travel funding for conferences. With these financial supports I could better concentrate on research.

My sincere thanks also go to Dr. Elizabeth Delmelle for her support as friend, officemate, and instructor of the courses for which I was a TA. I thank my former and present colleagues in the Urban and Regional System Analysis Laboratory (URSAL): Dr. Diep Dao, Zhaoya Gong, Mona Kashiha, Pooya Najaf, Ran Tao, Kailas

Venkitasubramanian, Daniel Yonto, and Yuhong Zhou, for the stimulating discussions, for the technical solutions for this research (especially Zhaoya and Mona), and for sharing the PhD life for the past six years.

In particular, I am grateful to Emeritus Professor Young-Woo Nam in Korea University for guiding me through the master's program in human geography and encouraging me to keep studying at UNCC for my PhD.

Last but not the least, I would like to thank my parents Gap-Soo Son and Seong-Sook Hong, my parents-in-law Young-Hoon Lee and Hyun-Ja Jang, for their continuous spiritual and financial support to me and my family. Also, I thank my daughters Serene and Arin for their endless love for me. They motivated me to complete PhD program.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xiii

LIST OF ABBREVIATIONS xv

CHAPTER I: INTRODUCTION 1

1.1. Statement of Research 3

1.2. Structure of the Dissertation 5

CHAPTER II: LITERATURE REVIEW 7

2.1. Globalization and Global Cities 7

2.1.1. Defining Globalization 7

2.1.2. Global City Where Globalization Occurs 13

2.2. Development of Measurements (Indices) for the Global City 22

2.2.1. Intrinsic Indices 25

2.2.2. Relational Indices 26

2.2.3. Qualitative Indices 28

CHAPTER III: IMAGE ON THE WEB FOR GLOBAL CITY STRATEGY 32

3.1. Image and Cities 33

3.2. The Web (the Internet) 35

3.3. Branding, Image, and Global City 37

3.4. Conceptualization of Image Extraction from Textual Contents 42

CHAPTER IV: RESEARCH QUESTIONS AND CONCEPTUAL FRAMEWORK 45

4.1. General Characteristics of Hyperlink Networks 47

4.2. Characteristics and Classification of Nodes 49

4.3. Characteristics of the Quantified Text of Webpages 51

4.4. Synthesis of the Hyperlink Network and the Quantified Contents 54

CHAPTER V: RESEARCH DESIGN 56

5.1. Cities under Study 57

5.2. Data 74

5.2.1. Linguistic Characteristics 74

5.2.2. Search Engines and Web Crawlers 80

5.2.3. Data Extraction and Pre-processing 83

5.3. Methods for Hyperlink Network and Contents of Webpages 94

5.3.1. Graph Theory, Social Network, and Hyperlink Network Analysis 94

5.3.2. Quantitative Content Analysis 97

5.4. Research Objective 4.1 Methodology 100

5.4.1. Network Connectivity 100

5.4.2. Cluster Analysis 105

5.5. Research Objective 4.2 Methodology 107

5.5.1. Network Centrality 107

5.6. Research Objective 4.3 Methodology 111

5.6.1. Frequency Analysis 112

5.6.2. Text Mining Analysis 114

5.7. Research Objective 4.4 Methodology 117

CHAPTER VI: GENERAL CHARACTERISTICS OF HYPERLINK NETWORKS 119

6.1. Hyperlink Network Measurements and Its Discrepancy 119

6.2. Comparison to Established Ranking of City Globality 121

6.3. Composite Global City Hyperlink Index 125

CHAPTER VII: CHARACTERISTICS AND CLASSIFICATION OF NODES 132

7.1. Overview of High Centrality Websites 132

7.2. Centrality of Nodes 135

7.3. Premium Nodes 138

CHAPTER VIII: CHARACTERISTICS OF QUANTIFIED TEXT OF WEBPAGES 142

8.1. Selection of Words for Code Scheme Design 143

8.2. Distribution of Frequencies 145

8.3. Similarity of Global Cities 147

CHAPTER IX: SYNTHESIS OF HYPERLINK NETWORKS AND THE QUANTIFIED CONTENTS 162

9.1. Correlation between Similarities of HNA and QCA 163

9.2. Distribution of Cities Considered Structure and Content 166

9.3. Contribution of the Composite Approach 171

CHAPTER X: CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH 172

REFERENCES 176

APPENDIX A: MYSQL SCRIPT FOR DMOZ DB AND SEED URL 188

APPENDIX B: MYSQL SCRIPT FOR RESEARCH DATA TABLE 190

APPENDIX C: LIST OF THE TOP CENTRALITY NODES 192

APPENDIX D: RANKINGS BASED ON PROPORTION OF EACH CATEGORY TO TOTAL WORDS 199

APPENDIX E: WORD FREQUENCY AND PROPORTION FOR QCA 206

APPENDIX F: CORRELATION COEFFICENT FOR EACH CITY 213

LIST OF TABLES

TABLE 2.1: World views of globalization 9

TABLE 2.2: Thematic classifications of global city researches from 1981 to 1998 14

TABLE 2.3: Conceptualizing inter-city linkages: a typology 17

TABLE 2.4: Examples of indices 23

TABLE 3.1: Five major components of webpages 36

TABLE 3.2: Six components of CBI 41

TABLE 3.3: Categories and factors of European City Brand Barometer 42

TABLE 3.4: Comparison between corporate branding and city branding 44

TABLE 5.1: Study cities from past global city research 58

TABLE 5.2: Distribution of population by continent 72

TABLE 5.3: Top ten languages used in the Web 76

TABLE 5.4: Number of final seed URLs 86

TABLE 5.5: Settings for crawling 93

TABLE 5.6: Descriptive analysis of networks 97

TABLE 5.7: Questions for refining research design of content analysis 100

TABLE 6.1: Correlation coefficient matrix among network measurements 121

TABLE 6.2: Correlation between 2012 GCI and hyperlink network measurements 122

TABLE 6.3: Descriptive statistics of selected network measurements 123

TABLE 6.4: City ranking based on new hyperlink index (GHI) 128

TABLE 6.5: Comparison between GHI and 2012 GCI 131

TABLE 7.1: Top websites by in-degree centrality and in-closeness centrality 136

TABLE 7.2: Top website by out-degree centrality, out-closeness centrality, and node betweenness centrality 137

TABLE 7.3: Descriptive statistics for top 30 URLs based on comprehensiveness 140

TABLE 8.1: Keywords for globalization and at its three dimensions 144

TABLE 8.2: Relative rankings based on the proportion of categorical words to total words 147

TABLE 9.1: Distribution of correlation coefficients between HNA and QCA similarities between cities 165

LIST OF FIGURES

FIGURE 2.1: Classifications of indices 24

FIGURE 4.1: Research components 47

FIGURE 4.2: Conceptual examples of the hyperlink networks 48

FIGURE 4.3: Conceptualization of image extraction from textual contents on the Web 52

FIGURE 5.1: Research objectives and methodologies 56

FIGURE 5.2: Trend of world regional composition of global city research 71

FIGURE 5.3: Distribution of study cities based on UN country grouping 73

FIGURE 5.4: Content languages for websites 77

FIGURE 5.5: Quarterly trends in the usage of content languages for websites 78

FIGURE 5.6: Concept of relationship between languages and webpages for research data 80

FIGURE 5.7: Simple architecture of the Web crawler 82

FIGURE 5.8: Breadth-first Search and Depth-first Search 83

FIGURE 5.9: Process of extracting pure text from the collected webpages 94

FIGURE 5.10: Relative position of hyperlink network 96

FIGURE 5.11: General model of content analysis 99

FIGURE 5.12: Conceptual diagram of methodology for research question 4.4 118

FIGURE 6.1: Classification of cities based on GHI 129

FIGURE 7.1: Distributions of website top-level domains (in-degree over 100) 134

FIGURE 7.2: Distribution of website types (in-degree over 100) 135

FIGURE 7.3: Average and maximum PageRank score of city networks, and the percentage of comprehensiveness 141

FIGURE 8.1: Procedure of word frequency extraction 143

FIGURE 8.2: Multidimensional scaling map for similarity of cities 149

FIGURE 8.3: Distributions of the proportion of keywords of each dimension 150

FIGURE 8.4: Multidimensional scaling map for 3 content dimensions and GCI 151

FIGURE 8.5: Distribution of African cities on the MDS map of globalization dimensions 157

FIGURE 8.6: Distribution of Asian cities on the MDS map of globalization dimensions 157

FIGURE 8.7: Distribution of Middle Eastern cities on the MDS map of globalization dimensions 158

FIGURE 8.8: Distribution of Oceania cities on the MDS map of globalization dimensions 158

FIGURE 8.9: Distribution of European cities on the MDS map of globalization dimensions 159

FIGURE 8.10: Distribution of North American cities on the MDS map of globalization dimensions 159

FIGURE 8.11: Distribution of Central American cities on the MDS map of globalization dimensions 160

FIGURE 8.12: Distribution of South American cities on the MDS map of globalization dimensions 160

FIGURE 8.13: Distribution of US cities on the MDS map of globalization dimensions 161

FIGURE 8.14: Distribution of Chinese cities on the MDS map of globalization dimensions 161

FIGURE 9.1: Comparison of MDS maps based on HNA and QCA similarities of cities 166

FIGURE 9.2: MDS map estimated on HNA and QCA attributes 169

FIGURE 9.3: Distribution of cities on the MDS property maps 170

LIST OF ABBREVIATIONS

APL	average path length
BFS	breadth-first search
CA	correspondence analysis
CBI	Anholt-Gfk Roper city brands index
CEO	chief executive officer
DCI	degree centralization index
DMOZ	directory Mozilla
FDI	foreign direct investment
FTA	free trade agreement
FTP	file transfer protocol
GATT	general agreement on tariffs and trade
GCI	global cities index
GDP	gross domestic product
GHI	global city hyperlink index
GIS	geographic information system
HITS	hyperlink-induced topic search
HNA	hyperlink network analysis
HTML	hypertext markup language
HTTP	hypertext transfer protocol
IGO	international governmental organization
IMF	international monetary fund
KWIC	key-word-in-context

LDC	less-developed country
MDS	multidimensional scaling
MIME	multipurpose internet mail extensions
MNC	multinational corporation
ODP	open directory project
PCA	principal component analysis
PDF	portable document format
QCA	quantitative (quantified) content analysis
SNA	social network analysis
SOM	self-organizing map
TNC	transnational corporation
UR	Uruguay round
URL	uniform resource locator
WB	World Bank
WTO	world trade organization
WWW	World Wide Web
XML	extensible markup language

CHAPTER I: INTRODUCTION

Today, the world has become a giant "village" where borders are increasingly irrelevant and people, capital, and technology flow in and out of the most remote corners of the globe. The Internet and satellite TV bring real-time news to billions of people around the world, deepening this globalization trend even more. "The Earth is one" is not a slogan for the Olympics or Greenpeace anymore. The world is linked together with political, economic, societal, and cultural relationships; and globalization is a diffused on-going process at the moment. Whether one likes it or not, globalization is an undeniable fact of life, unifying peoples of the world once isolated by geography. Nonetheless, there is little agreement as to what exactly globalization means to us; there are only incomplete- and competing- definitions based on partial understandings of the phenomenon.

Considering that globalization is an undeniable reality today, it is natural that researchers from various disciplines would study globalization. Although the frantic pace of development of transportation and telecommunications sets the stage for the end of geography, the role of geographers is still important to understand and explain globalization (Murray, 2006). Murray (2006, pp. 6-9) indicated that popular notions of globalization misapprehend what geography is, both as an entity and as an academic discipline; and also it fails to recognize how contemporary geographers define central components of their analyses such as space, place, scale, and location. In addition,

Murray stressed that the importance of relative distance has increased by the effect of the 'shrinking world'. Furthermore, full understanding of people and places (what geographers mostly do) with their histories, societies, and environments can explain differences in effects of globalization in different locales.

What is the geographers' view on globalization? In modern geography, the view of the world has evolved from mosaic, via network, to a system. How geographers see the world also affects how globalization is seen. Globalization is very well apprehended through the concept of system because globalization has multiple facets. In other words, different fields of interactions reciprocate with each unit (e.g. city, company, organization, etc.). Therefore, the systematic analysis of globalization is required. Geography is one of the academic fields that can deal with this kind of inquiries.

A distinctive trend of globalization research is the expansion of its domain in a manner that reflects the increasing reach of globalization matters and influences in our lives. Previously, globalization research focused on political and economic issues. In that stage, the concept of globalization was narrow and the unit of analysis was coarse. Research on world cities (or global cities) has broadened its horizons. As a result of the increasing interactions among people through the processes of globalization, not only the political and economic realms but also the cultural and societal realms have been considered as the area of globalization research. With cities being the primary conduits for globalization processes, they have also become the focus of attention in globalization studies. Commonly, the set of indices used to identify global cities has changed. In broad terms, research on indices has evolved from rather simple demographic data (e.g. population) to more complex, composite, and less tangible measures (e.g. quality of life).

As a city grows through interaction with other cities, various parts of the city experience the effect of globalization differently. Thus, we have a better chance to apprehend the complexity of the globalization phenomenon when we utilize a more complete set of indicators capturing the diversity of the urban environment.

1.1. Statement of Research

Global cities are the places from where globalization is driven. Although the effects of globalization are felt everywhere in the world, global cities are the actual focal points of this transformation. Global cities are also the nuclei in the global economic network. For nations and cities, one of the most important aspects of developing their economy and welfare is to set up an effective strategy to position oneself as a global city, like the Dubai Strategic Plan (Government of Dubai, 2007; Govers & Go, 2009, p. 88). Various indicators exist for defining and measuring global cities so that suitable benchmarks would be available, yet how the city is perceived by the public is hardly dealt with. The popular image of a global city influences its economy directly (e.g. tourism); it also affects subconscious awareness that touches economic decisions. It is our contention that one of the survival strategies of a global city is to enhance positive images and attenuate unfavorable images. Selling cities, city boosterism, and branding city are practical and effective tools to improve a city's image.

The World Wide Web has become the ultimate platform to create and disseminate information and images of the cities. The WWW connects people into one global network, accelerates the speed of communication, and integrates discourses, ideas and images. It is not surprising anymore how fast information is exchanged and how far it can reach out. Considering that the growth of the WWW started at the development of

browsers in the early 1990s and given the explosive increase of the WWW, it can be surmised that the WWW has influenced the image formation of global cities and its outcome. However, the image of cities on the WWW has not yet been researched in the context of global city networks. It is important for global city researchers to study information networks and the images that are conveyed on it because this analysis helps understand the contemporary complexion of global cities. It may also provide new opportunities for quantitatively apprehending the emerging achievements of cities on global city networks and provides the basis for the establishment of strategies of city image building. Moreover, it is critical for city strategists to know the structure and characteristics of city images on the Web.

Therefore, the goal of this research is to extend the concept of globalization and provide a new approach that can be used for global city research in order to reveal some hitherto hidden aspects of global cities. Specifically, this research proposes a methodology to transform the web data (i.e. webpages and hyperlinks) in the form of a network and extracts the textual content from webpages; it analyzes characteristics of the network and of the text and compares them to established views on global cities; it visualizes the distributional patterns of cities from the perspective of city hyperlink networks and of the textual contents that are shared on the WWW; and finally, we discuss the usefulness of this new approach to global city research.

While this research is in line with the current body of literature on the conceptualization and measurement of the global city, it also presents several significant points of departure. First, this research deals with the Web itself, as a complex set of features that would contribute to the differentiation of cities in terms of their global

standing. This means that this research innovates by regarding the Web and the data retrieved from it as a discriminating perspective for measuring the global significance of cities. Second, this research deals with structure and contents taken together. Analyzing the characteristics of both structure and contents is more helpful to understand the Web than examining only the structure or the contents because they both affect each other; and their joint analysis helps to understand each part. Third, this research deals with a non-material construct, the image of global cities. Although the image of global cities can be measured by a qualitative method like a survey, the quantified content analysis of this research provides a more data-driven approach to apprehending the image of cities. Through the quantification of Web contents, this research expands the area of indices for global city research to non-material indices.

To this end, the structure of hyperlink network based on extracted nodes (i.e. websites) and links (i.e. hyperlinks among websites) from the web crawling results of each global city will be analyzed with different measures of network analysis. Quantified content analysis (QCA) will examine the text from the related sets of crawled webpages. In addition, the distributional characteristics of the structure and the contents will be visualized.

1.2. Structure of the Dissertation

This dissertation consists of ten chapters. Chapter I contains the research purpose and the structure of the dissertation. Chapter II deals with important concepts and background, which includes definitions of globalization and of the global city, and the importance of measurement (i.e. indices) for global city research. Chapter III discusses the importance of the image of global cities on the Web for framing global city strategies.

Chapter IV contains four groups of research questions pertaining respectively to the general characteristics of cities from their hyperlink network, characteristics and classification of remarkable nodes of the city hyperlink networks, characteristics of the quantified text of webpages, and new integrated perspectives on global cities from the hyperlink networks and quantified contents. Chapter V provides justification for the selection of study cities, describes the data, and introduces the methodologies adopted for each research question. Chapter VI analyzes the result of various measurements of hyperlink network analysis (HNA), creates potential indices, and compares the new index to existing global city indices. Chapter VII contains the results from the analyses based on individual nodes and reveals nodal characteristics. Chapter VIII analyzes the quantified text based on QCA and classifies cities based on global dimensionalities. Chapter IX compares the results from the HNA to the results from the QCA and reveals the structural and the contextual characteristics of hyperlink network data in terms of the new aspect of global city research. Chapter X addresses the limitations of study and presents the conclusions of the dissertation and a discussion of future research.

CHAPTER II: LITERATURE REVIEW

The structure of the literature review is as follows. Firstly, the definition of globalization and the importance of global cities as focal points of globalization will be discussed. Especially, this first part will focus on the various aspects of globalization and the effort to set up the theory of globalization. Secondly, the development of measurements (i.e. indices) is reviewed. Lastly, we review the literature on the Web and city branding as elements contributing to building a Web-based image of the global city.

2.1. Globalization and Global Cities

2.1.1. Defining Globalization

Globalization is the on-going integration of humanity; it also influences all human-related activities. It exposes individuals, organizations, companies, and countries to one big stage. The boundary of news and information expands from the local to the global. At this moment, globalization influences everyday life directly and indirectly. Although globalization is a prevailing phenomenon, it is hard to find consensus on a definition of globalization. The reason is that the definition of globalization depends on the researcher's world view. At the same time, the academic background, personal experience, focusing point of research, and data availability also affect the definition of globalization. In other words, the definition of globalization is readily affected by what a researcher wants to show and emphasize based on their own scheme.

Ervin and Smith (2008) identified three major world views to interpret globalization: *The Neoliberals* as the free market enthusiasts, the *Institutionalists* who want to regulate globalization, and the *Critics* who see globalization as destructive imperialism. Table 2.1 shows how each world view apprehends globalization. The Neoliberals think that globalization is absolutely beneficial to humanity. The economic system of globalization is so perfect that intervention and regulation should be removed for the sake of 'free hand' working. Cultural globalization brings modern lifestyle and gives belief of the 'American dream' to developing countries. According to this view, global environmental problems will be solved by human innovation and technology. Institutionalists consider that globalization has positive benefits to humanity. However, they think that the global economic system is not perfect, so that intervention is required. The active role of government can decrease the negative impacts of the global economy. Furthermore, international agreements and cooperation help solve global environmental problems. Intercultural connections provide understanding among different cultures and help reduce global collective problems. Critics think globalization is a form of imperialism. In other words, developed countries exploit peripheral countries, and multinational corporation control globalization. Cultural globalization is a form of imperialism and cultural homogenization minimizes cultural rejection of foreign norms for subordination. The privileged few impoverish the masses through the degradation of global environment, and this destabilizes domestic and global economy. Critics believe economic democracy can solve the problems of globalization.

Table 2.1: World views of globalization (adapted from Ervin & Smith, 2008, pp. 29-59)

	The Neoliberals	Institutionalists	Critics
Globalization is	Beneficial to humanity	Positive benefit to humanity	Form of imperialism
Cultural globalization	• Positive process • Free markets diffuse the best ideas • Modern lifestyle • 'American dream' (out of poverty)	• Overall benefit to humanity • Disagree the actual impact and importance of cultural globalization • Negative or positive	• Cultural imperialism & homogenization • Minimize cultural rejection of foreign norms • MNCs' controlling
Economic globalization	• Perfect itself • Structural adjustment policies for developing countries by IMF and WB • Deregulation and liberalization	• Not perfect, interdependent • Active role of government to lessen the negative impacts from the global economy	• Exploitation • Global imperialism hurts the national interest
Environmental globalization	• Skeptical about the dangers of global environmental problems • Cornucopian, optimistic belief in human innovation and technology	• A healthy natural environment and global system stability cannot be separated • International organizations (IGOs) & regulation	• Environmental degradation triggers destabilizing economic globalization • Privileged few, impoverish many
Global Problem solved by	• 'Free hand' • No intervention and regulation	• Countries' own mechanisms • International agreements and cooperation	• Democracy for economy

Although these diverging world views make it hard to reach a consensual definition of globalization, a unique definition of globalization is a prerequisite because globalization is the object of scientific study. The common definition of globalization is a starting point to understanding the various aspects of globalization. To formulate the general definition of globalization, Al-Rodhan (2006) reviewed 114 definitions used by globalization researchers and organizations. His proposed definition is that globalization is "a process that encompasses the causes, course, and consequences of transnational and transcultural integration of human and non-human activities" (Al-Rodhan, 2006, p. 5). With the need for a general definition, he recognized that the notion of globalization

commonly has economic roots. While globalization is indeed multifaceted and complex, its premises are in economic considerations. Actors of globalization identified by Ervin and Smith (2008) also point to the economic roots of globalization. Actors of globalization are nations-states, central banks, international nongovernmental organizations, international governmental organizations, multinational corporations, international financial institutions, and free trade agreements. Among the seven actors of globalization, four actors (i.e. central banks, MNCs, international financial institutions, and FTAs) are directly or indirectly connected with the global economy. Their collective actions are certainly what propel the globalization process, rather than the actions of single actors. In addition, according to the on-line Oxford Dictionaries[1], globalization is "the process by which business or other organizations develop international influence or start operating on an international scale." In this simple definition, we can find two important words: 'business' and 'international'. These two words are indicative of the origin and direction of globalization.

The advent of globalization is the process through which the world economy adapts to the demand of the times. The oil shocks of the 1970s drove the world into a corner, and the hegemony of the United States was waning. Mass production and mass consumption of Fordism could not be the solution anymore. The world required a new economic regime. Therefore, the era of the post-Fordism in the 1980s is characterized by a flexible system of production. The world is connected closely for overcoming the economic crisis with the globalization of capital, finance, labor, etc. For the emergence of the new world-economy regime, three main actors (i.e. MNCs, banks, and governments) were required to adapt its role to the heightened economic competition (Thrift, 1988).

[1] http://oxforddictionaries.com/definition/globalization?q=globalization (retrieved on Jan. 31, 2012).

Especially, MNCs expanded their investment to the whole world. In fact, MNCs moved the manufacturing process to less-developed countries that have low wages, and tried to gain a certain portion of the local market thanks to local production facilities. Previous international division of labor was that the LDCs provide raw materials and resources, and the advanced countries produce commodities. In contrast, since the 1970s, LDCs have started to assume the production of commodities for the advanced countries because they provide cheap and plentiful labor. Compared to the previous international division of labor, this new phenomenon is called the new international division of labor (Fröbel, Heinrich, & Kreye, 1980).

In addition to the use of cheap and plentiful labor, there are other reasons behind the phenomenon that MNCs expand the power of influence to the LDCs. According to Thrift (1988), MNCs influence overseas subsidiaries by the export of capitalistic relations of production: the internationalization of capital. The export of capitalistic relations of production has three main forms: obtaining raw materials, penetrating the markets, and exploiting cheap labor. In short, multinational corporations intend to solve the lowering of the profit rate under the conventional capitalistic economy by advancing into the world market.

Because of the emergence of expanded MNCs, new international division of labor, and the internationalization of capital, manufacturing industries of the advanced countries gradually relocated to LDCs. At the same time, industrial cities of advanced countries held on to corporate headquarters and research and development facilities as they assumed the role of command and control. In other words, while existing manufacturing functions of the city moved to the LDCs, a new role emerged for cities, as

centers of command and control for world operations and centers of advanced producer services. This is the global city, which is the new form of the restructured city brought by globalization of the economy.

The rapid development of telecommunication and transportation has made it much easier to overcome the friction of distance. This phenomenon is expressed by different terms, but all of them point to the same reason, the development of technology. Janelle (1969) referred to time-space convergence for describing the rapid increase of overcoming the distances per time units, which is by technological innovations in transportation and communication. Harvey (1989) used the concept of time-space compression, which means that "the time horizons of both private and public decision-making have shrunk, while satellite communication and declining transport costs have made it increasingly possible to spread those decisions immediately over and ever wider and variegated space" (Harvey, 1989, p. 147). Giddens (1990) expressed it as time-space distanciation, which describes the increase of easy and speedy social interrelationship by modern technologies. It is easy to think that this shrinking world impairs the importance of the city where to socialize, exchange, and produce. However, the shrinking of the world imparted by the increase of interaction between cities means that cities are functionally connected to each other more. In other words, the city becomes the place where multiple functions overlay. It is important for the city to play a leading role in multiple functions, which attract human and financial resources. Thus, the city is highlighted as the unit for the analysis of urban systems.

King (1990) emphasized that the 1980s was a period of paradigmatic shift in research on cities. Although the first forays in the study of globalizing cities date back to

the 1970s, the effort was totally different in the 1980s. The major difference between the 1980s and previous periods is the globalization of the world economy. The creation of the World Trade Organization (WTO) shows the importance of the 8th round of the General Agreement on Tariffs and Trade, the so-called Uruguay Round[2], which was started in 1986. The Uruguay Round guided the creation of the WTO, which is the symbol of boundless economic competition; thus, the 1980s was the period that the new world economic order became the frame of reference. The integration of the world economy affected the city research agenda, which also accepted the concept of globalization.

In sum, globalization is the on-going integration of humanity from the integration of the world economy; it influences all human activities. Although authors have different opinions about globalization, there is common agreement that it stems from global economic integration. Apart from the likes and dislikes about globalization, one of the main motivations of the process of globalization is to secure economic benefits through economic globalization.

2.1.2. Global City Where Globalization Occurs

Historically, the city has been important to rulers, elites, merchants as the main place to proclaim, communicate, exchange, socialize, and rebel sometimes. Today, cities remain magnets, with people migrating from the countryside of developed countries and people also migrating from developing countries to the city of the developed world for various reasons: jobs (openings and income), education, politics, etc. While Holston and Appadurai (1996) mentioned the turmoil of the citizenship conflicts in the world's major cities due to the different groups of immigrants, the global city is now the melting pot of

[2] The Uruguay Round (1986-1994) decided 40% reduction in tariffs and agricultural subsidies, agreed full access for textiles and clothing from developing countries, and extended its area to intellectual property rights.

disparate peoples, cultures, thoughts, and value systems. In other words, the global city has become the focal point of people, business, culture, politics, and conflict.

Table 2.2: Thematic classifications of global city research from 1981 to 1998

Theme	Author	Title	Year
Historical background and theoretical basis	Cohen, R. B.	The new international division of labour, multi-national corporations and urban hierarchy	1981
	Friedmann, J. & Wolff, G.	World city formation: an agenda for research and action	1982
	Feagin, J. R. & Smith, M. P.	Cities and the new international division of labor	1987
	Thrift, N.	The geography of international economic disorder	1988
	King, A. D.	Global cities: post-imperialism and the internationalization of London	1990
	Sassen, S.	Cities in a world economy	1994
Global city system	Chase-Dunn, C. K.	The system of cities (A.D. 800-1975)	1985
	Smith, D. A & Timberlake, M. F.	Cities in global matrices: toward mapping the world-system's city system	1995
	Knox, P. L.	World cities in a world-system	1995
	Taylor, P. J.	Hierarchical tendencies amongst world cities	1997
Inequality	Massey, D.	Spatial divisions of labor: social structures and the geography of production	1984
	Harvey, D.	The limits to capital	1982
		The urbanization of capital: studies in the history and theory of capitalist urbanization	1985
	Soja, E. W.	Postmodern geographies: the reassertion of space in critical social theory	1989
	Hamnett, C.	Social polarization in global cities: theory and evidence	1994
Management and strategy of global city restructuring	Glickman, N. J.	Cities and international division of labor	1987
	Knight, R. V.	The emergent global society	1989
	Vonk, F. P. M.	Managing the metropolis	1989
	Gappert, G.	A management perspective on cities in a changing global environment	1989
	Friedmann, J.	World city futures: the role of public policies in the Asia-Pacific region	1997
	Nam, Y. W.	A study of linkage policy for downtown redevelopment	1998

Alongside research on the innate characteristics of global cities, there is also a large body of research on the networks formed among global cities. Given the broad-based consensus that globalization expands the economic network to the world and that it is based on global economic integration, it is natural for research to also shift its interest from single cities to networks of cities. Research on the global city can roughly be classified into the following four themes: historical background and theoretical basis of the global city, the global city system, inequality, and finally management and strategy of global city restructuring (Table 2.2). The first and the second themes are related to the characteristics of globalization and of the global city; the third theme is about the problem of globalization. This theme can be expanded to environmental and sustainability issues. The last theme is related to the competitive power of the global city.

Research on the historical background and theoretical basis has dealt with a number of topics including the relationship between the global city and other cities, world systems theory, the historical characteristics and uniqueness of the global city, and the role of the global city as the place of dispersion and concentration of capital (Choi, 1998). Cohen (1981) positioned the emergence of the global economy in the context of the change of international labor structure. Discourse about global cities by Friedmann and Wolff (1982) meaningfully pulled the thoughts of globalization together, and provided a chance to use the paradigm of globalization in the field of urban research. The conceptual foundation of the global city was established in Friedmann's 1986 paper[3]. Friedmann examined the changes experienced by world cities under the integration of the world economy. He also argued that world cities exhibit linkages and hierarchies and that the

[3] Hall (1984) and Markusen (1999) argued that Patrick Geddes coined the phrase 'world city' in 1915. However, the first use of 'world city' is by Goethe in the 18th century while praising the cultural renown of Paris and Rome.

increased economic and control function of major cities have affected the economic restructuring and polarization of social classes. King (1990) discussed the formation of global cities and argued that the global city is the historical result of the colonial expansion of developed countries to the LDCs. Sassen (1994) studied industrial sectors that are specific to global cities, particularly the rise of leading sectors such as financial services which need to be paid more attention with the advances of communication technologies. Commonly, research in this theme discusses the reason of the emergence of global cities, which is based on the global economic integration.

The line of research on the global city system is devoted to the functional and hierarchical network of cities worldwide, and the different roles and relationships between core, semi-core, and periphery in the system. When we examine the differentiation of global cities, global cities must be seen not only through a positional role in core, semi-core, and periphery but also through the internal mediation of the contingent conditions of local socio-economic and political structures as well as through the physical structure of the city (Beauregard, 1995; Knox, 1995; Rimmer, 1986). Chase-Dunn (1985) also insisted that functional differentiation and hierarchies of global city systems could be understood from political and economic structures of the city. With the examination of the global city along different relationships and scales, the research on global cities needs to embrace not only the characteristics of the city itself but also the characteristics of linkages (Beaverstock, Smith, & Taylor, 2000a). Smith and Timberlake (1995) suggested a conceptualization of linkages between cities (Table 2.3). As the range of linkages that bind global cities together has broadened, experts from more diverse fields have been drawn in. Naturally, research on the global city system requires an

explicit plan for data collection, credible interpretation of the data, and the analysis of hierarchical tendencies at the city level (Taylor, 1997). In sum, while research on the global city system provides the chance to expand the global city research agenda to adjacent academic fields, it also needs cooperative work from different academics.

Table 2.3: Conceptualizing inter-city linkages: a typology (adapted from Smith & Timberlake, 1995, table 5.1)

Function	Form		
	Human	Material	Information
Economic	Labor, Manager, Lawyers, Consultants	Capital, Commodities	Business phone calls, Faxes, telex messages, Technology transfer, Advertisements
Political	Troops, Diplomats, Social workers	Military hardware, Foreign aid	Treaties, Political threats
Cultural	Exchange students, Dance troupes, Rock concerts, Theatre	Paintings, Sculpture, Artifacts	Feature films, Videos, Phonograph albums (CDs)
Social reproduction	Families, Red Cross, Community organizers	Remittances, Foreign aid	Post cards, Night phone calls

Another research theme related to the global city system is the inequality of social groups within a city due to the hierarchization of the global city system. The social and spatial polarization is one of the main issues. There is a large hierarchical chasm between the elite and the blue-collar labor in a city; even in one country, there is a gap between the main cities following a global city strategy and the local cities that do not partake in this process. From a Marxist perspective, this problem is due to capital accumulation (Harvey, 1982, 1985). Researchers studying locality said that the restructuring of industries and the reformation of the city causes the inequality problem (Cooke, 1989; Massey, 1984). In the era of postmodernism, the change of regulation mode and the following emergence of new industrial districts are also pointed out as causes of deepening inequality (Harvey, 1989; Scott, 1988; Soja, 1989). Although scholars have

tried to develop a certain theory to explain the reason for the existence of inequalities in global cities, it is quite hard to say that inequality has a single cause. As Hamnett criticized Sassen's thesis of growing social polarization in global cities because it was based on the special cases of New York and Los Angeles (Hamnett, 1994), broad-based empirical analysis of global cities is needed to avoid hasty generalizations.

The final theme of research on global cities concerns the management strategies to deal with problems arising locally as a result of globalization. According to Nam et al. (2000), globalization transforms the existing economic, social, spatial structure of the city, and readily provokes political conflict. Thus, globalizing cities require new strategies to alleviate these problems. Historically, the city has adjusted its economic structure to the changing global economic situation continuing sustain its economic growth (Nam & Park, 1998). However, the globalizing city is required to provide proper strategies with full consideration of spatial organization, regional governance, social and environmental sustainability, migrant workers, the rise of civil society, and intercity networks (Friedmann, 1997). As once Thrift (1999, p. 283) indicated that the city would be managed like a business, this globalized world imposes business-mindedness on the city.

A suggested strategy for the management of the global city is to increase the accessibility to the resources of the city from other cities (and vice versa) through the expansion of intercity networks. The power of cities reflects their accessibility, which includes the range and quality of contacts (Knight, 1989, p. 40). Historically, the relationships between cities were established through colonial trade, then by the system of national states, and more recently by IGOs, NGOs, philanthropic and cultural

foundations, and transnational corporations (Knight, 1989, p. 39). Recently, governors and mayors have tried to advertise their cities to MNCs and IGOs. They already know that the era of globalization requires a business mindset for sound municipal management. However, the most important thing for a global city is to attract businesses to the city. High-tech-oriented industries and service sector activities (producer services, research & design, etc.) have more potential; and high-quality production environment (high-quality housing and infrastructure) is required to compete with other cities (Vonk, 1989). Another strategy is to establish a policy for reviving the city center, which has often deteriorated as a result of the urban restructuring process[4]. Like the cases of the London Docklands redevelopment project and of the Boston Waterfront redevelopment, the policy (especially linkage policy) can change space and revive the central urban area (Nam, 1998). These projects show the importance of strategic approaches for global cities.

We can better apprehend the potential strategies by investigating the strategies of the top three global cities: New York, London, and Tokyo. While nowadays, their global city status is often taken for granted, deliberate strategies for raising them to this standing were enacted. For New York, the task force of the Twentieth Century Fund (1980) reported the following strategic opinions. First, it begins with promoting international and white-collar industrial sectors because these sectors will receive an advantage from a continued growth in world trade and investment. Second, it should preserve the existing industrial base. This is not only sparing the cost of huge industrial park development, but also helping employing residents and the expansion of manufacturing enterprises. Third,

[4] Urban restructuring process is classified into economic restructuring, states restructuring, household restructuring (including migration), community restructuring (and community politics), and spatial restructuring (Feagin & Smith, 1987, p. 13).

people are more important than places, so that existing neighborhoods should be strengthened and protected. This has helped New York become a more family-friendly and livable place. Based on these strategies, the task force recommended the exemption from taxes to the properties of foreign governments and foreign portfolio investment. Furthermore, they advised improving public transportation for commuters and airport passengers, creating a special office for responding to foreign enterprises, removing legal distinctions between the domestic and the foreign, and even relocating the WB and IMF from Washington D. C. to New York.

The London Planning Advisory Committee (1991) suggested four visions for the future of London. For London to prosper as a global city, it should 1) Provide high-quality environment to citizens; 2) Become the center of international trade and business; 3) Guarantee equal opportunity to all; and 4) Develop sustainable neighborhoods. Based on these visions, they provided development strategies for financial services, manufacturing, cultural activities, and quality of life. Each strategy consists of practical goals. For the example of quality of life, it included the standards and regulations on sulfur dioxide emissions, noise levels, public space & urban design (ratio of park provision to office stock in commercial areas), personal safety (crime rate), and cultural provision (theater, the visual arts, film & audio industries, music, design, museums, and sports facilities). In addition, the committee dealt with infrastructure as major basis for global city. Transport systems (the role of government, integrated mass transit systems, fare system, highway networks, and environmental issues), international transport links (airport, international rail facilities, waterways, and waterfronts), communications, land use (commercial property), stock creation (choice and price), housing (prices, access, and

new building), education and training are considered another basis.

The Tokyo Metropolitan Government (1991) established the third long-term plan for the development of Tokyo in the 21st century. The governor of Tokyo mentioned four aims of this plan: 1) Comprehensive emergency plan for housing policy and recycling; 2) Affluent and comfortable lives with the nation's economic growth; 3) Solutions for overconcentration in Tokyo; and 4) Contribution to the development and peace in the world. This plan addressed the problems of Tokyo as a city, a regional center, a national capital, and as a world city. Although most parts of the plan presented how Tokyo can become a livable place like an advertisement, the discussion and planning for making a better place for residents are not different from those mentioned for New York and London. Interestingly, they mentioned the globalization of Tokyo and the need to work towards providing not only a comfortable environment for foreigners, but also a pleasant cityscape, better living environments including housing, and cultural facilities (Tokyo Metropolitan Government, 1991, p. 36). They already knew the importance of the environmental and the cultural aspects as well as sustained economic development.

All top three global cities articulated a vision for a global status. The path to global city status does not only mean the world center of business. It also encompasses the role of the premier city in a country, and the quality of life for its citizens. Although balance and harmony are what they want to realize in their cities, the reality is not easy to manage because of the complexity of factors, which is created by innate characteristics and multiple relationship with other cities.

From the review of global city research, it transpires that the trend has changed from recognizing the phenomenon (i.e. globalization) to providing strategies. Researchers

have found a growing number of internal and external factors that intersect with the globalization of the city. Considering that globalization has broadened its reach, the addition of new indices accounting for those factors is only natural.

2.2. Development of Measurements (Indices) for the Global City

Considering that globalization is the overall ‘process’ discussed above, we will use the term ‘globality’ to refer to the is the state of a city undergoing the effects of globalization. That is, indices measure the globality of a city in a certain area where globalization is on-going. Globality can be measured by various types of indices. An index captures the extent to which a city is inserted in a global network of cities from particular thematic perspective. Although one city may be renowned as a global city, it is hard to prove without an index (or indices). Thus, indices are necessary for measuring how globalization affects the city. In other words, indices are the basis to reveal the degree of the globalization, the ‘globality’.

Globalization started with the globalization of the world economy, and global cities are also understood as the focal points of economic relationships. However, globalization is multifaceted, and the various parts of the city are affected by globalization. Table 2.4 shows examples of indices proposed by Jo (1992) and Nam (2006); they have been used for defining global cities and measuring their level of prominence in the globalization movement. Economy indices are often preferred, such as the prominence of MNCs, banks, and other producer services. Social and cultural indices form the next preferred indices. Interestingly, the increased usage of indices from telecommunications and others reflects that researchers understand the multifaceted

characteristics of globalization, and all these indices can be used as indicators of the globalization.

Table 2.4: Examples of indices (modified from Jo, 1992; Nam, 2006)

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Population		×	×	×								×		4
Component ratio								×					×	2
MNC/TNC						×								1
Headquarters	×		×	×	×		×	×	×	×		×	×	10
Branch office			×	×	×		×	×						5
Producer Service														
Bank/finance	×	×	×	×	×	×	×	×	×	×	×	×	×	13
Business activity			×	×				×						3
Law	×	×			×		×				×			5
Insurance & securities		×			×	×	×				×	×	×	7
Accounts/consults/advertisements	×				×	×	×				×		×	6
Research & design		×			×		×							3
Transportation			×			×	×							3
Air traffic/passenger/freight/mail		×							×	×		×		4
Marine transport		×												1
Registered vehicles												×		1
Telecommunication							×	×	×				×	4
Internet utilization ratio												×		1
Social and Cultural Indices														
International organizations		×	×			×	×					×		5
International sports and meetings						×			×			×	×	4
Entertainment		×				×			×					3
Hospital		×				×								2
Education		×				×								2
Library		×												1
Museum		×				×								2
Theater												×		1
Others														
Stock market/exchange									×	×		×	×	4
Capital flow (FDI)										×			×	2
IT company headquarters												×		1
Major manufacturing center			×	×										2
Political stability							×						×	2
Livability/living cost												×	×	2
Doctor/dentist												×		1

Notes: 1 Cohen (1981), 2 Hall (1984), 3 Friedmann (1986), 4 Feagin & Smith (1987), 5 Thrift (1988), 6 Knight (1989), 7 King (1990), 8 Sassen (1994), 9 Short, Kim, Kuus, & Wells (1996), 10 Short & Kim (1999), 11 Taylor (2004), 12 Nam (2006), 13 Sassen (2011).

Based on the characteristics of the indices, one can roughly distinguish three categories: intrinsic indices, relational indices, and qualitative indices. Each category is

further divided into three common sub-categories: demographic indices, economic indices, and facility indices (Figure 2.1). That is, the placement of indices into the categories is decided by the mode of utilization, and their placement into sub-categories is based on the indices' own characteristics. For instance, total population and migration data are demographic indices; the former can be used for scaling the city, while the latter is used for explaining relationships among cities. This kind of classification is useful not only for analyzing the strengths and weaknesses of each city, but also for assisting a decision maker to invest efficiently on any single or combination of the internal part, the external part, and the qualitative part.

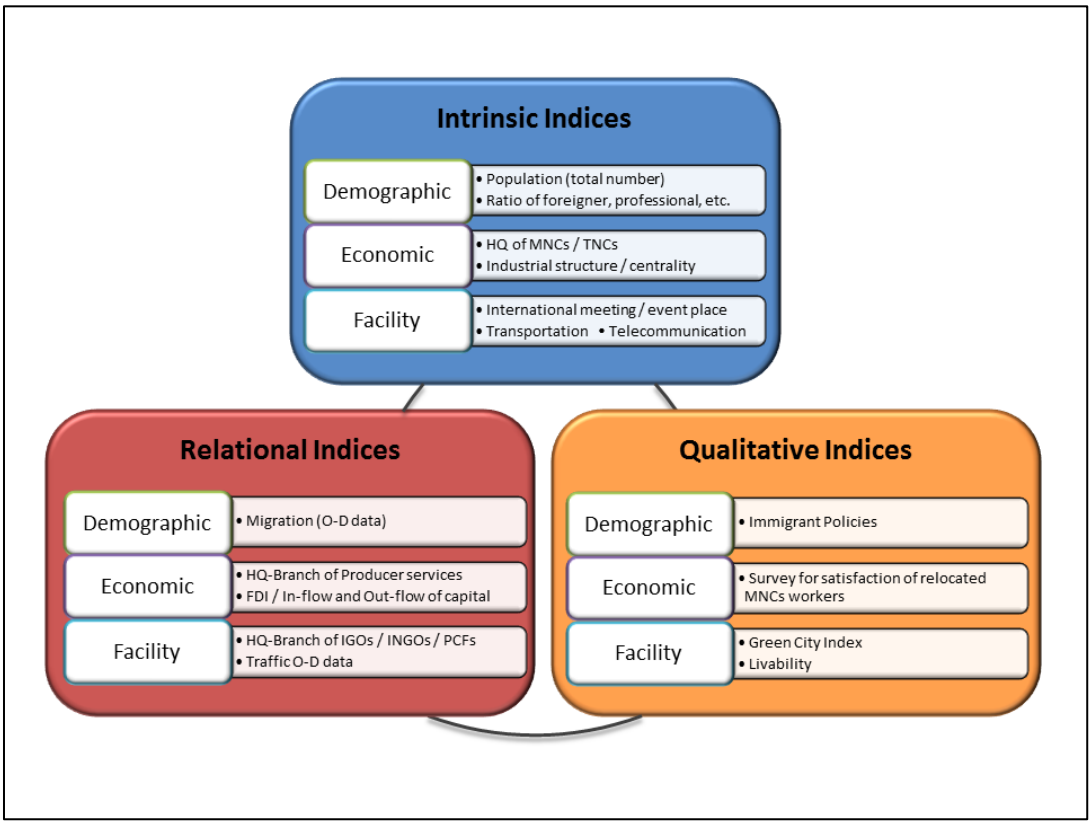

Figure 2.1: Classifications of indices

2.2.1. Intrinsic Indices

Intrinsic indices refer to properties that are specific to each city in and of itself. These indices include population, number of companies, and more generally the scale and scope of urban amenities. In early research, demographic indices, especially population, have been used as index to show the scale of the city. It is possible that a city of large population is also a major economic center; however, it is not necessarily the case. Many European cities are often regarded as global cities, yet their population is relatively smaller than cities in other world regions, especially than Asian cities. Thus, the composition of the population such as the ratio of foreigners, professionals, and educational attainment may be more important to reveal the distinguishing demographic characteristics of global cities than absolute population size. (Clark, 2003; Samers, 2002; Sassen, 2011). In addition, it is easy to think that the change of professional employees in workforce composition is only meaningful given the command and control function of global cities; however, day laborers or illegal immigrants also can be considered as indices showing the degree of globalization because one characteristic of the global city is polarization (Sassen, 2011), as discussed earlier.

Demographic characteristics are strongly connected with the economy of the city. When we review the development process of cities, usually the primary city of a country becomes the most populated at an early stage of economic development. This means that labor force and industries are concentrated in the primary city. Other cities rarely have a chance to grow their economy to become global cities due to unequal distribution of domestic resources. However, globalization gives other cities a chance to overcome this limitation. Legal (and illegal) immigrants fulfill the needs of business services with low

cost. The boundless competition of globalization requires that businesses lower their cost by using cheaper labors. High-skilled workers can move more easily to other cities. Thus, the analysis of demographic characteristics is in order when we examine the economic structure.

Infrastructure is also an important consideration for a global city. The existence and capacity of transportation facilities, including international airports and ports, is important to the comings and goings of people as well as to the traffic of international freight and mail. Transportation and telecommunication accelerate the interaction among people, money, properties, and information (Keeling, 1995). The international info-communication networks support the command and control function of the global city (Graham, 2002; Warf, 1995). However, the increase of global connections via telecommunications does not reduce the necessity of face-to-face meetings. Convention centers, hotels, stadiums, theaters, museums, concert halls, and opera houses are good examples for global indices because hosting international gatherings and entertaining visitors is almost impossible without these facilities. These indices catalyze global interactions. In addition, international governmental organizations (IGOs) and international nongovernmental organizations are also used for revealing complex characteristics of global cities.

2.2.2. Relational Indices

Relational indices capture the functional relationships that exist between cities. Usually, these indices are constructed from Origin-Destination (O-D) data. For the demographic aspect, flows of immigrants are useful to reveal the mobility of population between cities. With the demographic characteristics of immigrants, the link information

could be used for finding the reason of immigration. The linkages between headquarters and branch offices also can be used for finding the global connections. Unlike the absolute number of corporate headquarters in a city, the linkages explain the connectivity among a certain type of businesses; especially, it is suitable for producer services because producer services are highly concentrated in global cities where the benefits of accumulation and innovation materialize. Practically, Taylor (2004) analyzed the global network connectivity of global producer service companies, including finance, accounting, insurance, consulting, law, and advertising, based on the number of offices in multiple cities. From his network analysis and the map of multidimensional scaling, he argued that this analysis provides hints about how cities fit together in the world city network and a fresh way of looking at world cities and their inter-relations (Taylor 2004, p. 124). In addition, foreign direct investment is an important index for finding capital flows. The flow of global capital reveals the characteristics of global cities such as the concentration in international trade, the externality of urban economy, and the connectivity in the global hierarchy (Kim & Park, 2005).

O-D data of air passengers, freight, and mail are representative indices for showing the network of global cities. The International Civil Aviation Organization collects air traffic related data annually from member airports, which can be used for the analysis of air traffic. The relative centrality of air passenger traffic from 1977 to 1997 has been analyzed to reveal uneven development dynamic, which is characterized by a few global cities dominating the changes of the global city system (Smith & Timberlake, 2001). The flow pattern of international air freight has also been analyzed by using graph theory and factor analysis (Nam & Lee, 2004); it reveals the primacy of Tokyo and two

divided hinterlands, one spanning the Pacific Ocean and the other, the Atlantic Ocean. Flows of containers between ports also can be considered as an index that shows the connectivity. Maritime traffic is more freight centric than air traffic, so that reasoning out the command and control functions of global cities is more difficult. In addition, more assumptions are required for the traffic flows over air and maritime when we consider that airports and ports are not just the gateway of flows. Data will have additional values if it is collected with travel purpose; and other means of transportation (i.e. modes) also should be considered when we use air passenger data from nearby cities (Kim & Park, 2005).

2.2.3. Qualitative Indices

Qualitative indices refer to the perceived and descriptive conditions of the city rather than quantifiable and numerable conditions. The importance of these indices has not apprehended until recently, although they form critical factors that affect people's critical decision or comparison of cities. This group of indices is traditionally measured by qualitative methods like in-depth interviews and group discussions with open-ended questions. However, it shows limitations when researchers deal with a large number of cases and analyze the trend. One approach to overcome this limitation is to focus on the frequency of key words, phrases, or descriptors extracted from media material. Another is to proceed through indirect measurement by a combination of related indices. These two approaches are now further discussed.

The first approach involves the conversion of the qualitative data to a quantifiable form. It has been used for the analysis of newspaper content. The possibility of quantifying newspapers has been confirmed by several researchers (Beaverstock, Smith,

Taylor, Walker, & Lorimer, 2000b; Pred, 1980; Taylor, 1997). The authors mentioned that the frequency of keywords from newspapers would be a legitimate research method as long as we pay attention to the 'frequency' of keywords because high frequent keywords represent the tendency of (an) article(s). They insisted that this quantification approach can be used for classification of articles even when the newspaper editor infuses bias in the articles. Alternatively, finding characteristics and classifying global cities may be based on the number of unique articles in predefined categories, which would have appeared in major newspapers (Son, 2006b). This research shows abundant possibilities to do quantitative research on qualitative data if care is taken in the classification of raw textual data.

The second approach to developing qualitative indices is to combine one or several countable indices into one representative and composite indicator. For example, the number of events associated with hosting Olympic Games and world tour concerts can be used as a cultural indicator of global status (Short, Kim, Kuus, & Wells, 1996). Recently, this approach has developed into multileveled grouping and weighting for precise measurement. The methodology of the Economist Intelligence Unit for measuring global city competitiveness (EIU, 2012, pp. 29-35) is a good example. First, they set eight categories of criteria and assign weights to each category (economic strength, 30%; human capital, 15%; institutional effectiveness, 15%; financial maturity, 10%; global appeal, 10%; physical capital, 10%; environment and natural hazards, 5%; and social and cultural character, 5%). Each category consists of one to six sub-indices, and each index has different weights. For example, 'Economic strength' contains nominal gross domestic product (25%), GDP per capita (10%), households with annual consumption over

US$14,000 (10%), city real GDP growth rate (45%), and regional market integration (10%). This second approach offers the advantage that it is composite and each dimension is suitably weighted as part of a weighted average.

Although these two approaches have expanded the domain of usage of qualitative indices in global city research, their fundamental problem is that there are no standardized indices. One effort to overcome this problem is *The Global City Indicators Program*, which is supported by the World Bank. This program provides 22 themes and 94 indicators, which are organized into two categories: city services and quality of life (Global City Indicators Facility, n.d.). This program is impressive because it provides a great variety of indicators and shows the standardization effort undertaken to embrace all city data. However, the completion of this program is still far off because the data is provided by cities on a volunteer basis. So far most participating cities are hardly recognized as global cities, while the top global cities (e.g. London, New York, Tokyo, etc.) are not on the list. This case tells that it is important not only to create standardized indices but also to consider the way of collecting data.

Unlike the global city indicators program, the *Mastercard Worldwide Centers of Commerce Index* is a good example of standardized indices and of successful data collection. This report is the updated version of the research from the Globalization and World Cities Study Group of Peter J. Taylor and John Beaverstock. It has 7 dimensions consisting of 43 indicators and 74 sub-indicators (Mastercard Worldwide, 2008). The 2008 report covers 75 global cities, and it adds the new dimension of 'livability' that captures the attractiveness of the local environment for global business. Lastly, the series of *Green City Index* compiled by the Economist Intelligence Unit also shows the effort to

expand the area of underlying indicators. The series includes the *European Green City Index* in 2009 (30 indicators), the *Latin American Green City Index* in 2010 (31 indicators), the *Asian Green City Index* in 2011 (29 indicators), and *US and Canada Green City Index* in 2011 (31 indicators). Each indicator is grouped in CO2, energy, buildings, transport, water, waste and land use, air quality, and environmental governance (EIU, 2009, 2010, 2011a, 2011b). Although these reports mainly focus on the indicators that are related to livability only, this series is meaningful because it includes underlying indices that were once neglected. All these research efforts show that global city research attempts to avoid using a single index for a certain topic, while Wellbeing (i.e. Quality of Life) indicators are also as important as other traditional indicators.

As a conclusion, global city research has expanded the breadth of indices from purely economic indices to more qualitative indices. This trend demonstrates that global city researchers realize that these underlying indices meaningfully affect global cities.

CHAPTER III: IMAGE ON THE WEB FOR GLOBAL CITY STRATEGY

Having a positive public image is one of valuable assets for individuals, societies, and companies. The development of telecommunication technologies, especially the Internet, speeds up and expands the distribution of images and its effect. Cities are now compelled not only to advertise and market themselves but also to maintain an attractive and polished image. Notwithstanding the importance of image building for cities, there is a dearth of research on global cities that study their image quantitatively and seek their shared features.

As the Web has grown to become a new space for human activities, this space is a valuable source of information for global city research. Considering the flourishing of the Internet users, exchanged information and knowledge help us to figure out how people perceive a certain topic such as the state of city globality.

This chapter will firstly review the relationship between public image and cities. Second, the characteristics of the Web will be discussed to examine in more detail how the Web is embedded in this research. Third, the concept of branding and related research will be reviewed. Lastly, the discussion based on image, the Web, and branding provides the conceptual framework for the extraction of global city images from the Web textual contents. This chapter will help to frame how the concepts tied to the image of global cities contribute to pioneer new understandings of global cities.

3.1. Image and Cities

An image has different definitions. The image means the representation of the external form of a person or things in art; and it means a simile or metaphor. It also refers to the general impression that a person, organization, or product presents to the public. The latter view is in line with the purpose of this research because the aim here is to measure the ‘globality’ of cities based on the quantified textual information which is the representation of image of cities. Considering that the city is an integrated place for human interactions, the image of a city (or destination image) in the public is multifaceted since it results from the image formation of each individual (Gallarza, Saura, & Garc á, 2002). Especially, geographers have studied the process of human perception in relation to surrounding environments. Gould and White (1986) discussed the images of places with the concepts of spatial preference, perception, and interaction. They impressed on the importance of the image in that the image can affect changes of the physical environment. Legibility is crucial for the well-formed image of the city; and a city is regarded as distinct and remarkable when this city is highly imageable (apparent, legible, or visible) (Lynch, 1960).

The image of a city is the aggregate of the impressions that the city presents to the public. Many different types of sources affect the image one may have of a city such as reading a newspaper article, watching TV, personal visits, or word of mouth. The Internet is the fastest and most powerful information source nowadays, compared to the more traditional media. As the medium of information exchange, the contents of the Internet like news portals, Really Simple Syndication (RSS) feeds, and social network websites and mobile applications affect people while developing their image or view on a certain

topic. In addition, recent media based on the Internet allows information providers to react to the users immediately while the previous media are one-way mass communication. Thus, the process of information exchange between providers and users or among users reshapes the image that was once provided by the mass media.

From the perspective of the information providers, especially people who want to create a positive image for a city, the Internet is a double-edged sword. On the positive side, the Internet is an economical and speedy medium to distribute information. On the negative side, the efficiency and promptness of the medium deprive a city of a good image once false information and rumors have been spread. Fortunately, all the information on the Internet in various forms, like statistics, facts, news, and stories, affects the image formation of a city as a whole. In other words, it takes time to shape the image of a city. Thus, given the double-sided nature of the Internet, a strategy is needed to use this medium as a tool for building a positive image of a city.

Tourism has successfully adopted the Internet as the medium of information distribution. Anthropologically, the structure of tourism consists of tourists, residents of destination, and observers (Selwyn, 1996). These three components of tourism activities exchange city images; the Internet enhances the efficiency of information exchange and helps a city define its own image. One of the effective and economical means for enhancing the image of the city is the Internet, although there are many ways to deliver the image of the city to consumers. Although the Internet has transformed the distribution and marketing of tourism products (Buhalis & Spada, 2000), research on the Internet as an image formation agent is not yet widespread (Choi, Lehto, & Morrison, 2007). However, researchers have confirmed that the Internet is a prominent medium in tourism

marketing (O'Connor & Murphy, 2004; Oh, Kim, & Shin, 2004). It also leverages the powerful influence of the Internet on image formation, and has become an important issue for tourism researchers (Govers & Go, 2005). Recently, tourism researchers have conducted content analysis of a large number of websites with quantitative methods (Choi et al., 2007; Stepchenkova & Morrison, 2006). As far as global city research is concerned, tourism marketing has also shown the importance of the Internet as a conduit for revealing the image of global cities.

3.2. The Web (the Internet)

According to Internet World Stats (2011), there are over two billion[5] Internet users (32.7% of the World population) in the World. The growth of the World total between 2000 and 2011 is 528.1%. Although Africa (2,988.4%), the Middle East (2,244.8%), and Latin America (1,205.1%) show rapid growth, the percentage of Internet users in these parts of the world remains small in relation to the total of Internet users (Africa, 6.2%; Middle East, 3.4%; Latin America, 10.4%); it is quite smaller than the percentages of Asia (44.8%), Europe (22.1%), and North America (12.0%). This statistic shows the inequity of the Internet infrastructure, but it also shows that the Internet is an important tool in contemporary societies. The Internet was previously only possible through wired lines with the telecommunication port of desktop computers and laptops. The development of wireless technology and the popularization of smart phones have accelerated *Homo Interneticus*[6] (Barnes, 2010). Searching, collecting, and distributing information on the Web have become the inherence of humans like reading, listening, speaking, and writing.

[5] The number is 2,267,233,742 as of December 31, 2011.
[6] This is the title of 4th episode in *The Virtual Revolution*, which is the BBC documentary discussing the huge benefit and unforeseen downsides of the World Wide Web.

The Web consists of websites and links between websites (i.e. external links, but there are internal links between webpages in a website). A website is a set of related webpages in various forms of contents: text, audio, image, video, etc. A webpage is a hypertext document connected to the Web, which is typically written in Hypertext Markup Language (HTML). A web browser interprets HTML and visualizes the HTML to a user. On the end user's side, a webpage is usually seen as a simple page that is displayed through a web browser. However, the recognition that a webpage has a complex composition (Table 3.1) renders the situation less straightforward. One viewed webpage contains various types of contents and information technology.

Table 3.1: Five major components of webpages (revised from Thelwall, 2004, pp. 17-18)

Component	Illustrative Cases
File format	• An electronic file validly encoded in the language of the Web, HTML • Any file type accessible through a modern Web browser including non-HTML formats such as plain text, PDF and Microsoft Word
Access mechanism	• Requests made using the official 'port number' of the Web, 80 • Requests made using the official computer request language of the Web, the HTTP • Requests made using any mechanism available to a modern Web browser, including common non-web protocols such as FTP
Scope	• Public webpages that are available to all web users • Public and private webpages, including password protected pages and Intranet and Extranet pages
Permanence	• Static resources only • All resources, including dynamically-created webpages such as search engine results pages
Compound pages	• A single file is a single webpage • Compound documents, such as those built up from separate files using the HTML frameset feature also count as one single page

As of March 2012, there were 644,275,754 sites; the number of sites has increased without interruption, except during the period between December 2008 and December 2009, the period of the global economic downturn (Netcraft, 2012). The rapid proliferation of websites is further evidence that people accept the Web as a part of their

lives whether their purposes are personal, commercial (business), government, and nonprofit organization websites.

While other media (e.g. guidebooks, brochures, advertisements on TV and newspapers, etc.) exhibit comparative disadvantages such as "passive communication, expensive to produce, difficult to monitor effectiveness, and message is often not heard" (Kolb, 2006, p. 239), websites help marketing providers use the Internet as the solution for these disadvantages. The website of business and nonprofit organizations has three purposes: "direct selling, sales support and customer service, and advertising and public relations" (Kolb, 2006, p. 273); and "the main purpose of a website for tourism office is to advertise the city's features and benefits" (Kolb, 2006, p. 273). Considering that websites are visited by people who are interested in specific information, a website that is easily accessed, well organized, and has newly updated information is the most desirable website for website visitors. In addition, the reason why people use the Internet for getting information is that materials and contents on the Internet are more thorough and richer than conventional promotional agents (Govers & Go, 2003; Heung, 2004). This is a good opportunity for website providers to enhance the image of the city as well as the satisfaction of website visitor's needs.

3.3. Branding, Image, and Global City

A 'brand' is the word used for a type of product manufactured by a particular company under a particular name. 'Branding' is the promotion of a particular product or company by means of advertising and distinctive design. According to Govers and Go (2009), "Place branding refers to branding and building brand equity in relation to national, regional, and/or local (or city) identity" (p. 16); and brand equity can be built

through “brand loyalty; name awareness; perceived quality; brand associations in addition to perceived quality; and other proprietary brand assets – trademarks, channel relationship” (p. 17). In short, city branding is intended to make people believe a city has the positive unique something. We can call it city marketing because it induces the economic activities of people. Marketing activities for places (or cities) are well defined by Blain et al. (2005) as:

> The marketing activities that (1) supporting the creation of a name, symbol, logo, word mark or other graphic that both *identifies* and *differentiates* the destination; (2) that convey the *promise* of a memorable travel *experience* that is uniquely associated with the destination; and (3) that serve to *consolidate* and *reinforce* the recollection of pleasurable *memories* of the destination experience, all with the intent purpose of creating an *image* that influences consumers’ decisions to visit the destination in question, as opposed to an *alternative* one. (pp. 331-332)

It is easy to think about place marketing (or selling) as a recent phenomenon. However, the marketing and promotion of towns and cities has existed since towns and cities have required the inflow of population and investment for its development. Before the advances of telecommunication and transportation, marketing information about a city took a long time to spread out. However, the development of technology has allowed much easier and faster delivery of information. From the era of the Frontier to the post-industrial city, the posters, the advertisement columns and pages of newspapers, and any type of advertising medium has sold the image of towns and cities as well as the catchphrase, which reflects the state of the period (Ward, 1998). For the development of towns and cities, imaging is important because the image and the perception of cities

become active components of economic success or failure (Ashworth & Voogd, 1990). In the age of global competition, countries, cities and regions require to market themselves; and they need the art of selling – good advertising (Anholt, 2010).

The starting point of city branding is to ascertain the identity of the city. Whether the city likes or dislikes this identity, they form the image of the city. More exactly, the interaction between identity and people makes the image. For marketing purposes, it is useful to know those images in order to set up a branding strategy. Information on the Web becomes important in this era of digital information. Text, audio, and visuals on the Web contribute to the image that people form of a destination (or any type of things). As the influence of online digital information on image formation has become an important issue (Govers & Go, 2004), information on the Web is important for branding city.

Branding departments of cities (e.g. visitors bureau, chamber of commerce, etc.) should be interested in the projected image for the city. However, it is not easy to achieve that a desired image is projected on the internet without a strategic approach. Govers and Go (2005) used pictures and text from tourism-related websites in Dubai for finding projected images of Dubai. They concluded that information provided by the web is fragmented, lacks creativity and coherence, and just offers limited products by few business sectors (i.e. only dining and shopping, no consumption of place). In addition, Choi et al. (2007) used the narrative and visual information on a sample of websites in Macau. They pointed out that the image of Macau projected online varies according to the information sources due to the different communication objectives and targeted audiences. Both studies show the need for a master plan, including identities of the city,

desired images, projected images, and solution for discrepancy. In any case, the priority is to know the identity of the city.

To develop a branding strategy, it is firstly required to analyze the city's identity. Although there are many branding case studies of cities, their diversity and fragmentation make it difficult to conclude on the shared image of global cities. In fact, the image of a global city is the mixture of global traits and the city's own identity. While it may be possible to induce the shared characteristics of globalization, it is hard to generalize each global city's identity. In addition, considering that the image of a city is how people perceive the city, the problem is how to generalize people's perception. The perceived image is not static, and it is altered by one's emotion, experience, and knowledge. Some research has tried to solve this issue by following an alternative research venue.

Instead of finding the common image of global cities directly from the generalization of people's perception, researchers often try to measure the city brand with relative rankings. In other words, they focus on the relative value of the city based on surveys. Representative research along this line is the Anholt-Gfk Roper City Brands Index (CBI, Gfk Custom Research North America, 2011). In this research, the image is equal to the brand power of the city. Anholt (2006) insisted that city's brand power can be analyzed through six components: the presence, the place, the potential, the pulse, the people, and the prerequisites (Table 3.2). The online survey was conducted among 15,255 people aged 18-64 from a wide range of income groups in 20 countries as well as the global panel interview (Anholt, 2006, p. 3-4). The rankings of 60 global cities are annually announced as CBI since 2005.

Table 3.2: Six Components of CBI (from Anholt, 2006, p. 3)

Components	Exploring	Asking
The Presence	How familiar people are with each city	• Whether they have actually visited them or not • What the cities are famous for • Whether each city has made an important contribution to the world in culture, science, or in the way cities are governed, during the last 30 years.
The Place	People's perceptions about the physical aspect of each city	• How pleasant or unpleasant they imagine it to be outdoors and to travel around the city • How beautiful it is • What the climate is like.
The Potential	The economics and educational opportunities that each city is believed to offer visitors, businesses and immigrants.	• How easy to find a job in the city • How good of a place to do business • Whether each city would be a good place to get a higher educational qualification
The Pulse	How exciting people think the cities are	• How easy they think it would be to find interesting things to do, both as a short-term visitor and a long-term resident.
The People	How about the people related issues	• Whether the inhabitants would be warm and friendly, or cold and prejudiced against outsiders • Where it would be easy to find and fit into a community that shares language and culture • How would it feel in the city
The Prerequisites	How people perceive the basic qualities of the city	• What they think it would be like to live there • How easy they think it would be to find satisfactory, affordable accommodation • What they believe the general standard of public amenities is like schools, hospitals, public transport, sports facilities, and so on.

The Saffron European City Brand Barometer (Hildreth, 2008) is another popular index based on surveys, which is similar to the CBI, but it only covers 72 European cities. However, the survey consists of questions similar to those of the CBI (Table 3.3). The broad categories are city asset strength and city brand strength. The first category measures how strong a city's brand could be; the second category asks how strong the city's brand is right now. Then, the X-Y plot of two categories visualizes the gap between the potential and the actual.

Table 3.3: Categories and factors of European City Brand Barometer (from Hildreth, 2008, p. 7-8)

Category	Factor	Asking (or Meaning)
City asset strength	Cultural	When considering a city for a break, which of the following things are most important to you?
	Amenity	If you were considering a city break, what kind of city might appeal to you the most?
City brand strength	Pictorial recognized	Many people could recognize the city from a postcard without having to read the description on the back
	Quantity / strength of positive / attractive qualities	What prompted and unprompted associations do people have of the city
	Conversational value	How interesting it would be at a cocktail party to say, "Hey, I just got back from A."
	Media recognition	Determined statistically by counting media references to the city over a set period

The above indices have two common characteristics. Firstly, these city brand indices are based on a survey instrument that prompts people for the relative ranking among study cities. That is, they rely on a kind of popularity vote based on the perception that respondents have of each city. Secondly, both indices utilize predefined categories regarded as the important dimensions to measure the strength of city brand. The second characteristic (i.e. the utilization of predefined categories) can be applied to analyze the text from webpages. In this case, the predefined categories can be the dimensions of global cities such as economic, political, cultural, and infrastructural characteristics. Instead of using a survey method, the set of related words for each dimension can be used for quantifying the text from each city's dataset.

3.4. Conceptualization of Image Extraction from Textual Contents

When we consider the integration of the global economy as one of the core drivers of globalization, the effort of corporations to globalize their business cannot be left out of the discussion. Since every organization is now global, it is important to have one solution, one price, one distribution structure, and one billing system (Schultz &

Kitchen, 2004, p. 361). Among different types of efforts, the corporate image needs to be updated for global recognition and cost savings by promoting all offerings under one corporate brand (Balmer, 2010; Erdogmus, Bodur, & Yilmaz, 2010; Hatch & Schultz, 2009). In other words, branding of the corporation and of its products is an inevitable consequence of globalization. Globalization makes cities also try to brand themselves for economic benefits. Successfully branded cities have an image like an attractive and pleasant place to live, work, travel, and invest. Successful branding guarantees their continuing economic success. Like the corporate entity is more important than the products and services it produces or sells (Mitchell, Agle, & Wood, 1997; Schultz & Kitchen, 2004), the image of a city is more important than what the city offers in reality.

What is required for successful city branding? According to Hatch and Schultz (2008), it is important for a corporation to have its organizational identity for successful branding, which can be achieved through the harmony of strategic vision, organizational culture, and stakeholder's images. In addition, they emphasized that organizational culture (inside) and stakeholder's images (outside) should be integrated under a strategic vision. If we adapt the management strategy of corporate branding to city branding initiatives, a comparative table can be produced (Table 3.4). City branding can be successful as long as a city has its own identity, which is affected by strategic vision, residents' culture, and external people's images. As a matter of course, the harmony of the three axes is important. Among the three axes, this research handles the external people's images. More specifically, it tries to analyze how the cities are discussed on the Web.

In this research, nodes (i.e. websites) contain city images in the form of texts. Hyperlinks among websites connect nodes pairwise. The connection between the local image of a city (i.e. residents' culture) and the global image of a city (i.e. external people's images) produces the final image of a city. Interaction among nodes increases the globality of nodes in a city through the addition of new hyperlinks to global webpages and the adoption of global contents which contain external people's images. Among the analyses of this research, hyperlink network analysis measures the effectiveness of these connections (i.e. faster and easier access) which is the infrastructure for city image formation. Quantified content analysis measures the status of image based on dimensions of globalization.

Table 3.4: Comparison between corporate branding and city branding

Core	Corporate Branding*	City Branding
Identity	Strategic Vision	Strategic Vision
	Organizational Culture	Residents' Culture
	Stakeholder's Images	External People's Images

* Source: Hatch & Schultz (2008)

CHAPTER IV: RESEARCH QUESTIONS AND CONCEPTUAL FRAMEWORK

Considering the trends in global city research and the recent development of information networks, the Internet is worth being closely considered as a relevant source of information. As people increasingly rely on the Internet as an information source due to its efficiency and accessibility, the Internet has taken on the critical and active role of building the image of the global city. Thus, it is befitting to study global cities and globality through the lens of the WWW so as to extend the concepts to new perspectives that may contribute to their redefinition.

What kind of data is best suited for global city research from the perspective of the Internet? The Internet network consists of webpages and hyperlinks. One possible conjecture is that a larger number of webpages pertain to more globalized cities and that these pages are better connected through hyperlinks. If we measure the connectivity among webpages relative to a certain city, we can use this measure to define a new typology of cities. Also, we can bring out the image of cities believed to have global status by profiling relevant webpages through content analysis. Thus, the research questions of this study are mainly organized in two parts, namely those pertaining to the linkages of hyperlinks and those pertaining to the contents of webpages. In other words, the first part consists in the analysis of the structure of hyperlink networks, and the second part is to quantify the contents from related webpages.

The structure of the research consists of four components (Figure 4.1). The first component is related to the general characteristics of each city's hyperlink network; the second component focuses on the remarkable nodes (i.e. websites) in the hyperlink network of each city. In other words, the questions in the first component focus on connectivity, while the questions in the second more specifically focus on the characteristics and classification of the nodes themselves. Methodologically, both components use hyperlink network analysis (HNA). It should be clarified that the unit of analysis for these components is a website and an external link, not a webpage and all links. The website is more meaningful as the unit of analysis than webpages because the website represents a group of webpages that belong to the website for a specific purpose. In other words, understanding the purpose of the website is easier than understanding webpages. In addition, many links between webpages are created primarily for navigation within the website and replicated for connecting to the portal service or advertisement (Thelwall, 2009). This fact supports why we focus on the external links. The third component is to find the characteristics of the text content of webpages by quantitative content analysis (QCA). The last component is synthetic in nature as it brings together the results of the HNA and the QCA.

Each of these components supports a conceptual expansion of globalization through the utilization of two different types of the Web data separately and together.

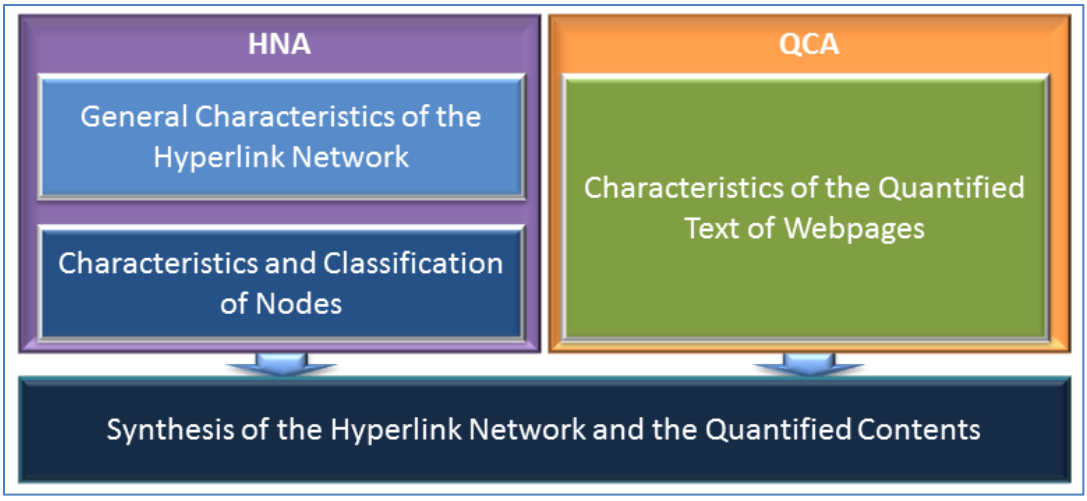

Figure 4.1: Research components

4.1. General Characteristics of Hyperlink Networks

The hyperlink network traces the flow of knowledge. In other words, it is the route for information exchange. When we consider that the image of a global city is formed through information exchange, it is important to know the structural characteristics of the network. Specifically, a city with a well (or tightly)-connected hyperlink network is also a city that exchanges its information efficiently. This efficiency could be a characteristic of global city. This conceptual approach is a novel way to conceive of the web data in the context of global city research.

In this research, the hyperlink network of each city is defined as a set of nodes and edges, where the nodes are the websites and the edges are the external hyperlinks between the webpages. The edges are directional because we have information where the hyperlinks point from. The directional connectivity of city networks is different on a city by city basis. In other words, the hyperlink network (i.e. the set of nodes and directional edges) of each city is different according to how the nodes and edges are organized (i.e. how they are topologically connected). Figure 4.2 shows conceptual examples of the

hyperlink networks for this research. If we measure the connectivity of each city's hyperlink network, we can depict the city in a way that adds a hitherto unexplored perspective and therefore enhance our comprehension of the multidimensionality of global cities. Thus, the questions in this component aim to differentiate cities on the basis of the average connectivity of their hyperlink-generated network. In addition, it aims to confirm that this characteristic (the average connectivity of the hyperlink network) can be used for indexing global cities by conducting a comparison to a global city list established by other researchers.

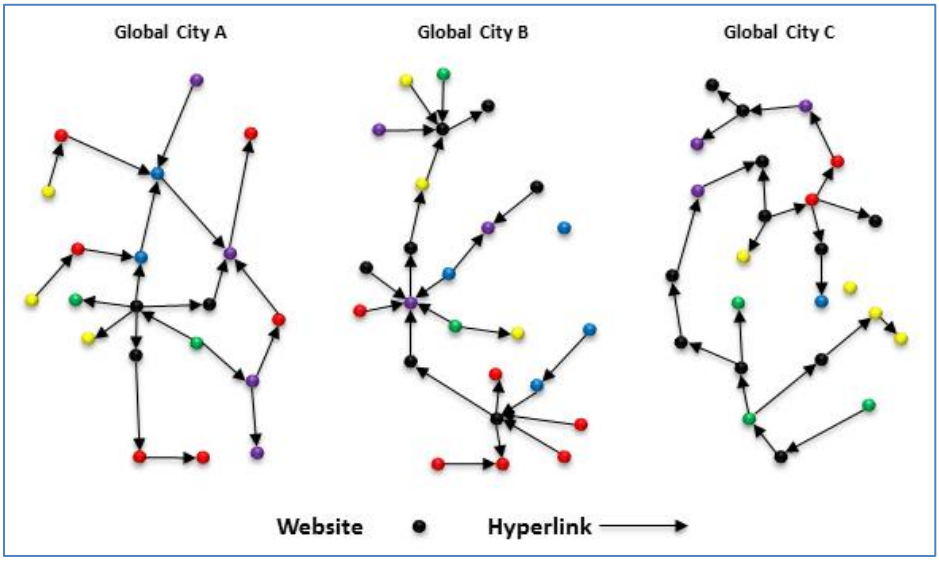

Figure 4.2: Conceptual examples of the hyperlink networks

Note: Nodes are color-coded to represent the value (degree or centrality) and the type (domain) of the nodes for a specific purpose.

Once the hyperlink network of each global city has been defined, we can ask the following questions:

1) *Which cities have a hyperlink network system with a higher (or lower) connectivity?*

   There are several methods for measuring network connectivity. The

connectivity of the hyperlink network of each city will be calculated and systematically compared between methods.

2) *Does city network connectivity support defining a typology of global cities? If it does, what is the classification based on the measures of connectivity?*

   The results of the first sub-question are used for answering this question. Through this question, we can establish the suitability of network connectivity as a criterion for the classification of cities according to their degree of globality.

3) *Does the distribution of global cities according to hyperlinking connectivity parallel that obtained on the basis of tangible relationships and measurements? If discrepancies happen, what can they be attributed to? (that is, what do conventional indices overlook?)*

   This question involves a comparison between a classification of cities based of hyperlink network measures on the one hand, and a classification based on a more conventional approach. As discussed before, existing research has used total counts of population, global companies, immigrants, or are based on linkages and relationships such as headquarter to branches relations, air passenger flows, etc. Our comparison will be with the Global Cities Index (Kearney, 2012).

4.2. Characteristics and Classification of Nodes

While the first component of the research focuses on the characteristics of the whole hyperlink network of each city, this second component focuses on the characteristics and typology of the remarkable nodes (i.e., websites) in the city hyperlink

network. As the number of links incident at each node is different, each node has a distinctive value. In other words, if one node receives many incoming links (i.e. pointed by other nodes), it has a high chance to have valuable information (i.e. important location in the network). The importance of the node in the hyperlink network can be measured by the centrality of the node. Through the centrality of nodes, we can identify the premium nodes in the hyperlink network. Moreover, we can find the characteristics of those nodes as well as the share of nodes in the hyperlink network.

After calculating the centrality of each node in each hyperlink network, the following questions deal with more specific characterizations of network components:

1) *What kind of nodes shows higher centrality?*

   The nodes that have higher centrality are the hubs of the hyperlink network. The hub in the network plays a critical role for the distribution of information. The hubs are helpful to find the 'globalizer' of information network. In other words, these hubs can be used for strategic points to expand a city's information network. To find the basic information (e.g. type of domain) is helpful to understand the characteristics of the node.

2) *What are the network characteristics of these premium nodes?*

   The network characteristics are also important because this information helps to understand the role of premium nodes through the comparison to one of the hyperlink networks of other cities. In other words, the common network characteristics of the premium nodes (e.g. degree of the premium node, the distance to other nodes and other premium nodes) help us generalize the

characteristics of the premium nodes in the hyperlink networks of global cities.

3) *Do specific websites or domains dominate the city hyperlink network? If yes, what is the share of the premium nodes? Is it highly concentrated or competitive?*

   The Internet provides users with useful information and also constitutes a virtual space well suited for business. Arousing users' interest in the webpage is directly connected with the sales of dot-coms. Thus, it is meaningful to find that each hyperlink network is information-centric or business-centric or both.

4.3. Characteristics of the Quantified Text of Webpages

Figure 4.3 shows the process of conceptualization in this component. This conceptualization is based on the definition of globalization (i.e. economic integration) and the importance of city branding. The importance of branding of a city is derived from the theory of corporate branding in the academic field of marketing. Successful branding of a city (i.e. intangible asset) increases the value of the city, and it builds up city competitiveness not only against domestic cities but also foreign cities. One of the important axes of successful branding is how a city is projected to the outside, that is to say external people's image (in terms of cities, table 3.4).

When we consider that the Web is the most popular space to share images nowadays, the quantified textual contents of the Web can give a semantic meaning to the images of each city in relation to the recognized dimensions of globality. The dimensions that collectively constitute the concept of globalization (i.e. categories) are individually used to guide the filtering of textual contents from the Web. Based on the thematic

definitions (i.e. the dimensions) of globalization, the match between each dimension and textual data can help to find how each dimension is projected on the textual contents of the Web. In other words, the result provides the closeness of each city to each dimension of globality. In turn, this can be used for measuring the similarity among cities. In addition, it can be used as the basis to set up the strategy to enhance the global image of a city.

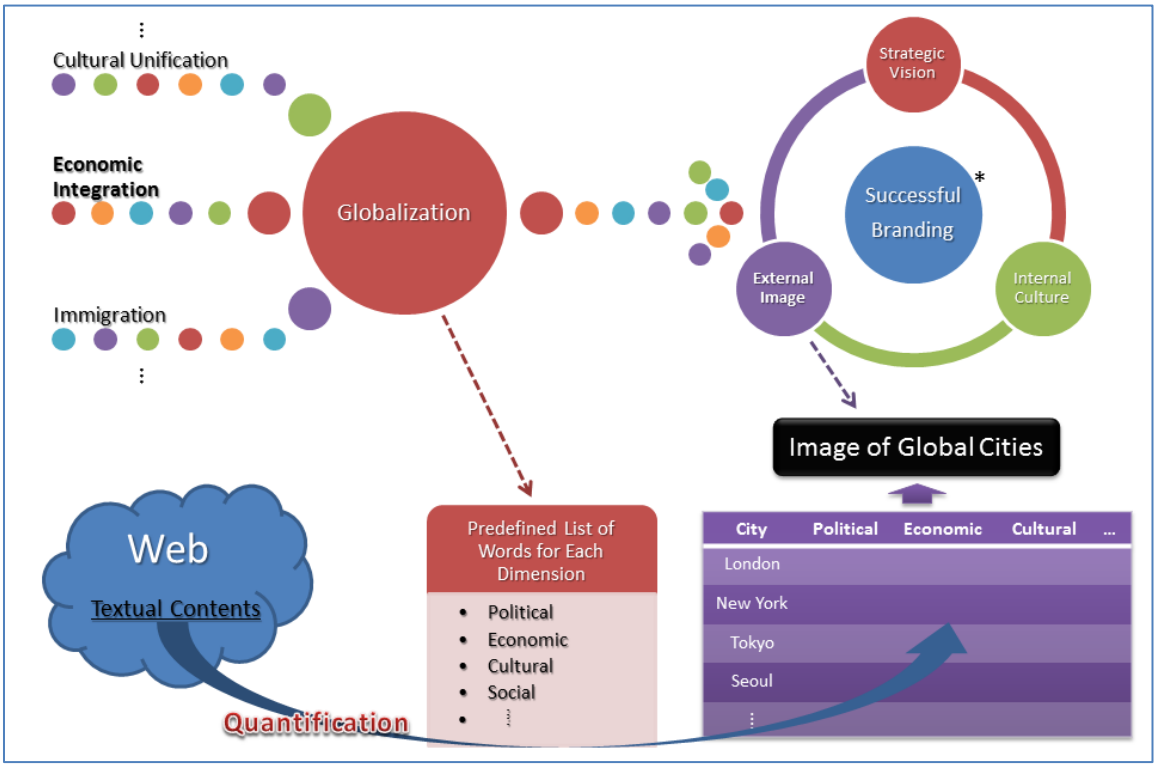

Figure 4.3: Conceptualization of image extraction from textual contents on the Web

Note: Successful branding is adapted from Hatch & Schultz (2008).

The text can be the real description or experience of the city or possibly the aspired image, like advertisements. While other types of media in the webpages are hard to analyze by quantification, textual content is relatively easy to process. The processing of the text content involves tokenization, parsing, stemming, and lemmatization.

After these procedures we have a set of keywords for each city. These sets are used for the calculation of frequency and categorization, which are based on a predefined code scheme. The following questions drive our research agenda here:

1) *What keywords can be used to construct a predefined code scheme of the latent dimensions of globality?*

   This research question tries to provide keyword references for the predefined dimensions of global city. The general definitions of global city can be collected from different sources such as Wikipedia, the extant academic literature, dictionaries, and other research references. These sources identify the commonly recognized dimensions of global cities. The expected keyword list for each dimension (e.g. 'company' for economical dimension, 'school' for cultural dimension, and 'democracy' for political dimension) will be provided for measuring how many keywords belong in each dimension. That is, these keywords become the markers of each dimension.

2) *Is there a difference in the usage (or the frequency) of the keywords between the documents pertaining to different cities? If yes, what is the classification of cities based on the difference of the frequencies?*

   This research question entails the classification of cities based on the frequency of the filtered keywords of each city. The predefined code scheme helps filter the more relevant keywords from massive amounts of textual data. The frequency of these filtered keywords is the textual profile of each global city. The results of this question help characterize global cities in the context of textual data.

4.4. Synthesis of the Hyperlink Network and the Quantified Contents

While the previous components dealt with the webpage data individually based on the type of the Web data, the last component pertains to the synthesis of the hyperlink network with the quantified contents. In other words, this component integrates the results from the first component (i.e. the distribution of global cities based on general typological characteristics) and from the third component (i.e. the distribution of cities based on predefined global city code schemes). Methods like multidimensional scaling (MDS) or self-organizing map (SOM) can be used for visualizing the common structures exhibited by cities based on a large number of indicators that are reduced a more manageable set of components. The synthesis of typological characteristics with textual characteristics provides an integrative view based on the whole data of the webpages. It can be used for assess the usefulness of the hyperlink approach as a new and complementary perspective in global city research. The following questions are examined:

1) *What is the distributional classification of cities based on both typological characteristics and textual characteristics?*

   This question tries to reveal the distributional characteristics of cities based on both typological characteristics and textual characteristics. The visualization can include the distribution of global cities based on typological characteristics, textual characteristics, and both.

2) *How does this new approach contribute to global city research with respect to the analysis of the Web data?*

The answer articulates the structure (i.e. hyperlink network) and the content (i.e. text) of the Web in a composite way. Through a new measurement to compare the textual distribution to the structural distribution, we can discover the difference and the commonality of two distributions. This comparison informs on how hyperlink networks are related with global city context as well as the validation of the hyperlink network for global city research.

CHAPTER V: RESEARCH DESIGN

This chapter consists of several sections dealing with cities under study, the data, and methodologies for each research objective. First, the section on the cities under study explains the selection process of global cities, which is the basis for collecting URLs. Second, the data section contains the linguistic characteristics of the data, the explanation of search engines and web crawlers, and pre-processing procedures. Lastly, the methodology section contains the basic description of the methods used in this research. Specific methods are used for each research objective. Figure 5.1 describes the relationship between research objectives and methods.

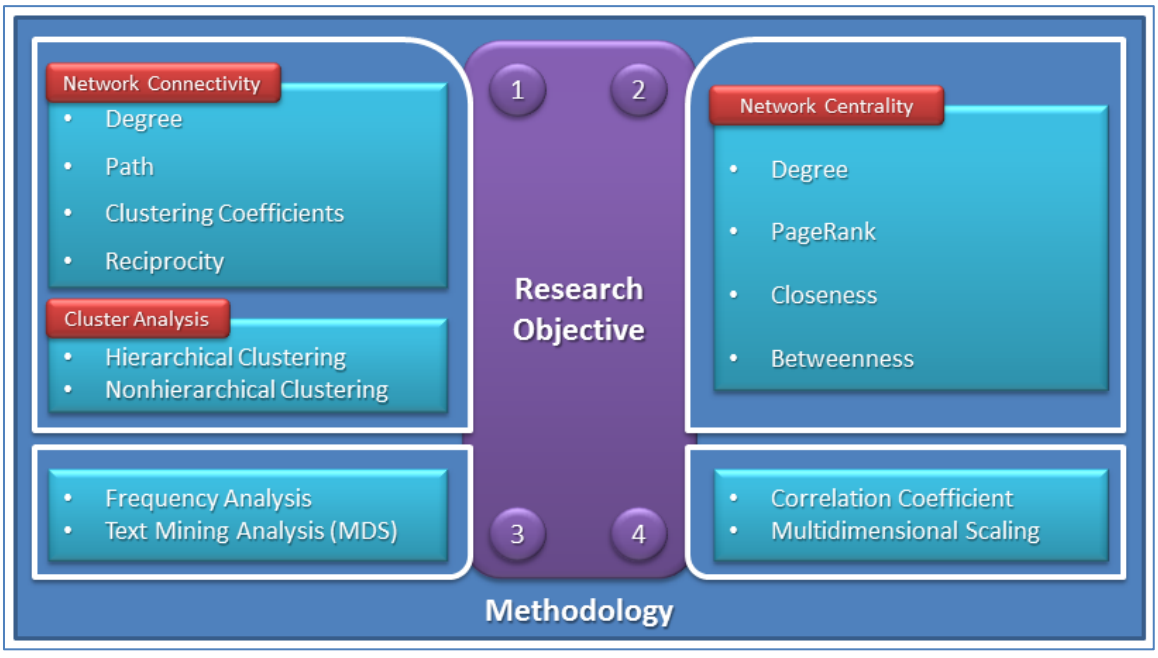

Figure 5.1: Research objectives and methodologies

Note: Methods in network connectivity and centrality can be used for both research objectives 1 and 2.

5.1. Cities under Study

Selecting study cities for globalization research is not a trivial task. Data availability has often constrained researchers in their selection of cities just as it has limited the breadth of global city research. A practical way to choose study cities is to use the list of global cities used by previous researchers. Although their definitions, data, and methodologies are diverse, a large corpus of studies that complement each other has accumulated. In other words, taken collectively, global city studies contain the various areas of investigation that focus on global cities.

The number of webpages discovered for each city could be used to identify cities to study; however, it is very hard to count the 'real' total number of webpages that relate with a study city. We cannot measure the exact number of webpages on a certain city because the Web changes constantly, and we can only roughly estimate the number of the webpages. Thus, following a comprehensive review of the literature on global cities, we have compiled a list of 264 cities that have been labeled as global cities by two studies at least, including statistics from the United Nations (Table 5.1). Relying on a list of global cities from previous studies brings to bear considerations of historical trend and coverage in global city research.

Table 5.1: Study cities from past global city research (Cities in bold are selected for this research)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Aarhus	Denmark																×									1
Aberdeen	United Kingdom																×									1
Abidjan	Cote d'lvoire																×						×	×		3
Abu Dhabi	United Arab Emirates																×						×		×	3
Accra	Ghana																×						×			2
Addis Ababa	Ethiopia																×						×			2
Adelaide	Australia																×						×			2
Ahmadabad	India																×						×	×	×	4
Aleppo	Syria																						×	×		2
Alexandria	Egypt																×						×	×	×	4
Algiers	Algeria																×						×			2
Almaty	Kazakhstan																×						×		×	3
Amman	Jordan																×						×			2
Amsterdam	Netherlands		×	×	×	×	×	×	×				×	×	×	×	×	×	×	×	×	×	×		×	19
Ankara	Turkey																×						×	×	×	4
Antwerp	Belgium																×						×			2
Asunción	Paraguay																×						×			2
Athens	Greece																×		×				×	×	×	5
Atlanta	United States, GA			×			×									×	×	×	×	×	×	×	×	×	×	12
Auckland	New Zealand																×						×		×	3
Bandar Seri Begawan	Brunei																×									1
Baghdad	Iraq																×						×	×		3
Bahrain	Bahrain													×												1
Baku	Azerbaijan																×						×			2
Baltimore	United States, MD																×						×			2
Bandung	Indonesia																×						×		×	3
Bangalore	India																×		×		×	×	×	×	×	7
Bangkok	Thailand	×			×											×	×	×	×	×	×	×	×	×	×	12

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Barcelona	Spain		×						×							×	×	×	×	×		×	×	×	×	11
Basel	Switzerland													×			×									2
Batam	Indonesia																×						×			2
Beijing	China						×			×					×	×	×	×	×	×	×	×	×	×	×	13
Beirut	Lebanon																×		×				×		×	4
Belfast	United Kingdom																×									1
Belgrade	Serbia																×						×			2
Belo Horizonte	Brazil																×						×	×	×	4
Bergen	Norway																×									1
Berlin	Germany												×			×	×	×	×	×	×	×	×	×	×	11
Bern	Switzerland																×									1
Bilbao	Spain																×									1
Birmingham	United Kingdom																×						×		×	3
Bogotá	Colombia																×		×		×	×	×	×	×	7
Bologna	Italy																×									1
Bonn	Germany									×			×				×									3
Bordeaux	France																×						×			2
Boston	United States, MA		×												×	×	×	×	×	×	×	×	×	×	×	12
Brasília	Brazil																×						×	×		3
Bratislava	Slovakia																×									1
Brazzaville	Congo (Rep. of)																×						×			2
Brisbane	Australia																×						×			2
Bristol	United Kingdom																×									1
Brussels	Belgium	×		×			×	×					×	×	×	×	×	×	×	×	×	×	×		×	16
Bucharest	Romania																×						×		×	3
Budapest	Hungary															×	×	×	×	×			×		×	7
Buenos Aires	Argentina	×			×									×		×	×	×	×	×	×	×	×	×	×	13
Buffalo	United States, NY																×						×			2

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Bulawayo	Zimbabwe																×									1
Busan	Korea (Rep. of)																×						×	×	×	4
Cairo	Egypt				×												×		×		×	×	×	×	×	8
Calgary	Canada																×						×			2
Canberra	Australia																×									1
Cape Town	South Africa								×								×						×	×	×	5
Caracas	Venezuela	×														×	×	×	×	×	×	×	×	×		10
Cardiff	United Kingdom																×									1
Casablanca	Morocco																×						×	×		3
Changchun	China, Jilin																						×	×		2
Changsha	China, Hunan																						×	×		2
Charlotte	United States, NC														×		×						×			3
Chengdu	China, Sichuan																		×				×	×	×	4
Chennai	India																×						×	×	×	4
Chicago	United States, IL	×	×	×			×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	21
Chittagong	Bangladesh																						×	×		2
Chongqing	China																		×		×	×	×	×	×	6
Christchurch	New Zealand																×									1
Cincinnati	United States, OH																×						×			2
Ciudad Juarez	Mexico																×						×			2
Cleveland	United States, OH																×						×			2
Cologne	Germany		×	×							×						×						×			5
Colombo	Sri Lanka																×								×	2
Columbus	United States, OH																×						×			2
Conakry	Guinea																×						×			2
Copenhagen	Denmark								×				×			×	×	×	×	×	×	×	×		×	11
Curitiba	Brazil																×						×	×		3
Dakar	Senegal																×						×			2

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Dalian	China, Liaoning																×						×	×	×	4
Dallas	United States, TX			×								×				×	×	×	×	×			×	×	×	10
Damascus	Syria																×						×			2
Dar es Salaam	Tanzania																×						×	×		3
Denver	United States, CO																×						×			2
Detroit	United States, MI														×		×						×	×		4
Dhaka	Bangladesh																×				×	×	×	×	×	6
Djibouti	Djibouti																×									1
Doha	Qatar																×								×	2
Dongguan	China, Guangdong																						×	×		2
Dortmund	Germany																×									1
Douala	Cameroon																×						×			2
Dresden	Germany																×									1
Dubai	United Arab Emirates																×		×		×	×	×		×	6
Dublin	Ireland																×		×		×	×	×		×	6
Durban	South Africa																×						×		×	3
Düsseldorf	Germany		×	×				×	×		×			×	×	×	×	×	×	×						12
East Rand	South Africa																						×	×		2
Edinburgh	United Kingdom																×		×							2
Edmonton	Canada																×						×			2
Essen	Germany																×									1
Fortaleza	Brazil																						×	×		2
Foshan	China, Guangdong																						×	×		2
Frankfurt	Germany	×	×	×	×	×	×	×	×		×		×	×	×	×	×	×	×	×	×	×			×	20
Freetown	Sierra Leone																×						×			2
Fukuoka	Japan																								×	1
Gaborone	Botswana																×									1
Geneva	Switzerland							×	×							×	×	×	×	×		×			×	9

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Genoa	Italy																×									1
Georgetown	Guyana																×									1
Glasgow	United Kingdom																×						×			2
Gothenburg	Sweden																×									1
Grenoble	France																×									1
Guadalajara	Mexico																×						×	×	×	4
Guangzhou	China, Guangdong																×				×	×	×	×	×	6
Guatemala City	Guatemala																×						×			2
Guayaquil	Ecuador																×						×			2
Harbin	China, Heilongjiang																						×	×		2
Hamburg	Germany			×				×						×		×	×	×	×	×			×		×	10
Hamilton	Canada																×									1
Hangzhou	China, Zhejiang																						×	×	×	3
Hannover	Germany																×									1
Hanoi	Vietnam																×						×		×	3
Harare	Zimbabwe																×						×			2
Hartford	United States, CT			×													×						×			3
Havana	Cuba																×						×			2
Helsinki	Finland																×						×			2
Ho Chi Minh City	Vietnam																×				×	×	×	×	×	6
Hobart	Australia																×									1
Hong Kong	China	×	×		×	×	×	×	×			×	×	×		×	×	×	×	×	×	×	×	×	×	20
Honolulu	United States, HI											×					×						×			3
Houston	United States, TX	×	×	×				×								×	×	×	×	×		×	×	×	×	13
Hyderabad	India																×						×	×	×	4
Incheon	Korea (Rep. of)																								×	1
Indianapolis	United States, IN																×						×			2
Islamabad	Pakistan																×						×			2

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Istanbul	Turkey								×							×	×	×	×	×	×	×	×	×	×	11
Jaipur	India																×						×	×		3
Jakarta	Indonesia								×							×	×	×	×	×	×	×	×	×	×	11
Jeddah	Saudi Arabia																×						×	×		3
Jerusalem	Israel																×						×			2
Jinan	China, Shandong																						×	×		2
Johannesburg	South Africa	×			×		×		×							×	×	×	×	×	×	×	×	×	×	14
Kabul	Afghanistan																×						×	×		3
Kampala	Uganda																×						×			2
Kano	Nigeria																						×	×		2
Kansas City	United States, MO																×						×			2
Karachi	Pakistan																×				×	×	×	×	×	6
Kawasaki	Japan																×									1
Khartoum	Sudan																×						×	×		3
Kiev	Ukraine																×						×		×	3
Kingston	Jamaica																×									1
Kinshasa	Congo (Dem. Rep. of)																×						×	×		3
Kobe	Japan													×			×						×			3
Kolkata (Calcutta)	India																×				×	×	×	×	×	6
Krakow	Poland																×								×	2
Kuala Lumpur	Malaysia						×		×							×	×	×	×	×	×	×	×		×	11
Kunming	China, Yunnan																						×	×		2
Kuwait City	Kuwait																×						×		×	3
Kyoto	Japan																×						×			2
La Paz	Bolivia																×						×			2
Labuan	Malaysia																×									1
Lagos	Nigeria																×				×	×	×	×	×	6
Lahore	Pakistan																×						×	×		3

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Las Vegas	United States, NV																×						×			2
Lausanne	Switzerland																×									1
Leeds	United Kingdom																×									1
Leipzig	Germany																×									1
Liége	Belgium																×									1
Lille	France																×						×			2
Lima	Peru																×						×	×	×	4
Limasol	Cyprus																×									1
Linz	Austria																×									1
Lisbon	Portugal												×				×		×				×		×	5
Liverpool	United Kingdom																×						×			2
Ljubljana	Slovenia																×									1
Lomé	Togo																×									1
London	United Kingdom	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	24
Los Angeles	United States, CA	×	×	×	×	×	×	×	×			×		×	×	×	×	×	×	×	×	×	×	×	×	21
Luanda	Angola																×						×	×		3
Lucknow	India																×						×			2
Lusaka	Zambia																×						×			2
Luxembourg	Luxembourg													×			×									2
Lyons	France		×						×								×						×			4
Macau	China																×									1
Madrid	Spain	×	×	×									×	×	×	×	×	×	×	×	×	×	×	×	×	16
Mainz	Germany																×									1
Malacca	Malaysia																×									1
Malmö	Sweden																×									1
Managua	Nicaragua																×						×			2
Manama	Bahrain																×									1
Manaus	Brazil																×						×			2

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Manchester	United Kingdom																×						×			2
Manila	Philippines	×						×								×	×	×	×	×	×	×	×	×	×	12
Mannheim	Germany																×									1
Maputo	Mozambique																×						×			2
Marseille	France																×						×			2
Medan	Indonesia																×						×			2
Medellin	Colombia																×						×	×	×	4
Melbourne	Australia							×							×	×	×	×	×	×			×	×	×	10
Mexico City	Mexico	×	×			×	×							×		×	×	×	×	×	×	×	×	×	×	15
Miami	United States, FL	×	×		×	×			×			×				×	×	×	×	×	×	×	×	×	×	16
Milan	Italy	×	×	×			×	×					×	×	×	×	×	×	×	×	×	×	×		×	17
Minneapolis	United States, MN														×	×	×	×		×			×			6
Minsk	Belarus																×						×			2
Mombasa	Kenya																×						×			2
Monaco	Principality of Monaco																								×	1
Monrovia	Liberia																×						×			2
Monterrey	Mexico																×						×	×	×	4
Montevideo	Uruguay																×						×			2
Montreal	Canada		×	×				×	×					×	×	×	×	×	×	×		×	×	×	×	15
Moscow	Russia				×		×			×						×	×	×	×	×	×	×	×	×	×	13
Mumbai	India				×									×			×		×		×	×	×	×	×	9
Munich	Germany		×	×					×						×	×	×	×	×	×	×	×	×			12
Muscat	Oman																								×	1
Nagoya	Japan			×													×						×	×	×	5
Nairobi	Kenya																×					×	×	×	×	5
Nanjing	China, Jiangsu																×						×	×		3
Naples	Italy																×						×			2
Nassau	Bahamas																×									1

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
New Delhi	India																×		×		×	×	×	×	×	7
New Orleans	United States, LA																×						×			2
New York	United States, NY	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	24
Newcastle	United Kingdom																×						×			2
Nicosia	Cyprus																×									1
Norwich	United Kingdom																×									1
Nottingham	United Kingdom																×									1
Nuremberg	Germany																×									1
Omaha	United States, NE																×						×			2
Osaka	Japan		×	×					×		×			×	×	×	×	×	×	×	×	×	×	×	×	16
Oslo	Norway																×						×		×	3
Ottawa	Canada																×						×			2
Palermo	Italy																×						×			2
Palo Alto	United States, CA																×									1
Panama City	Panama													×			×						×		×	4
Paris	France	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	24
Penang	Malaysia																×									1
Perth	Australia																×						×			2
Philadelphia	United States, PA			×				×									×		×				×	×	×	7
Phoenix	United States, AZ																×						×	×		3
Pittsburgh	United States, PA																×						×			2
Plymouth	United Kingdom																×									1
Port Louis	Mauritius																×									1
Port Moresby	Papua New Guinea																×									1
Port of Spain	Trinidad and Tobago																×									1
Port-au-Prince	Haiti																×						×			2
Portland	United States, OR							×									×						×			3
Porto Alegre	Brazil																×						×	×	×	4

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Prague	Czech Republic															×	×	×	×	×			×		×	7
Pretoria	South Africa																×						×			2
Pune	India																						×	×	×	3
Pyongyang	Korea (D. P. R. of)																×						×			2
Qingdao	China, Shandong																						×	×	×	3
Quebec City	Canada																×									1
Quito	Ecuador																×						×			2
Rabat	Morocco																×						×			2
Rawalpindi	Pakistan																×						×			2
Recife	Brazil																×						×	×		3
Reykjavik	Iceland																×									1
Richmond	United States, VA																×						×			2
Riga	Latvia																×									1
Rio de Janeiro	Brazil	×			×									×			×		×		×	×	×	×	×	10
Riyadh	Saudi Arabia																×		×				×	×	×	5
Rochester	United States, NY																×						×			2
Rome	Italy			×									×	×	×	×	×	×	×	×	×	×	×	×	×	14
Rotterdam	Netherlands	×							×								×			×			×			5
Ruwi	Oman																×									1
Sacramento	United States, CA																×						×			2
Saint Petersburg	Russia																×		×				×	×	×	5
Salvador	Brazil																×						×	×		3
San Diego	United States, CA																×						×	×		3
San Francisco	United States, CA	×	×	×			×	×				×		×	×	×	×	×	×	×	×	×	×	×	×	18
San Jose	United States, CA																×						×			2
San José	Costa Rica																×						×			2
San Salvador	El Salvador																×						×			2
Sanaa	Yemen																×									1

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Santiago	Chile							×								×	×	×	×	×			×	×	×	9
Santo Domingo	Dominican Republic																×						×			2
São Paulo	Brazil	×	×		×	×	×	×	×					×	×	×	×	×	×	×	×	×	×	×	×	19
Sarajevo	Bosnia and Herzegovina																×									1
Seattle	United States, WA		×					×									×						×	×	×	6
Seoul	Korea (Rep. of)	×	×					×						×	×	×	×	×	×	×	×	×	×	×	×	15
Seville	Spain																×									1
Shanghai	China								×							×	×	×	×	×	×	×	×	×	×	11
Shantou	China, Guangdong																						×	×		2
Sheffield	United Kingdom																×									1
Shenyang	China, Liaoning																						×	×		2
Shenzhen	China, Guangdong																×		×		×	×	×	×	×	7
Singapore	Singapore	×	×		×		×	×	×			×		×		×	×	×	×	×	×	×	×	×	×	18
Sofia	Bulgaria																×						×			2
Southampton	United Kingdom																×									1
St. Louis	United States, MO																×						×			2
Stockholm	Sweden			×			×	×								×	×	×	×	×	×	×	×		×	12
Strasbourg	France																×									1
Stuttgart	Germany								×						×		×									3
Surabaya	Indonesia																								×	1
Surat	India																						×	×		2
Suva	Fiji																×									1
Suzhou	China, Jiangsu																						×	×	×	3
Sydney	Australia	×	×		×	×	×	×	×			×		×	×	×	×	×	×	×	×	×	×	×	×	20
Taipei	China (Rep. of; Taiwan)	×												×		×	×	×	×	×	×	×	×		×	11
Taiyuan	China, Shanxi																						×	×		2
Tallinn	Estonia																×									1
Tampa	United States, FL																×						×			2

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Tashkent	Uzbekistan																×						×			2
Tbilisi	Georgia																×						×			2
Tegucigalpa	Honduras																×						×			2
Tehran	Iran																×						×	×	×	4
Tel Aviv	Israel						×										×		×		×	×	×	×	×	8
The Hague	Netherlands														×		×									2
Tianjin	China																×						×	×	×	4
Tijuana	Mexico																×						×			2
Tirana	Albania																×									1
Tokyo	Japan	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	24
Toronto	Canada	×	×	×		×	×	×	×					×	×	×	×	×	×	×	×	×	×	×	×	19
Trieste	Italy																×									1
Tripoli	Libya																×						×			2
Tunis	Tunisia																×						×			2
Turin	Italy																×						×			2
Ulan Bator	Mongolia																×						×			2
Utrecht	Netherlands																×									1
Valencia	Spain																×						×			2
Vancouver	Canada		×					×	×								×		×				×		×	7
Venice	Italy																×									1
Vienna	Austria	×						×						×			×		×		×	×	×		×	9
Vilnius	Lithuania																×									1
Warsaw	Poland															×	×	×	×	×			×		×	7
Washington	United States, DC			×												×	×	×	×	×	×	×	×	×	×	11
Wellington	New Zealand																×									1
Wilmington	United States, NC																×									1
Windhoek	Namibia																×									1
Winnipeg	Canada																×									1

Table 5.1: (continued)

City	Country	Source																								Total
		F1	F2	PK	DK	SS	FG	NO	RP	HY	CO	TH	LP	RE	SK	BS	GW	PB	MI	KA	K1	K2	UN	U3	EU	
Wuhan	China, Hubei																						×	×		2
Wuxi	China, Jiangsu																						×	×		2
Xi'an	China, Shaanxi																						×	×		2
Xiamen	China, Fujian																×						×			2
Yangon	Myanmar																×						×	×		3
Yaoundé	Cameroon																×						×			2
Yerevan	Armenia																×						×			2
Yokohama	Japan																×									1
Zagreb	Croatia																×									1
Zhengzhou	China, Henan																						×	×		2
Zurich	Switzerland	×	×	×	×	×	×	×	×		×		×	×	×	×	×	×	×	×	×	×	×		×	21

Sources:
F1: Friedmann (1986, table 1)
F2: Friedmann (1995, table 2.1)
PK: Knox (1995, figure 1.1)
DK: Keeling (1995, table 7.1)
SS: Sassen (1994, from chapters 1 and 2)
FG: Finnie (1998), adapted in Graham (1999, Figure 1)
NO: Nomura (in Rimmer, 1991, figure 4.1)
RP: Petrella (1995, p. 21)
HY: Hymer (1972, p. 124)
CO: Cohen (1981, p. 308)
TH: Thrift (1999, p. 70)
LP: London Planning Advisory Council (1991, figure 1.2)
RE: Reed (1981, pp. 59-60)
SK: Short & Kim (1999, table 3.10)
BS: Beaverstock, Smith, & Taylor (1999, table 2)
GW: Taylor & Catalano (2000, GaWC data set 11)
PB: Taylor, Walker, & Beaverstock (2002, table 1)
MI: Mastercard (2008, pp. 20-21)
KA: Knox, Agnew, & McCarthy (2008, figure 7.4)
K1: Kearney (2008, p. 65; 2010, figure 1)
K2: Kearney (2010, figure 1)
UN: UN DESA (2012, file 12)
U3: UN DESA (2012, file 12, over 3 millions)
EU: EIU (2012, appendix 1)

The distribution of cities retained for study across world regions[7] shows interesting trends (Figure 5.2). Among early contributing studies, the world regional complexion of global cities was quite Eurocentric and North America-centric. Although the geographic diversity increased over time, the tendency of Eurocentric and North America-centric selection of global cities remained strong until 2000. Since 2002, the

[7] Global cities from all published research, except the statistics from the United Nations, which is used for the selection of study cities in this research, are reclassified based on the United Nations country grouping from the list of the Internet World Stats (http://www.internetworldstats.com/list1.htm; retrieved on August 27, 2012).

proportion of global cities found in Europe and North America has decreased and, conversely, the percentage of global cities in Asia has increased. In contrast to earlier studies (Son, 2006a; Taylor, 2004), the number of Asian cities has increased with studies published in recent years. This parallels the accelerated insertion of Asia in global economic, social and cultural trends, and the incorporation of Asian cities to global city networks. This trend corresponds to the forecast that Asian cities have high potential for global influence (Hales & Pena, 2012).

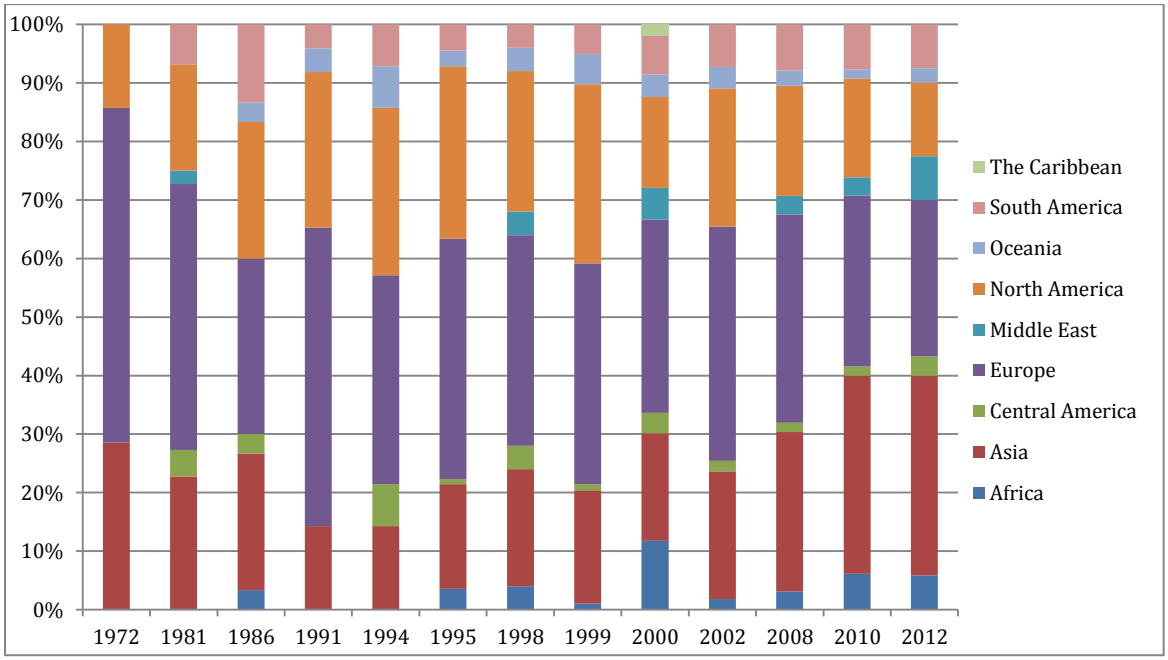

Figure 5.2: Trend of world regional composition of global city research

Note: Number of global cities mentioned in the following pieces of research. The average was used for calculating the number of study cities if there is more than one publication in the same year.

Sources: Hymer (1972), Cohen (1981), Reed (1981), Friedmann (1986), Rimmer (1991), LPAC (1991), Sassen (1994), Friedmann (1995), Knox (1995), Keeling (1995), Petrella (1995), Finnie (1998), Thrift (1999), Short & Kim (1999), Beaverstock et al. (1999), Taylor & Catalano (2000), Taylor et al. (2002), Mastercard (2008), Knox et al. (2008), Kearney (2008, 2010), EIU (2012).

The final set of study cities used in this research has more world regional diversity than previous research has considered. 73 Asian cities (28% of total) are selected for this study. The proportion of Asian cities is larger than that of North American cities (44

cities, 17%) and European cities (58 cities, 22%) taken separately. However, the combination (102 cities, 39%) of European and North American cities still exceeds the proportion of Asia. The combination of Asian cities and the Middle Eastern cities tallies 87 (33%); and this combination is similar to the combination of North American, Central American, South American, and Caribbean cities (79, 30%). In contrast to Asian cities, cities in Africa (34, 13%) and Oceania (6, 2%) still form a low proportion in the compiled list of global cities. Considering continental population, it seems that European and North American cities may be overrepresented (Table 5.2). However, this is understandable if we consider that the population of individual cities in those areas is much lower than in other regions. Especially, the cities in China and India heavily affect the total population of Asia. The final distribution of study cities is shown in Figure 5.3.

Table 5.2: Distribution of population by continent (source: UN DESA, 2011)

Continents	Number of Study Cities	Percentage	Study City Population	Percentage	Continent Population	Percentage
Asia	73	27.6	294,706,182	45	4,164,252,000	57
Europe	58	22.0	90,327,049	14	738,199,000	10
North America	44	16.7	44,590,300	7	344,529,000	5
Africa	34	12.9	87,575,985	13	1,022,234,000	14
South America	21	8.0	75,391,682	11	392,555,000	5
Middle East	14	5.3	29,501,044	4	385,331,000	5
Central America	11	4.2	20,944,336	3	155,881,000	2
Oceania	6	2.3	12,022,059	2	36,593,000	1
The Caribbean	3	1.1	5,600,507	1	41,646,000	1
Total	264	100	660,659,144	100	7,281,220,000	100

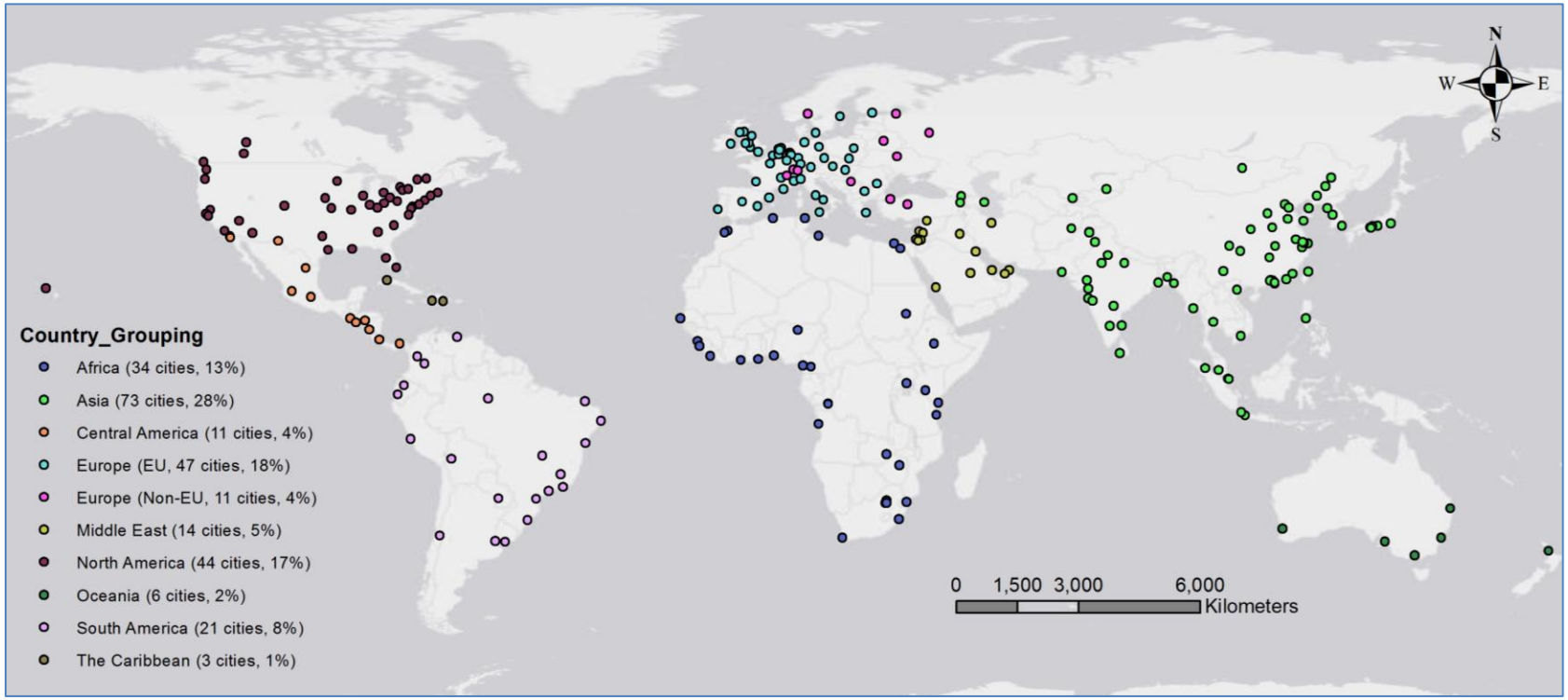

Figure 5.3: Distribution of study cities based on the UN country grouping

5.2. Data

Data for this research come from webpages. While webpages can be downloaded manually, web crawling is more efficient to download a massive amount of webpages in a limited amount of time. The use of crawlers requires a list of seed URLs. For each city under study, one part of seed URLs comes from the database of Open Directory Project (ODP), while another part comes from the results of Google Search. These seeds were collected from February to July 2012. Collected seed URLs were uploaded to a commercial-grade crawling website (80legs.com). Downloading the main data set of webpages was conducted from August to September 2012. The specific information about data collection will be discussed later in the section on procedures.

5.2.1. Linguistic Characteristics

The Web consists of a multitude of websites. Each website is a webpage or a group of webpages. Each webpage is connected to other webpages by hyperlinks. From a structural perspective, the whole system of the Web forms a network composed of nodes and links. However, it is not simple to analyze the content in webpages because the content is written in different languages. Although we can analyze the structure of webpages through HTML, contents cannot be interpreted if you do not know the language of the page contents. Thus, language is one of the aspects to consider before analyzing webpages unless the focus is just on the structure of your webpages.

How many languages are used on the Web? What is the most frequently used language? Just as it is hard to measure the size of the Web, to know the exact variety of languages is not easy. In this World, there are 6,909 ethnological spoken languages (Paul, 2009). Of all these ethnological languages, only 5~10% have an alphabet (Prado, 2012, p.

38); and 300~500 languages are represented on the Internet (SIL International, 2011, p. 2). It is not sufficient for a language to have an alphabet for this language to be used on the Internet. To be used on the Internet, it should meet conditions such as the number of users, technical support, suitability of the language, agreement to use, etc. Thus, the actual number of languages used on the Web is very limited. When we assumes that an Internet user uses one language, the top 10 languages based on the Internet users account for about 82.2% of all used languages online (Table 5.3). Interestingly, Internet users speaking Chinese have almost caught up with the number of English users. Spanish users follow Chinese users; however, in percentage of speakers, Spanish users are far behind Chinese users. Japanese users (78.4%) and German users (79.5%) show the high Internet penetration rate (see note 2 in Table 5.3). Arabic, Russian, and Chinese users exhibit the three highest rates of growth in the Internet between 2000 and 2011.

If we look at the contents of the Internet, the situation is a little different from the results based on the number of speakers. Only 13 languages represent more than 1.0% of the total content of the Web; and the sum of all other languages is no match to English (Figure 5.4). The dominance of English is related to the invention of the Internet. The first users were people who speak English. In 1998, the proportion of English contents was about 75% (Pimienta, 2005, p. 31). However, the share of English contents has decreased due to the relative increase of other languages. Although the 20% decrease of English usage seen over 14 years seems to suggest that English would lose its dominance, the recent trend shows that English would not lose its privileged status easily.

Table 5.3: Top ten languages used in the Web (adapted from Internet World Stats, 2011)

Rank of Language by Users		Internet Users by language	Internet Users (% of Total)	Internet Penetration by language (%)	Growth in Internet (2000-2011, %)	World Population for language (2011 Estimate)
1	English	565,004,126	26.8	43.4	301.4	1,302,275,670
2	Chinese	509,965,013	24.2	37.2	1478.7	1,372,226,042
3	Spanish	164,968,742	7.8	39.0	807.4	423,085,806
4	Japanese	99,182,000	4.7	78.4	110.7	126,475,664
5	Portuguese	82,586,600	3.9	32.5	990.1	253,947,594
6	German	75,422,674	3.6	79.5	174.1	94,842,656
7	Arabic	65,365,400	3.3	18.8	2501.2	347,002,991
8	French	59,779,525	3.0	17.2	398.2	347,932,305
9	Russian	59,700,000	3.0	42.8	1825.8	139,930,205
10	Korean	39,440,000	2.0	55.2	107.1	71,393,343
Top 10 languages		1,615,957,333	82.2	36.4	421.2	4,442,056,069
Rest of the languages		350,557,483	17.8	14.6	588.5	2,403,553,891
Total		2,099,926,965	100.0	30.3	481.7	6,930,055,154

Notes: 1 Statistics were updated for May 31, 2011; 2 Internet Penetration is the ratio between the sum of Internet users speaking *a language* and the total population estimate that speaks that specific language; 3 Internet usage information comes from data published by Nielson Online, International Telecommunications Union, Gfk, etc.; 4 World population comes from the U.S. Census Bureau.

Sources: Internet World Stats, http://internetworldstats.com/stats7.htm (retrieved on April 2, 2012)

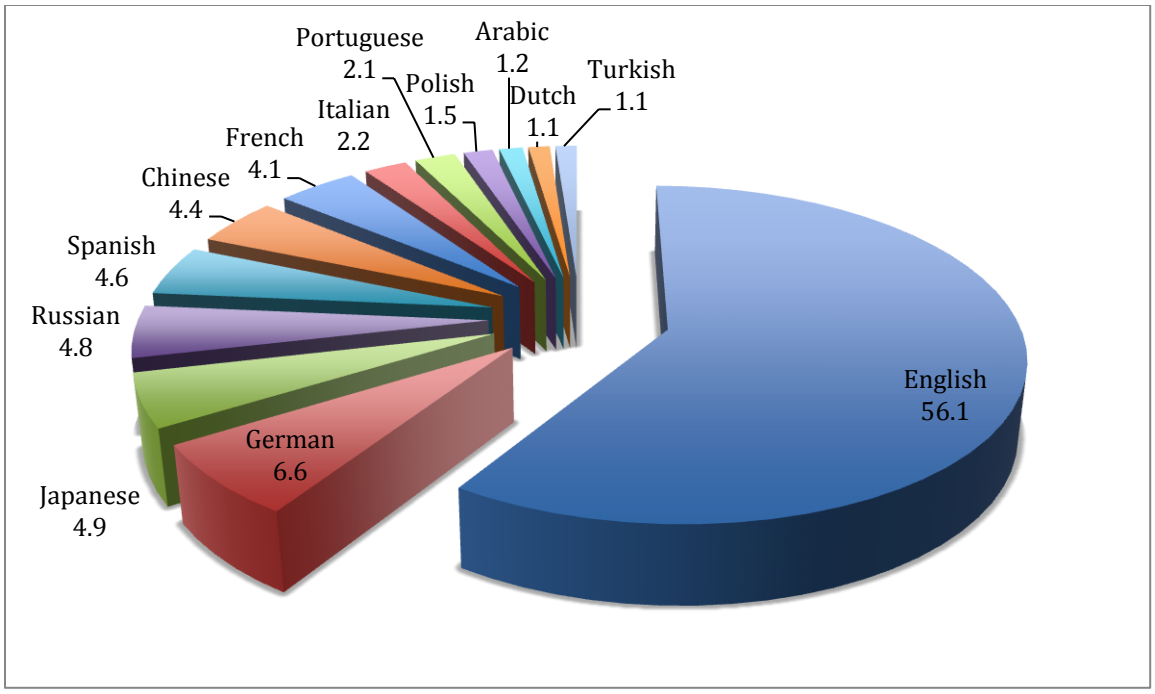

Figure 5.4: Content languages for websites (on April 1, 2012)

Sources: W3Techs, http://w3techs.com/technologies/overview/content_language/all, retrieved on April 2, 2012

Note: A website may use more than one content language.

Based on historical quarterly trends from W3Techs.com (Figure 5.5), the share of English websites has remained over 55% of all websites from July 1, 2010 to April 1, 2012. German and Japanese websites decreased by 3.4% (2.8% and 0.6% for each); and Russian, Spanish, Chinese, French, and Italian websites increased by a total of 3.7%. It is hard to estimate the share of English websites in the future. However, given the recent two-year trend, the share of English web sites may not decrease as rapidly as before.

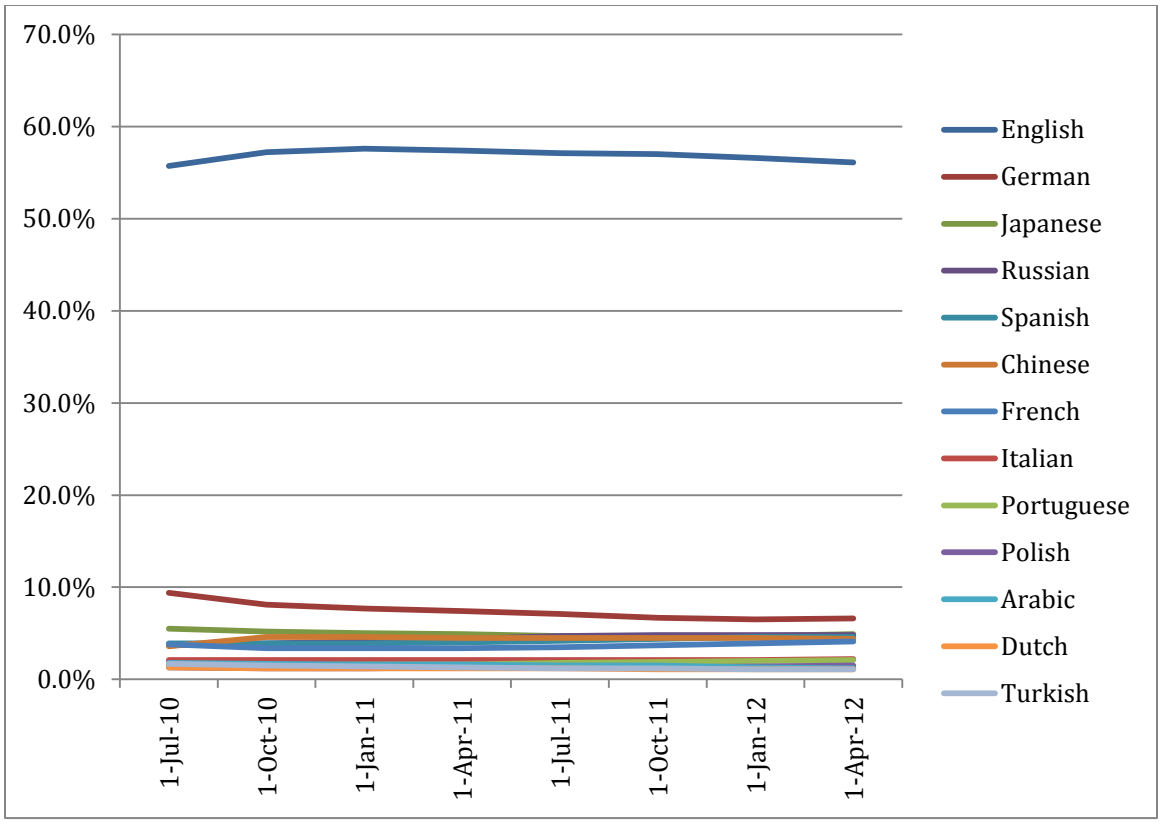

Figure 5.5: Quarterly trends in the usage of content languages for websites from July 1, 2010 to April 1, 2012

Sources: W3Techs, http://w3techs.com/technologies/history_overview/content_language/ms/q, retrieved on April 2, 2012

English has been recognized as the first of the World's 10 most influential languages (Weber, 1997). Weber (1997) classified countries into three categories: Core, Outer Core, and Fringe countries. Core countries are where the language enjoys full legal and official status. Outer Core countries are where the language has some official or legal status like English in India, French in Algeria. Fringe countries are where the language has limited official or legal status and is understood and spoken by a minority like English in Japan, French in Romania. Through this classification, he found that English is used in 115 countries (Weber, 1997, Fig. 7); and population of those countries is about four times more than Chinese (Weber, 1997, Fig. 8); and the sum of GNP in countries

using English (for Core, Outer Core, and Fringe countries) overwhelmed other languages (Weber, 1997, Fig. 10). According to his comments added in 2008 to the 1997 online version of his paper, the growth of English as a world language has not slowed down as anticipated, because of the increasingly large number of citizens of many other countries such as China who learn English (Weber, 1997, notes in Fig. 10). Throughout the world, learning English is seen as a sign of success by parents, local governments, and nations; and many countries have English education programs in their official curriculum (Weber, 1997, supplementary information). English will not resign from the top rank of global languages as long as people think English is the most useful language for their success.

English has been used for communicating internationally (lingua franca) for many decades, and it has powerful influences. On the one hand, to select only English as the field of data may narrow the meaning of this research. However, it is meaningful to know the thought of people who lead globalization of the world because the usage of English has be instrumental in the process of globalization and may well continue to do so in the future.

In sum, this research uses webpages written in English as data source. Figure 5.6 schematically shows the relationship between the linguistic distribution in the real world and the webpages used for this research. Again, it may be a limitation of this research; different conclusions could be reached if webpages written in other languages were also considered. However, English is not only the most-used language in the real world and on the Web, but it is also the acknowledged vehicle of globalization; thus, it does not restrict us to recognize the image of global cities.

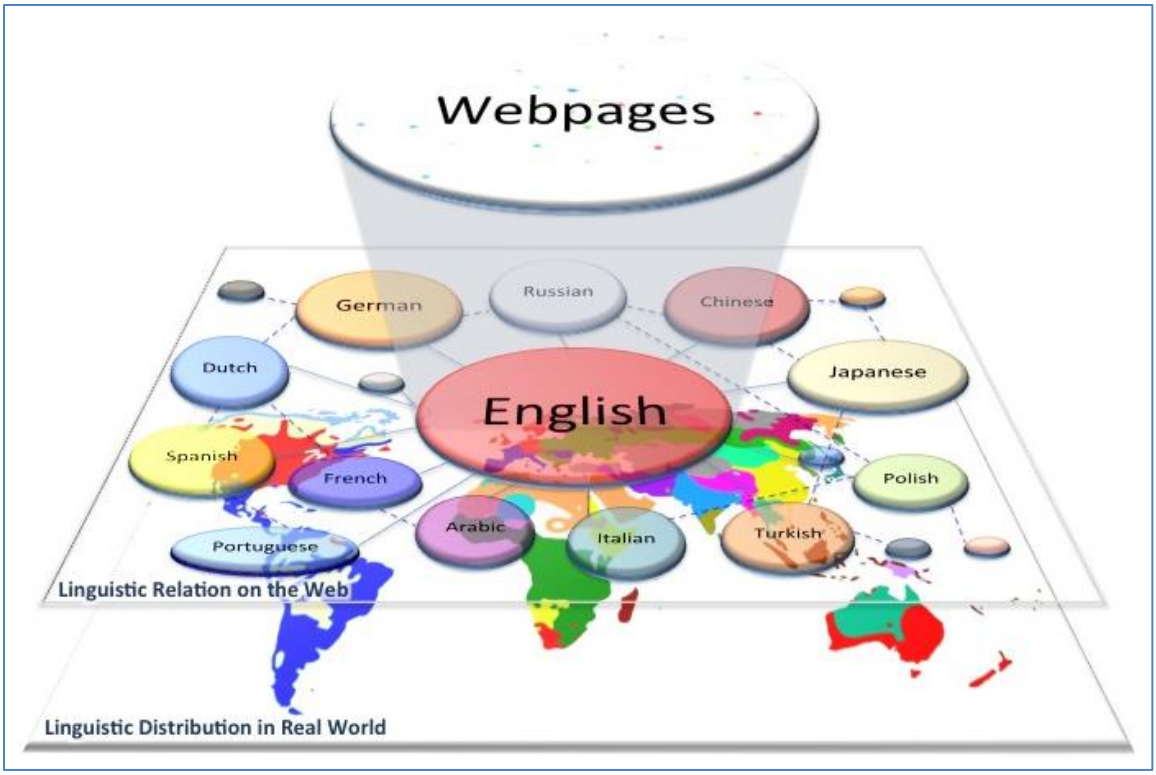

Figure 5.6: Concept of relationship between languages and webpages for research data

5.2.2. Search Engines and Web Crawlers

Search engines are the medium between users and websites. Search engines include web search engines and mobile/tablet search engines. Based on the user's keywords, a search algorithm returns best matching webpages, images, documents, and other ancillary information from indexed and crawled data. Most commercial search engines also have a complicated query function to return search results that can be tailored to meet individual requirements.

Search engines usually comprise three parts: crawlers, data stores, and processors (Thelwall, 2004). The basic process is as follows: crawlers collect webpages periodically; collected webpages are stored with indexing; finally, processors retrieved the information from the webpages according to users' requests. Search engines vary the cycle of

collecting webpages; and some algorithms are used for setting the collecting cycle based on preset updating schedules, the number of visitors, the freshness of the webpages, etc.

Crawled data are stored as full data or partial data (e.g. URLs or links); however, data should be stored in a way that is efficient for information retrieval. A specific request by users retrieves information from data storage through a processor. For example, if a user wants to receive only images for the result, an image processor works well for this purpose. Search engines have known biases such as linguistic problems for site indexing, multilingual text querying, and differentiated site coverage (Vaughan & Thelwall, 2004). In addition, the order of returned pages is totally decided by the ranking algorithm specific to the search engine. Thus, relevant results depend highly on the ranking algorithm; this is one of the reasons that Google dominates almost 77% of the search engine market[8].

A web crawler (also known as a robot, a bot, or a spider) is a system or a program that downloads the bulk of webpages automatically. In addition to being a part of search engines, it is also used for updating and maintaining the web archive (e.g. the Internet archive), web data mining, and web monitoring services (Olston & Najork, 2010). The basic architecture of web crawling is simple (Figure 5.7). First, a set of URLs is collected for seeding and fed into a crawling system. Second, crawlers download the webpages based on the set of seed URLs. Third, hyperlinks are extracted from downloaded webpages and changed into new URLs. Lastly, the crawler downloads the webpages again from webpages based on new URLs. Literally, the crawler collects the hyperlinks simultaneously with downloading webpages.

[8] Google's market share was 76.66% in February 2012. More than thirty other search engines shares the rest (http://marketshare.hitslink.com).

Crawling exhibits some inherent challenges, which are caused by the natural characteristics of the Web. Olston and Najork (2010) indicated scale, content selection tradeoffs, social obligations, and adversaries as the main challenges. The Web is huge and continuously evolving; but a crawler needs to keep broad coverage and freshness of webpages. However, this is almost impossible; thus, crawling collects highly related contents for providing speedy results; but keeping the balance between coverage and freshness is not easy. Also, crawling should not burden the Web with the creation of high traffic. Finally, a pernicious problem is injecting useless or misleading contents into the corpus for promoting the rank of certain webpages.

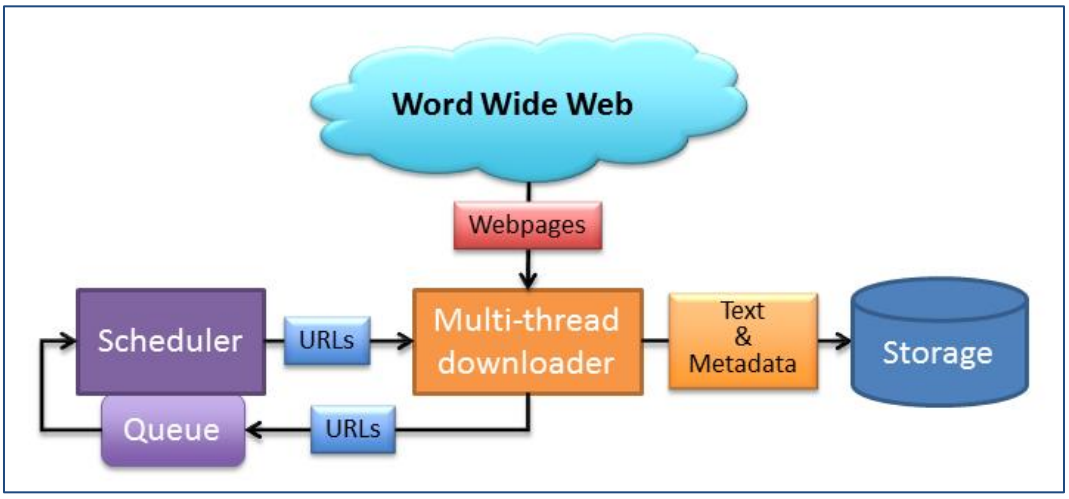

Figure 5.7: Simple architecture of the Web crawler

Comprehensive crawling and scoped crawling are two common types of crawling. The goal of comprehensive crawling is to collect high-quality content of all varieties. For this purpose, the searching order is set by the ordering policy, the Breadth-first Search (BFS) and the Depth-first Search (DFS). Figure 5.8 explains the difference of prioritization between BFS and DFS. Compared to comprehensive crawling, scoped crawling proceeds by limiting crawling activities in a certain categories: topic, geography, format, genre, language, etc. Topical (or focused) crawling is one of the

scoped crawling; it gets contents faster and cheaper than comprehensive crawling (Olston & Najork, 2010).

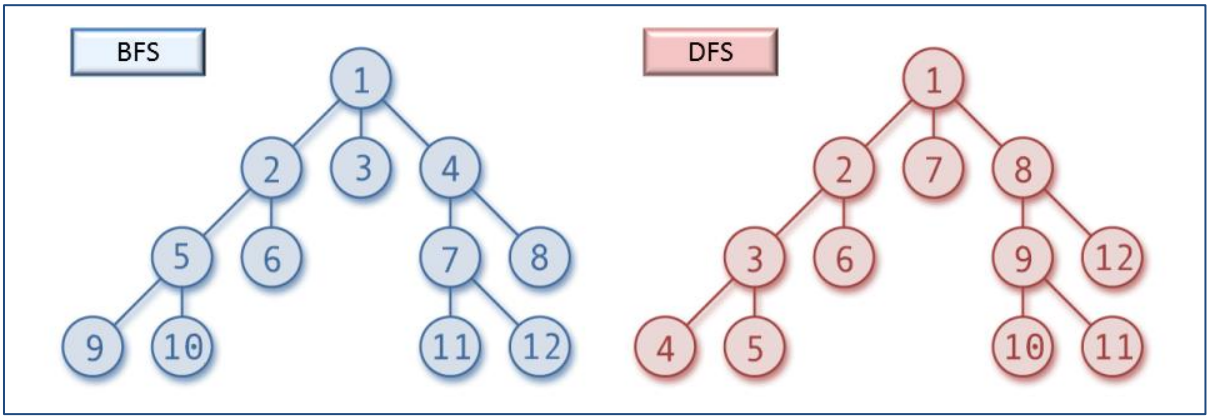

Figure 5.8: Breadth-first Search and Depth-first Search (adapted from the Wikimedia Commons file "Breadth-first-tree.png" http://commons.wikimedia.org/wiki/File:Breadth-first-tree.png and "Depth-first-tree.png" http://commons.wikimedia.org/wiki/File:Depth-first-tree.png)

5.2.3. Data Extraction and Pre-processing

5.2.3.1. Collecting Seed URLs

The collection of webpages on global cities starts with web crawlers that identify seed URLs. Although there are different types of crawling methods, the list of seed URLs is commonly important. Literally, a seed URL becomes a 'seed' for crawling the Web for further useful information. The seeding can involve a single URL or multiple URLs. From seed URLs, a web crawler starts to collect webpages from the Web. It downloads webpages automatically and moves to other webpages via hyperlinks. If the list of seed URLs is far from the center of the network (it means that the seed URLs are isolated or not well-connected), collected webpages are easily out of focus. With a focus on global cities, data collection is here done specifically in webpages related to global cities, not all webpages. This means that the method of crawling in this research is focused crawling.

Two sources are used for collecting seed URLs for each city. One source is to use registered URLs of ODP (http://www.dmoz.org/). Using ODP is a common and simple approach because its links cover a broad range of topics (Olston & Najork, 2010). ODP is a multilingual open content directory of web links, which is classified by a hierarchical ontology scheme; this structure provides the core of directory service for popular search engines and portals, including Google. Because this directory has hierarchical categories, URLs for each city are also classified under the city name. For example, London is under the category of Top/Regional/Europe/United Kingdom/England/London. 14,343 URLs for London are registered as of April 9, 2012.

The other source is the results of the Google search engine. Google's PageRank algorithm is based on the prestige of URLs, which is decided by how many incoming links it has. It means that the top URLs of Google search results have high prestige, and it is possible to find the central hub or prime node (URL) there. Combinations of city name and region like "Amsterdam, Netherlands" are provided for search keyword. After limiting the search result to terms written in English, the top 500 URLs are retrieved from the results of Google search engines (Google.com).

Table 5.4 contains information about the number of ODP URLs, the number of URLs from Google Search, the number of duplicate URLs between ODP and Google, and the final number of seed URLs for each study city. The number of URLs from Google Search is not exactly 500 because the top 500 results from Google contain invalid [9] URLs or duplicate URLs or sub-links (URLs) (i.e. sub-links indexed independently). Seed URLs from ODP are extracted from the regional listing file of the

[9] The reasons are that the webpage does not allow external access for crawling, the term of validity has expired since Google's indexing, or there is failure of server connection (i.e. timed-out).

ODP database (regional_listing.txt, 266 MB) through MySQL. Seed URLs from the results of Google Search are downloaded through Visual Web Spider 7.2, which is the commercial web crawler software developed by Newprosoft (www.newprosoft.com, retrieved on Apr 17, 2012).

Table 5.4: Number of final seed URLs

City	Country	ODP URLs (Feb 2012)	Google URLs (Apr/Jul 2012)	Duplicate URLs	Final Seed URLs
Abidjan	Cote d'lvoire	0	462	0	462
Abu Dhabi	United Arab Emirates	23	474	1	496
Accra	Ghana	0	481	0	481
Addis Ababa	Ethiopia	2	459	2	459
Adelaide	Australia	202	473	17	658
Ahmedabad	India	471	470	19	922
Aleppo	Syria	75	460	11	524
Alexandria	Egypt	15	461	4	472
Algiers	Algeria	0	473	0	473
Almaty	Kazakhstan	4	461	1	464
Amman	Jordan	27	474	8	493
Amsterdam	Netherlands	319	456	22	753
Ankara	Turkey	76	472	4	544
Antwerp	Belgium	47	462	4	505
Asunci ón	Paraguay	3	444	1	446
Athens	Greece	602	483	51	1034
Atlanta	United States, GA	1606	491	104	1993
Auckland	New Zealand	260	487	16	731
Baghdad	Iraq	10	457	3	464
Baku	Azerbaijan	83	474	8	549
Baltimore	United States, MD	1300	489	11	1778
Bandung	Indonesia	27	466	2	491
Bangalore	India	418	490	35	873
Bangkok	Thailand	1032	478	75	1435
Barcelona	Spain	390	483	29	844
Basel	Switzerland	40	438	0	478
Batam	Indonesia	0	472	0	472
Beijing	China	160	487	9	638
Beirut	Lebanon	179	478	14	643
Belgrade	Serbia	24	483	4	503
Belo Horizonte	Brazil	0	440	0	440
Berlin	Germany	480	492	16	956
Birmingham	United Kingdom	1240	490	13	1717
Bogot á	Colombia	33	477	2	508
Bonn	Germany	15	477	3	489
Bordeaux	France	13	475	2	486
Boston	United States, MA	2541	500	123	2918
Bras ília	Brazil	0	413	0	413
Brazzaville	Congo (Rep. of)	0	422	0	422
Brisbane	Australia	323	479	15	787
Brussels	Belgium	172	481	20	633
Bucharest	Romania	86	480	7	559
Budapest	Hungary	126	495	29	592
Buenos Aires	Argentina	150	483	17	616
Buffalo	United States, NY	284	476	11	749
Busan	Korea (Rep. of)	32	401	1	432
Cairo	Egypt	48	479	11	516
Calgary	Canada	1629	496	26	2099

Table 5.4: (continued)

City	Country	ODP URLs (Feb 2012)	Google URLs (Apr/Jul 2012)	Duplicate URLs	Final Seed URLs
Cape Town	South Africa	613	484	46	1051
Caracas	Venezuela	4	487	0	491
Casablanca	Morocco	9	431	2	438
Changchun	China, Jilin	5	470	0	475
Changsha	China, Hunan	16	471	0	487
Charlotte	United States, NC	1052	507	60	1499
Chengdu	China, Sichuan	7	471	0	478
Chennai	India	695	490	32	1153
Chicago	United States, IL	3778	493	137	4134
Chittagong	Bangladesh	19	454	4	469
Chongqing	China	11	477	0	488
Cincinnati	United States, OH	2083	492	26	2549
Ciudad Juarez	Mexico	25	458	3	480
Cleveland	United States, OH	1130	467	7	1590
Cologne	Germany	51	488	5	534
Colombo	Sri Lanka	63	470	8	525
Columbus	United States, OH	1234	489	10	1713
Conakry	Guinea	0	445	0	445
Copenhagen	Denmark	252	469	4	717
Curitiba	Brazil	0	439	0	439
Dakar	Senegal	0	451	0	451
Dalian	China, Liaoning	17	481	1	497
Dallas	United States, TX	2278	489	118	2649
Damascus	Syria	99	459	2	556
Dar es Salaam	Tanzania	32	463	8	487
Denver	United States, CO	1642	485	8	2119
Detroit	United States, MI	333	495	37	791
Dhaka	Bangladesh	10	458	1	467
Doha	Qatar	0	477	0	477
Dongguan	China, Guangdong	14	490	0	504
Douala	Cameroon	0	440	0	440
Dubai	United Arab Emirates	188	485	13	660
Dublin	Ireland	583	493	52	1024
Durban	South Africa	74	475	4	545
Düsseldorf	Germany	62	451	5	508
East Rand	South Africa	0	491	0	491
Edinburgh	United Kingdom	1899	387	21	2265
Edmonton	Canada	1052	491	25	1518
Fortaleza	Brazil	0	412	0	412
Foshan	China, Guangdong	61	478	0	539
Frankfurt	Germany	59	473	4	528
Freetown	Sierra Leone	0	452	0	452
Geneva	Switzerland	159	485	13	631
Glasgow	United Kingdom	1802	487	10	2279
Guadalajara	Mexico	42	461	8	495
Guangzhou	China, Guangdong	34	477	0	511
Guatemala City	Guatemala	20	474	1	493
Guayaquil	Ecuador	0	452	0	452
Harbin	China, Heilongjiang	0	438	0	438

Table 5.4: (continued)

City	Country	ODP URLs (Feb 2012)	Google URLs (Apr/Jul 2012)	Duplicate URLs	Final Seed URLs
Hamburg	Germany	88	457	7	538
Hangzhou	China, Zhejiang	30	467	0	497
Hanoi	Vietnam	57	479	9	527
Harare	Zimbabwe	0	461	0	461
Hartford	United States, CT	634	502	75	1061
Havana	Cuba	49	470	5	514
Helsinki	Finland	59	478	3	534
Ho Chi Minh City	Vietnam	52	482	8	526
Hong Kong	China	2197	486	78	2605
Honolulu	United States, HI	256	492	46	702
Houston	United States, TX	3798	499	145	4152
Hyderabad	India	143	487	13	617
Indianapolis	United States, IN	1729	494	28	2195
Islamabad	Pakistan	15	477	0	492
Istanbul	Turkey	393	484	56	821
Jaipur	India	193	483	21	655
Jakarta	Indonesia	101	484	23	562
Jeddah	Saudi Arabia	25	456	2	479
Jerusalem	Israel	213	483	18	678
Jinan	China, Shandong	3	467	0	470
Johannesburg	South Africa	157	488	8	637
Kabul	Afghanistan	10	464	1	473
Kampala	Uganda	0	478	0	478
Kano	Nigeria	0	451	0	451
Kansas City	United States, MO	439	494	6	927
Karachi	Pakistan	98	486	9	575
Khartoum	Sudan	1	466	0	467
Kiev	Ukraine	61	483	6	538
Kinshasa	Congo (Dem. Rep. of)	0	349	0	349
Kobe	Japan	18	477	2	493
Kolkata (Calcutta)	India	641	482	34	1089
Krakow	Poland	100	481	7	574
Kuala Lumpur	Malaysia	712	493	26	1179
Kunming	China, Yunnan	10	452	4	458
Kuwait City	Kuwait	0	482	0	482
Kyoto	Japan	105	490	15	580
La Paz	Bolivia	0	427	0	427
Lagos	Nigeria	170	473	5	638
Lahore	Pakistan	44	472	10	506
Las Vegas	United States, NV	1035	467	5	1497
Lille	France	0	469	0	469
Lima	Peru	130	433	16	547
Lisbon	Portugal	69	467	5	531
Liverpool	United Kingdom	709	485	8	1186
London	United Kingdom	14409	493	115	14787
Los Angeles	United States, CA	1882	493	86	2289
Luanda	Angola	1	462	1	462
Lucknow	India	35	462	7	490
Lusaka	Zambia	12	468	3	477

Table 5.4: (continued)

City	Country	ODP URLs (Feb 2012)	Google URLs (Apr/Jul 2012)	Duplicate URLs	Final Seed URLs
Luxembourg	Luxembourg	274	486	42	718
Lyons	France	20	470	0	490
Madrid	Spain	169	472	14	627
Managua	Nicaragua	40	462	5	497
Manaus	Brazil	0	406	0	406
Manchester	United Kingdom	3125	499	20	3604
Manila	Philippines	45	481	9	517
Maputo	Mozambique	2	403	1	404
Marseille	France	9	484	0	493
Medan	Indonesia	7	462	1	468
Medellin	Colombia	7	448	0	455
Melbourne	Australia	387	485	27	845
Mexico City	Mexico	49	477	3	523
Miami	United States, FL	601	489	39	1051
Milan	Italy	244	465	32	677
Minneapolis	United States, MN	785	510	62	1233
Minsk	Belarus	6	459	3	462
Mombasa	Kenya	47	455	8	494
Monrovia	Liberia	0	446	0	446
Monterrey	Mexico	41	467	1	507
Montevideo	Uruguay	14	434	0	448
Montreal	Canada	1560	464	37	1987
Moscow	Russia	86	475	8	553
Mumbai	India	1000	482	40	1442
Munich	Germany	186	460	13	633
Nagoya	Japan	28	477	4	501
Nairobi	Kenya	141	472	17	596
Nanjing	China, Jiangsu	29	455	2	482
Naples	Italy	68	475	4	539
New Delhi (Delhi)	India	884	486	34	1336
New Orleans	United States, LA	690	500	20	1170
New York	United States, NY	5030	488	122	5396
Newcastle	United Kingdom	435	500	3	932
Omaha	United States, NE	1064	478	19	1523
Osaka	Japan	99	484	8	575
Oslo	Norway	12	477	2	487
Ottawa	Canada	4285	486	74	4697
Palermo	Italy	19	469	2	486
Panama City	Panama	40	492	3	529
Paris	France	421	484	28	877
Perth	Australia	464	478	24	918
Philadelphia	United States, PA	828	486	51	1263
Phoenix	United States, AZ	1559	509	9	2059
Pittsburgh	United States, PA	654	481	5	1130
Port-au-Prince	Haiti	0	408	0	408
Portland	United States, OR	1578	493	120	1951
Porto Alegre	Brazil	0	392	0	392
Prague	Czech Republic	367	488	22	833
Pretoria	South Africa	82	481	8	555

Table 5.4: (continued)

City	Country	ODP URLs (Feb 2012)	Google URLs (Apr/Jul 2012)	Duplicate URLs	Final Seed URLs
Pune	India	269	485	17	737
Pyongyang	Korea (D. P. R. of)	8	368	2	374
Qingdao	China, Shandong	56	483	1	538
Quito	Ecuador	0	448	0	448
Rabat	Morocco	7	434	1	440
Rawalpindi	Pakistan	7	461	0	468
Recife	Brazil	0	433	0	433
Richmond	United States, VA	593	485	9	1069
Rio de Janeiro	Brazil	88	459	9	538
Riyadh	Saudi Arabia	34	480	7	507
Rochester	United States, NY	923	476	9	1390
Rome	Italy	411	471	46	836
Rotterdam	Netherlands	43	441	3	481
Sacramento	United States, CA	472	490	6	956
Saint Louis	United States, MO	894	492	8	1378
Saint Petersburg	Russia	237	466	24	679
Salvador	Brazil	0	410	0	410
San Diego	United States, CA	2685	500	25	3160
San Francisco	United States, CA	1897	494	122	2269
San Jose	United States, CA	772	485	13	1244
San Jos é	Costa Rica	345	447	0	792
San Salvador	El Salvador	16	459	0	475
Santiago	Chile	25	477	5	497
Santo Domingo	Dominican Republic	25	431	3	453
S ão Paulo	Brazil	41	462	4	499
Seattle	United States, WA	4393	489	155	4727
Seoul	Korea (Rep. of)	177	477	2	652
Shanghai	China	235	459	1	693
Shantou	China, Guangdong	14	485	0	499
Shenyang	China, Liaoning	3	482	0	485
Shenzhen	China, Guangdong	61	483	2	542
Singapore	Singapore	2419	486	105	2800
Sofia	Bulgaria	62	472	8	526
Stockholm	Sweden	112	462	4	570
Stuttgart	Germany	47	453	4	496
Surat	India	58	473	2	529
Suzhou	China, Jiangsu	13	435	0	448
Sydney	Australia	925	482	54	1353
Taipei	China (Rep. of; Taiwan)	41	479	4	516
Taiyuan	China, Shanxi	0	464	0	464
Tampa	United States, FL	576	500	6	1070
Tashkent	Uzbekistan	7	455	2	460
Tbilisi	Georgia	67	473	11	529
Tegucigalpa	Honduras	12	439	3	448
Tehran	Iran	8	462	0	470
Tel Aviv	Israel	184	478	17	645
The Hague	Netherlands	41	479	5	515
Tianjin	China	10	453	0	463
Tijuana	Mexico	25	448	1	472

Table 5.4: (continued)

City	Country	ODP URLs (Feb 2012)	Google URLs (Apr/Jul 2012)	Duplicate URLs	Final Seed URLs
Tokyo	Japan	614	489	50	1053
Toronto	Canada	5369	481	95	5755
Tripoli	Libya	0	450	0	450
Tunis	Tunisia	44	477	1	520
Turin	Italy	60	460	1	519
Ulan Bator	Mongolia	0	443	0	443
Valencia	Spain	13	490	2	501
Vancouver	Canada	2165	453	67	2551
Vienna	Austria	353	453	23	783
Warsaw	Poland	30	452	5	477
Washington	United States, DC	3026	481	55	3452
Wuhan	China, Hubei	8	460	0	468
Wuxi	China, Jiangsu	5	483	0	488
Xi'an	China, Shaanxi	15	450	0	465
Xiamen	China, Fujian	25	456	1	480
Yangon	Myanmar	7	460	4	463
Yaoundé	Cameroon	0	435	0	435
Yerevan	Armenia	30	455	2	483
Zhengzhou	China, Henan	6	470	0	476
Zurich	Switzerland	145	433	6	572

5.2.3.2. Collecting Webpages to Structured Data

After the preparation of seed URLs, a crawling program is used for collecting webpages. One important consideration is the criterion used to stop crawling. With commercial search engines (e.g. Google Search), crawling stops when it meets certain conditions such as: 1) scheduled end time; 2) completion of crawling; 3) hard limits like the degree of freshness or the ratio of new and old documents (Google, 2007, 2011). However, Google's stop-crawling conditions are not suitable for an individual researcher because these conditions are subject to indexed archive, and it requires mass storage and high-speed Internet connection. Furthermore, consideration should be given to how many nodes and edges the network analysis program can handle. For example, UCINET 6 for Windows can handle a maximum of 32,767 nodes officially. NetMiner 4 can handle an unlimited number of nodes for the enterprise license; however, student use (research) license can handle a maximum of 100,000 nodes. Neither UCINET nor NetMiner

guarantees the full speed of analysis when the number of nodes is close to the maximum. It totally depends on the hardware system of a user. Furthermore, the number of possible webpages for each city varies between popular cities and relatively unknown cities. What is the proper number of crawling webpages for the study of cities? Frankly, we do not know how many webpages there are for each city. Although the target number of crawls is set to a certain small number of webpages (nodes or URLs) for each city, it is possible that some cities cannot meet the goal. Thus, the number of URLs for the analysis of the research should be decided after crawling and data clean-up for each city. However, a target number of crawls should be set; for this study, it is 100,000 URLs (nodes) per city.

A commercial crawling portal, 80legs.com is selected for the collection of the main data after reviewing multiple commercial and open-source programs. When it is compared to other crawling programs run on the operating systems, the Web portal crawling service 80legs.com has the following advantages. Firstly, it is not affected by the local machine. This means that the local machine is only used for downloading the crawling results from the portal service because all the processes occur on the server of the portal site. Secondly, it is fast because this portal service uses 50,000 computers to collect URLs by using parallel crawls. In addition, it is possible to submit multiple crawling jobs. Lastly, it can customize the crawling jobs and export data to different formats. Specific crawling settings for collecting the main data are given in Table 5.5.

Table 5.5: Settings for crawling

Items	Settings
Outgoing Links to Crawl	Crawl all links
Depth Level	6[10]
Crawl Type	Breadth first
Max Number URLs to Crawl	100,000
Max URLs per Page	No Limit
Max Pages per Domain	No Limit
MIME Types to Crawl	Text

Literally, all kinds of contents can be extracted from webpages through crawling unless a server manager puts a robots.txt file in the top-level directory, although this is not a perfect method for the disallowance of crawling because the crawler can ignore the robots.txt file. Among the various types of webpage information such as text, audio, video, animation, and so on, the basic deliverable and searchable type is text because text is the basis of HTML and extensible markup language. HTML is designed to display data, and XML is designed to transfer and store data. HTML and XML consist of a pair of brackets: opening tags and closing tags. The tags[11] are a key for the extraction of text. Possible types of attributes from webpages depend on what kinds of tags the webpage has. In other words, tags show the role and the visual of the in-between text. Through the tags, we can know the purpose of the text, and we can classify the text into the structured database. Unless the tag is used wrongly (or missing) tags can be used for the classification of webpage information. Because the collected webpages are a mixture of tags and text, it needs to be separated by tags. After the separation by the function of tags,

[10] This is based on the theory of "Six degrees of separation", which explains that everything is linked in six or fewer steps (Barabási, 2003).

[11] HTML tags refer to http://www.w3schools.com/tags/default.asp (retrieved on Mar 20, 2012). XML has no predefined tags, and the author can define their own tags and document structure (http://www.w3schools.com/xml/xml_whatis.asp; retrieved on August 26, 2012).

there is no need to keep the tags. Truncation operates the removal of the tags from the data. Finally, we get the indexed table, which has the pure text separated by tags and tags' function for its attributes (Figure 5.9).

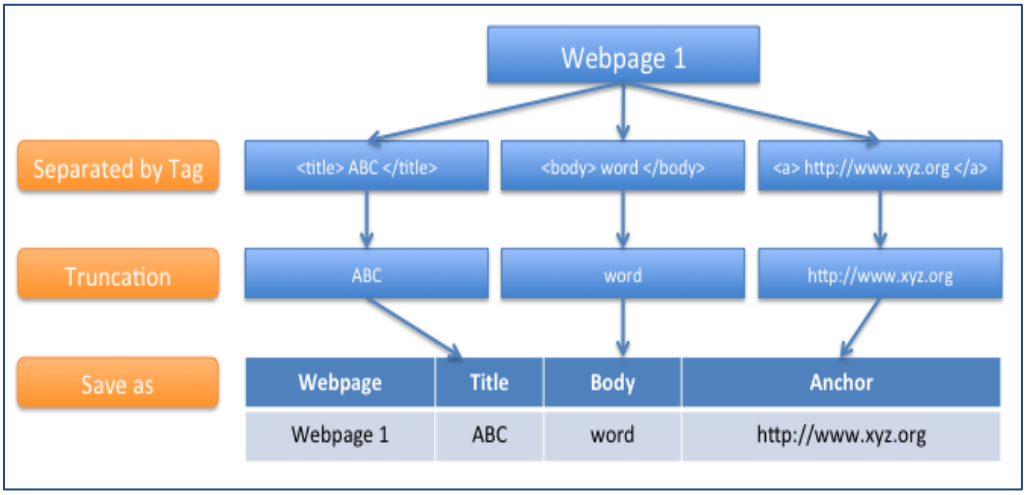

Figure 5.9: Process of extracting pure text from the collected webpages

5.3. Methods for Hyperlink Network and Contents of Webpages

5.3.1. Graph Theory, Social Network, and Hyperlink Network Analysis

The Web has become the treasure house of knowledge in recent years. Although researchers from different fields have studied properties of the Web from various perspectives, there is consensus that the Web has the characteristics of a link. The link is the backbone for the communication among individuals, groups, affiliations, and nations. In the network, nodes and edges (or arcs) can describe everything theoretically as long as there is some relation between actors (or givers) and recipients (or receivers). Historically, some 'structural' sociologists focused on the structure of the individual's social network. In other words, they wanted to find answers on social relationships through the structural characteristics of the network. The adaptation of graph theory to sociology provides the basis for the emergence of Social Network Analysis (SNA).

SNA is one of the popular fields of research nowadays. Figure 5.10 presents the relative position of hyperlink networks with respect to social networks (Park, 2003, p. 52). Based on Park's classification, a social network includes any kind of social relation. A communication network is the network that is composed of interconnected individuals linked by patterned flows of information. This network includes the computer-mediated communication network, the Internet network, and the hyperlink network. The hyperlink network is the smallest of the networks depicted, which is based on the hyperlinks among webpages. However, it is the most critical network because hyperlinks are the fundamental elements that enable the Web to act as the 'web'. In other words, the hyperlink network shows the most basic relations between webpages. The webpage creator inputs URLs into their own webpage for the purpose of linking this webpage to others. Continuous and simultaneous listing of the hyperlinks by webpage creators forms the network. As the social network expands relations like the web of a spider, the hyperlink network also shows the same expansion and characteristics. This means that the hyperlink network can also be analyzed by graph theory.

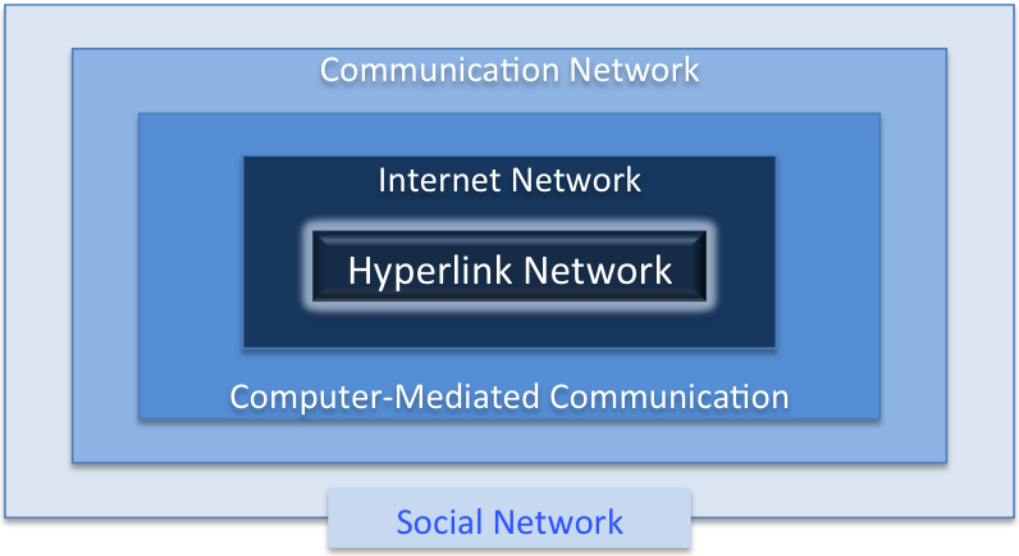

Figure 5.10: Relative position of hyperlink network (modified from Park, 2003, pp. 51-52)

As the Internet usage increased and computing power improved, the number of researchers using hyperlink data has exploded. According to Park and Thelwall (2003), two different research approaches, hyperlink network analysis (HNA) and Webometrics, are commonly used for the examination of hyperlinks among webpages. These approaches complement each other as their started from different aspects. The former considers the hyperlinks as a new social channel through the formalized connections between the authors of hyperlinked webpages. The latter focuses on the reliability and validity of hyperlink data itself, which is mostly researched in academia (e.g. Ranking Web of World Universities, http://www.webometrics.info). While research fields of HNA include e-commerce, social movements, interpersonal communication, inter-organizational communication, and international communication, Webometrics include journal articles and websites, different levels of collections of universities, academic departmental websites, and commercial websites. In addition, HNA can benefit from collecting data and its processing and validation in Webometrics, and the extensive social

network analysis tool of HNA helps Webometrics interpret data from the perspective of social or communication ties. Although the two approaches have different focal points, they share that the hyperlink is the path of information exchange.

What is the most important finding for a network? As ‘structural’ sociologists claimed one time, the structure of the network can help understand the reason of a phenomenon. In other words, the structural characteristics of the network (i.e. the pattern of the network) give a hint to explain what inherence is behind the relationship. Several descriptive statistics, more complex statistical classification, and sub-groupings of the network are applied to reveal the characteristics of the network (Table 5.6). The advance of computing technology nowadays allows researchers to calculate the characteristics of a large network having a huge number of nodes and links. Nevertheless, it is still complex and problematic to explain the phenomenon from the extracted pattern by hyperlink analysis (Park & Thelwall, 2003).

Table 5.6: Descriptive analysis of networks (modified from Kolaczyk, 2009, pp. 79-122)

Characterization	Type	Descriptive Analysis
Vertex and Edge	Degree	Distributions
		Correlation
	Centrality	Closeness
		Betweenness
		Eigenvector
		PageRank algorithm
		HITS algorithm
Network Cohesion	Local Density	
	Connectivity	
	Graph Partitioning	Hierarchical Clustering
		Spectral Partitioning
	Assortativity and Mixing	

5.3.2. Quantitative Content Analysis (QCA)

Content analysis is a method to find valid inferences from text by classifying textual material and reducing text to more relevant and manageable bits of data (Weber,

1990, pp. 5-9). This methodology has been used for analyzing different types of textual data such as lyrics, addresses, newspaper articles, dialogues, etc. The methodologies for quantitative text analysis have been developed in five types, which are (a) the simple analysis of frequency, (b) the evaluated load (i.e. pros and cons) of text through valence analysis, (c) the relative strength of words through intensity analysis, (d) contingency (or associations) analysis with statistics, and (e) computer-assisted analysis (Popping, 2000). With the adoption of computer technology, scholars can find the common intrinsic attribute from a large set of textual data. Recently, data mining processes also have helped classify a set of data into reasonable clusters. As the development of computing technologies leads scholars to focus on semantic analysis, natural language processing, and data mining, computer programs for QCA also follow these big computing trends.

Research on textual web-based media can be classified in three types. First, mass media researchers conduct research on how media contents affect people. They have tried to reveal the effect of media on subjects, whether it is powerful, or limited, or contingent (Riffe, Lacy, & Fico, 2005). Second, informatics researchers conduct research on the content itself: they try to find the proper classification of the content and hidden coherence among mass contents. Third, sociologists or behavioral scientists conduct research to find lifestyles and gender portrayals by using a large set of contents. Content analysis is a common analytical method to analyze the content of textual data. As the volume of information has dramatically increased, computer-aided content analysis has emerged as the methodology of choice (Riffe et al., 2005, p. 215).

For content analysis, the Internet provides new areas of research and contributes to improving the quality and validity of content analysis with new sources of data and a

reduction in the costs of data collection (Weare & Lin, 2000, p. 289). However, as a data source, the internet also presents challenges. Stemple and Stewart (2000) mentioned four problems: sampling problem, inconsistent indexing from website to website, missing information, and expensive cost for content. Sampling problems (samples from the Internet are convenience samples rather than representative samples) could be overcome by having a very large sample (Riffe et al., 2005). However, ignoring small sites that are not indexed, missing information with frequent change, and expensive cost of acquiring specific data remain problematic due to the characteristics of the Web itself.

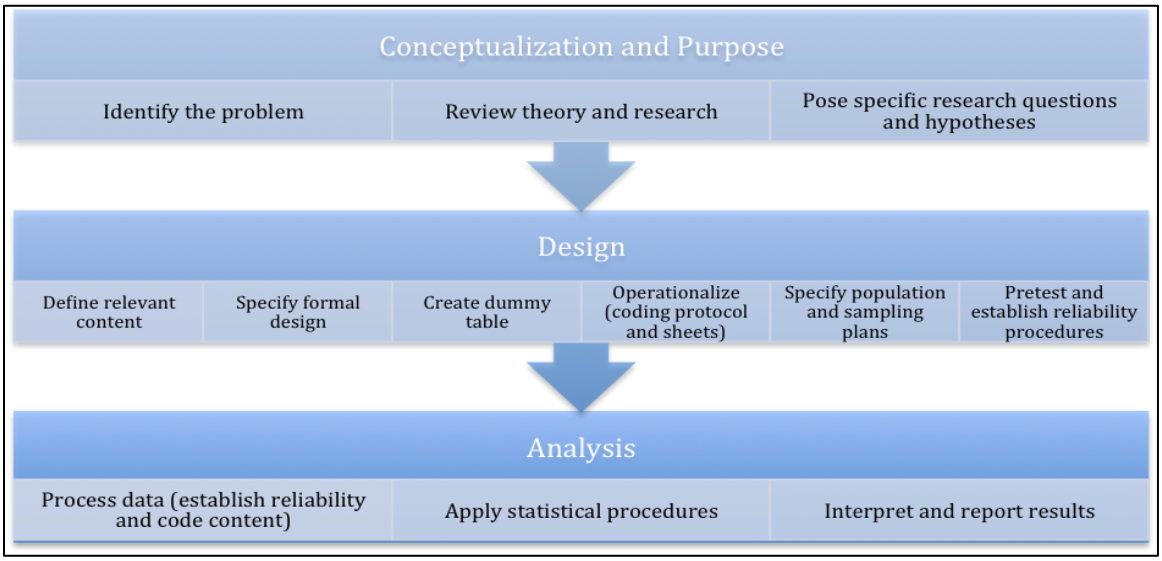

Figure 5.11: General model of content analysis (modified from Riffe et al., 2005, table 3.1)

A general model of content analysis consists of three steps: conceptualization and purpose, design, and analysis (Figure 5.11). As for the process of quantitative research, recognizing a problem, reviewing the literature, and providing hypotheses form the first step. The design process consists of several steps to provide relevant data and procedures. Finally, the full dataset is processed, statistical procedures are applied, and conclusions are derived from the interpretation of the results. The whole procedure is recursive so as to refine the theory framing (Riffe et al., 2005, p. 55). Table 5.7 explains possible

questions for each level of procedure in the content analysis. Through these questions, researchers build the blueprint; and finally, a researcher can answer the research question and fulfill the study's purpose.

Table 5.7: Questions for refining research design of content analysis (modified from Riffe et al., 2005, pp. 56-62)

Procedure	Questions
Conceptualization and Purpose	• What is the phenomenon or event to be studied? • How much is known about the phenomenon already? • What are the specific research questions or hypotheses?
Design	• What will be needed to answer the specific research question or test the hypothesis? • What is the formal design of the study? • How will coders know the data when they see it? • How much data will be needed to test the hypothesis or answer the research question? • How can the quality of the data be maximized?
Analysis	• What kind of data analysis will be used? • Has the research question been answered or the research hypothesis tested successfully?

5.4. Research Objective 4.1 Methodology

The first research objective is to find the relational system of global cities from the hyperlink network. In other words, it is to find the characteristics of the network of each city; then those characteristics are used to create a typology of networks. For this purpose, several measures of network connectivity are used and methods of clustering analysis are applied to them to derive a meaningful typology.

5.4.1. Network Connectivity

5.4.1.1. Degree

In graph theory, the *degree* of a vertex (or node) is the number of edges connected to the vertex. Degrees are easy to compute, but this simple measure is quite informative (Wasserman & Faust, 1994, p. 100). The minimum of the degree of a node is equal to 0,

which is that the node does not have any edge to other nodes; and it is called an *isolate*[12]. The mean degree is informative to show the connectivity of a network. If we denote the degree of vertex i as d_i, the number of edges as e, and the number of vertex as n, the mean degree $\bar{d}$ is like (5.1) for an undirected graph.

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{2e}{n} \tag{5.1}$$

For a directed graph, the degree of a vertex should be considered separately as in-degree and out-degree. In-degree is the number of incoming edges to a vertex; and out-degree is the number of outgoing edges from a vertex. The number of edges e is the same as the total number of incoming edges, or equivalent to the total number of outgoing edges at all vertices, because one edge has two roles, that is as an incoming edge for one vertex and as an outgoing edge for another vertex. Thus, the mean degree of a directed graph is

$$\bar{d}_{in} = \frac{\sum_{i=1}^{n} d_i^{in}}{n} = \frac{\sum_{i=1}^{n} d_i^{out}}{n} = \bar{d}_{out} = \bar{d} = \frac{e}{n} \tag{5.2}$$

Another important index for the degree is the variance of the degree because it can be used as a measure of uniformity (*d-regularity*). A graph is said to be regular if all vertices have the same degree. An infinite square lattice is an example of a 4-regular graph (Newman, 2010, p. 135). If a graph is not a regular graph, the nodes differ in degree; and it can be used as a measure of similarity or "activity" of a network (Wasserman & Faust, 1994, p. 101). The variance of the degree S_D^2 for an undirected graph is calculated as:

[12] Graph theorists provide a vocabulary for the typology of nodes based on their degree because the degree can be used for finding the role of the node in a directed graph. A node is a(n): *Isolate* if $\bar{d}_{in} = \bar{d}_{out} = 0$, *Transmitter* if $\bar{d}_{in} = 0$ and $\bar{d}_{out} > 0$, *Receiver* if $\bar{d}_{in} > 0$ and $\bar{d}_{out} = 0$, and *Carrier* or *ordinary* if $\bar{d}_{in} > 0$ and $\bar{d}_{out} > 0$ (as cited in Wasserman & Faust, 1994, p. 128).

$$S_D^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n} \quad (5.3)$$

For a directed graph, the variance of the degree should be calculated separately for in-degree (5.4) and out-degree (5.5).

$$S_{D_{in}}^2 = \frac{\sum_{i=1}^{n}(d_i^{in} - \bar{d}_{in})^2}{n} \quad (5.4)$$

$$S_{D_{out}}^2 = \frac{\sum_{i=1}^{n}(d_i^{out} - \bar{d}_{out})^2}{n} \quad (5.5)$$

In addition, the density of a graph can be used for measuring a network. The density of a graph means the ratio of the number of actual edges in a graph to the maximum number of possible edges in this graph. The number of vertices determines the maximum number of possible edges. Since an edge is an ordered pair of vertices, the possible number of edges is equivalent to $n(n-1)$. Thus, the density Δ, is:

$$\Delta = \frac{e}{n(n-1)} \quad (5.6)$$

The density is a fraction valued from a minimum of 0, if no edges exist, to a maximum of 1, if all edges are present.

5.4.1.2. Path

A *path* is a *walk* in which all vertices and edges are distinct. A walk in a network means a sequence of vertices and edges starting and ending with nodes. Some vertices and edges are included more than once. The length of a walk is the number of edges between the starting vertex and the ending vertex in a walk. In addition, a *trail* is a walk in which all edges are distinct, but some vertices could be included more than once. Trails and paths are special cases of walks with stricter conditions. All paths are trails, and all trails are walks (Wasserman & Faust, 1994, p.107). For a directed graph, a walk is

a *directed walk* in which all edges are pointing in the same direction. A trail and a path are also a *directed trail* and a *directed path*. Among paths, the shortest path (*geodesic path* or *geodesic distance*) is important to define the distance between two vertices and the diameter of a network. In a directed graph, the distance between vertex *i* and vertex *j* (the *length*[13] of the geodesic path) may be different in the *i* to *j* direction and in the reverse direction. Since the length of a path indicates how many steps are required to deliver information from v_i to v_j, the average path length can be used for measuring the 'efficiency' of the network (Kawamura, Otake, & Suzuki, 2009, p. 1162). Except for the situation where there is no route from v_i to v_j, the average path length $\bar{l}$ is calculated, where $d(v_i, v_j)$ is the shortest distance between vertices, and n is the number of vertices, as:

$$\bar{l} = \frac{\sum_i^j d(v_i, v_j)}{n(n-1)} \tag{5.7}$$

Since $n(n-1)$ is the maximum number of possible edges, the average path length $\bar{l}$ becomes 1 if all vertices are linked each other. Thus, the range of the average path length is from 0 to 1.

The *diameter* is the longest geodesic distance between any pair of vertices in a graph. For each vertex, the longest geodesic distance of a vertex is called the eccentricity of a vertex. Thus, the distance of a graph is the maximum eccentricity between a pair (or else some pairs) of vertices. With the average path length, the diameter is important to examine the efficiency of a network. Except for the extreme cases of small networks that we can analyze visually, the diameter provides us the sense of how large a network is. By

[13] The length of a path does not mean the number of vertices, but the number of edges traversed along the path.

looking at the average path length and the diameter together, it can be determined whether a graph is connected strongly in comparison with the size of the network.

5.4.1.3. Clustering Coefficients

The clustering coefficient measures the density of triangles[14] (i.e. the frequency of three loops) in a network. Therefore, it measures the average probability that two neighbors of a vertex are themselves neighbors (Newman, 2010, p. 262). Thus, a higher clustering coefficient means that there is a higher probability of having a relationship between vertices. Thus, it could be called *transitivity*. It is easy to be led to believe that the links among actors are randomly distributed. However, most social networks do not follow the random distribution. Specifically, social networks such as the collaboration network of physicists and academic co-works exhibit higher clustering coefficients. On the other hand, Internet networks like peer-to-peer networks show far lower clustering coefficients than one would expect by chance. Calculating clustering coefficients for the same type of networks can be used for comparison between networks. The clustering coefficient C is calculated as:

$$C = \frac{(\text{number of triangles}) \times 3}{(\text{number of connected triples})} \tag{5.8}$$

5.4.1.4. Reciprocity

Compared to an undirected graph, a meaningful characteristic of a directed graph is the presence of reciprocal edges. Reciprocal edges are actually two edges, which have opposite directions between a pair of vertices. Although reciprocal edges have a loop of

[14] Social network analysts call this *triadic closure*, which is an "open" triad of vertices (i.e. one vertices has two edges to neighbors, but there is no edge between neighbors) becomes "closed" (turning to a triangle) by adding an edge between two vertices that did not have an edge (Newman, 2010, p. 263).

length two, it only has two vertices. This characteristic is different from the clustering coefficient. For example, when one webpage has a hyperlink to another webpage, the linked webpage also has a hyperlink to the former webpage (i.e. co-links in the WWW). *Reciprocity* is the fraction of edges that are reciprocated. This property is helpful to understand the characteristics of the network because it shows how likely webpage B links to webpage A if webpage A links to webpage B. The reciprocity can be calculated easily with the product of adjacency matrix elements. The product of adjacency matrix elements $A_{ij}A_{ji}$ is 1 if there is an edge from i to j and an edge from j to i; otherwise it is 0. Thus, when we note m is the total number of directed edges, the calculation for reciprocity r is:

$$r = \frac{1}{m}\sum_{ij} A_{ij}A_{ji} = \frac{1}{m}\mathrm{Tr}A^2 \tag{5.9}$$

5.4.2. Cluster Analysis

Each single network characteristic introduced above would produce an alternative typology of city networks. However, given the multifaceted nature of city hyperlink networks, it is preferable to classify networks on multiple characteristics together. Thus, we present here the techniques for the classification of networks based on their characteristics. Data are in the form of a geographical matrix, 264 cities by the number of characteristics. Attributes must be normalized before clustering.

5.4.2.1. Hierarchical Clustering

Hierarchical clustering[15] starts with as many clusters as the number of records. The calculation of distance between records leads to the merger of clusters; and this

[15] Hierarchical clustering means agglomerative hierarchical clustering here. The explanation about divisive hierarchical clustering is omitted because it is not be used in this research.

process is repeated until a single cluster remains (Milligan & Cooper, 1987). Hierarchical clustering techniques can be classified in four types, which are differentiated by how the distance is calculated. The single linkage method (Sneath, 1957) calculates the shortest Euclidean distance between two clusters in the attribute space, while the complete linkage method (McQuitty, 1960) calculates the longest distance between two clusters. The average linkage method (Sokal & Michener, 1958) calculates the average distance of records between two clusters. Lastly, Ward's method uses the sum of the squared deviation for the distance between two clusters instead of the Euclidean metric (Ward, 1963). This can be written as:

$$d = \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{5.10}$$

Here, d is the distance of Ward's method, n is the number of records for each cluster, X_i is the value of i^{th} record, and $\bar{X}$ is the average of records in the cluster. At every stage, it calculates the combination of clusters; and the combination having the shortest distance becomes a new cluster.

5.4.2.2. Nonhierarchical Clustering

While hierarchical clustering reduces the number of clusters one by one, nonhierarchical clustering classifies records by the predefined number of clusters. In other words, a six-cluster solution in hierarchical clustering is the union of any two clusters in seven-cluster solutions; however, a six-cluster solution in nonhierarchical clustering is the best of all possible six-cluster solutions. The representative method of nonhierarchical clustering is *k*-means clustering, where *k* is the number of clusters. The process of *k*-means clustering starts with a number *k* of randomly selected cluster seeds.

Each record is clustered into the cluster seed with the nearest mean; thus, the centroid of each cluster changes at each iteration. The new centroid becomes the new mean. Based on the new centroid, the nearest mean is recalculated; and a record having the nearest mean falls into the cluster. The calculation of the nearest mean and the new mean by new centroid are repeated until convergence has been reached.

Nonhierarchical clustering has some advantages compared to hierarchical clustering. In hierarchical clustering, records cannot change the cluster once it is clustered into a certain cluster. Also, it assigns all outliers to clusters. However, one consideration of nonhierarchical clustering is that we should predefine the number of clusters. To find the proper number of clusters can be done by several trials. However, the more practical way is to use hierarchical clustering to determine the proper number of clusters. Then, we can use that number of clusters for nonhierarchical clustering. In addition, outliers can be removed after hierarchical clustering before nonhierarchical clustering.

5.5. Research Objective 4.2 Methodology

The second research objective focuses on *centrality* and *prestige* of network nodes. To know the vertices located in the center of a network is important because this process serves to find the best actors and critical points in the network. We present below a series of measures of centrality and a simple measure for calculating the share of a vertex or group.

5.5.1. Network Centrality

5.5.1.1. Degree

Degree centrality is the simplest method to find the central vertex of a network. For a directed graph, this can be calculated as in-degree and out-degree. The in-degree of

a vertex v_i is the number of vertices that are incident at v_i. Since this is the same number of coming edges from adjacent vertices, the in-degree is the total number of edges terminating at v_i. The out-degree of a vertex v_i is the number of vertices originating from v_i. A linked edge has two directionalities between two linked vertices. In other words, an edge is for in-degree of a vertex and for out-degree of an adjacent vertex at the same time. The relationship between degrees d^{in} / d^{out} and the number of edges m can be described as:

$$\sum_{i=1}^{n} d_i^{in} = \sum_{j=1}^{n} d_j^{out} = m \tag{5.11}$$

This relationship is important to calculate degrees by the adjacency matrix of a directed network.

5.5.1.2. PageRank

PageRank is a variant of eigenvector centrality. The eigenvector centrality is a natural extension of the simple degree centrality, and the eigenvector centrality of a vertex depends on the neighboring weight of the vertex which is calculated by the centrality of its neighbors (Newman, 2010, p. 169). It is also an improved version of the Katz centrality, which gives a small amount of basic centrality to all vertices to prevent the problem of eigenvector centrality that the centrality of a vertex becomes 0 if the vertex has only one incoming edge with its 0 centrality. However, Katz centrality has a problem that the high centrality of one neighbor has too much effect on the centrality of a vertex. For instance, one personal website may have high centrality, maybe due to having an incoming link from Google. Although this example is quite rare, it is possible. One

possible solution is to get proportional centrality from neighbors, which is calculated by their out-degree. This concept can be incorporated in the formula of Katz centrality as:

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta \tag{5.12}$$

A problem happens if a vertex has a value 0 for the out-degree. However, it can be solved by setting $k_i^{out} = 1$ for that type of vertices[16].

Let us use matrix notation, where $\mathbf{1}$ means the vector (1, 1, 1, …), and $\mathbf{D}$ is the diagonal matrix with elements with elements $D_{ii} = \max(k_i^{out}, 1)$. Then the centrality $\mathbf{x}$ can be expressed as:

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{D}^{-1} \mathbf{x} + \beta \mathbf{1} = \beta (\mathbf{I} - \alpha \mathbf{A} \mathbf{D}^{-1})^{-1} \mathbf{1} \tag{5.13}$$

In this expression, β just plays the role of overall multiplier for the centrality. Here, we can follow the convention of setting $\beta = 1$; thus the final formulation is:

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A} \mathbf{D}^{-1})^{-1} \mathbf{1} = \mathbf{D} (\mathbf{D} - \alpha \mathbf{A})^{-\mathbf{1}} \mathbf{1} \tag{5.14}$$

5.5.1.3. Closeness

Closeness centrality captures the average distance between a vertex and every other vertex in the network (Hansen, Shneiderman, & Smith, 2010, p. 40; Newman, 2010, p. 181). The idea behind the closeness centrality is that a vertex (or an actor) is central if it can quickly interact with all others (Wasserman & Faust, 1994, p. 183). In other words, a vertex having the minimum steps to others is more efficient to deliver information (or whatever) in a network. When we consider the length d_{ij} of the geodesic path from i to j, the mean geodesic path l_i can be written as:

[16] It does not affect the result of the centrality as long as the artificial setting (i.e. giving k_i^{out} a certain value) is not zero (Newman, 2010, p. 176).

$$l_i = \frac{1}{n}\sum_j d_{ij} \tag{5.15}$$

However, the mean geodesic path l_i is the opposite of the centrality in other measures. This means that a high value of l_i corresponds to a low centrality. Thus, the inverse of l_i is used for the closeness centrality C_i:

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}} \tag{5.16}$$

5.5.1.4. Betweenness

While the closeness centrality focuses on the steps between a vertex and all other vertices, *betweenness centrality* focuses on the vertices that play a critical role (i.e. the control of passing information) in a path between vertices. When we consider the path of news diffusion among friends for instance, several paths may be possible. In first instance, it is easy to conceive that the transmission of the news follows the shortest path. However, it is also meaningful to consider who is on the transmission path because intermediate nodes can stop or delay the passage of the item. In addition, the centrality of the intermediate friend is high if he has a lot of links to other friends. Based on this basic idea, the betweenness centrality x_i of a vertex i can be defined as the number of paths n_{st}^i that pass through i from s to t:

$$x_i = \sum_{st} n_{st}^i \tag{5.17}$$

Here, formula (5.17) assumes that there is only one geodesic path from s to t. However, there may be multiple geodesic paths between two vertices. For this case, we can weigh geodesic paths by the inverse of the total number of paths from s to t. Considering g_{st} as

the total number of geodesic paths from s to t, we can define the betweenness centrality x_i as:

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}} \tag{5.18}$$

Conventionally, $n_{st}^i / g_{st} = 0$ if both n_{st}^i and g_{st} are zero.

While the variance of closeness centralities of a graph is typically small (Newman, 2010, p. 183), the betweenness centrality shows large variations. Although this characteristic of the betweenness centrality is an advantage compared to the closeness centrality, it also needs normalization for comparison purposes. Let us look at the way to normalize in SNA programs such as Pajek and UCINET. They normalize the path count by dividing by the total number of vertex pairs (i.e. the fraction of paths); thus the range of values lies between zero and one. The normalized betweenness centrality can be written as:

$$x_i = \frac{1}{n^2} \sum_{st} \frac{n_{st}^i}{g_{st}} \tag{5.19}$$

5.6. Research Objective 4.3 Methodology

The purpose of this research objective is to reveal the image of the global city based on the analysis of the textual content of webpages. For this purpose, the smallest unit of meaning, the word, is used for the analysis. Through the frequency of words in the text, we can identify common words for the webpage that contains information about global cities; then, the list of common words can be used for inducing the image of the global cities. However, it is hard to induce the image of a global city based on the list of all common words because the word rank in the list is not a short list. Thus, an alternative methodology, the pre-defined keyword list based on the recognized dimensions of the

global city can be used for the extraction and the quantification of the text. This approach can provide an efficient data processing; and it is helpful to abstract the image of global cities in a context of global city research. Finally, multiple statistical methods and visualizations are used for the text mining analysis.

5.6.1. Frequency Analysis

Frequency analysis is not a complex methodology. However, it requires lots of pre-processing steps before the main analysis. During the pre-processing, the data passes through the following processes. *Stemming* is the process that removes common English suffixes and prefixes. *Lemmatization* is the process that sort out words by grouping inflected or variant forms of the same word. *Spelling correction* helps to correct common misspellings. The *Exclusion* process removes unnecessary words from the documents (i.e. the name of the global city in the documents). Although these processes are automated, intervention of the analyst is necessary for judging the selection and removal of specific words. Since this methodology uses the count of words in the text, it is a kind of univariate keyword frequency analysis. Therefore, it mostly serves descriptive purposes, including basic statistics and frequency distribution.

The way to count keywords from the text is not complicated. The main part of the programming is to make a dictionary that contains the counts of words. Once the word tally is completed, the sorted results are returned for analysis. A rather straightforward application, such as the Python code below, can serve this purpose (Zelle, 2004, pp. 373-374):

```
# wordfreq.py
import string

def compareItems ((w1, c1), (w2, c2)):
       if c1 > c2:
              return -1
       elif c1 == c2:
              return cmp (w1, w2)
       else:
              return 1

def main():
       print "This program analyzes word frequency in a file"
       print "and prints a report on the n most frequent words. \n"

       # get the sequence of words from the file
       fname = raw_input("File to analyze: ")
       text = open(fname, 'r').read()
       text = string.lower(text)
       for ch in '!"#$%()*+-./:;⇔?@[\\]^_'{|}~':
              text = string.replace(text, ch, ' ')
       words = string.split(text)

       # construct a dictionary of word counts
       counts = {}
       for w in words:
              counts[w] = counts.get(w,0) + 1

       # output analysis of n most frequent words.
       n = input("Output analysis of how many words? ")
       items = counts.items()
       items.sort(compareItems)
       for I in range(n):
              print "%-10s%5d" % items[i]

if __name__ == '__main__': main()
```

Additional useful information is the co-occurrence of words. The co-occurrence refers to words that are used together in documents. Word co-occurrence can be displayed in a matrix form; this matrix is helpful to induce the classification of keywords. In addition, we can roughly estimate the image of global cities based on the frequency and the co-occurrence of the words.

The pre-defined word list for the frequency analysis will be based on different sources such as Wikipedia, literatures, research, etc. Firstly, the dimensions of global city

are defined. Secondly, possible words for each dimension are collected and reviewed. Finally, the word list (i.e. codebook) will be used for the extraction and the quantification of the text.

5.6.2. Text Mining Analysis

Text mining is a knowledge-intensive process in which the researcher interacts with the collection of documents over time by using a suite of analysis tools (Feldman & Sanger, 2007, p. 1). While the first methodology of research objective 4.3 focuses on the data processing, the basic analysis, and categorization, the third methodology is the final process of inducing the image of the global city. This process includes hierarchical and non-hierarchical clustering, MDS, SOM, heat maps, and correspondence analysis (CA). In other words, this methodology is the compilation of statistical and visual approaches to find the meaningful knowledge from huge sets of documents.

MDS is a statistical technique that extracts a limited number of dimensions from a multidimensional dataset based on similarities or dissimilarities exhibited by observations on a number of original characteristics. It creates a space of low-dimensionality, which can be visualized to explore patterns. MDS estimates the coordinates in a space of specified dimensionality that comes from the distance between pairs of objects (Deun & Delbeke, 2000; Clark, 2004, p. 2470). Generally, MDS refers to a group of models by which information contained in a set of data is represented by a set of points in a space. The MDS procedure provides the visualization that allows researchers to analyze the relationship among variables more easily. For measuring distances among variables, the basic method consists in using Euclidean distances. The distance d_{ij} between two points i and j in a m-dimensional Euclidean space can be calculated as:

$$d_{ij} = \left[\sum_{a=1}^{m} (x_{ia} - x_{ja})^2\right]^{1/2} \quad (5.20)$$

After the calculation of distance between two points, the result can be drawn on the graph that has two or more dimensions. In this research, MDS maps are used for the graphical representation of the concept maps. In other words, a point represents items (i.e. keyword or content category); and the distances between pairs of items indicate how likely those items are to appear together.

The Kohonen map or self-organizing map is one of the techniques in artificial neural networks based on unsupervised learning (Skupin & Agarwal, 2008; Vesanto, 1999; Vesanto et al., 2000). The characteristic of this method is to provide both the reduction of attributes' dimension and the reduction of records' dimension at the same time. Plus, it keeps the topological relationships throughout the process (Furukawa, 2009). In other words, the SOM can be used for understanding geometric proximity in space and (dis)similarity between attributes (Skupin & Agarwal, 2008). In addition, the SOM is a visualization method that projects multidimensional data to a 1-D space or a 2-D space as a feature map.

The simple structure of the SOM is divided into three parts: the input vector, output arrays, and weights between input and output. The SOM can be used for generalizing the characteristics because the SOM maps the *n*-dimensional attribute space on a lower dimensional feature map. The overall process of the SOM consists of initialization, training, and testing. The learning algorithm follows the winner-take-all strategy, which selects the best matched unit (BMU), and the BMU forces close nodes (neurons) to accept the result of learning by weighted distance. During the training of a SOM, the component plane changes over the iterations until the convergence criterion is

reached. Similar items in the attribute space are located close to each other on the feature map through learning-competitive-cooperative processes. For this research, the result from topological analysis (i.e. centrality and connectivity indices) and the result from textual analysis (i.e. the number of words for each category based on the common definition of global city) will be used for the input vectors of the SOM. The unified distance matrix (U-matrix) presents the distance of cells, which can be used for finding clusters. Component planes provide the comparison of input variables.

Heat maps (i.e. heatmap plots) are the way to represent cross-tables where relative frequencies are shown by different color brightness or tones and on which a clustering is applied to reorder rows and/or columns (Provalis Research, 2010, p.114). Although this plotting technique is usually used in molecular biological researches to identify gene expressions, it can be implemented in text data mining as an exploratory data analysis tool. Especially, it is efficient to examine the relationship between keywords (rows) and values of an independent variable (columns); and it also can be used for displaying the relationship between keyword frequency and occurrences within cases.

CA is a multivariate statistical technique, which is similar to principal component analysis. It is also called correspondence mapping, perceptual mapping, social space analysis, correspondence factor analysis, principal component analysis of qualitative data, and dual scaling (Garson, 2011). While the PCA uses continuous data, the CA uses categorical data. The CA is an exploratory and (multivariate) descriptive data analytic technique. Here, ‘exploratory’ means there is no ‘priori’ hypothesis about relations between variables; and ‘descriptive’ means it can simplifies complex data and provides a detailed description simultaneously (UNESCO, 1999). Researchers can use the CA to

reduce the multidimensional frequency data into a 2-dimensional map (Choi et al., 2007, p. 125). The CA uses a definition of chi-square distance rather than Euclidean distance between variables. The process of the CA starts with a cross tabulation of two discrete variables. Then, the calculation is conducted for row profiles with average row profiles and column profiles with average column profiles. The chi-square distances between points are computed, and finally the correspondence matrix. In this research, the result of the CA is used for examining the relationship between words or categories (i.e. row points), the relationship among subgroups (i.e. column points), and the relationship between words/categories and subgroups (i.e. row and column points).

5.7. Research Objective 4.4 Methodology

For the last objective, we aim to reduce the dimensions of the data. The visualization process of the reduced data set can include the distributional maps based on topological characteristics (from the result of research objective 4.1), textual characteristics (from the result of research objective 4.3), and both. The MDS and/or SOM can be used for the visualization because the methods require reducing the dimensions of data. Discussion of the principles of MDS and SOM can be found in Section 5.6.2. (Text Mining Analysis).

In addition to visualization, a new methodology is needed to measure differences and commonality between the topological and textual distributions. One possible method is to calculate the correlation coefficient between distance matrices (i.e. similarities among cities from the MDS). Although it is a simple method, it provides the scores for measuring relationship between the two different types of Web data. This process is depicted in Figure 5.12.

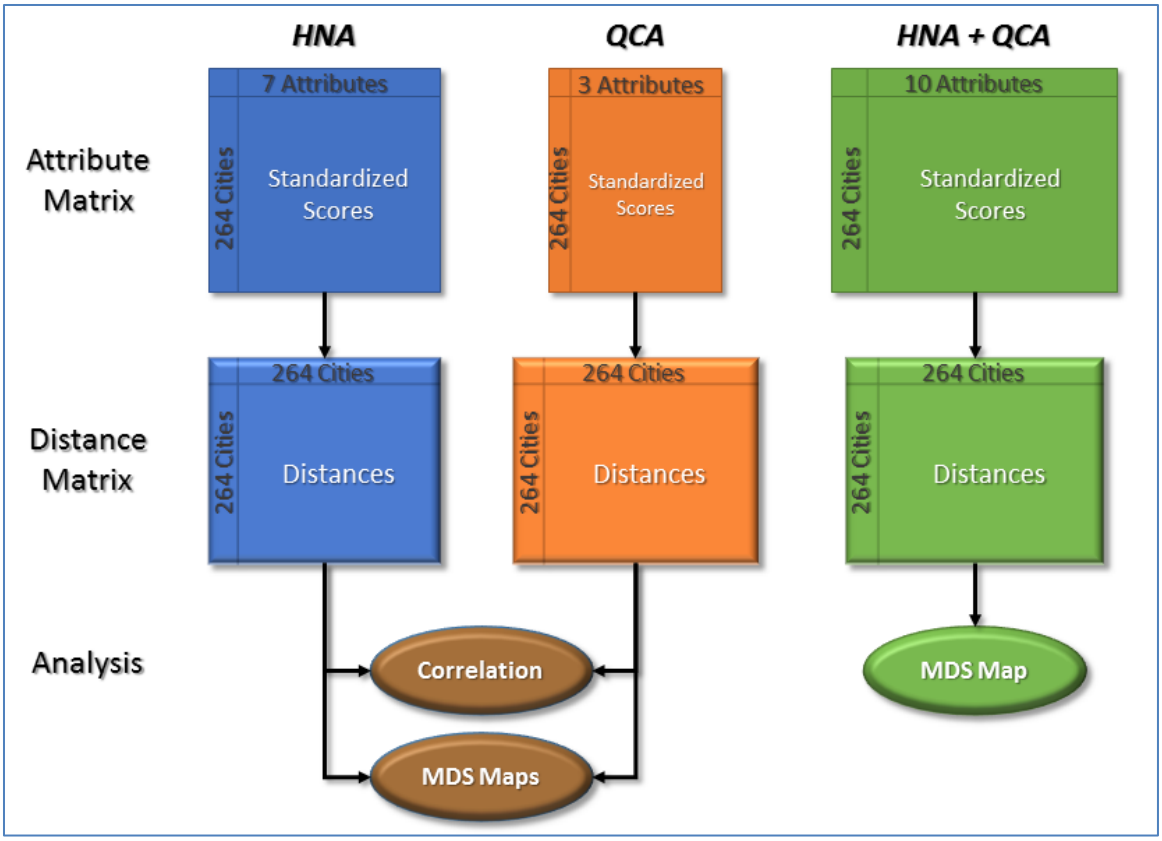

Figure 5.12: Conceptual diagram of methodology for research question 4.4

CHAPTER VI: GENERAL CHARACTERISTICS OF HYPERLINK NETWORKS

The purpose of this chapter is to reveal the usefulness of the hyperlink network data to better comprehend emerging forms of city globality through the comparison of different new measurements to existing approaches and the creation of a new hyperlink network index. The collected web data were processed in response to the problems posed in the first research question (4.1). The directed hyperlinks among webpages (i.e. nodes) serve as the basic unit of analysis, and various measurements are adjusted to calculate the characteristics of hyperlink networks. Several measurements are then used to form a new composite index that captures emerging forms of city globality.

Hyperlink network measurements are useful to design a new index for measuring globality in terms of the informational structure on the Web. The new index reveals this new perspective on cities embedded in global city networks through the internet and hyperlink networks, thus complementing the conventional perspectives. The analysis concludes that the new index based on hyperlink network analysis can be used for global city research as an index of the efficiency of sharing information and knowledge.

6.1. Hyperlink Network Measurements and Its Discrepancy

A critical point for using hyperlink network data is to determine what the relationship is between the characteristics of hyperlink networks and globality. The hyperlink between webpages is the passage for information transference, just as the roads for transportation. The efficiency of information transference on the hyperlink network is

determined by how these hyperlinks are connected to each other. This means that a city of high-connectivity and of high-centrality is in an advantageous position to secure network efficiencies to transfer information. In today's information age, access to an efficient information network is important to acquire, process, and distribute information and knowledge. Therefore, the connectivity and centrality of each city's hyperlink network may be a strong factor of a city's efficiency of information and knowledge exchange, and may therefore be an indicator of the city's globality.

As a starting point, we compute various HNA measures, including maximum in-degree, share of transmitter, share of receiver, share of carrier, share of ordinary, Assortativity, number of communities, modularity, weakly connected components, strongly connected components, share of weakly connected components, average path length, average clustering coefficient, and reciprocity. Some correlation exists among the measures (Table 6.1) and high correlations (more than |0.3|: colored cells in the table) are observed among measures computed on the basis of the same elements, that is degree-based or path-based measures.

Closer examination of the hyperlink network measures suggests that some may not be good indicators on the ability of a city to leverage its hyperlink network to position itself advantageously on global city networks. For this purpose, we undertake to compare each of the computed hyperlink network measures to a well-established measure of globality so as to identify and filter out measures that are ill-suited to the purpose of this study because empirical relationships are inconsistent with theorized expectations and should therefore not be retained. Also, retained measures should be representative of the diverse aspects of network connectivity. Finally, since the various HNA measures are

rather complementary than redundant, we propose to construct a composite measure that encompasses all the characteristics of connectivity and centrality together.

Table 6.1: Correlation coefficient matrix among network measurements

	MID	ST	SR	SC	SO	ATY	NCM	MOD	WCC	SCC	SWCC	APL	ACC	RCY
MID		0.215	-0.406	0.221	0.365	0.421	0.343	-0.678	0.323	0.615	0.227	0.060	0.092	-0.271
ST	0.215		-0.812	0.682	0.519	0.341	0.740	-0.077	0.761	-0.201	0.363	0.389	-0.247	-0.058
SR	-0.406	-0.812		-0.837	-0.877	-0.500	-0.654	0.105	-0.654	0.204	-0.344	-0.441	0.134	-0.020
SC	0.221	0.682	-0.837		0.740	0.348	0.668	0.084	0.651	-0.162	0.310	0.562	-0.270	0.057
SO	0.365	0.519	-0.877	0.740		0.564	0.410	-0.067	0.405	-0.259	0.157	0.374	0.046	0.179
ATY	0.421	0.341	-0.500	0.348	0.564		0.273	-0.364	0.247	-0.029	0.041	0.059	-0.012	0.005
NCM	0.343	0.740	-0.654	0.668	0.410	0.273		-0.154	0.977	0.176	0.381	0.495	-0.346	-0.173
MOD	-0.678	-0.077	0.105	0.084	-0.067	-0.364	-0.154		-0.140	-0.437	-0.171	0.087	0.024	0.476
WCC	0.323	0.761	-0.654	0.651	0.405	0.247	0.977	-0.140		0.151	0.427	0.494	-0.347	-0.156
SCC	0.615	-0.201	0.204	-0.162	-0.259	-0.029	0.176	-0.437	0.151		0.094	0.032	0.046	-0.449
SWCC	0.227	0.363	-0.344	0.310	0.157	0.041	0.381	-0.171	0.427	0.094		0.073	-0.347	-0.201
APL	0.060	0.389	-0.441	0.562	0.374	0.059	0.495	0.087	0.494	0.032	0.073		-0.197	-0.146
ACC	0.092	-0.247	0.134	-0.270	0.046	-0.012	-0.346	0.024	-0.347	0.046	-0.347	-0.197		0.329
RCY	-0.271	-0.058	-0.020	0.057	0.179	0.005	-0.173	0.476	-0.156	-0.449	-0.201	-0.146	0.329	

Note: MID, maximum in-degree; ST, share of transmitter; SR, share of receiver; SC, share of carrier; SO, share of ordinary; ATY, assortativity; NCM, number of communities; MOD, modularity; WCC, weakly connected components; SCC, strongly connected components; SWCC, share of weakly connected components; APL, average path length; ACC, average clustering coefficient; RCY, reciprocity

6.2. Comparison to Established Ranking of City Globality

Comparison of our hyperlink network measures to an established ranking of global cities provides a reference to select suitable hyperlink measurements for the creation of a new index. As the reference, the 2012 GCI is selected, which is based on 25 metrics categorized in 5 dimensions with different weights: business activity (30%), human capital (30%), information exchange (15%), cultural experience (15%), and political engagement (10%). From the highlights of the 2012 GCI (Kearney, 2012, p. 2), we find that New York, London, Tokyo, and Paris form the top-ranked cities; Hong Kong, Los Angeles, and Chicago are the following group of cities; and Seoul, Brussels, and Washington complete the top 10. The top 3 (New York, London, and Tokyo) or top 4

(adding Paris) cities are regularly listed as global cities in other lists based on other globalization indices as well.

For the purpose of ranking comparison, we work with the 66 cities for which the 2012 GCI is computed (all of which are also in our dataset of 264 cities), and the rank of each measurement is compared to the 2012 GCI with Spearman's rank correlation. The selected hyperlink measurements are maximum in-degree, share of receivers, share of carriers, assortativity, weakly connected components, reciprocity, and modularity (bold in Table 6.2).

Table 6.2: Correlation between 2012 GCI and hyperlink network measurements

Hyperlink Network Measurements	Correlation Coefficient with 2012 GCI	Basis	Reason for Exclusion
Maximum In-Degree	0.303	Degree	
Share of Transmitter	0.145	Degree (Type)	Low correlation coefficient
Share of Receiver	-0.268	Degree (Type)	
Share of Carrier	0.264	Degree (Type)	
Share of Ordinary	0.264	Degree (Type)	High correlation with Carrier (0.740)
Assortativity	0.269	Degree (Attribute)	
Number of Communities	0.343	Degree (Modularity)	High correlation with Weakly Connected Components (0.977)
Modularity	-0.283	Degree	
Weakly Connected Components	0.346	Path	
Strongly Connected Components	0.203	Path	Low correlation coefficient
Share of WCC	0.216	Path	Low correlation coefficient
Average Path Length	0.309	Path	Average score
Average Clustering Coefficient	-0.260	Neighboring links	Average score
Reciprocity	-0.271	Neighboring links	

Note: the listed measurements are filtered from the correlation matrix of the whole measurements. If one measurement is highly correlated with other measurements and it can be explained by other measurements, it is removed from this list.

Table 6.3: Descriptive statistics of selected network measurements

	Maximum In-Degree	Share of Receivers	Share of Carriers	Assortativity	Weakly Connected Components	Reciprocity	Modularity
Count	264	264	264	264	264	264	264
Mean	1441.117	0.910	0.005	-0.126	144.523	0.062	0.767
Median	1404.000	0.913	0.004	-0.125	135.000	0.058	0.773
Standard Deviation	387.239	0.021	0.002	0.024	93.856	0.025	0.034
Range	3117.000	0.190	0.010	0.139	1299.000	0.201	0.304
Minimum	450.000	0.756	0.002	-0.218	42.000	0.017	0.549
Maximum	3567.000	0.947	0.012	-0.079	1341.000	0.218	0.853
Skewness	1.154	-2.089	0.947	-0.551	8.384	2.513	-1.489

The descriptive statistics of the measurements are listed in Table 6.3. The following explains the characteristics of each measurement one by one.

Maximum in-degree: The maximum in-degree represents the degree of the node with the largest number of edges in a hyperlink network. A positive relationship with the GCI means that a city with a more centered node tends to be ranked higher. Having a more connected central node has an advantage to distribute and search information and knowledge.

Share of receivers and **share of carriers**: A *Receiver* is a type of nodes that has only incoming edges and no outgoing edges. The *Carrier* node type has more than one incoming edge and outgoing edge. The receiver is the dominant type of node by about 91% on average; the carrier takes a smaller portion (about 0.5% on average). This denotes that most websites are the terminal of information and knowledge. Because the share of receivers has a negative relationship with the GCI, more terminal websites in a hyperlink network means more websites only consume information and knowledge. However, the share of carriers has a positive relationship. Although it has a small portion

of website types, more information-connective websites catalyze the flow of information and knowledge.

Assortativity: Assortativity (or assortative mixing) measures the preference of a node for connecting to similar nodes based on an attribute value of the node (here, degree). In other words, high-degree nodes tend to be connected to other high-degree nodes; low-degree nodes are connected to low-degree nodes. Assortativity has a positive relationship with the GCI, which means that connection of similar degree websites makes synergy to the flow of information and knowledge.

Weakly connected components: Weakly connected components are each pair of nodes in a sub-graph that is connected by a semi-path. Because a semi-path considers only connection between nodes, and not the direction of this connection, websites in weakly connected components act like a bridge between components compared to directional deliverers in strongly connected components. A positive relationship with the GCI denotes that *Weak Ties* (Granovetter, 1973) is also important in hyperlink networks as informational connections.

Reciprocity: The reciprocity here is based on the arc method, which measures the ratio of the number of reciprocated relational links to the total number of links. A high ratio means that the network has many reciprocated links compared to the link size of the network. That is, websites with high reciprocity have a chance to connect each other more closely compared to websites with low reciprocity. Regarding information diffusion, low reciprocity is better to deliver information from one website to another.

Modularity: It measures how well a network decomposes into modular communities. A high modularity score indicates a sophisticated internal structure.

Hyperlink networks with high modularity have dense connections between websites within modules but sparse connections between websites in different modules. A negative relationship with the GCI indicates that a low modularity score is relatively better for informational connection of the whole hyperlink network.

In summary, the following characteristics of hyperlink networks are positively related with the CGI ranking: having a powerful center of informational hub (maximum in-degree), carrier-type of websites instead of receiver-type (share of carrier and share of receiver), connective preference of similar degree (assortativity), weakly connected components as informational bridges, low reciprocity for information diffusion, and dense connection between modules (i.e. low modularity). The relationship between the GCI and the selected measurements supports that hyperlink networks are closely related with effectiveness of informational connections and diffusions in its network.

6.3. Composite Global City Hyperlink Index

When we consider the selected hyperlink network measurements corresponding to the effectiveness of informational distribution in hyperlink networks, a city having the better hyperlink network for informational connection has an advantage over other cities in terms of visibility and informativity of the city on the Web. We propose to construct a new composite Global City Index for this purpose. A high score on this new index indicates that the city has a more connected and centralized hyperlink network so that it can be more exposed, searched, and informed to web users.

The process of creating the new composite index involves the following steps: 1) normalization of each hyperlink network measure; 2) factor analysis of normalized scores, defining factors, and calculating factor scores; 3) multiplying each Eigenvector

and each factor score for weighting factor scores; 4) summation of weighted factor scores; 5) conversion of the raw scores into relative scores based on the 1st ranker (index score: 100) and the 264th ranker (index score: 0).

Hyperlink network measurements selected in the previous section are used in this process. Factor analysis is used to manage the correlation between measures and to extract the hidden common factors. Here, 4 factors are retained to account for a sufficient proportion of the total variance (80.7%). The final score of the index is calculated by the summation of these 4 factors.

Table 6.4 contains the results of the global city hyperlink index (GHI). If we focus on the top 30 percent (i.e. about 79 cities), North American cities are found to be the top of continental distribution (about 34 percent, 27 cities). Asian cities are the next group with about 30 percent (24 cities). Next, 14 European cities are listed (about 18 percent). 5 cities from each of Africa and Oceania, and 2 cities of Central America are listed; and 1 city from each of the Middle East and South America is listed. The distribution of cities at country level shows: the United States (23 cities), China (11), India (9), the United Kingdom (6), Canada (4), South Africa (4), Australia (4), 1 city from each of Ethiopia, Singapore, Indonesia, Malaysia, Thailand, Guatemala, Nicaragua, France, Germany, Greece, Hungary, Ireland, Netherlands, Romania, Russia, the United Arab Emirates, New Zealand, and Brazil.

The distribution of high ranking cities by country underscores the deep relationship between the results and the basic characteristics of the data. When we consider the number of internet users in the world, China, the United States, and India are placed in the top 3 in 2012 (Internet World Stats, 2012). Not only the number of internet

users, but the effect of English language also can be found. Except for China, all other countries listed more than 2 cities (the US, India, the UK, Canada, South Africa, and Australia) recognize English language as official language. The classification based on the GHI also supports the above analysis (Figure 6.1). We can observe the concentration of the top group of cities in North America, the United Kingdom, India, China, South Africa, and Oceania.

Table 6.4: City ranking based on the new hyperlink index (GHI)

Rank	City	GHI	Rank	City	GHI	Rank	City	GHI	Rank	City	GHI
1	London	100.00	67	Cleveland	48.81	133	Kunming	37.37	199	The Hague	24.35
2	Glasgow	98.53	68	Saint Louis	48.60	134	Asunción	37.27	200	Guangzhou	24.25
3	Shanghai	81.25	69	Honolulu	48.45	135	Seattle	37.27	201	Durban	24.22
4	Liverpool	80.45	70	Cape Town	47.87	136	Minsk	36.58	202	Seoul	24.00
5	Johannesburg	80.32	71	Budapest	47.62	137	Nairobi	36.28	203	Naples	23.95
6	Edinburgh	80.28	72	Foshan	47.51	138	Lima	35.73	204	Kobe	23.83
7	Chennai	79.64	73	Managua	46.64	139	Surat	35.55	205	Quito	23.71
8	Denver	77.45	74	Fortaleza	46.61	140	Dakar	35.14	206	Lahore	23.62
9	Mumbai	76.89	75	Saint Petersburg	46.51	141	Maputo	34.66	207	Freetown	23.39
10	Singapore	72.99	76	Shantou	45.84	142	Shenyang	34.61	208	Algiers	23.27
11	Hyderabad	72.71	77	Suzhou	45.30	143	Shenzhen	34.51	209	Brasília	23.23
12	Guatemala City	72.32	78	Detroit	45.16	144	Richmond	34.24	210	Guadalajara	23.15
13	Xiamen	71.24	79	Addis Ababa	45.04	145	Beirut	34.12	211	San Jose	22.91
14	Bangalore	69.77	80	Nagoya	44.63	146	Munich	34.11	212	Karachi	22.72
15	Sydney	68.60	81	Baltimore	44.56	147	Copenhagen	34.00	213	Cologne	22.55
16	Atlanta	66.41	82	Vancouver	44.19	148	Belgrade	33.68	214	Havana	22.54
17	Charlotte	65.54	83	Kuwait City	44.06	149	Lucknow	33.37	215	Santiago	22.46
18	New Delhi (Delhi)	65.46	84	Prague	44.01	150	Buffalo	33.32	216	Kabul	22.41
19	Edmonton	63.41	85	Dar es Salaam	43.85	151	Rome	33.13	217	Mombasa	21.93
20	Pune	62.98	86	Batam	43.45	152	Sofia	33.09	218	Oslo	21.82
21	Calgary	62.56	87	Xi'an	43.06	153	Luxembourg	32.97	219	Antwerp	21.78
22	Houston	61.14	88	New Orleans	43.05	154	Busan	32.80	220	Hamburg	21.74
23	Dalian	60.96	89	Nanjing	42.61	155	Chengdu	32.51	221	Sacramento	21.31
24	Chicago	60.95	90	Warsaw	42.53	156	Valencia	32.24	222	Yangon	20.92
25	Wuxi	60.73	91	Lisbon	42.44	157	Stuttgart	32.22	223	Kinshasa	20.84
26	Dongguan	60.72	92	Beijing	42.40	158	Bordeaux	32.11	224	Qingdao	20.76
27	Perth	60.20	93	Amsterdam	41.98	159	Chittagong	31.66	225	Ulan Bator	20.11
28	Ahmedabad	59.66	94	Tbilisi	41.73	160	Istanbul	31.42	226	Kiev	20.09
29	Hong Kong	59.62	95	Stockholm	41.39	161	Cairo	31.40	227	Rabat	20.04
30	Omaha	59.03	96	Portland	41.34	162	Moscow	31.34	228	Bogotá	19.98
31	Boston	59.00	97	Barcelona	41.11	163	Tegucigalpa	31.12	229	Ciudad Juarez	19.49
32	East Rand	58.85	98	Tokyo	41.10	164	Basel	31.06	230	Ankara	19.35
33	Dublin	58.79	99	Krakow	41.07	165	Frankfurt	30.76	231	Dhaka	19.25
34	Harbin	57.97	100	Tashkent	40.99	166	Geneva	30.67	232	São Paulo	18.75
35	Bangkok	57.78	101	Vienna	40.92	167	San Francisco	30.63	233	Monterrey	18.53
36	Phoenix	57.66	102	Changchun	40.72	168	Lusaka	30.23	234	Islamabad	18.09
37	Minneapolis	56.48	103	Belo Horizonte	40.43	169	Riyadh	29.73	235	Medellin	18.08
38	Cincinnati	55.92	104	Dallas	40.28	170	Yerevan	29.61	236	Casablanca	18.06
39	Columbus	55.76	105	Colombo	40.15	171	Milan	29.36	237	Hanoi	17.77
40	Tampa	55.52	106	Accra	39.91	172	Marseille	29.24	238	Chongqing	17.70
41	Jinan	55.26	107	Zurich	39.86	173	Tunis	29.19	239	Montevideo	17.47
42	Rochester	54.82	108	San José	39.85	174	Caracas	28.78	240	Changsha	17.38
43	Jaipur	54.74	109	Newcastle	39.84	175	Brussels	28.76	241	Mexico City	16.05
44	Las Vegas	54.49	110	Salvador	39.67	176	Rio de Janeiro	28.46	242	Lyons	14.80
45	Toronto	53.82	111	Recife	39.59	177	San Salvador	27.85	243	Tijuana	14.44
46	Philadelphia	53.82	112	Manila	39.58	178	Washington	27.64	244	Pyongyang	13.79
47	Kolkata (Calcutta)	53.76	113	Douala	39.24	179	Bonn	27.46	245	Conakry	13.69
48	Pretoria	53.64	114	Wuhan	39.10	180	Hangzhou	27.39	246	Tehran	13.47
49	Bucharest	53.02	115	Tel Aviv	38.94	181	Taipei	27.31	247	Almaty	13.23
50	Ottawa	52.86	116	New York	38.92	182	Abu Dhabi	27.30	248	Curitiba	13.14
51	Athens	52.74	117	Kansas City	38.90	183	Lagos	27.11	249	Kyoto	12.83
52	Miami	52.21	118	Jeddah	38.85	184	Medan	27.07	250	Amman	12.53
53	Birmingham	51.71	119	Palermo	38.62	185	Santo Domingo	27.06	251	Porto Alegre	11.80
54	Manchester	51.46	120	Düsseldorf	38.62	186	Kampala	27.05	252	Baghdad	11.44
55	Auckland	51.16	121	Adelaide	38.44	187	Taiyuan	26.80	253	Los Angeles	11.05
56	Brisbane	51.04	122	La Paz	38.41	188	Turin	26.61	254	Harare	10.60
57	Paris	50.88	123	Bandung	38.37	189	Buenos Aires	26.60	255	Port-au-Prince	9.89
58	Melbourne	50.82	124	Madrid	38.19	190	Abidjan	26.58	256	Brazzaville	9.54
59	Indianapolis	50.75	125	Doha	38.17	191	Osaka	26.32	257	Kano	9.09
60	Dubai	50.73	126	Hartford	38.03	192	Tianjin	26.29	258	Damascus	8.94
61	San Diego	50.42	127	Zhengzhou	37.76	193	Guayaquil	26.14	259	Aleppo	8.17
62	Pittsburgh	50.29	128	Yaoundé	37.66	194	Panama City	26.07	260	Monrovia	7.24
63	Kuala Lumpur	50.05	129	Ho Chi Minh City	37.65	195	Baku	25.63	261	Rawalpindi	3.69
64	Berlin	49.95	130	Luanda	37.64	196	Helsinki	25.48	262	Montreal	2.78
65	Rotterdam	49.45	131	Manaus	37.56	197	Lille	25.43	263	Tripoli	1.63
66	Jakarta	49.38	132	Jerusalem	37.49	198	Alexandria	24.88	264	Khartoum	0.00

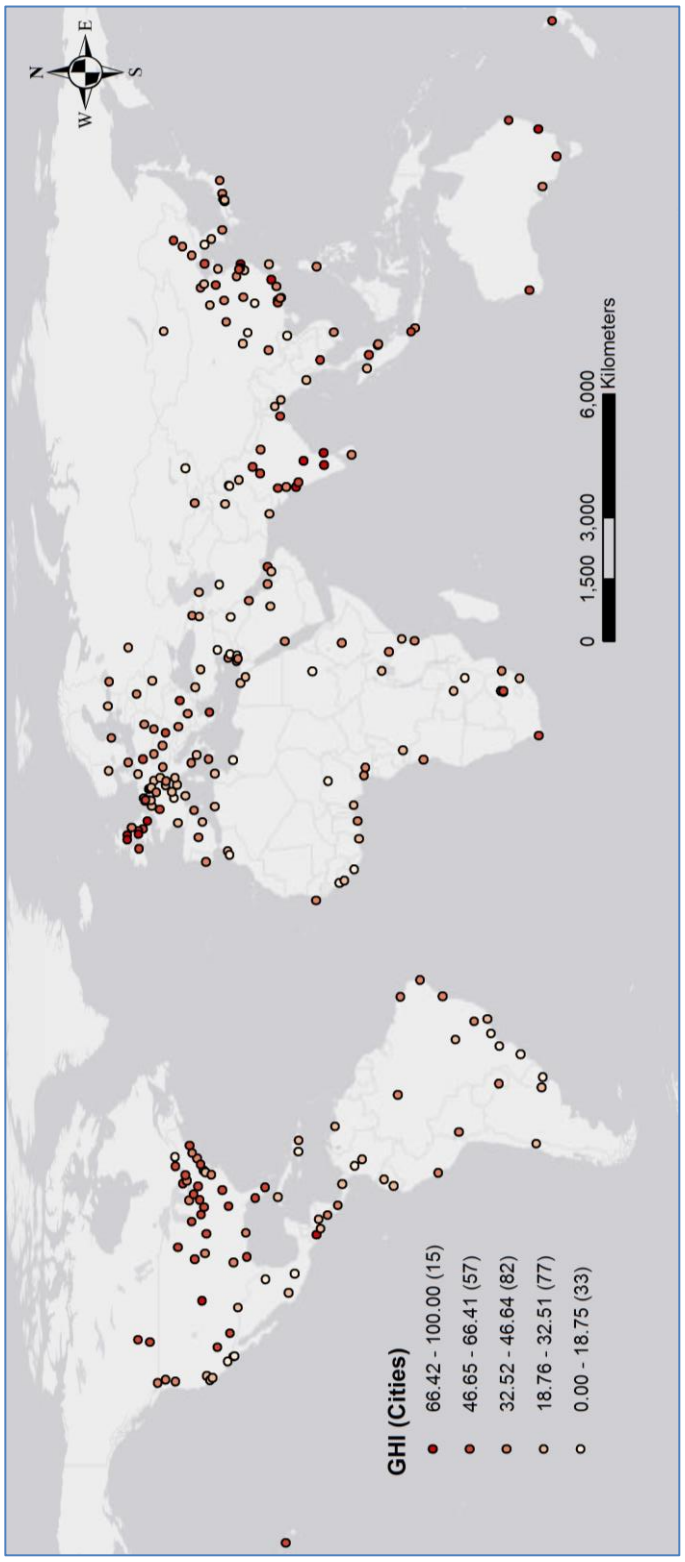

Figure 6.1: Classification of cities based on GHI

Table 6.5 is a comparison between the GHI and 2012 GCI, where 66 cities are grouped into three classes on each of the indices (High, Medium, and Low; each class has the same number of cities, 22 cities). The relations of High & High, Medium & Medium, and Low & Low (i.e. diagonal cells shaded in gray) account for a total of 26 cities (39.4%), which denotes the correspondence between the two indices. Cities marked by high discordance between the GHI and GCI scores (i.e. High GHI & Low GCI and Low GHI & High GCI) are useful to reveal the difference between two indices. First, High GHI & Low GCI cities are all Indian cities except for Johannesburg. This result supports the own characteristics of the Web data (i.e. the effect of the number of the Internet users) on the hyperlink-based index. It also explains that cities in this group are more visible and searchable on the Web than the values based on other indicators that compose the GCI. Second, Low GHI & High GCI cities (Brussels, Buenos Aires, Los Angeles, San Francisco, Seoul, and Washington) denote that these cities are high in other parts of globalization and low in hyperlink-based globality. That is, these cities are less appreciated on the Web than the values of indicators in the GCI. Other relations (High GHI & Medium GCI, Medium GHI & Low GCI, Medium GHI & High GCI, and Low GCI & Medium GCI) explain the relative degree of this discrepancy between the GHI and the GCI. In sum, the discrepancy provided in this table tells the existing global city index overlooks the effect of the Web. The new hyperlink-based global city index helps the detection of the relative visibility of cities on the Web which could not be measured by the existing global city index.

Table 6.5: Comparison between GHI and 2012 GCI

	High 2012 GCI	Medium 2012 GCI	Low 2012 GCI
High GHI	Boston, Chicago, Hong Kong, London, Paris, Shanghai, Singapore, Sydney, Toronto	Atlanta, Bangkok, Dubai, Dublin, Houston, Melbourne, Miami	Bangalore, Johannesburg, Kolkata (Calcutta), Kuala Lumpur, Mumbai, New Delhi (Delhi)
22	9	7	6
Medium GHI	Beijing, Berlin, Madrid, Moscow, New York, Tokyo, Vienna	Amsterdam, Barcelona, Copenhagen, Istanbul, Munich, Rome, Stockholm, Zurich	Cairo, Ho Chi Minh City, Jakarta, Manila, Nairobi, Shenzhen, Tel Aviv
22	7	8	7
Low GHI	Brussels, Buenos Aires, Los Angeles, San Francisco, Seoul, Washington	Frankfurt, Geneva, Mexico City, Milan, Montreal, São Paulo, Taipei	Bogotá, Caracas, Chongqing, Dhaka, Guangzhou, Karachi, Lagos, Osaka, Rio de Janeiro
22	6	7	9
Total	22 cities	22 cities	22 cities
66			

CHAPTER VII: CHARACTERISTICS AND CLASSIFICATION OF NODES

The measurement of the hyperlink networks in the previous chapter brings the question of the origin of these network characteristics. Characteristics of each node influence the whole hyperlink network. Therefore, the study of the characteristics of individual nodes helps to understand the characteristics of the whole hyperlink network as well as the general characteristics of hyperlink data.

This chapter consists of three parts. The first part is a general overview of the characteristics of nodes with high centrality through the distributions of website domains and types. The second part reports on the analysis of the top centrality websites based on the different centrality measures of city hyperlink networks. The last part defines premium nodes based on PageRank scores and comprehensiveness.

We find the distinguishing characteristics of the data in this research through the classification of websites. The hyperlink network of each city shares common websites for its top centrality website. The 'premium' websites throughout the hyperlinks of global cities are confirmed through the comprehensiveness of the high-PageRank score websites. We conclude that having linkages to those premium websites are important to enhance the efficiency of the hyperlinks for the exchange of information and knowledge.

7.1. Overview of High Centrality Websites

The distributional characteristics of high-centrality websites provide an overview of the data. Here, the analysis is based on the websites of cities whose in-degrees (i.e.

incoming links) are larger than 100 (6,438 websites in total), which is the knee point of the in-degree distribution. Figure 7.1 shows the distribution of the top-level domains of these websites. This distribution is compared to world averages as reported in *Usage of top level domains for websites* (W3Techs, 2014). *W3Techs* provides daily reports based on the top 10 million websites from 3 month average rankings of *Alexa* (Web analytic company of *Amazon.com*). Many websites are inactive or have duplicated and auto-created content for URL grabbers and spammers. This creates problems that can be circumvented by focusing on the top ranking websites; in addition, the report post-processes data to reduce bias. The distribution of domains from *W3Techs* can be summarized in decreasing order as .com (52.6%), country-specific domain (34.3%), .net (5.5%), .org (4.1%), .info (1.5%), .biz (1.0%), and .gov (less than 0.1%).

Our data shows a larger percentage in .com, .org, .gov, and .biz than in the population of websites, as reported by *W3Techs*. Country-specific domains, .net, and .info occur in relatively smaller percentages than in the population at large. Many .org and .gov websites in our data can be understood as the providers of city information are organizations and governments. A high share of .com (or .biz) and a low share of country-specific domains indicate that the websites containing city information prefer global domains to local domains.

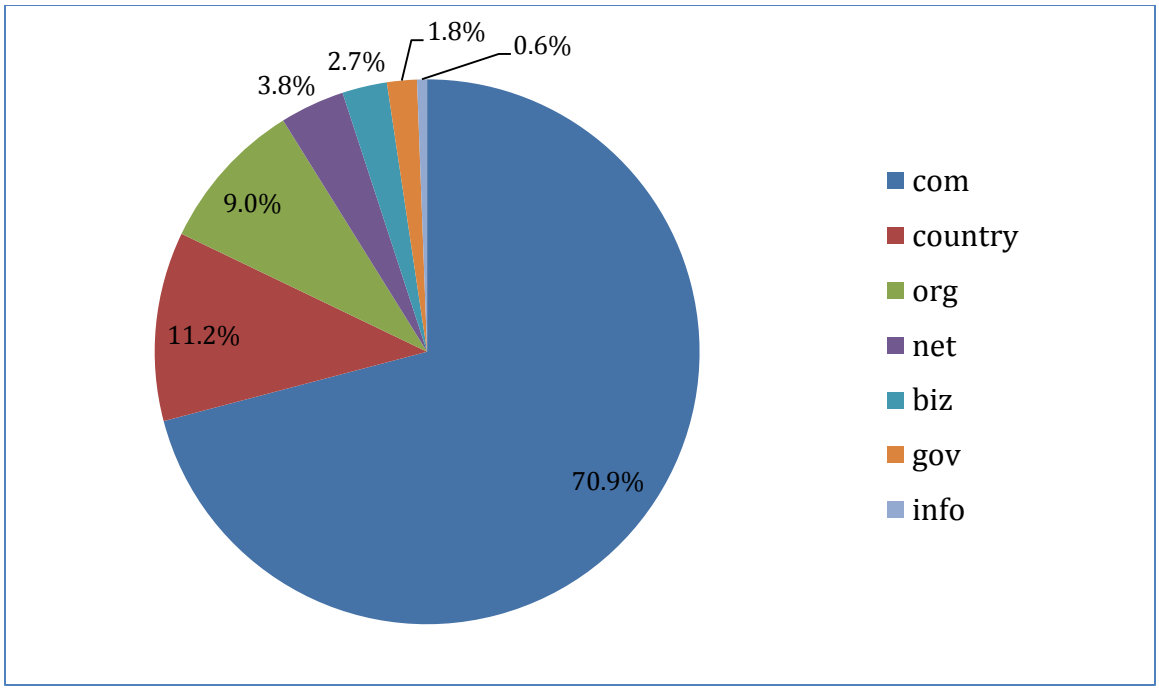

Figure 7.1: Distributions of website top-level domains (in-degree over 100)

The type of websites provides a deeper understanding of our data than the domain of websites because the classification of website types is based on the purpose and the contents of the website. It is true that a certain website can be hard to classify because of the multiple purposes and the blurred boundary of its contents. In spite of that, the types of websites help us understand the distributional characteristics of high centrality websites pertaining to our selected cities. From figure 7.2 we can find that corporate websites, electronic commerce sites, and news sites form 82% of the total. This means that these three types are dominant types of websites. Government sites (4%), information sites (4%), and media sharing sites (2%) follow, but the percentage is much lower than for the top 3 types; and other types are even less important. The strong presence of corporate, e-commerce, and news sites does not guarantee that these websites have a critical role in the network they belong to because the role depends on the network

properties of the node, not its type. The following sections will seek to address this question further.

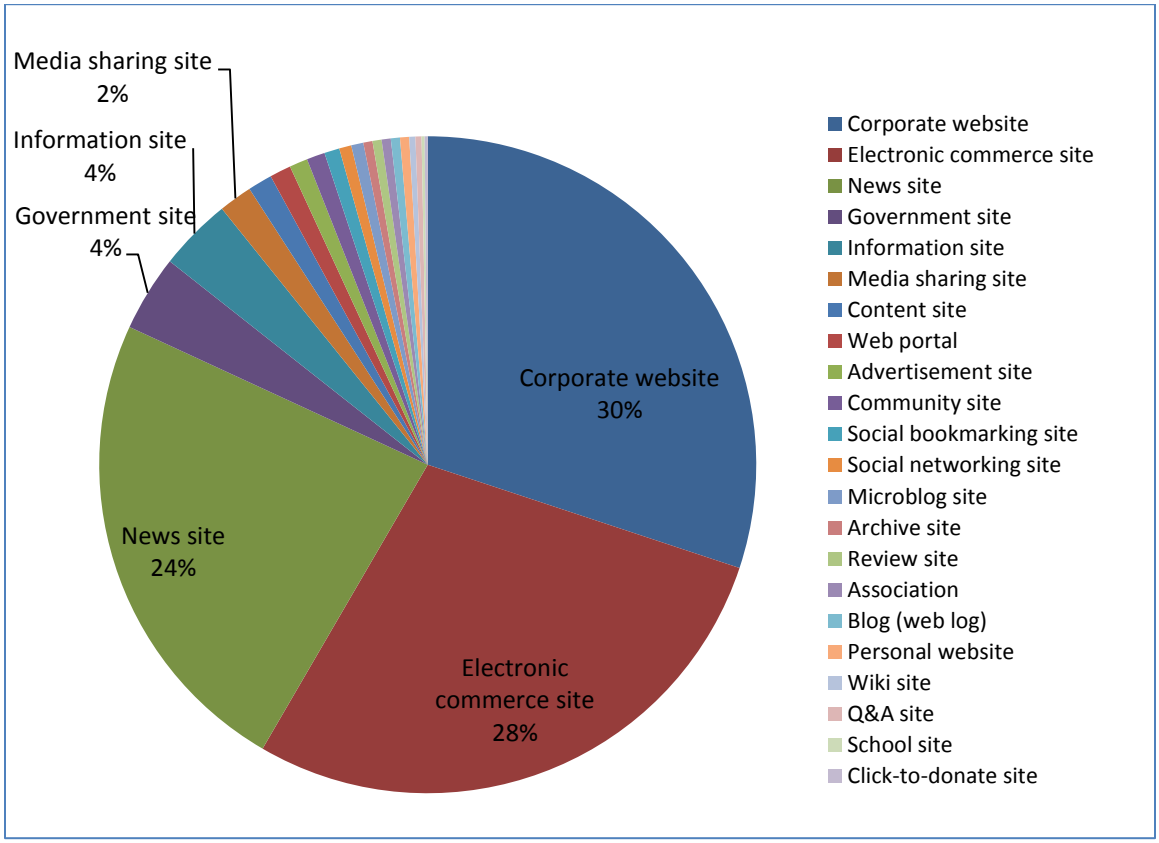

Figure 7.2: Distribution of website types (in-degree over 100)

7.2. Centrality of Nodes

The centrality of a node can be measured in various ways, including degrees, the shortest paths, the role as a bridge, etc. Table 7.1 lists the frequency of nodes (i.e., web sites) with the single highest centrality[17] based on in-degree centrality and in-closeness centrality. Only two distinct nodes are found to be the node with highest centrality in any of the 264 cities under study, namely *twitter.com* and *facebook.com*. In terms of in-degree centrality, these two nodes are the nodes most often pointed to from other nodes in a network. In terms of in-closeness centrality, these two nodes have the shortest distance to

[17] A complete list of the top centrality nodes of each city is in Appendix C.

other nodes in the network. About 85.6% (226 cities, in-degree centrality) and about 72.7% (192 cities, in-closeness centrality) of cities have *twitter.com* as the top-centrality node. The remaining cities have *facebook.com* as their most central node. This result denotes that representative social networking websites, *twitter.com* and *facebook.com* are located at the center of all the networks. This also represents the dominance of information and knowledge exchange in the present era of the Internet.

Table 7.1: Top website by in-degree centrality and in-closeness centrality

Node (i.e. website)	Count by In-Degree Centrality	Count by In-Closeness Centrality
twitter.com	226 (85.6%)	192 (72.7%)
facebook.com	38 (14.4%)	72 (27.3%)
Total	264 (100%)	264 (100%)

Considering that data in this research are directional, we can also analyze the reverse direction: out-degrees. Table 7.2 shows the top website by out-degree centrality, out-closeness centrality, and node betweenness centrality. In short, high out-degree centrality means that the node has many out-going edges. That is, the listed websites have many hyperlinks to connect themselves to other websites. Nowadays, so-called knowledge treasure box, *Wikipedia.org* has the top centrality in out-degree based centrality measurements. This is understandable because *Wikipedia* contains a lot of hyperlinks for cited references and other information. In addition, websites for news (*huffingtonpost.com*, *topix.com*, *globalvoicesonline.org*, and *wn.com*), weather (*weatherforecastmap.com*), Q&A (*ask.com*), and other information portals (*enotes.com* and *mycountdown.org*) are the top websites identified for more than 2 cities.

Node betweenness centrality is useful to find the important node in a network

because it tallies the number of shortest paths that pass through that node. The node with high node betweenness centrality is here the website with high loading of shortest paths; the role of this website is important to connect pairs of nodes. On the basis of node betweenness centrality, *Wikipedia.org* is the top ranked website in more than 70% of the cases, and *twitter.com* is the top website for 25%. Two news websites (*huffingtonpost.com* and *washingtonpost.com*) are the next highest ranked.

Table 7.2: Top websites by out-degree centrality, out-closeness centrality, and node betweenness centrality

Node (i.e. website)	Count by Out-Degree Centrality
en.wikipedia.org	252 (95.5%)
weatherforecastmap.com	3 (1.1%)
huffingtonpost.com	3 (1.1%)
travigator.com	1 (0.4%)
enotes.com	1 (0.4%)
usnpl.com	1 (0.4%)
blogger.com	1 (0.4%)
schema-root.org	1 (0.4%)
qfkd.com	1 (0.4%)
Total	264 (100%)

Node (i.e. website)	Count by Out-Closeness Centrality
en.wikipedia.org	201 (76.1%)
huffingtonpost.com	13 (4.9%)
topix.com	11 (4.2%)
ask.com	6 (2.3%)
globalvoicesonline.org	3 (1.1%)
wn.com	3 (1.1%)
enotes.com	2 (0.8%)
mycountdown.org	2 (0.8%)
southafrica.info	1 (0.4%)
ottawastart.com	1 (0.4%)
theproudfranchise.com	1 (0.4%)
december.com	1 (0.4%)
scam.com	1 (0.4%)
fabiocaparica.com	1 (0.4%)
thecomingcrisis.blogspot.com	1 (0.4%)
faustasblog.com	1 (0.4%)
nationsonline.org	1 (0.4%)
avoelectronic.blogspot.com	1 (0.4%)
qfkd.com	1 (0.4%)
bangkokcompanies.com	1 (0.4%)
schema-root.org	1 (0.4%)
usnpl.com	1 (0.4%)
sxl.net	1 (0.4%)
abc-directory.com	1 (0.4%)
thefullwiki.org	1 (0.4%)
blogs.wsj.com	1 (0.4%)
absoluteastronomy.com	1 (0.4%)
calcna.ab.ca	1 (0.4%)
mypetjawa.mu.nu	1 (0.4%)
lawpundit.com	1 (0.4%)
livingingreece.gr	1 (0.4%)
Total	264 (100%)

Node (i.e. website)	Count by Node Betweenness Centrality
en.wikipedia.org	186 (70.5%)
twitter.com	66 (25.0%)
huffingtonpost.com	7 (2.7%)
washingtonpost.com	2 (0.8%)
blogger.com	1 (0.4%)
youtube.com	1 (0.4%)
itunes.apple.com	1 (0.4%)
Total	264 (100%)

Overall, we can identify the shared characteristics of the websites with the highest centrality. The city hyperlink networks are highly affected by a few common websites, which are for social networking, news watching, and knowledge retrieving. These

common websites with high centrality play the important role of connecting other websites.

7.3. Premium Nodes

Here, premium nodes distinguish themselves from other high-centrality websites by comprehensiveness, which is calculated by how many times the websites ranked as the top PageRankers. The recent history of network analysis intersects with algorithmic advances of search engines. PageRank is at the core of Google Search, the most popular search engine algorithm, which is based on the number and the quality of links. A website with a high PageRank score indicates the websites has a lot of high quality links. That is, it denotes the important and central websites in a hyperlink network. Therefore, using PageRank is reasonable to detect and analyze the premium node in this research.

Table 7.3 lists the top 30 websites by its comprehensiveness. This list is based on the website list of each city, which ranked all websites by PageRank score. As far as comprehensiveness is concerned, social networking websites (*twitter.com* and *facebook.com*) and media sharing websites (*youtube.com*) take the highest comprehensiveness. This means that every hyperlink network indicates these three website are premium nodes. Social bookmarking service (*addthis.com*), computer software company (*adobe.com*), social networking websites (*linkedin.com*, *plus.google.com*, *nytm.org*), and web search engine (*google.com*) are listed for more than 30% of cities as central nodes.

Similar to the results of the in-degree based centrality analysis in the previous section, *twitter.com* and *facebook.com* are ranked as premium nodes found in every city. These two websites also emerge as the sites with the highest average and maximum

PageRank scores. Compared to *twitter.com* and *facebook.com*, *youtube.com* has much lower maximum and average PageRank scores (Figure 7.3). This denotes that *youtube.com* can be found in every city network although it is not the top PageRank scorer for all cities. Therefore, we should consider these three websites as the final premium nodes in all the hyperlink city networks. Especially, *twitter.com* and *facebook.com* should be considered as the most powerful premium nodes.

The meaning of the discovered common premium nodes from PageRank is that the hyperlink networks are controlled by these premium nodes. In other words, the informational hyperlink network of each city is strongly affected by the possession of connections to these premium nodes. Therefore, for a city to garner a high score in information exchange depends on whether local websites have many connections to these premiums nodes. A city would have more chances to enhance it informational network if it is exposed in the premium websites. Thus, this research is important in identifying each city's central nodes (not only *twitter.com* and *facebook.com*) and where they should invest their resources to enhance their visibility on the web.

Table 7.3: Descriptive statistics of top 30 URLs based on comprehensiveness (only contain the websites having PageRank score over 0.0001)

Rank by Comprehensiveness	URL	n	Max	Min	Mean	SD	Comprehensiveness (n/264, %)
1	twitter.com	264	0.002977	0.000444	0.000813	0.000245	100.0
1	facebook.com	264	0.002565	0.000352	0.000739	0.000263	100.0
1	youtube.com	264	0.001131	0.000138	0.000307	0.000123	100.0
4	addthis.com	186	0.000396	0.000100	0.000136	0.000035	70.5
5	adobe.com	149	0.000403	0.000100	0.000157	0.000060	56.4
6	linkedin.com	123	0.000449	0.000100	0.000154	0.000053	46.6
7	plus.google.com	104	0.000598	0.000100	0.000140	0.000064	39.4
8	nytm.org	83	0.002180	0.000101	0.000627	0.000496	31.4
9	google.com	81	0.000232	0.000100	0.000127	0.000024	30.7
10	fourmilab.ch	60	0.000381	0.000102	0.000277	0.000049	22.7
10	blogger.com	60	0.000304	0.000102	0.000141	0.000037	22.7
12	astroviewer.com	59	0.000381	0.000171	0.000278	0.000043	22.3
12	lunaf.com	59	0.000381	0.000171	0.000277	0.000041	22.3
14	maps.google.com	45	0.000520	0.000101	0.000193	0.000084	17.0
15	go.microsoft.com	43	0.000829	0.000100	0.000158	0.000112	16.3
16	yaml.de	41	0.000201	0.000101	0.000126	0.000021	15.5
16	matussek.com	41	0.000167	0.000101	0.000125	0.000017	15.5
16	ad.doubleclick.net	41	0.000215	0.000101	0.000123	0.000023	15.5
19	itunes.apple.com	39	0.000152	0.000100	0.000117	0.000015	14.8
20	airbnb.com	34	0.002895	0.000104	0.000330	0.000463	12.9
21	wwp.greenwichmeantime.asia	32	0.000212	0.000113	0.000170	0.000028	12.1
22	weatherforecastmap.com	28	0.000257	0.000102	0.000134	0.000032	10.6
22	flickr.com	28	0.000194	0.000100	0.000120	0.000021	10.6
24	miibeian.gov.cn	22	0.000348	0.000111	0.000177	0.000059	8.3
25	t.co	21	0.000319	0.000100	0.000137	0.000048	8.0
26	get.adobe.com	17	0.000261	0.000111	0.000139	0.000037	6.4
27	bit.ly	14	0.000318	0.000101	0.000134	0.000053	5.3
28	focuschina.com	13	0.000758	0.000100	0.000348	0.000249	4.9
29	wordpress.org	11	0.000173	0.000100	0.000115	0.000020	4.2
30	en.wikipedia.org	10	0.000223	0.000120	0.000149	0.000030	3.8

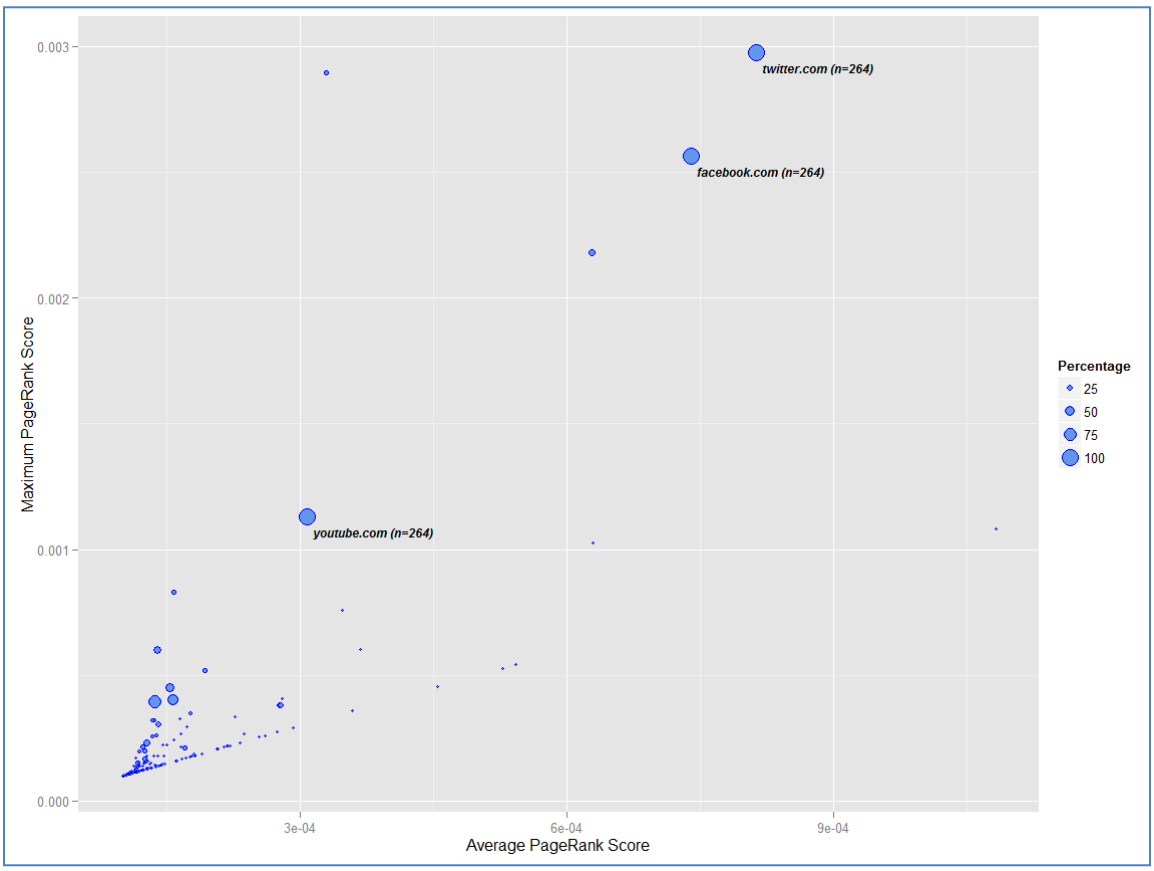

Figure 7.3: Average and maximum PageRank score of city networks, and the percentage of comprehensiveness (only contain the website having PageRank score over 0.0001)

CHAPTER VIII: CHARACTERISTICS OF QUANTIFIED TEXT OF WEBPAGES

The purpose of this chapter is to study a new approach for global city research based on the quantified text of webpages. Considering that webpages contain a lot of information, knowledge, and description about cities, it is argued that the quantified textual contents can be used for extracting the characteristics of cities depicted on the Web and for comparing a city's characteristics to other cities. The huge amount of textual information on the Web and its fast accumulation hinders utilizing this information for global city research. However, the approach proposed in this chapter provides a possible method for global city research through the quantification of textual data from webpages and dimensional reduction technique.

For this purpose, we follow the general process of quantitative content analysis. The QCA includes the preparation of code schemes (i.e. the creation of dictionaries), frequency analysis, and the interpretation of the frequency based on code schemes. The conceptual process of frequency extraction is depicted in Figure 8.1. Webpages are limited by the inclusion of globalization-related words. These webpages are parsed and lemmatized into sets of words. These sets are then compared to predefined categories of keywords that match dimensions of globalization in order to calculate keyword frequencies.

This chapter consists of the selection of words for designing code schemes, distribution of frequency, and the calculation of similarity among global cities. The first

part explains how to filter the textual data and how to create code schemes for three selected thematic categories of keywords. The second part analyzes the tendency of global cities based on the frequency of keywords in each category. Lastly, the similarity among cities is analyzed through MDS maps.

The creation of predefined word lists for filtering webpages from raw data serves to focus content analysis on globalization-related data. Three categories of keywords are distinguished to reflect representative dimensions of globalization and to help differentiate cities. Frequency analysis reveals the tendencies of cities with respect to each thematic dimension. Finally, the geographic characteristics of textual contents are analyzed through the MDS maps based on similarities of word category profiles.

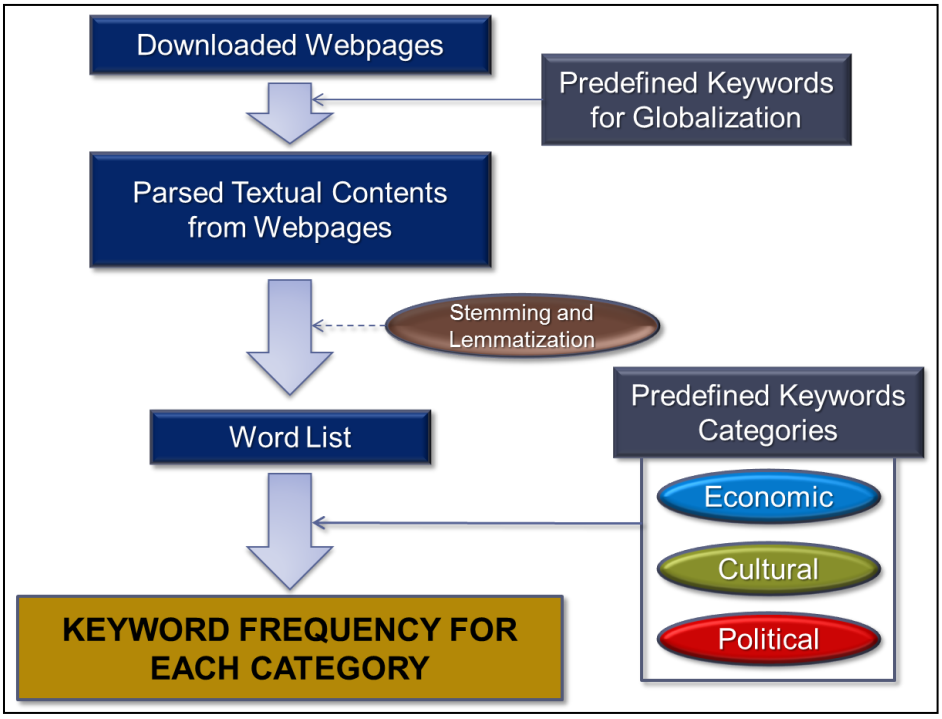

Figure 8.1: Procedure of word frequency extraction

8.1. Selection of Words for Code Scheme Design

The quantification of textual data is a task as challenging as webpage crawling because it handles a huge volume of data. Because a lot of webpages contain information

unrelated to globalization, it is appropriate to select and analyze the webpages that have a content related to globalization only. A critical question is how to sort out webpages with globalization content from the others.

Similarly, categorical code schemes also need proper reference. Keywords which are used for extracting globalization-related webpages and for calculating the frequency of each category are collected from articles relevant to globalization from The Global Policy Forum (2014a, 2014b, 2014c, 2014d). The Global Policy Forum archives articles under the following classification: definition of globalization, globalization of economy, globalization of cultures, globalization of the politics, among others. The final selection of keywords extracted from these articles is made after a close review of the list of frequent keywords from each set of articles. The final list is given in Table 8.1. In the first row of the table, keywords for Globalization are used for extracting globalization-related webpages. Keywords in other rows are used for calculating the frequency.

Table 8.1: Keywords for globalization and at its three dimensions

	Keywords
Globalization (16)	world, universal, ubiquitous, transnational, systemic, solidarity, overseas, national, localizing, international, growth, global, develop, deglobal, cross-border, abroad
Economic (44)	bank, business, capital, company, corporate, currency, debt, demand, develop, dollar, economy, employer, export, finance, firm, flow, fund, gdp, goods, income, industry, inequality, interests, investment, investors, labor, market, monetary, money, pay, poverty, price, product, profit, rates, sectors, services, subsidies, supply, taxes, trade, union, wealth, work
Cultural (25)	church, communication, community, culture, diversity, education, English-language, heritage, humanitarian, indigenous, information, intellectual, language, media, Olympic, religion, revolution, school, society, sports, subculture, television, UNESCO, war, web
Political (50)	activist, administration, capitalism, capitalist, commission, conflict, congress, cooperation, corruption, council, crisis, davos, democracy, democratic, deregulation, diplomatique, elections, g20, governance, government, humanitarian, institutions, intervention, justice, leadership, leftwing, liberal, military, minister, nation-state, neoliberal, ngo, organiz(s)ation, peace, policy, policymakers, politics, president, protection, protests, protocol, public, reform, security, socialist, sovereignty, treaty, un-ngo, vote, war

Note: the stems of these words are used for frequency analysis.

8.2. Distributions of Frequencies

Table 8.2 contains the short rankings (the full rankings are in Appendix D) of cities on each dimension, where rankings are derived from the each city's share of each categorical frequency to the total number of keywords. These rankings can be used for comparison among global cities because they account for the fact that keywords associated with a particular dimension are more prevalent in the webpages of a highly ranked city. For instance, the top rankers on the economic dimension share a strong tendency of the usage of economic keywords compared to the low rankers on this dimension. Therefore, we can use these rankings to define a typology of cities in terms of their tendency for each of the three dimensions of globalization.

Let us look first at the top 15 rankings on each dimension (the left side of Table 8.2). Except for Brazzaville, which is ranked highly on both the economic and political dimensions, no city makes the top 15 rankings on more than a single dimension. This denotes that the top rankers on each dimension have a strong tendency for one of the three dimensions, while ignoring the others. For the economic dimension, we find that the economy is an important topic for Chinese cities, as 6 Chinese cities are ranked in the top 15 cities. Luxembourg is the only European city that stands out on the economic dimension, which underscores the low relative tendency of other European cities for economy. The top 15 cities on the cultural dimension are all North American cities. This denotes that these cities have strong tendency for cultural information. In other words, the webpages for North American cities have much information related with culture (i.e. keywords in the cultural category). The existence of many webpages related with cultural topics means that people's interest in cultural aspects of these cities is stronger than one

of other cities. Top rankers in the political dimension are well known cities for political issues (i.e. conflicts). Considering the keywords of predefined categories for political dimension, it is natural for these cities to be ranked in the top list on the political dimension.

In the bottom 15 rankings (the right side of Table 8.2), we can find that La Paz, Lahore, and Tijuana are listed in all three categories. This denotes that these cities have a weak tendency for all dimensions, and these cities have low level of polarity in terms of all three dimensions. Kyoto, Batam, Bandung, Medan, Palermo, San Salvador, Lagos, and Medellin are listed on two dimensions. These cities also have low level of polarity, but they have at least one dimension with relatively stronger tendency than the other two dimensions.

One important characteristic of the rankings is that top rankers on the cultural dimension consist of English-speaking cities, but top rankers and bottom rankers on the other dimensions are all non-English-speaking cities. This distribution reminds us that there is a big thematic difference between English-speaking cities and others. In other words, English-speaking cities exchange information which is highly related with the cultural theme of globalization, while information for other cities is more strongly focused on other dimensions.

Table 8.2: Relative rankings based on the proportion of categorical words to total words

Ranking	Economic	Cultural	Political	Ranking	Economic	Cultural	Political
1	Dalian	Portland	Kinshasa	250	Medan	Fortaleza	Medellin
2	Shanghai	Houston	Islamabad	251	Ulan Bator	Lagos	Valencia
3	East Rand	Minneapolis	Pyongyang	252	Rabat	Maputo	Panama City
4	Zhengzhou	Atlanta	Baghdad	253	Guayaquil	San Salvador	Helsinki
5	Luxembourg	Chicago	Kabul	254	San Salvador	Montevideo	Tijuana
6	Chittagong	Indianapolis	Damascus	255	Porto Alegre	Palermo	Lagos
7	Sao Paulo	Dallas	Khartoum	256	Jerusalem	Medellin	Bordeaux
8	Brazzaville	Charlotte	Aleppo	257	Tijuana	Kuwait City	Tokyo
9	Chennai	Boston	Kano	258	Recife	Medan	Fortaleza
10	Foshan	Cincinnati	Brazzaville	259	Antwerp	Tijuana	Palermo
11	Shenzhen	Saint Louis	Tripoli	260	Lyons	Bandung	Batam
12	Dubai	Ottawa	Tehran	261	Bandung	Foshan	Harbin
13	Kuala Lumpur	Philadelphia	Addis Ababa	262	La Paz	Batam	Lahore
14	Riyadh	Detroit	Monrovia	263	Kyoto	Lahore	Kyoto
15	Xiamen	Hartford	Rawalpindi	264	Lahore	La Paz	La Paz

8.3. Similarity of Global Cities

While the frequency of keywords tells us about the strong and weak tendency of a city towards each dimension of globalization, it does not provide the overall position of a city, considering all three dimensions together. MDS is helpful to reveal the similarity of global cities based on the relative share of the three categories of keywords. It provides the position of a city in the space of similarity based on the three aspects of globalization.

Figure 8.2 is a 2-dimensional MDS map, which depicts similarity between cities. The proportions of each of the three categories of keywords are used as input attributes. The similarities (i.e. distances) are calculated by metric MDS in *R* statistic package. Goodness of fit is 1 for 2 dimensions. The relative strength of each dimension is also represented on the MDS map in Figure 8.3 to facilitate the interpretation of the positioning of cities in the space of globalization dimensions. The direction to the top-left of the plot points to a higher ratio of economic keywords; the direction to the top-right

corresponds to a high ratio of political keywords, and the direction to bottom-right a high ratio of cultural keywords. Thus, the distribution of cities fits on a triangle whose three corners represent the highest ratio of each dimension.

Chinese cities including Foshan, Dalian, Zhengzhou, Dongguan, Shenzhen, Shantou, Xiamen, and Shanghai in the top-left of the plot form a cluster of high ratio of economic words with Luxembourg and East Rand. While East Rand is similar to Zhengzhou and has a strong economic tendency, Luxembourg shows political tendency as well. Khartoum, Damascus, Islamabad, Kinshasa, Addis Ababa, Pyongyang, Baghdad, Tripoli, Kabul, Aleppo, Kano, Ciudad Juarez, and Monrovia in the top-right of the plot are similar to each other with the strong political tendency of their textual contents. We can also include Brazzaville, Maputo, and Conakry into this group, but these cities have more economic tendency compared to the first group of political cities. Porto Alegre in the bottom-right has the strongest cultural tendency.

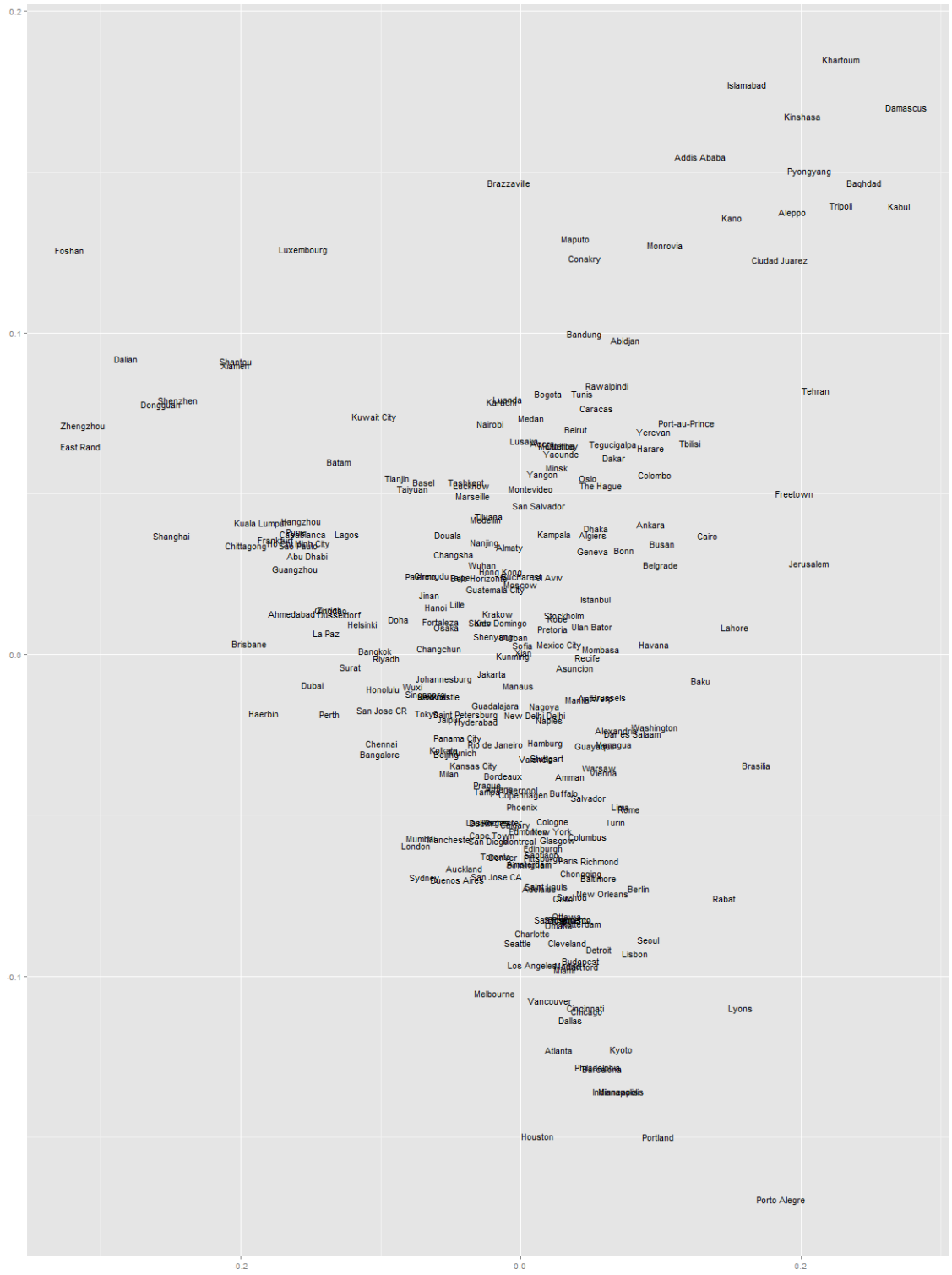

Figure 8.2: Multidimensional scaling map for similarity of cities

Figure 8.3: Distributions of the proportion of keywords of each dimension

Related with the analysis of this polarity, the existence of a relationship between the polarity of textual contents and globality is an intriguing question that deserves to be examined more closely. The MDS map in Figure 8.4 is based on 66 cities from our data set which are matched to the cities for which the GCI has been estimated. Input attributes are same as in the previous MDS map for the entire set of 264 cities, but the size of labels is shown in proportion to the GCI scores. City names in large font are high ranker on the GCI. We can find the cities are located at the pole of each dimension (i.e. the economic in

the left, the politics in the top-right, and the cultural in the bottom-right) tend to have lower globality; cities found around the origin of the coordinates exhibit a greater diversity of GCI values, but all cities with high GCI are exclusively found in this area of the plot. This point to the tendency for cities of high-globality to have more balanced contents compared to the cities of low-globality. In other words, less-global cities are likely to have high polarity in the contents of their webpages. It is clear however that the relationship between content polarity and globality is not straightforward. This result supports contention that contents are related with globality of the cities. Furthermore, it shows the possibility to apply the content analysis to measure globality although this research does not deal with this question further.

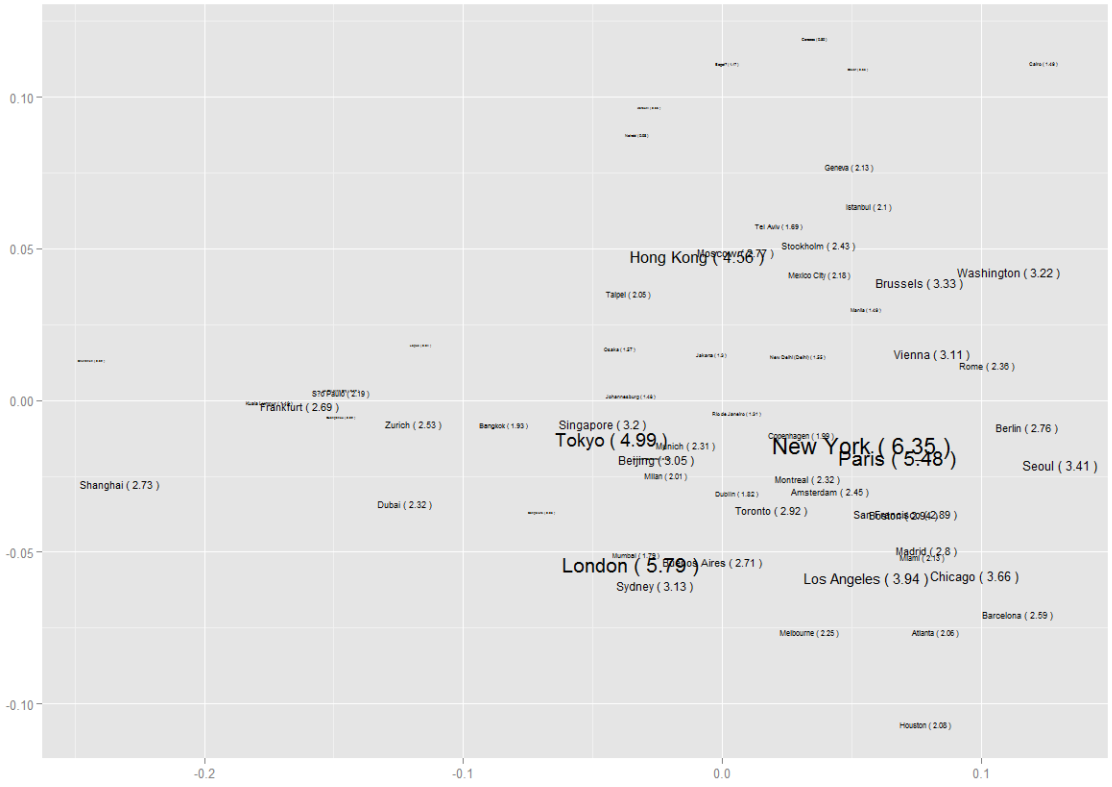

Figure 8.4: Multidimensional scaling map for 3 content dimensions and GCI

Cities of each continent can also be separately plotted on the MDS map. This results in eight distinct maps (Figure 8.5 to Figure 8.12). These maps provide evidence for the identification of geographic differences among cities on the basis of textual data. Continents are classified into 8 ensembles: Africa, Asia, Middle East, Oceania, Europe, North America, Central America, and South America.

African cities (Figure 8.5): We can draw two hypothetical lines on the plot: one is the line of the economy-culture connection, from the bottom-right to the top-left; another is the line of the culture-politics connection from the bottom-right to the top-right. The former includes Cape Town, Johannesburg, Casablanca, Lagos, and East Rand; the latter includes other cities. East Rand shows the strongest tendency for the economic, and Casablanca and Lagos follow. Rabat shows the strongest tendency for the cultural, and Cape Town, Alexandria, and Dar es Salaam are next. Khartoum, Kinshasa, Addis Ababa, Tripoli, and Kano show strong tendency for the political. Other cities are distributed between the political and the cultural poles. Overall, African cities show a high tendency towards the political and the cultural. This is highly related with the political issues that plague African cities and are routinely discussed and commented on the internet.

Asian cities (Figure 8.6): Three groups of cities have the strongest tendency towards each pole. For the economic, 7 Chinese cities including Foshan, Dalian, Zhengzhou, Dongguan, Shenzhen, Shantou, and Xiamen are grouped together. Islamabad, Pyongyang, and Kabul are associated with a strong political tendency. As for the cultural dimension, we find Kyoto, Seoul, Suzhou, and Chongqing being grouped together. More cities are distributed along the hypothetical line connecting the economic

and cultural poles. Cities towards the political poles are more dispersed. Many cities are distributed around the origin of the coordinates, which denotes the low polarity of these cities. In sum, three small groups of cities associated with each content dimension and the low polarity of other cities are the distributional characteristics of Asian cities.

Middle Eastern cities (Figure 8.7): Damascus, Baghdad, and Aleppo are associated with a strong political tendency; Kuwait City and Abu Dhabi show tendency towards the economic. Doha, Riyadh, Dubai, Jeddah, and Tel Aviv are distributed between the economic and cultural poles. Tehran and Jerusalem are located between the political and cultural poles. The distributional characteristics of Middle Eastern cities are marked by overlapping tendencies among the three dimensions and the absence of cities having a strong tendency for economic and cultural considerations.

Oceania cities (Figure 8.8): All 6 cities are distributed on the hypothetical line between the economic and cultural poles. Brisbane and Perth exhibit some tendency towards the economic. Other cities including Sydney, Auckland, Adelaide, and Melbourne form a group having more of a tendency for culture.

European cities (Figure 8.9): European cities are distributed around the center of the map and extend towards the cultural pole, except for Luxembourg. This denotes that European cities have fairly balanced images. Cities distributed in the top-left side (Luxembourg, Frankfurt, Zurich, Düsseldorf, Helsinki, etc.) exhibit a tendency for economic themes while cities on the bottom-right side have a tendency for culture. Mink, Oslo, The Hague, Ankara, Geneva, Bonn, and Belgrade form a group with political tendency. Luxembourg and Frankfurt have the strongest tendency for the economic. Zurich, Düsseldorf, Helsinki, Palermo, Basel, and Marseille can be identified as the next

group of strong tendency for the economic. Barcelona and Lyons have the strongest tendency for culture. Rotterdam, Berlin, Lisbon, Budapest, and Madrid can be the next group of strong tendency for culture.

North American cities (Figure 8.10): North American cities commonly have a strong tendency for culture so that they are distributed predominantly in the bottom-right of the plot. Honolulu shows relatively strong tendency for the economic; Washington shows relatively strong tendency for the political. Strong similarity among North American cities means that constituent ratio of web contents for these cities are highly similar to each other, and these contents are highly identified with the cultural theme.

Central American cities (Figure 8.11): Central American cities are distributed at the center of the plot except for Ciudad Juarez. Ciudad Juarez has a strong tendency for the political. Cities are similar to each other in terms of the weakness of tendency for any specific dimension. Some differentiation among these cities can be detected. For instance, we can tell that Managua has a relatively stronger tendency for culture than Tegucigalpa and Monterrey; and Tegucigalpa and Monterrey have a relatively stronger tendency for politics than Managua. However, a shared characteristic of Central American cities is that they exhibit images that are fairly balanced between the economic, political, and cultural dimensions.

South American cities (Figure 8.12): Porto Alegre has a strong tendency for culture. São Paulo and La Paz form a group of weak economic tendency. Recife, Asunción, Manus, Rio de Janeiro, Guayaquil, Salvador, Lima, Santiago, Buenos Aires, Quito, and Brasília can be the next group of cultural tendency. Bogotá and Caracas show

a relatively strong tendency towards the political. Other cities are overlapping both the economy and political categories.

In addition, the MDS maps for US cities (Figure 8.13) and Chinese cities (Figure 8.14) are provided as special cases. The MDS maps of these countries are of interest because of all the countries represented in our dataset, these countries encompass the largest numbers of cities. The comparison of distributions is also useful to compare distributional characteristics between countries, which confirm the existence of geographic difference in textual web-data.

US cities (Figure 8.13): The MDS map for US cities is not much different from the map for North American cities (figure 8.9) because most North American cities are in the US. US cities are highly concentrated towards the cultural dimension even though Honolulu and Washington exhibit some weak tendency towards the economic and the political, respectively. Overall, US cities have a strong similarity to the cultural aspect of global cities.

Chinese cities (Figure 8.14): Chinese cities are distributed along an axis between the economic and cultural poles. Foshan, Dalian, Zhengzhou, Dongguan, Shenzhen, Shantou, Xiamen, and Shanghai are distributed in the top-left of the plot, which shows a strong tendency for the economic. Tianjin, Hangzhou, Guangzhou, Qingdao, and Harbin form the next group of strong tendency for the economic. On the other hand, a small number of Chinese cities show a strong cultural tendency, which includes Chongqing (the first order of strong tendency for culture) and Beijing (the second order of strong tendency for culture). Others fall between these two extremes.

A comparison between US cities and Chinese cities reveals several interesting points. First, we note the absence of a city having strong tendency for the political in either country. Cities in both countries are situated on a hypothetical axis between the economic and cultural poles. Second, several differences are noticeable. The first difference is the different level of similarity among cities of the same group. US cities are more homogeneous than Chinese cities. This denotes that the characteristics of US cities are more similar to each other than among Chinese cities. The second difference is the different polar tendency. While US cities are mostly situated in the plot area of strong cultural emphasis, Chinese cities are distributed near the plot area of strong economic polarity.

In sum, the analysis makes it possible to list cities having strong tendency towards each thematic pole from the MDS map. Geographic tendency matching the similarity of cities is also confirmed through continental MDS maps and the comparison of US cities and Chinese cities.

Figure 8.5: Distribution of African cities on the MDS map of globalization dimensions

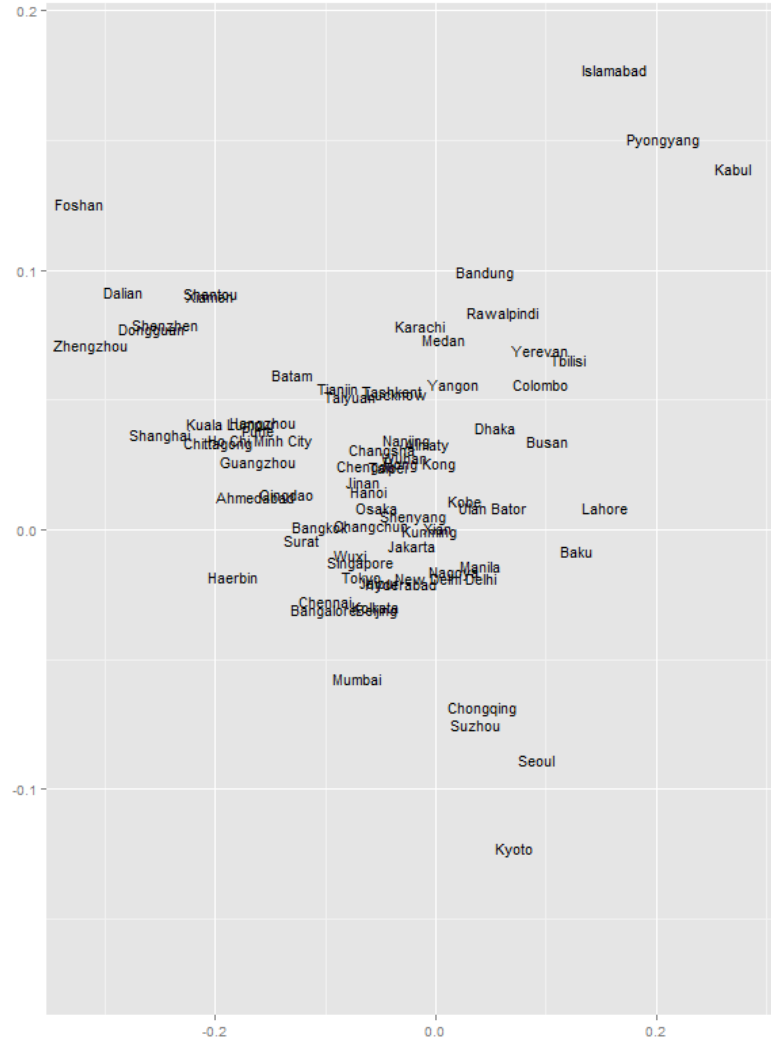

Figure 8.6: Distribution of Asian cities on the MDS map of globalization dimensions

Figure 8.7: Distribution of Middle Eastern cities on the MDS map of globalization dimensions

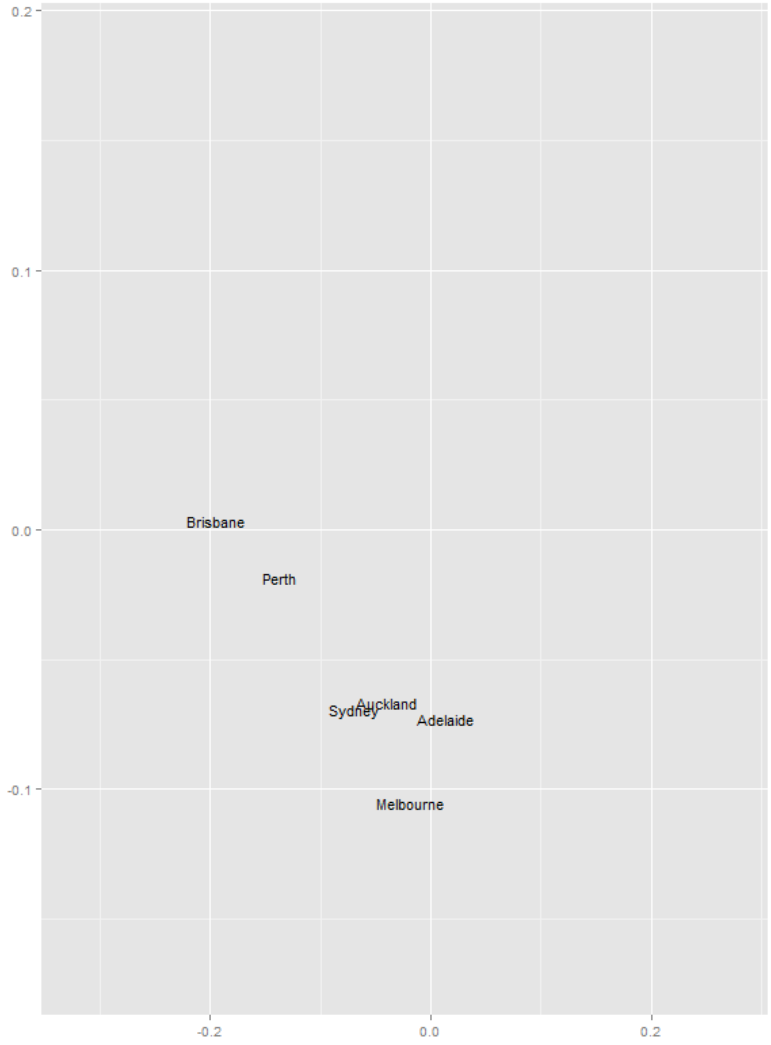

Figure 8.8: Distribution of Oceania cities on the MDS map of globalization dimensions

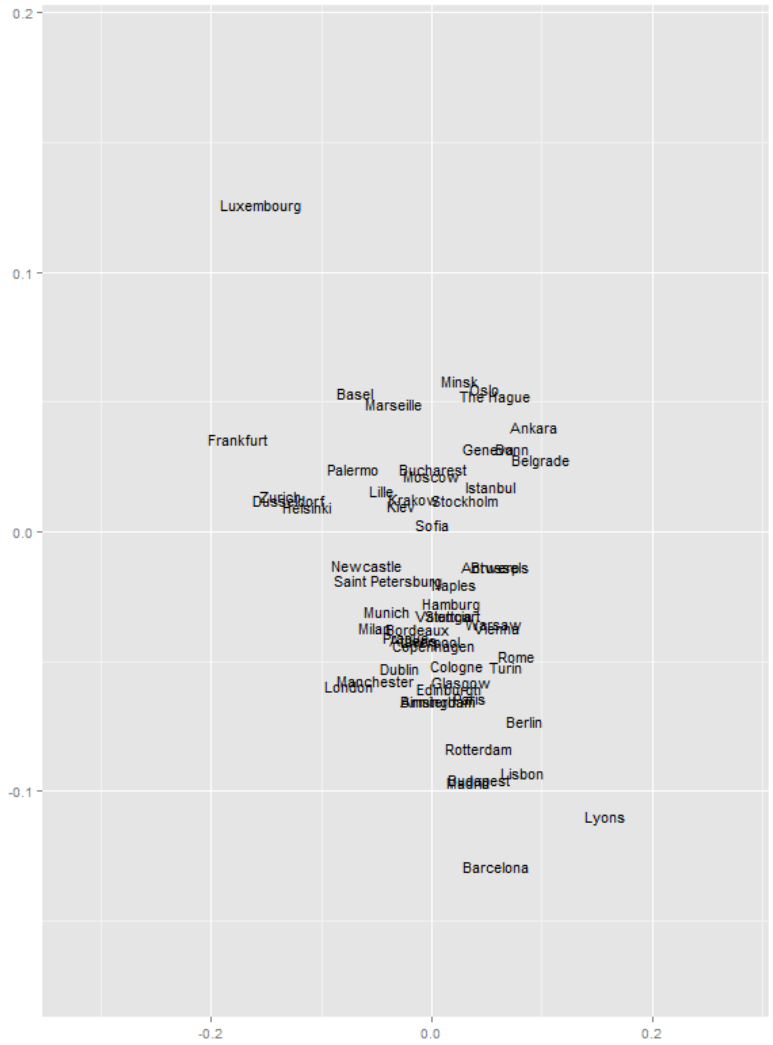

Figure 8.9: Distribution of European cities on the MDS map of globalization dimensions

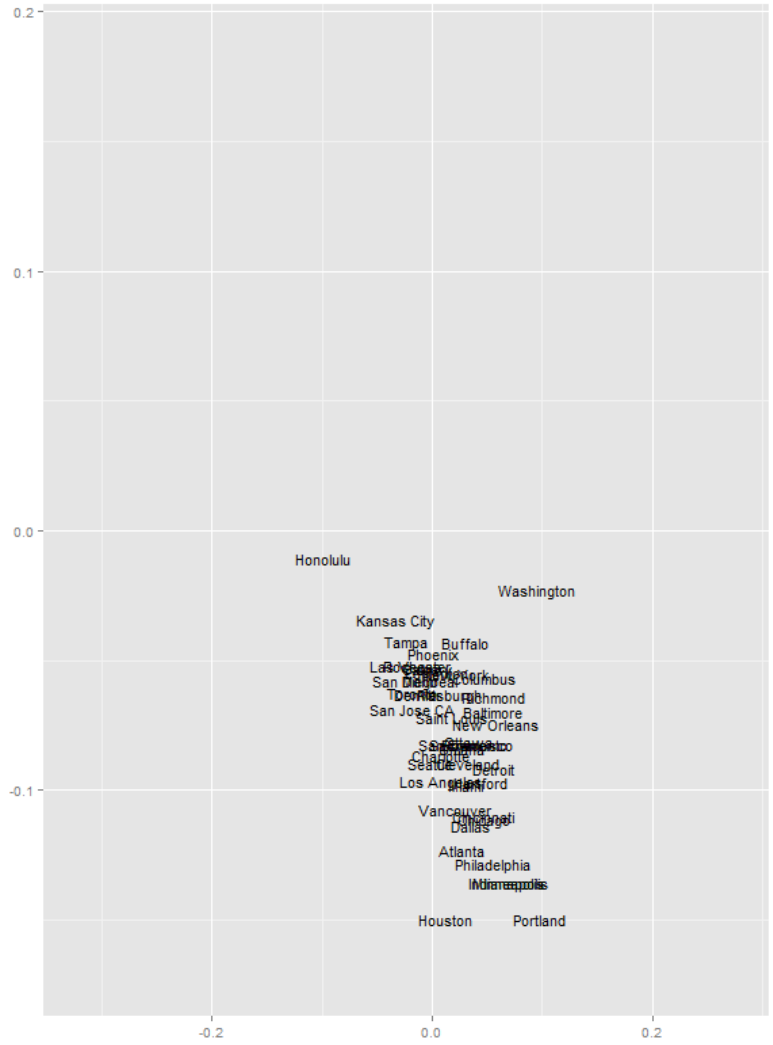

Figure 8.10: Distribution of North American cities on the MDS map of globalization dimensions

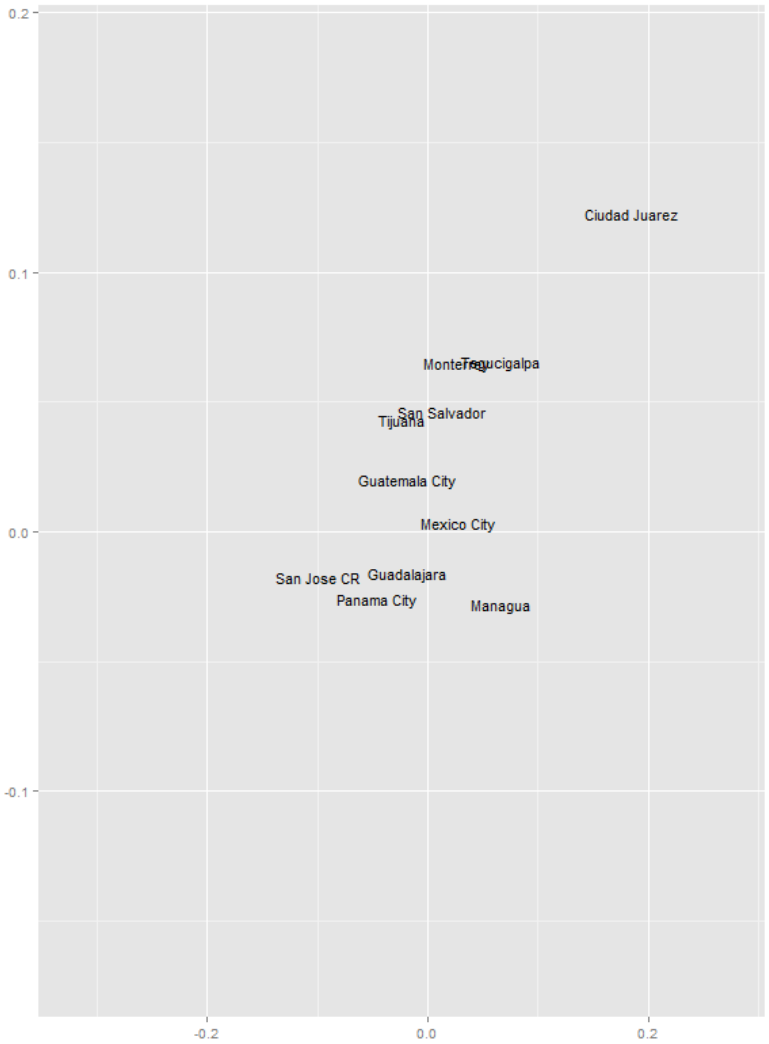

Figure 8.11: Distribution of Central American cities on the MDS map of globalization dimensions

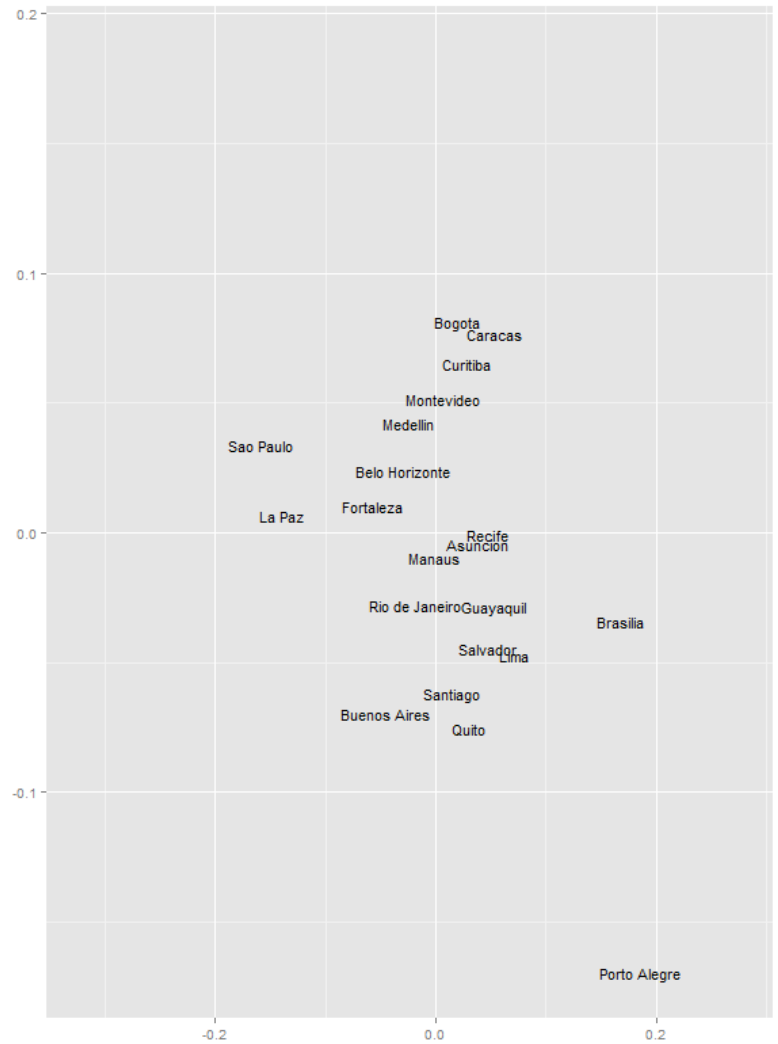

Figure 8.12: Distribution of South American cities on the MDS map of globalization dimensions

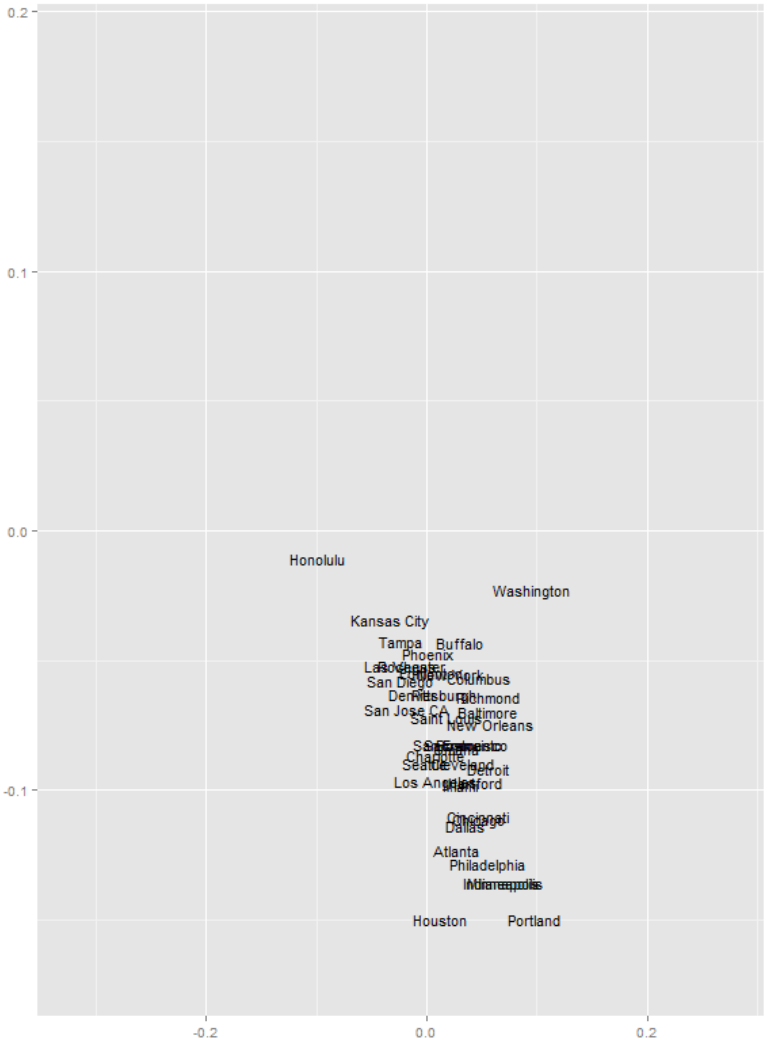

Figure 8.13: Distribution of US cities on the MDS map of globalization dimensions

Figure 8.14: Distribution of Chinese cities on the MDS map of globalization dimensions

CHAPTER IX: SYNTHESIS OF HYPERLINK NETWORKS AND THE QUANTIFIED CONTENTS

The purpose of this chapter is to explore the contribution of the study of city hyperlink networks through an approach that integrates hyperlink network analysis and webpage content analysis to global city research. Although HNA and QCA have each been found meaningful to analyze the Web-oriented data on their own right, the analysis in this chapter seeks a deeper understanding of the relationship between structural characteristics and textual characteristics. It suggests an approach to analyze the Web data for global city research considering the utilization of different types of the Web data together.

The methodology followed in this part of the research is outlined in Figure 5.12. The basic relationship between the results of HNA and QCA is analyzed through correlation and MDS maps. The correlation and the MDS maps are based on distances (i.e. similarities) among cities. The former provides the statistical results and the latter provides the visual comparison. Then, the 2-dimensional map of the MDS based on all the attributes from the HNA and the QCA is used for the examination of the synthetic results based on all attributes. This MDS map is used so as to derive the salient structures of city globality that emerge from the attributes of hyperlink network measurements and the attributes of quantified content analysis. Similarities among cities are explored from the perspective of all these dimensions taken together. Finally, the contribution of the new approach is discussed.

9.1. Correlation between Similarities of HNA and QCA

The distance matrices based on hyperlink network measurements and textual analysis are used to estimate the correlation between the two sets of measures on a city by city basis. Table 9.1 shows the distribution of the correlation coefficients computed for each of the 264 cities. Only 11 cities (4.1%) have a correlation coefficient over 0.4. Although we consider the marginal correlation is more than 0.3, only one-third of the cities (82 cities) are included in this margin. This denotes that the general correlation between similarities of cities based on the HNA and QCA measures is very low, and these two measures capture rather distinct realities of the Web.

When we visualize the relatively highly correlated cities (more than 0.4), we can also notice the differences between HNA and QCA. Figure 9.1 is the comparison of the MDS maps: (a) is based on the distances of HNA; (b) is based on the distances of QCA. Metric MDS is used for each MDS map, and the axes of each MDS map correspond to dimensions 1 and 2, respectively. Blue squared cities are US cities (Chicago, Atlanta, Boston, Houston, and Charlotte), and red squared cities are non-US cities (Kano, Brazzaville, Conakry, Monrovia, Ciudad Juarez, and Tripoli). US cities in (a) are scattered (i.e. dissimilar) while they are more clustered (i.e. similar) in (b). In contrast, non-US cities are close to each other (i.e. similar) in both (a) and (b). If we recall the polarities of the content analysis in the previous chapter, here US cities are associated with the cultural dimension and non-US cities are associated with the political dimension. We can find that political cities (non-US cities) are similar to each other on both MDS maps while the similarity of cultural cities (US cities) on the distances of HNA is less similar than the one on the distances of QCA. That is, this inconsistency of distributions

between the two types of web data denotes that one type can capture a partial characteristic of the cities on the Web.

The low correlation between similarities of HNA and QCA denotes that each analysis focuses on different parts of the Web. The inconsistency between distributions of relatively highly correlated cities suggests the use of one analysis could miss other characteristics of cities on the Web. Thus, it is worth to use the two types of data together and try to reveal how each part can support the analysis of the characteristics of cities on the Web.

Table 9.1: Distribution of correlation coefficients between HNA and QCA similarities between cities

Correlation Coefficient	Cities (High to Low)	No. of Cities (%)
0.4 (0.43~0.35)	Chicago, Atlanta, Boston, Houston, Kano, Brazzaville, Conakry, Monrovia, Ciudad Juarez, Tripoli, Charlotte	11 (4.1)
0.3 (0.34~0.25)	Portland, Tehran, Pune, Port-au-Prince, Khartoum, Bogotá, Denver, Detroit, Rawalpindi, Baghdad, The Hague, Tunis, Islamabad, Edinburgh, Mumbai, Calgary, Shantou, Philadelphia, Toronto, Kabul, Ottawa, Damascus, Chennai, Xiamen, Harare, Ahmedabad, Indianapolis, Maputo, Minneapolis, Glasgow, New York, Columbus, Miami, Cincinnati, Foshan, Tegucigalpa, Sydney, Medan, Dublin, Saint Louis, Kinshasa, Pittsburgh, Curitiba, Dhaka, Kuwait City, Abidjan, Liverpool, Baltimore, New Orleans, Washington, Pyongyang, London, Dongguan, Tashkent, Yangon, Vancouver, Oslo, Bonn, Dakar, Karachi, Las Vegas, Seattle, Busan, Luanda, Caracas, Dalian, Phoenix, Lusaka, Yerevan, Edmonton, Yaoundé	71 (26.8)
0.2 (0.24~0.15)	Ankara, Bandung, Rome, Athens, Batam, Nairobi, Lucknow, San Diego, Algiers, Omaha, Taiyuan, San Salvador, Minsk, Montevideo, Kolkata (Calcutta), Aleppo, Rochester, Monterrey, Kampala, Singapore, Baku, East Rand, Marseille, Beirut, Bangkok, Hangzhou, Kuala Lumpur, Cairo, Accra, Ho Chi Minh City, Tampa, Moscow, Shanghai, Tijuana, Geneva, Cleveland, Bangalore, Jakarta, Jinan, Palermo, Belgrade, Basel, Brisbane, Wuxi, Richmond, Melbourne, Hanoi, Dallas, Havana, Brussels, Perth, Stockholm, Freetown, Hyderabad, Frankfurt, Surat, Harbin, Berlin, Johannesburg, Zurich, Manila, San Francisco, Dubai, Tianjin, Nanjing, Fortaleza, Krakow, Hong Kong, Ulan Bator, Douala	70 (26.5)
0.1 (0.14~0.05)	Tel Aviv, Recife, Sofia, Istanbul, Mexico City, Lahore, Xi'an, Belo Horizonte, Bucharest, Jaipur, Düsseldorf, Changsha, Lagos, Vienna, Zhengzhou, Lille, Colombo, Paris, Buffalo, Wuhan, Lyons, Tbilisi, Guangzhou, Chengdu, Shenzhen, Kobe, Medellin, Chittagong, Manchester, Montreal, Jerusalem, Abu Dhabi, Pretoria, Saint Petersburg, Seoul, Nagoya, Beijing, New Delhi (Delhi), Casablanca, Kunming, Salvador, Warsaw, Santo Domingo, Changchun, Luxembourg, Munich, Mombasa, Guatemala City, Honolulu, Shenyang, Barcelona, Dar es Salaam, Auckland, Jeddah, Helsinki, Alexandria, Birmingham, Kiev, Manaus, Rio de Janeiro, Managua, Durban, Asunción	63 (23.9)
0 (0.04~-0.04)	Qingdao, Tokyo, Taipei, La Paz, Osaka, Newcastle, Amsterdam, Prague, Lima, Hartford, Brasília, Cape Town, Valencia, São Paulo, Guadalajara, Doha, Copenhagen, Naples, Hamburg, Stuttgart, Addis Ababa, Madrid, Lisbon, Milan, Antwerp, Kansas City, San José, Budapest, Almaty, Turin, Rotterdam, Bordeaux	32 (12.1)
-0.1 (-0.05~-0.13)	Buenos Aires, Riyadh, Adelaide, Rabat, Chongqing, Sacramento, Guayaquil, Panama City, Amman, Cologne, Suzhou, Los Angeles	12 (4.5)
-0.2 (-0.15~-0.24)	San Jose, Porto Alegre, Santiago, Quito, Kyoto	5 (1.9)
Total		264

Note: Correlation coefficient for each city is listed in Appendix F.

Figure 9.1: Comparison of MDS maps based on HNA and QCA similarities of cities

Note: only cities of which correlation coefficient is over 0.4 are colored.

9.2. Distribution of Cities Considered Structure and Content

Metric MDS is used for creating a 2-dimensional MDS map, whose goodness of fit is 0.6. The MDS map is derived from similarities (distances) computed on all 10 attributes of HNA and QCA (Figure 9.2). The 10 attributes consist of 7 HNA measurements (maximum in-degree, share of receiver, share of carrier, assortativity, weakly connected component, reciprocity, and modularity) and 3 attributes of QCA (economic proportion, cultural proportion, and political proportion). The distance (similarity) matrix is calculated from the standardized scores of these attributes. Based on the MDS property (attribute) map (Figure 9.3), we can interpret more fully the general areas on the final MDS map. That is, the right side of the MDS map shows the

distribution of cities with positive attributes of HNA (which is marked as orange-colored plus sign) while the left side of the MDS map shows the negative attributes of HNA (which is marked as orange-colored minus sign). The orange-colored vertical dashed line represents just a hypothetical boundary between the positive area and the negative area for easy understanding. We can thus interpret the '+' and '-' panels as the general level of the efficiency of information flow. For textual attributes, the top left of the MDS map is associated with the tendency towards political texts, while the bottom left shows the tendency for economic texts; and the center-right shows the tendency for cultural texts. Blue-colored circles are hypothetical circles for marking dominant category of keywords; and the arrows attached to the circles represent the directionality of those tendencies.

Compared to the MDS maps estimated separately on the QCA and HNA criteria in Figure 9.1, the final MDS map shows the combination of the two MDS maps based on HNA and QCA. However, this map preserves the general distribution of cities on the basis of the HNA as well as the separation of categorical groups derived from QCA. While the result of the HNA shows many cities around the origin of coordinate, the final MDS map moves those clustered cities in the top-left (i.e. the politic), bottom-left (i.e. the economic), and right (i.e. the cultural) directions.

We can consider the efficiency of information flows (i.e. positive or negative panel formed on HNA-related attributes) and the thematic clusters for the interpretation of the MDS map. The political cities located in the top-left portion of the final MDS map show low efficiency of information flows. Economic cities in the bottom-left are distributed between high efficiency and low efficiency. Cultural cities are located in the right side of the map, which is an area of the MDS map associated with high efficiency of

information flows.

This result denotes that there is a relationship between content type and the efficiency of information flows. Although we cannot define the causality between the two, the relationship indicates that hyperlink networks handling cultural texts have more efficiency in handling information flows compared to hyperlink networks handling economic texts or political texts. We can also interpret this in a different way. That is, the cities with high efficiency of information flows are exposed, discussed, and shared on the Web with cultural text. However, the politically described cities tend to be portrayed on websites with low efficiency of information flow (i.e. the cities are discussed in limitation of information flows).

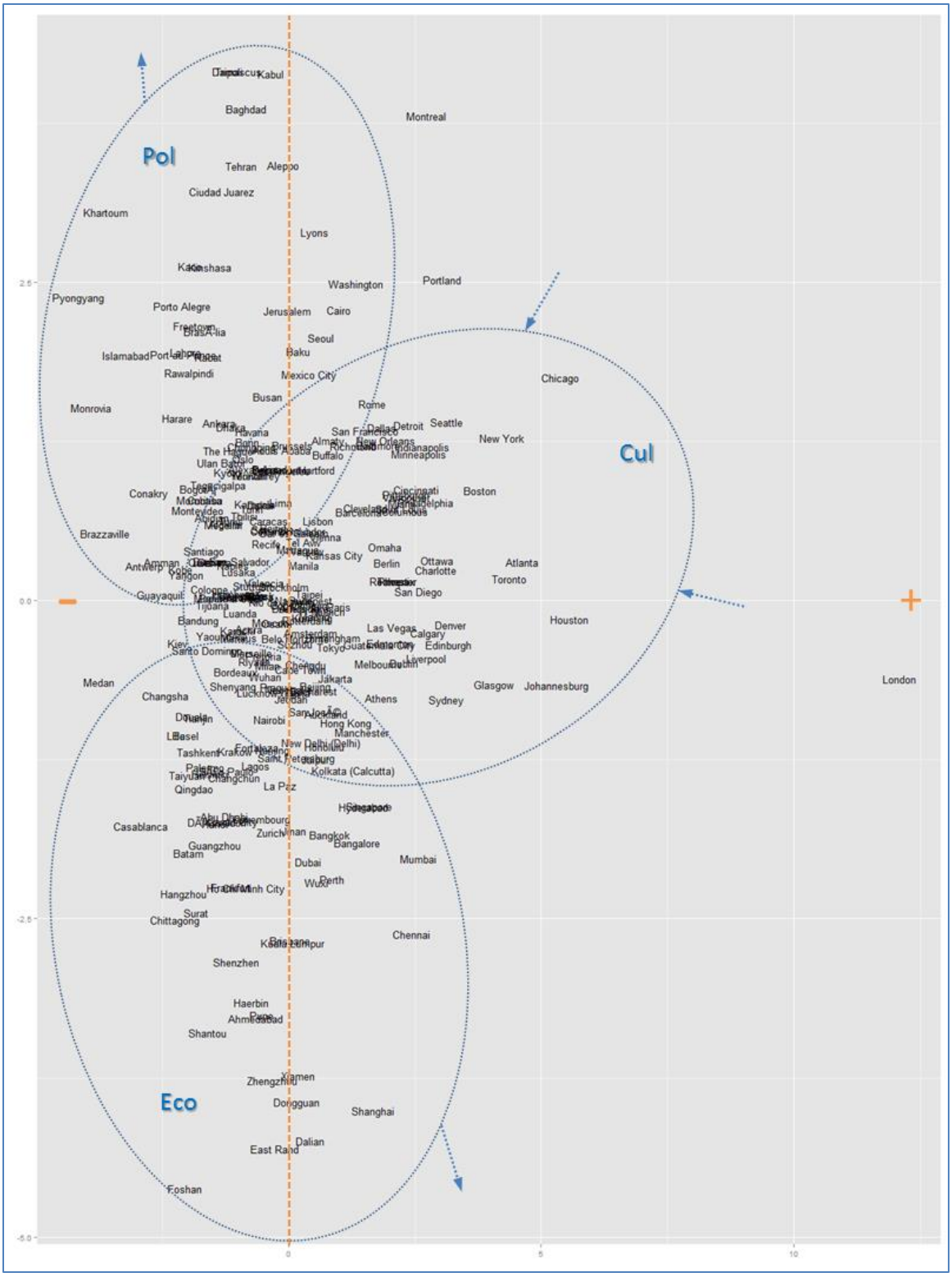

Figure 9.2: MDS map estimated on HNA and QCA attributes

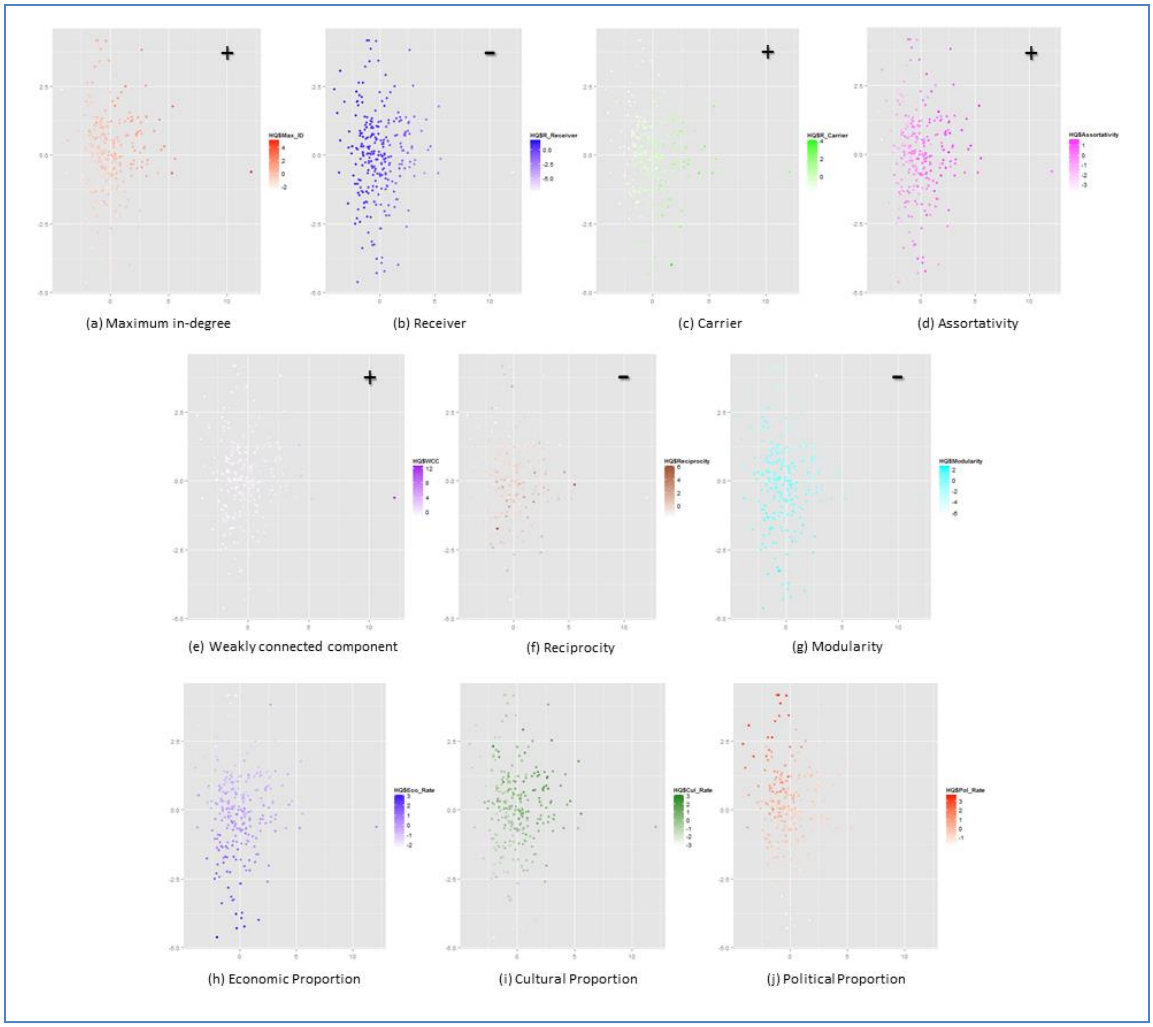

Figure 9.3: Distribution of cities on the MDS property maps

Note: the plus (+) sign and minus (-) sign from (a) to (g) on the attributes of the HNA mean the relationship of those attributes to GHI (global city hyperlink index) of the Chapter 6. Other attributes from (h) to (j) do not have marks because these attributes indicate textual composition.

9.3. Contribution of the Composite Approach

The analysis of this chapter shows the benefit of using two different types of the Web data together. Although the separate analyses provide results focused on structural characteristics and textual characteristics, the use of a broader set of attributes of the Web data helps us detect the relationship between the types of the Web data like the case of political cities and its low efficiency of information flows. Accounting for both the perspectives of structural analysis and content analysis provides complementarity in their mutual interpretation.

This new approach confirms the usefulness of the Web data to measure the 'globality' of cities. Research based on either the structure or the contents of the Web gain from this approach. The result of the textual analysis can bring explanation rooted in the context of global city to inform the structural analysis. Conversely, the result of the structural analysis brings a hierarchical perspective on global cities to the textual analysis for the understanding of distributional characteristics. Consequently, this new approach provides a richer approach to global city research that leverages the new perspective brought by Web-based data.

CHAPTER X: CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

Web data have increased their importance in various academic fields including geography. As people spend more time on the Web and use it as a tool for overcoming geographic limitations, Web data possess valuable information and knowledge which can be applied to the analyses of different aspects of human interactions. Globalization is a complex process that is most obviously manifested by of world-wide human interaction. With the advance of the Web, globalization meets a new phase for the speed of the integration and the large extent of the penetration. This research was a new approach to study the relationship between the Web and globality. Specifically, this research expanded the concept of globality to this hitherto untouched area with the study of a number of new measurements. The application of hyperlink network analysis and quantified content analysis allowed the utilization of Web-oriented data in the field of global city research.

The general characteristics of hyperlink networks provided a new index which is based on hyperlink network measurements. The new global city hyperlink network index (GHI) was created by the critical assessment of possible measurements. This index denotes the efficiency of the hyperlink network in information flow, exchange, search, etc. The effectiveness of this efficiency represents the new globality of cities on the Web.

The analysis of nodes (i.e. websites) provided the distribution of website domain and types. The distribution of the top-level domains on organizational and governmental

websites explained the sources of the city-related information. The commonality of the website types indicated the information is exchanged mainly on specific website types. From the analysis of premium nodes, it was confirmed that the center of the websites is social networking service (SNS) such as *Twitter* and *Facebook*. This result represents the importance of the SNS nowadays as the main hub of information exchange.

Quantified content analysis was useful to extract the meaningful dimensions of globalization from the Web contents and to assess textual data. The distribution of cities in the map of MDS showed tendencies of cities for each dimension of globalization. Each dimension represented the image of cities how they are expressed, discussed, and shared on the Web. Comparisons by continent and between US cities and Chinese cities supported the geographic distributional differences and similarities on the textual data.

A new approach to utilize both types of Web-based data was helpful to understand the characteristics of these data in a complete way. Each analysis based on HNA and QCA captures distinct realities of the Web. The analysis based on both HNA and QCA enabled relational interpretation for deeper understanding of the Web data. The mutual complementarity for the interpretation of the results reveals the usefulness of this new approach.

After all the analysis, we can enumerate the contributions of this research as follows. First, this research proved the usefulness of Web-based data for global city research. This means that this research opens a gate to the new world of information research, which can support global city research in the conceptual expansion of globalization and the inclusion of new data sources to overcome data inconsistency and scarcity. Second, this research confirmed the inherent characteristics of the Web data,

which should be considered for future research. The common premium websites from the hyperlink networks explain that the information of cities is exchanged through the SNS-centered hyperlink network. This finding supports why the SNS data are so focused nowadays as well as where global city research based on the Web should concentrate. Third, this research confirmed that the quantification of the textual data from the Web is useful to reveal the image of cities on the topical space. The position of a city on the topical space may suggest the direction for future investment to enhance the image of the city. Lastly, this research provided a research framework for handling two types of the Web data: hyperlink networks and contents. The series of processes such as the creation of structured database from the crawled webpages, the automation of text processing and frequency calculation, the use of MDS for data reduction technique, et cetera allowed applying the latest techniques from other fields to global city study.

In spite of the above contributions, there are limitations. The main limitation of this research is due to the dynamic situation of the Web. It is hard to capture the structure of the Web in real-time. In other words, by the time data has been collected, processed, and analyzed, it already reflects an outdated reality. Another limitation is that this research is only based on the webpages which are written in English language. Considering the different languages provide different perceptions of globalization, it is unfortunate not to deal with other languages. In addition, this research aims at finding the possibility of using hyperlink network data as a new index for global city research, not to define the global city. The selection of study cities in this research is based on existing studies, which have different definitions of global cities. In other words, this research uses the compilation of predefined global cities. From a methodological perspective, this

research is meaningful to expand possible measures of globality to the Web data. From a theoretical perspective on globalization, however, further development and testing is required for using this index for defining global cities, which is to find the causal relationship between hyperlink networks and globality.

Considering the contributions and limitations of this research, the direction for future research includes the following. In order to overcome the temporal quality of the dynamic Web data, shorter time span or time series analysis can be considered. If the research were repeated with the same seed URLs in the future, it would be helpful to figure out how city images change, and how the ability of cities to leverage hyperlink network structures changes. Another direction is to use the webpages written in other languages for comparative analysis. It could reveal the linguistic characteristics of the Web data and human perception of which people use these languages.

REFERENCES

Al-Rodhan, N. (2006). *Definitions of Globalization: A Comprehensive Overview and a Proposed Definition*. Geneva: Centre for Security Policy.

Anholt, S. (2006). *The Anholt city brand index: how the world views its cities* (2nd ed.). Seattle, WA: Simon Anholt and GMI (Global Market Insite, Inc.).

Anholt, S. (2010). *Places: identity, image and reputation*. New York, NY: Palgrave Macmillan.

Ashworth, G. J., & Voogd, H. (1990). *Selling the city: marketing approaches in public sector urban planning*. London, UK: Belhaven Press.

Balmer, J. M. T. (2010). Explicating corporate brands and their management: Reflections and directions from 1995. *Journal of Brand Management*, 18, 180-196.

Barabási, A. –L. (2003). *Linked: how everything is connected to everything else and what it means for business, science, and everyday life*. New York, NY: Plume.

Barnes, R. (Producer). (2010, February 20). *The virtual revolution (episode 4: Homo Interneticus?)* [Television broadcast]. London: BBC.

Beauregard, R. A. (1995). Theorizing the global-local connection. In P. L. Knox & P. J. Taylor (Eds.), *World Cities in a World-System* (pp. 232-248). Cambridge, UK: Cambridge University Press.

Beaverstock, J. V., Smith, R, G., & Taylor, P. J. (2000a). World-City Network: A New Metageography? *Annals of the Association of American Geographers*, 90(1), 123-134.

Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (1999). A roster of world cities. *Cities*, 16, 445-458.

Beaverstock, J. V., Smith, R. G., Taylor, P. J., Walker, D. R. F., & Lorimer, H. (2000b). Globalization and world cities: some measurement methodologies. *Applied Geography*, 20, 43-63.

Blain, C., Levy, S. E., & Brent Ritchie, J. R. (2005). Destination branding: insights and practices from destination management organizations. *Journal of Travel Research*, 43(4), 328-338.

Buhalis, D., & Spada, A. (2000). Destination management systems: criteria for success-an exploratory research. *Information Technology & Tourism*, 3(1), 41-58.

Chase-Dunn, C. K. (1985). The system of cities (A.D. 800-1975). In M. Timberlake (Ed.), *Urbanization and world-economy* (pp. 269-292). New York, NY: Academic Press, Inc.

Choi, J. H. (1998). Research agenda for the global paradigm in an urban geography in the era of globalization. *Korean Journal of Urban Geographical Society*, 1(1), 31-46.

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: content analysis of Macau travel related websites. *Tourism Management*, 28, 118-129.

Clark, D. (2003). *Urban world/global city* (2nd Ed.). New York, NY: Routledge.

Clark, V. (2004). *SAS/STAT 9.1: user's guide*. Cary, NC: SAS Publishing.

Cohen, R. B. (1981). The new international division of labour, multi-national corporations and urban hierarchy. In M. Dear & A. J. Scott (Eds.), *Urbanization and Urban Planning in Capitalist Society* (pp. 287-317). New York, NY: Methuen.

Cooke, P. (1989). Locality, economic restructuring and world development. In P. Cook (Ed.), *Localities: the changing face of urban Britain* (pp. 1-44). London, UK: Unwin Hyman.

Deun, K. V., & Delbeke, L. (2000). Multidimensional scaling. Open and Distance Learning. Leuven, Belgium: University of Leuven. Available at: http://www.mathpsyc.uni-bonn.de/doc/delbeke/delbeke.htm; accessed 1 May 2012.

Economist Intelligence Unit (EIU). (2009). *European green city index: assessing the environmental impact of Europe's major cities*. Munich, Germany: Siemens AG.

Economist Intelligence Unit (EIU). (2010). *Latin American green city index: assessing the environmental performance of Latin America's major cities*. Munich, Germany: Siemens AG.

Economist Intelligence Unit (EIU). (2011a). *Asian green city index: assessing the environmental performance of Asia's major cities*. Munich, Germany: Siemens AG.

Economist Intelligence Unit (EIU). (2011b). *US and Canada green city index: assessing the environmental performance of 27 major US and Canadian cities*. Munich, Germany: Siemens AG.

Economist Intelligence Unit (EIU). (2012). *Hot spots: benchmarking global city competitiveness*. A report from the Economist Intelligence Unit. Available at: http://www.managementthinking.eiu.com/hot-spots.html; accessed 12 July 2012.

Erdogmus, I. E., Bodur, M., & Yilmaz, C. (2010). International strategies of emerging market firms. *European Journal of Marketing*, 44(9-10), 1410-1436.

Ervin, J., & Smith, Z. A. (2008). *Globalization: a reference handbook*. Santa Barbara, CA: ABC-CLIO.

Feagin, J. R., & Smith, M. P. (1987). Cities and the new international division of labor: an overview. In M. P. Smith & J. R. Feagin (Eds.), *The Capitalist City: global restructuring and community politics* (pp. 3-34). Oxford, UK: Basil Blackwell.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. New York, NY: Cambridge University Press.

Finnie, G. (1998). Wired cities. *Communications Week International*, 18 May, 19-22.

Friedmann, J. (1986). The world city hypothesis. *Development and Change*, 17, 69-83.

Friedmann, J. (1995). Where we stand: a decade of world city research. In P. L. Knox & P. J. Taylor (Eds.), *World Cities in a World-System* (pp. 21-47). Cambridge, UK: Cambridge University Press.

Friedmann, J. (1997). World City Futures: the role of urban and regional policies in the Asia-Pacific Region. *Occasional Paper*, 56. Shatin, Hong Kong: Hong Kong Institute of Asia-Pacific Studies, The Chinese University of Hong Kong.

Friedmann, J., & Wolff, G. (1982). World city formation: an agenda for research and action. *International Journal of Urban and Regional Research*, 6(3), 309-344.

Fröbel, F., Heinrich, I., & Kreye, O. (1980). *The new international division of labor: structural unemployment in industrial countries and industrialization in developing countries*. Cambridge, MA: Cambridge University Press.

Furukawa, T. (2009). SOM of SOMs. *Neural Networks*, 22, 463-478.

Gallarza, M. G., Saura, I. G., & García, H. C. (2002). Destination image: towards a conceptual framework. *Annals of Tourism Research*, 29(1), 56-78.

Gappert, G. (1989). A management perspective on cities in a changing global environment. In R. V. Knight & G. Gappert (Eds.), *Cities in a Global Society* (pp. 312-325). London, UK: SAGE Publications.

Garson, G. D. (2011). Correspondence analysis. StatNotes: topics in multivariate analysis. Available at: http://faculty.chass.ncsu.edu/garson/PA765/correspondence.htm; accessed 1 May 2012.

Gfk Custom Research North America. (2011, July 18). Tokyo Earns Strong Image, Despite Earthquake and Tsunami, in 2011 Anholt-GfK Roper City Brands Index. Available at http://www.gfk.com/north_america/htdocs/newsroom/press_releases/single_sites/008385/index.en.html; accessed 29 March 2013.

Giddens, A. (1990). *The consequences of modernity*. Stanford, California: Stanford University Press.

Glickman, N. J. (1987). Cities and International Division of Labor. In M. P. Smith & J. R. Feagin (Eds.), *The Capitalist City: global restructuring and community politics* (pp. 66-86). Oxford, UK: Basil Blackwell.

Global City Indicators Facility. (n.d.). List of Indicators. Retrieved from http://www.cityindicators.org/ProjectDeliverables.aspx

Global Policy Forum. (2014a). Defining globalization. Retrieved from http://www.globalpolicy.org/globalization/defining-globalization.html

Global Policy Forum. (2014b). General analysis on globalization of the economy. Retrieved from http://www.globalpolicy.org/globalization/globalization-of-the-economy-2-1/general-analysis-on-globalization-of-the-economy.html

Global Policy Forum. (2014c). Globalization of culture. Retrieved from http://www.globalpolicy.org/globalization/globalization-of-culture.html

Global Policy Forum. (2014d). General analysis on globalization of politics. Retrieved from http://www.globalpolicy.org/globalization/globalization-of-politics/general-analysis-on-globalization-of-politics.html

Google, (2007). Administering crawl for web and file share content: introduction. Google Search Appliance software version 5.0. Retrieved from http://code.google.com/apis/searchappliance/documentation/50/admin_crawl/Introduction.html#whenhd1

Google, (2011). Administering crawl: running a crawl. Google Search Appliance software version 6.12. Retrieved from http://code.google.com/apis/searchappliance/documentation/612/admin_crawl/StartingCrawl.html#dbstart

Gould, P., & White, R. (1986). *Mental maps* (2nd ed.). Boston, MA: Allen & Unwin.

Government of Dubai, (2007). *Highlights Dubai Strategic Plan (2015)*. Dubai, UAE. Available at: http://www.deg.gov.ae/SiteCollectionImages/Content/pubdocs/Dubai_Strategic_Plan_2015.pdf; accessed 4 January 2013.

Govers, R., & Go, F. M. (2003). Deconstructing destination image in the information age. *Information Technology & Tourism*, 6(1), 13-29.

Govers, R., & Go, F. M. (2005). Projected destination image online: website content analysis of pictures and text. *Information Technology & Tourism*, 7(2), 73-89.

Govers, R., & Go, F. M. (2009). *Place branding: glocal, virtual and physical identities, constructed, imagined and experienced*. New York, NY: Palgrave Macmillan.

Graham, S. (1999). Global grids of glass: on global cities, telecommunications and planetary urban networks. *Urban Studies*, 36(5-6), 929-949.

Graham, S. (2002). Communication grids: cities and infrastructure. In S. Sassen (Ed.), *Global network, linked cities* (pp. 71-91). New York, NY: Routledge.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.

Hales, M., & Pena, A. M. (2012). *2012 Global cities index and emerging cities outlook*. Chicago, IL: A.T. Kearney, Inc.

Hall, P. (1984). *The world cities* (3rd ed.). New York, NY: St. Martin's Press.

Hamnett, C. (1994). Social polarization in global cities: theory and evidence. *Urban Studies*, 31(3), 401-424.

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: insights from a connected world*. Burlington, MA: Elsevier Inc.

Harvey, D. (1982). *The limits to capital*. Oxford, UK: Blackwell.

Harvey, D. (1985). *The urbanization of capital: studies in the history and theory of capitalist urbanization*. Baltimore, MD: Johns Hopkins University Press.

Harvey, D. (1989). *The condition of postmodernity: an enquiry into the origin of cultural change*. London, UK: Basil Blackwell.

Hatch, M. J., & Schultz, M. (2008). *Taking Brand Initiative: How to Align Strategy, Culture and Identity through Corporate Branding*. San Francisco, California: Jossey-Bass/Wiley.

Hatch, M. J., & Schultz, M. (2009). Of Bricks and Brands: From corporate to enterprise branding. *Organizational Dynamics*, 38(2), 117-130.

Heung, V. C. S. (2004). Internet usage by international travelers: reasons and barriers. *International Journal of Contemporary Hospitality Marketing*, 15(7), 370-378.

Hildreth, J. (2008). *The Saffron European city brand barometer*. Saffron Brand Consultants. Available at: http://www.macrame.tv/storage/Saff_CityBrandBarom.pdf; accessed 29 March 2013.

Holston, J., & Appadurai, A. (1996). Cities and citizenship. *Public Culture*, 8, 187-204.

Hymer, S. (1972). The multinational corporation and the law of uneven development. In J. Bhagwati (Ed.), *Economics and World Order from the 1970s to the 1990s* (pp. 113-140). New York, NY: Macmillan.

Internet World Stats. (2011). Internet Usage Statistics – The Big Picture: World Internet Users and Population Stats. Internet World Stats, 31st December. Available at: http://www.internetworldstats.com/stats.htm; accessed 6 March 2012.

Internet World Stats. (2012). Top 20 countries in the Internet by number of users on June 30, 2012. Miniwatts Marketing Group. Available at: www.internetworldstats.com/top20.htm; accessed 4 June 2014.

Janelle, D. G. (1969). Spatial reorganization: a model and concept. *Annals of the Association of American Geographers*, 59, 348-364.

Jo, S. J. (1992). *The world city hierarchy and the city of Seoul: operationalizing and assessing the 'world city hypothesis'*. Unpublished PhD thesis, University of Delaware.

Kawamura, S., Otake, Y. -H., & Suzuki, T. (2009). The structure of the hyperlink network formed by the web pages of Japanese public libraries. *Journal of the American Society for Information Science and Technology*, 60(6), 1159-1167.

Kearney, A. T. (2008). *The 2008 Global Cities Index. Executive Agenda*. Chicago, IL: A.T. Kearney, Inc.

Kearney, A. T. (2010). *The Urban Elite: The A.T. Kearney Global Cities Index 2010*. Chicago, IL: A.T. Kearney, Inc.

Kearney, A. T. (2012). *2012 Global Cities Index and Emerging Cities Outlook*. Chicago, IL: A.T. Kearney, Inc. Available from http://www.atkearney.com/documents/10192/dfedfc4c-8a62-4162-90e5-2a3f14f0da3a; accessed 18 February 2014.

Keeling, D. J. (1995). Transport and the world city paradigm. In P. L. Knox & P. J. Taylor (Eds.), *World Cities in a World-System* (pp. 115-131). Cambridge, UK: Cambridge University Press.

Kim, H. M., & Park, J. Y. (2005). Segyedosiseong jipyo boonryuwa sisajeome gwanhan yeongu [A study on world-cityness indicators and their implications]. *Journal of Korea Planners Association*, 40(6), 23-37.

King, A. D. (1990). *Global cities: post-imperialism and the internationalization of London*. London, UK: Routledge.

Knight, R. V. (1989). The emergent global society. In R. V. Knight & G. Gappert (Eds.), *Cities in a Global Society* (pp. 24-43). Newbury Park, California: SAGE Publications.

Knox, P. L. (1995). World cities in a world-system. In P. L. Knox & P. J. Taylor (Eds.), *World cities in a world-system* (pp. 3-20). Cambridge, UK: Cambridge University Press.

Knox, P. L., Agnew, J., & McCarthy, L. (2008). *The geography of the world economy* (5th ed.). London, UK: Hodder Education.

Kolaczyk, E. D. (2009). *Statistical analysis of network data: methods and models*. New York, NY: Springer.

Kolb, B. M. (2006). *Tourism marketing for cities and towns: using branding and events to attract tourists*. Burlington, MA: Butterworth-Heinemann.

London Planning Advisory Council. (1991). *London: world city moving into the twenty first century*. London, UK: HMSO.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: The MIT Press.

Markusen, A. (1999). Fuzzy concepts, scanty evidence, policy distance: the case for rigour and policy relevance in critical regional studies. *Regional Studies*, 33, 869-884.

Massey, D. (1984). *Spatial divisions of labor: social structures and the geography of production*. New York, NY: Methuen.

Mastercard. (2008). *Insights: Worldwide Centers of Commerce Index 2008*. Purchase, NY: Mastercard Worldwide.

McQuitty, L. L. (1960). Hierarchical linkage analysis for the isolation of types. *Educational and Psychological Measurement*, 20(1), 55-67.

Milligan, G. W., & Cooper, M. C. (1987). Methodology review: clustering methods. *Applied Psychological Measurement*, 11(4), 329-354.

Mitchell, R. K., Agle, B., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, 22(4), 853-886.

Murray, W.E. (2006). *Geographies of globalization*. New York, NY: Routledge.

Nam, Y. W. (1998). A study of linkage policy for downtown redevelopment. *Journal of Korea Planners Association*, 33(6), 49-65.

Nam, Y. W. (2006). *Globalsidaeeui segyedosiron* [The global city in the era of globalization]. Seoul, Korea: Bobmunsa.

Nam, Y. W., & Lee, H. S. (2004). Hanggonghwamoolryuedongryangeuro bon segyedosisystemeui byeonhwa [Changes of global urban system reflected in air freight flows]. *Journal of Korea Planners Association*, 39(1), 129-143.

Nam, Y. W., & Park, S. K. (1998). World city strategy and urban restructuring; a case study of Fukuoka, Japan. *Korean Journal of Urban Geographical Society*, 1(1), 15-30.

Nam, Y. W., Lee, H. Y., & Choi, J. H. (2000). *Gyeongje, geumyung, dosieui segyehwa* [Economic, financial, and urban globalization]. Seoul, Korea: Darakbang.

Netcraft. (2012). Web Server Survey. Netcraft, 5th March. Available at: http://news.netcraft.com; accessed 7 March 2012.

Netcraft. (2014). June 2104 Web Server Survey. Netcraft, 6th June. Available at: http://news.netcraft.com; accessed 17 June 2014.

NetMiner (Version 4.2) [Software and user manual]. (2014). Available from http://www.netminer.com/index.php

Newman, M. E. J. (2010). *Networks: an introduction*. New York, NY: Oxford University Press.

Newprosoft. (2012). Visual Web Spider 7.2. Available from http://newprosoft.com/web-spider.htm

O'Connor, P., & Murphy, J. (2004). Research on information technology in the hospitality industry. *International Journal of Hospitality Management*, 23(5), 473-484.

Oh, H., Kim, B. Y., & Shin, J. H. (2004). Hospitality and tourism marketing: recent developments in research and future directions. *International Journal of Hospitality Management*, 23(5), 425-447.

Olston, C., & Najork, M. (2010). Web crawling. *Foundation and Trends® in Information Retrieval*, 4(3), 175-246.

Park, H. W. (2003). Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web. *Connections*, 25(1), 49-61.

Park, H. W., & Thelwall, M. (2003). Hyperlink analyses of the World Wide Web: a review. *Journal of Computer-Mediated Communication*, 8(4), 0. doi: 10.1111/j.1083-6101.2003.tb00223.x

Paul, L. M. (Ed.). (2009). *Ethnologue: Languages of the World* (16th ed.). Dallas, TX: SIL International. Available from http://www.ethnologue.com/

Petrella, R. (1995). A global agora vs. gated city-regions. *New Perspectives*, Winter, 21-22.

Pimienta, D. (2005). Linguistic diversity in cyberspace – models for development and measurement. In UNESCO institute of statistics (Ed.), *Measuring linguistic diversity on the Internet* (pp. 13-34). Paris, France: the United Nations Educational, Scientific and Cultural Organization.

Popping, R. (2000). *Computer-assisted Text Analysis*. London, UK: SAGE Publications Ltd.

Prado, D. (2012). Language presence in the real world and cyberspace. In L. Vannini, & H. L. Crosnier (Eds.), *NET.lang: towards the multilingual cyberspace* (pp. 34-51). Caen, France: C&F èditions.

Pred, A. (1980). *Urban Growth and City Systems in the United States 1840-1860*. London, UK: Hutchinson.

Provalis Research. (2010). *User's guide. WordStat 6: Content Analysis Module for QDA Miner & SimStat*. Montreal, Canada. Available from http://www.provalisresearch.com/Documents/WordStat6.pdf

Provalis Research. (2010). WordStat 6: Content Analysis Module for QDA Miner & Simstat. Montreal, Canada. Available from http://provalisresearch.com/

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available from http://www.R-project.org/

Reed, H. C. (1981). *The Pre-eminence of International Financial Centers*. New York, NY: Praeger.

Riffe, D., Lacy, S., & Fico, F. G. (2005). *Analyzing media messages: using quantitative content analysis in research* (2nd ed.). New York, NY: Taylor & Francis Group.

Rimmer, P. J. (1986). Japan's world cities: Tokyo, Osaka, Nagoya, or Tokaido megalopolis? *Development and Change*, 17(1), 121-158.

Rimmer, P. J. (1991). The global intelligence corps and world cities: engineering consultancies on the move. In P. W. Daniels (Ed.), *Services and Metropolitan Development: International Perspectives* (pp. 66-106). London, UK: Routledge.

Samers, M. (2002). Immigration and the global city hypothesis: towards an alternative research agenda. *Journal of Urban and Regional Research*, 26(2), 389-402.

Sassen, S. (1994). *Cities in a world economy*. Thousand Oaks, California: Pine Forge Press.

Sassen, S. (2011). *Cities in a world economy* (4th ed.). Thousand Oaks, California: Pine Forge Press.

Schultz, D. E., & Kitchen, P. J. (2004). Managing the changes in corporate branding and communication: Closing and re-opening the Corporate Umbrella. *Corporate Reputation Review*, 6(4), 347-366.

Scott, A. J. (1988). *Metropolis: from the division of labor to urban form*. Berkeley, California: University of California Press.

Selwyn, T. (1996). Introduction. In T. Selwyn (Ed.), *The Tourist Image: Myths and Myth Making in Tourism* (pp. 1-32). New York, NY: John Wiley & Sons.

Short, J. R., & Kim, Y. H. (1999). *Globalization and the City*. London, UK: Longman.

Short, J. R., Kim, Y., Kuus, M., & Wells, H. (1996). The dirty little secret of world cities research: data problems in comparative analysis. *International Journal of Urban and Regional Research*, 20(4), 697-717.

SIL International. (2011). First-Language: Bridge to New Possibilities, 2011 Update, pp. 2, 8. Available from http://www.sil.org/sil/annualreport/2011-sil-update.pdf

Skupin, A., & Agarwal, P. (2008). Introduction: What is a Self-Organizing Map? In P. Agarwal & A. Skupin (Eds.), *Self-Organizing Maps: Application in Geographic Information Science* (pp. 1-20). Chichester, UK: John Wiley & Sons, Ltd.

Smith, D. A., & Timberlake, M. F. (1995). Cities in global matrices: toward mapping the world-system's city system. In P. L. Knox & P. J. Taylor (Eds.), *World Cities in a World System* (pp. 79-97). Cambridge, UK: Cambridge University Press.

Smith, D. A., & Timberlake, M. F. (2001). World city networks and hierarchies, 1977-1997: an empirical analysis of global air travel links. *American Behavioral Scientist*, 44(10), 1656-1678.

Sneath, P. H. A. (1957). The application of computers to taxonomy. *Journal of General Microbiology*, 17(1), 201-226.

Soja, E. W. (1989). *Postmodern Geographies: the reassertion of space in critical social theory*. London, UK: Verso.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 38(22), 1409-1438.

Son, J. S. (2006a). *Jooyo global sinmooneuro bon segyedosieui teukseonggwa yuhyeonghwa* [The global cities' characteristics and their classification in the world's prestigious newspapers] (Master's thesis). Korea University, Seoul, Korea.

Son, J. S. (2006b). Jooyo global sinmooneuro bon segyedosieui teukseonggwa yuhyeonghwa [The global cities' characteristics and their classification in the world's prestigious newspapers: gleaned from The Wall Street Journal, The Financial Times, and Sankei Shimbun]. *Korean Journal of Urban Geographical Society*, 9(2), 101-111.

Stempel, G. H., III, & Stewart, R. K. (2000). The Internet provides both opportunities and challenges for mass communication researchers. *Journalism and Mass Communication Quarterly*, 77(3), 541-548.

Stepchenkova, S., & Morrison, A. (2006). The destination image of Russia: from the online induced perspective. *Tourism Management*, 27(5), 943-956.

Taylor, P. J. (1997). Hierarchical tendencies amongst world cities: a global research proposal. *Cities*, 14(6), 323-332.

Taylor, P. J. (2004). *World city network: a global urban analysis*. New York, NY: Routledge.

Taylor, P. J., & Catalano, G. (2000). World city network: the basic data (data set 11). Globalization and World Cities (GaWC). Available from http://www.lboro.ac.uk/gawc/datasets/da11.html; accessed 26 June 2012.

Taylor, P. J., Walker, D. R. F., & Beaverstock, J. V. (2002). Firms and their global service networks. In S. Sassen (Ed.), *Global network, linked cities* (pp. 93-115). New York, NY: Routledge.

Thelwall, M. (2004). *Link analysis: an information science approach*. San Diego, CA: Elsevier Inc.

Thelwall, M. (2009). *Introduction to webometrics: quantitative web research for the social sciences*. Lexington, KY: Morgan & Claypool.

Thrift, N. (1988). The geography of international economic disorder. In D. Massey & J. Allen (Eds.), *Uneven re-development: cities and regions in transition* (pp. 6-46). London, UK: Hooder & Stoughton.

Thrift, N. (1999). Cities and economic change: global governance?. In J. Allen, D. Massey, & M. Pryke (Eds.), *Unsettling Cities* (pp. 283-338). New York, NY: Routledge.

Tokyo Metropolitan Government. (1991). *The 3rd long-term plan for the Tokyo Metropolis (outline): "My town Tokyo," for the dawn of the 21st century*. Tokyo, Japan: TMG Municipal Library.

Twentieth Century Fund. (1980). *New York-word city: report of the Twentieth Century Fund task force on the future of New York City*. Cambridge, MA: Oelgeschlager, Gunn & Hain.

UN DESA, Population Division. (2011). *World Population Prospects: the 2010 Revision, Volume I: Comprehensive Tables*. United Nations, Department of Economic and Social Affairs, Population Division.

UN DESA, Population Division. (2012). *World urbanization prospects: the 2011 revision* (CD-ROM ed.). United Nations, Department of Economic and Social Affairs, Population Division.

UNESCO. (1999). *Guide to advanced data analysis using IDAMS software*. United Nations Educational, Scientific and Cultural Organization. Available at: http://www.unesco.org/webworld/idams/advguide/TOC.htm; accessed 1 May 2012.

Vaughan, L., &Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 40(4), 693-707.

Vesanto, J. (1999). SOM-based data visualization methods. *Intelligent Data Analysis*, 3, 111-126.

Vesanto, J., Himberg, J., Alhonienmi, E., & Parhankangas, J. (2000). *SOM toolbox for Matlab 5*. Helsinki, Finland: Helsinki University of Technology.

Vonk, F. P. M. (1989). Managing the metropolis. In R. V. Knight & G. Gappert (Eds.), *Cities in a Global Society* (pp. 181-194). Newbury Park, California: SAGE Publications.

W3Techs. (2012a). Usage of content languages for websites. Available at: http://w3techs.com/technologies/overview/content_language/all; accessed 2 April 2012.

W3Techs. (2012b). Historical quarterly trends in the usage of content languages for websites. Available at: http://w3techs.com/technologies/history_overview/content_language/ms/q; accessed 2 April 2012.

W3Techs. (2014). Usage of top level domains for websites. Available at: http://w3techs.com/technologies/overview/top_level_domain/all; accessed 4 June 2014.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.

Ward, S. V. (1998). *Selling places: the marketing and promotion of towns and cities 1850-2000*. New York, NY: Spon Press.

Warf, B. (1995). Telecommunications and the changing geographies of knowledge transmission in the late 20th century. *Urban Studies*, 32(2), 361-378.

Wasserman, S., & Faust, K. (1994). *Social network analysis*. New York, NY: Cambridge University Press.

Weare, C., & Lin, W. Y. (2000). Content analysis of the World Wide Web: opportunities and challenges. *Social Science Computer Review*, 18(3), 272-292.

Weber, G. (1997). The World's 10 most influential languages. *Language Monthly*, 3, 12-18.

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-049. Newbury Park, CA: Sage.

Zelle, J. (2004). *Python programming: an introduction to computer science*. Wilsonville, OR: Franklin, Beedle & Associates, Inc.

APPENDIX A: MYSQL SCRIPT FOR DMOZ DB AND SEED URL

1. Loading data into DB

```
LOAD DATA LOCAL INFILE 'c:/Temp/Regional_categ.txt' into table category;
LOAD DATA LOCAL INFILE 'c:/Temp/Regional_listing.txt' into table list;
CREATE TABLE category1 (
category_id MEDIUMINT(9) NOT NULL,
path VARCHAR(255) NOT NULL,
name VARCHAR(255) NOT NULL
);
LOAD DATA LOCAL INFILE 'c:/Temp/Regional_categ.txt' into table category1
FIELDS TERMINATED BY '\t';
CREATE TABLE listing1 (
category_id MEDIUMINT(9) NOT NULL,
url VARCHAR(255) NOT NULL DEFAULT '',
title VARCHAR(255) NOT NULL DEFAULT '',
description TEXT NOT NULL,
path VARCHAR(255) NOT NULL DEFAULT ''
);
LOAD DATA LOCAL INFILE 'c:/Temp/Regional_listing.txt' into table listing1
FIELDS TERMINATED BY '\t';
```

2. Example of extracting URLs from DB

```
SELECT url FROM listing1
WHERE path LIKE "%/North_Holland/Amsterdam/%"
```

```
INTO OUTFILE 'c:/temp/dmozurls/Amsterdam.txt'
LINES TERMINATED BY '\r\n';
```

3. Example of adding seed URLs to table

```
LOAD DATA LOCAL INFILE 'c:/Temp/dmozurls/Amsterdam.txt' into TABLE
amsterdam;
```

4. Example of exporting seed URLs to text file

```
SELECT seedurls FROM dissert.amsterdam
INTO OUTFILE 'c:/temp/seedurls/Amsterdam.txt'
LINES TERMINATED BY '\r\n';
```

APPENDIX B: MYSQL SCRIPT FOR RESEARCH DATA TABLE

1. Example of creating bulk tables for importing crawled data

```
CREATE TABLE abudhabi_440788.bulk_doc (
raw MEDIUMTEXT
);
LOAD DATA LOCAL INFILE
'F:/FN_MergedCSV/440788/440788_document.csv' INTO TABLE
abudhabi_440788.bulk_doc;
CREATE TABLE abudhabi_440788.bulk_link (
raw MEDIUMTEXT
);
LOAD DATA LOCAL INFILE 'F:/FN_MergedCSV/440788/440788_link.csv'
INTO TABLE abudhabi_440788.link_doc;
CREATE TABLE abudhabi_440788.bulk_total (
raw MEDIUMTEXT
);
LOAD DATA LOCAL INFILE 'F:/FN_MergedCSV/440788/440788_total.csv'
INTO TABLE abudhabi_440788.bulk_total;
```

2. Example of parsing information from bulk tables and creating structured table

```
USE abudhabi_440788;
CREATE TABLE total (
t_id INT PRIMARY KEY AUTO_INCREMENT,
bodyurl TEXT
```

```
)

SELECT SUBSTRING_INDEX(SUBSTRING_INDEX(raw,"",2),"",-1) as url

FROM abudhabi_440788.bulk_total

WHERE raw like "%http%";

UPDATE total SET bodyurl=SUBSTRING_INDEX(url,'/',3);
```

APPENDIX C: LIST OF THE TOP CENTRALITY NODES

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Abidjan	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Abu Dhabi	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Accra	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Addis Ababa	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Adelaide	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Ahmedabad	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Aleppo	twitter.com	en.wikipedia.org	twitter.com	twitter.com	huffingtonpost.com
Alexandria	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Algiers	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Almaty	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Amman	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Amsterdam	twitter.com	en.wikipedia.org	twitter.com	twitter.com	topix.com
Ankara	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Antwerp	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Asunción	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Athens	twitter.com	en.wikipedia.org	twitter.com	facebook.com	livingingreece.gr
Atlanta	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Auckland	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Baghdad	twitter.com	en.wikipedia.org	twitter.com	twitter.com	ask.com
Baku	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Baltimore	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Bandung	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Bangalore	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Bangkok	facebook.com	en.wikipedia.org	twitter.com	facebook.com	en.wikipedia.org
Barcelona	twitter.com	en.wikipedia.org	twitter.com	twitter.com	globalvoicesonline.org
Basel	twitter.com	travigator.com	twitter.com	twitter.com	wn.com
Batam	twitter.com	blogger.com	blogger.com	twitter.com	scam.com
Beijing	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Beirut	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Belgrade	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Belo Horizonte	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	topix.com
Berlin	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Birmingham	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Bogotá	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	huffingtonpost.com
Bonn	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	huffingtonpost.com
Bordeaux	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	topix.com
Boston	facebook.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Brasília	twitter.com	enotes.com	en.wikipedia.org	twitter.com	enotes.com

APPENDIX C: (continued)

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Brazzaville	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Brisbane	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Brussels	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Bucharest	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Budapest	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Buenos Aires	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Buffalo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Busan	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Cairo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Calgary	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	calcna.ab.ca
Cape Town	twitter.com	weatherforecastmap.com	en.wikipedia.org	twitter.com	en.wikipedia.org
Caracas	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	faustasblog.com
Casablanca	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Changchun	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Changsha	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Charlotte	facebook.com	usnpl.com	twitter.com	facebook.com	usnpl.com
Chengdu	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Chennai	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Chicago	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Chittagong	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Chongqing	twitter.com	en.wikipedia.org	youtube.com	twitter.com	blogs.wsj.com
Cincinnati	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Ciudad Juarez	twitter.com	weatherforecastmap.com	en.wikipedia.org	twitter.com	huffingtonpost.com
Cleveland	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Cologne	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Colombo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Columbus	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Conakry	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Copenhagen	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Curitiba	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Dakar	twitter.com	weatherforecastmap.com	washingtonpost.com	twitter.com	topix.com
Dalian	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Dallas	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Damascus	facebook.com	en.wikipedia.org	itunes.apple.com	facebook.com	en.wikipedia.org
Dar es Salaam	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Denver	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Detroit	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Dhaka	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Doha	twitter.com	en.wikipedia.org	twitter.com	twitter.com	huffingtonpost.com

APPENDIX C: (continued)

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Dongguan	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Douala	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	topix.com
Dubai	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	avoelectronic.blogspot.com
Dublin	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Durban	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Düsseldorf	twitter.com	en.wikipedia.org	huffingtonpost.com	twitter.com	en.wikipedia.org
East Rand	twitter.com	en.wikipedia.org	twitter.com	twitter.com	topix.com
Edinburgh	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Edmonton	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Fortaleza	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	topix.com
Foshan	twitter.com	en.wikipedia.org	twitter.com	facebook.com	en.wikipedia.org
Frankfurt	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Freetown	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Geneva	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	thefullwiki.org
Glasgow	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Guadalajara	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Guangzhou	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	ask.com
Guatemala City	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Guayaquil	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Harbin	twitter.com	en.wikipedia.org	twitter.com	facebook.com	nationsonline.org
Hamburg	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Hangzhou	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Hanoi	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Harare	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Hartford	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	ask.com
Havana	twitter.com	en.wikipedia.org	twitter.com	twitter.com	southafrica.info
Helsinki	twitter.com	en.wikipedia.org	washingtonpost.com	twitter.com	ask.com
Ho Chi Minh City	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Hong Kong	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	sxl.net
Honolulu	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Houston	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Hyderabad	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Indianapolis	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Islamabad	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Istanbul	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Jaipur	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Jakarta	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Jeddah	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Jerusalem	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org

APPENDIX C: (continued)

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Jinan	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Johannesburg	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Kabul	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Kampala	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Kano	twitter.com	en.wikipedia.org	twitter.com	twitter.com	huffingtonpost.com
Kansas City	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Karachi	twitter.com	en.wikipedia.org	twitter.com	facebook.com	en.wikipedia.org
Khartoum	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Kiev	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Kinshasa	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Kobe	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Kolkata (Calcutta)	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Krakow	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Kuala Lumpur	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	globalvoicesonline.org
Kunming	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Kuwait City	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Kyoto	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
La Paz	twitter.com	en.wikipedia.org	twitter.com	twitter.com	huffingtonpost.com
Lagos	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Lahore	twitter.com	en.wikipedia.org	twitter.com	twitter.com	globalvoicesonline.org
Las Vegas	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Lille	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Lima	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Lisbon	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Liverpool	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
London	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Los Angeles	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Luanda	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Lucknow	facebook.com	en.wikipedia.org	twitter.com	facebook.com	en.wikipedia.org
Lusaka	twitter.com	en.wikipedia.org	twitter.com	twitter.com	ask.com
Luxembourg	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Lyons	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Madrid	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Managua	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Manaus	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Manchester	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Manila	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Maputo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Marseille	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org

APPENDIX C: (continued)

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Medan	twitter.com	en.wikipedia.org	twitter.com	twitter.com	topix.com
Medellin	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Melbourne	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Mexico City	twitter.com	en.wikipedia.org	twitter.com	twitter.com	theproudfranchise.com
Miami	twitter.com	en.wikipedia.org	twitter.com	facebook.com	huffingtonpost.com
Milan	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Minneapolis	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Minsk	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Mombasa	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Monrovia	twitter.com	en.wikipedia.org	twitter.com	twitter.com	topix.com
Monterrey	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Montevideo	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Montreal	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Moscow	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Mumbai	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Munich	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Nagoya	twitter.com	en.wikipedia.org	twitter.com	twitter.com	wn.com
Nairobi	facebook.com	en.wikipedia.org	twitter.com	facebook.com	topix.com
Nanjing	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Naples	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
New Delhi (Delhi)	twitter.com	qfkd.com	en.wikipedia.org	facebook.com	qfkd.com
New Orleans	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
New York	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	december.com
Newcastle	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Omaha	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Osaka	twitter.com	en.wikipedia.org	twitter.com	twitter.com	fabiocaparica.com
Oslo	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Ottawa	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	ottawastart.com
Palermo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Panama City	twitter.com	en.wikipedia.org	huffingtonpost.com	twitter.com	huffingtonpost.com
Paris	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Perth	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Philadelphia	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Phoenix	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Pittsburgh	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Port-au-Prince	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Portland	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Porto Alegre	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	enotes.com
Prague	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org

APPENDIX C: (continued)

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Pretoria	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Pune	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Pyongyang	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Qingdao	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Quito	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Rabat	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	thecomingcrisis.blogspot.com
Rawalpindi	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	mypetjawa.mu.nu
Recife	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	absoluteastronomy.com
Richmond	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Rio de Janeiro	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Riyadh	twitter.com	huffingtonpost.com	huffingtonpost.com	twitter.com	en.wikipedia.org
Rochester	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Rome	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	lawpundit.com
Rotterdam	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Sacramento	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Saint Louis	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Saint Petersburg	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Salvador	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	mycountdown.org
San Diego	twitter.com	en.wikipedia.org	huffingtonpost.com	twitter.com	huffingtonpost.com
San Francisco	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
San Jose	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
San Jos é	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	wn.com
San Salvador	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Santiago	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Santo Domingo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	huffingtonpost.com
S ão Paulo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	mycountdown.org
Seattle	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Seoul	twitter.com	huffingtonpost.com	en.wikipedia.org	twitter.com	abc-directory.com
Shanghai	twitter.com	schema-root.org	twitter.com	facebook.com	schema-root.org
Shantou	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Shenyang	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Shenzhen	twitter.com	en.wikipedia.org	huffingtonpost.com	twitter.com	huffingtonpost.com
Singapore	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Sofia	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Stockholm	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	topix.com
Stuttgart	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Surat	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Suzhou	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Sydney	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org

APPENDIX C: (continued)

City	In-Degree	Out-Degree	Node Betweenness	In-Closeness	Out-Closeness
Taipei	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Taiyuan	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Tampa	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Tashkent	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Tbilisi	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Tegucigalpa	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Tehran	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Tel Aviv	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
The Hague	twitter.com	huffingtonpost.com	huffingtonpost.com	facebook.com	huffingtonpost.com
Tianjin	twitter.com	en.wikipedia.org	huffingtonpost.com	twitter.com	en.wikipedia.org
Tijuana	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Tokyo	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Toronto	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Tripoli	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	bangkokcompanies.com
Tunis	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Turin	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Ulan Bator	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Valencia	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Vancouver	twitter.com	en.wikipedia.org	twitter.com	twitter.com	en.wikipedia.org
Vienna	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Warsaw	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Washington	facebook.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Wuhan	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	ask.com
Wuxi	twitter.com	en.wikipedia.org	en.wikipedia.org	facebook.com	en.wikipedia.org
Xiamen	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Xi'an	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Yangon	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Yaoundé	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Yerevan	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Zhengzhou	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org
Zurich	twitter.com	en.wikipedia.org	en.wikipedia.org	twitter.com	en.wikipedia.org

APPENDIX D: RANKINGS BASED ON PROPORTION OF EACH CATEGORY TO TOTAL WORDS

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
1	Dalian	0.023666	1	Portland	0.010752	1	Kinshasa	0.012899
2	Shanghai	0.019999	2	Houston	0.010577	2	Islamabad	0.011307
3	East Rand	0.015876	3	Minneapolis	0.009979	3	Pyongyang	0.011134
4	Zhengzhou	0.015717	4	Atlanta	0.009463	4	Baghdad	0.010998
5	Luxembourg	0.014039	5	Chicago	0.009260	5	Kabul	0.010666
6	Chittagong	0.014035	6	Indianapolis	0.009104	6	Damascus	0.010471
7	Sao Paulo	0.013967	7	Dallas	0.009057	7	Khartoum	0.010232
8	Brazzaville	0.013369	8	Charlotte	0.009035	8	Aleppo	0.009243
9	Chennai	0.013336	9	Boston	0.008895	9	Kano	0.009206
10	Foshan	0.013114	10	Cincinnati	0.008870	10	Brazzaville	0.009179
11	Shenzhen	0.012334	11	Saint Louis	0.008820	11	Tripoli	0.008949
12	Dubai	0.012116	12	Ottawa	0.008810	12	Tehran	0.008405
13	Kuala Lumpur	0.012056	13	Philadelphia	0.008805	13	Addis Ababa	0.007812
14	Riyadh	0.011897	14	Detroit	0.008773	14	Monrovia	0.007656
15	Xiamen	0.011794	15	Hartford	0.008674	15	Rawalpindi	0.007517
16	Ahmedabad	0.011726	16	Cleveland	0.008457	16	Abidjan	0.007002
17	Brisbane	0.011722	17	Melbourne	0.008364	17	Karachi	0.006671
18	Dongguan	0.011501	18	Washington	0.008318	18	Ciudad Juarez	0.006594
19	Singapore	0.011290	19	Brussels	0.008142	19	Conakry	0.006494
20	London	0.011178	20	San Francisco	0.007965	20	Washington	0.006477
21	Karachi	0.011154	21	Porto Alegre	0.007941	21	Freetown	0.006420
22	Doha	0.011090	22	Los Angeles	0.007911	22	Busan	0.006403
23	Hong Kong	0.010952	23	Toronto	0.007779	23	Dakar	0.006375
24	Mumbai	0.010926	24	Chennai	0.007735	24	Brussels	0.006370
25	Kolkata	0.010735	25	Adelaide	0.007671	25	Cairo	0.006143
26	Perth	0.010484	26	Omaha	0.007548	26	Nairobi	0.005981
27	Nairobi	0.010425	27	London	0.007495	27	Hong Kong	0.005872
28	Basel	0.010388	28	Miami	0.007357	28	The Hague	0.005817
29	Qingdao	0.010353	29	Mumbai	0.007337	29	Dhaka	0.005736
30	Houston	0.010299	30	Vancouver	0.007300	30	Kampala	0.005701
31	Frankfurt	0.010298	31	Columbus	0.007285	31	Geneva	0.005636
32	Bangalore	0.010281	32	Seoul	0.007282	32	Luxembourg	0.005622
33	Shantou	0.010247	33	Seattle	0.007276	33	Harare	0.005598
34	Charlotte	0.010230	34	Lisbon	0.007240	34	Jerusalem	0.005554
35	Beijing	0.010198	35	New York	0.007231	35	Tbilisi	0.005538
36	Ho Chi Minh City	0.010150	36	Baltimore	0.007226	36	Beirut	0.005417
37	Bangkok	0.010113	37	San Diego	0.007106	37	Brasilia	0.005411

APPENDIX D: (continued)

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
38	Abu Dhabi	0.010087	38	Kolkata	0.007000	38	Ankara	0.005342
39	Saint Louis	0.010079	39	Denver	0.006980	39	Port-au-Prince	0.005326
40	Toronto	0.009976	40	Richmond	0.006954	40	Lusaka	0.005313
41	Sydney	0.009856	41	San Jose CA	0.006898	41	Baku	0.005219
42	Pune	0.009832	42	Sydney	0.006879	42	Asuncion	0.005192
43	Zurich	0.009801	43	Kinshasa	0.006833	43	Dalian	0.005127
44	Guangzhou	0.009783	44	Phoenix	0.006802	44	Caracas	0.005052
45	Melbourne	0.009713	45	Rochester	0.006790	45	Yerevan	0.005040
46	Surat	0.009711	46	Singapore	0.006777	46	Bonn	0.004924
47	Boston	0.009625	47	Hong Kong	0.006777	47	Almaty	0.004919
48	Brussels	0.009538	48	Brasilia	0.006774	48	Saint Louis	0.004858
49	Rawalpindi	0.009530	49	Budapest	0.006721	49	Basel	0.004809
50	Ottawa	0.009465	50	Beijing	0.006697	50	Dar es Salaam	0.004759
51	Atlanta	0.009387	51	Chongqing	0.006695	51	Ottawa	0.004751
52	San Diego	0.009354	52	Pittsburgh	0.006636	52	Belgrade	0.004744
53	Wuxi	0.009302	53	Dublin	0.006629	53	Boston	0.004743
54	Taipei	0.009252	54	Warsaw	0.006610	54	Pretoria	0.004707
55	Manchester	0.009132	55	Asuncion	0.006551	55	Portland	0.004693
56	Casablanca	0.008986	56	Athens	0.006549	56	Detroit	0.004690
57	Kampala	0.008976	57	Shanghai	0.006517	57	Colombo	0.004684
58	Dallas	0.008966	58	New Delhi (Delhi)	0.006486	58	Monterrey	0.004647
59	Dublin	0.008962	59	Manchester	0.006476	59	Warsaw	0.004625
60	Islamabad	0.008946	60	Sacramento	0.006473	60	New Delhi (Delhi)	0.004616
61	Chicago	0.008938	61	Riyadh	0.006454	61	Tunis	0.004589
62	Kinshasa	0.008906	62	Tehran	0.006411	62	Minsk	0.004560
63	Cleveland	0.008899	63	Barcelona	0.006407	63	Bogota	0.004557
64	Hangzhou	0.008890	64	Baghdad	0.006366	64	Columbus	0.004550
65	Adelaide	0.008872	65	Busan	0.006364	65	Charlotte	0.004528
66	Rochester	0.008852	66	Pyongyang	0.006361	66	Manila	0.004509
67	Denver	0.008830	67	Kabul	0.006347	67	Chicago	0.004427
68	Harbin	0.008819	68	Berlin	0.006296	68	Singapore	0.004421
69	Kansas City	0.008815	69	Dar es Salaam	0.006285	69	Minneapolis	0.004419
70	Athens	0.008809	70	Auckland	0.006278	70	Hartford	0.004416
71	Almaty	0.008802	71	Santiago	0.006180	71	Taipei	0.004412
72	Jinan	0.008782	72	Kansas City	0.006136	72	Cleveland	0.004381
73	Los Angeles	0.008753	73	Rawalpindi	0.006111	73	New York	0.004375
74	Hartford	0.008729	74	Cape Town	0.006095	74	Accra	0.004369
75	San Jose CA	0.008699	75	Baku	0.006093	75	Moscow	0.004355
76	New Delhi (Delhi)	0.008662	76	Madrid	0.006083	76	Maputo	0.004334

APPENDIX D: (continued)

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
77	Johannesburg	0.008649	77	Bangalore	0.005998	77	Chennai	0.004327
78	Detroit	0.008644	78	Paris	0.005995	78	Baltimore	0.004293
79	Phoenix	0.008638	79	Kampala	0.005966	79	Doha	0.004280
80	Washington	0.008636	80	Manila	0.005942	80	Riyadh	0.004274
81	Cincinnati	0.008609	81	Doha	0.005923	81	Sao Paulo	0.004267
82	San Francisco	0.008598	82	Edinburgh	0.005922	82	Cincinnati	0.004254
83	Lusaka	0.008570	83	Dubai	0.005911	83	San Francisco	0.004253
84	Minneapolis	0.008540	84	Geneva	0.005898	84	Lucknow	0.004250
85	Hyderabad	0.008533	85	Cairo	0.005895	85	Richmond	0.004249
86	New York	0.008513	86	Glasgow	0.005893	86	Kolkata	0.004236
87	Auckland	0.008436	87	New Orleans	0.005860	87	Toronto	0.004212
88	Asuncion	0.008417	88	Karachi	0.005844	88	Seoul	0.004189
89	Seattle	0.008396	89	Buenos Aires	0.005838	89	Adelaide	0.004170
90	Jeddah	0.008353	90	Dhaka	0.005795	90	Dallas	0.004170
91	Abidjan	0.008291	91	Sao Paulo	0.005783	91	Phoenix	0.004128
92	Portland	0.008275	92	Hyderabad	0.005760	92	Algiers	0.004079
93	Tianjin	0.008268	93	Amman	0.005759	93	Istanbul	0.004076
94	Dakar	0.008241	94	Dakar	0.005751	94	Atlanta	0.004066
95	Conakry	0.008194	95	Tampa	0.005743	95	Beijing	0.004040
96	Omaha	0.008179	96	Rabat	0.005725	96	Shanghai	0.004032
97	San Jose CR	0.008120	97	Birmingham	0.005716	97	Lisbon	0.004020
98	Jakarta	0.008113	98	Montreal	0.005711	98	Indianapolis	0.003998
99	Lucknow	0.008094	99	Islamabad	0.005707	99	Athens	0.003980
100	Geneva	0.008088	100	Lyons	0.005706	100	Omaha	0.003973
101	Dhaka	0.008065	101	Pretoria	0.005697	101	Stockholm	0.003961
102	Philadelphia	0.008011	102	Calgary	0.005674	102	Jakarta	0.003933
103	Cape Town	0.008001	103	Kano	0.005667	103	Chongqing	0.003928
104	Dusseldorf	0.007983	104	Freetown	0.005613	104	Rochester	0.003914
105	Columbus	0.007976	105	Aleppo	0.005604	105	Philadelphia	0.003913
106	Kano	0.007908	106	Dalian	0.005566	106	Havana	0.003913
107	Pretoria	0.007899	107	Edmonton	0.005537	107	San Diego	0.003905
108	The Hague	0.007888	108	Brazzaville	0.005502	108	Tel Aviv	0.003900
109	Monrovia	0.007885	109	The Hague	0.005500	109	Amman	0.003865
110	Buenos Aires	0.007882	110	Damascus	0.005479	110	Tegucigalpa	0.003858
111	Indianapolis	0.007879	111	Nairobi	0.005462	111	Berlin	0.003851
112	Tampa	0.007878	112	Wuxi	0.005425	112	Yaoundé	0.003842
113	Pittsburgh	0.007784	113	Almaty	0.005422	113	Yangon	0.003841
114	Moscow	0.007765	114	Perth	0.005418	114	Curitiba	0.003830
115	Pyongyang	0.007680	115	Jakarta	0.005407	115	Luanda	0.003815

APPENDIX D: (continued)

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
116	Munich	0.007664	116	Prague	0.005359	116	Hyderabad	0.003813
117	Miami	0.007644	117	Tripoli	0.005351	117	Pittsburgh	0.003812
118	Vancouver	0.007630	118	Bangkok	0.005338	118	Denver	0.003808
119	Beirut	0.007616	119	Taipei	0.005326	119	Mexico City	0.003800
120	Busan	0.007581	120	Amsterdam	0.005313	120	Managua	0.003796
121	Changchun	0.007567	121	Johannesburg	0.005304	121	Jinan	0.003788
122	Baltimore	0.007507	122	Chittagong	0.005288	122	Houston	0.003767
123	Honolulu	0.007480	123	Abidjan	0.005279	123	Los Angeles	0.003763
124	Warsaw	0.007462	124	Rotterdam	0.005276	124	Rabat	0.003759
125	Manila	0.007402	125	Jerusalem	0.005256	125	Kansas City	0.003758
126	Prague	0.007386	126	Las Vegas	0.005217	126	Oslo	0.003754
127	Santiago	0.007300	127	Ankara	0.005202	127	Porto Alegre	0.003735
128	Richmond	0.007295	128	Munich	0.005180	128	Dublin	0.003717
129	Chengdu	0.007278	129	Managua	0.005170	129	Mumbai	0.003653
130	Taiyuan	0.007258	130	Ahmedabad	0.005145	130	London	0.003653
131	Douala	0.007242	131	Buffalo	0.005134	131	Miami	0.003637
132	Chongqing	0.007223	132	Lima	0.005126	132	Johannesburg	0.003634
133	Calgary	0.007191	133	Turin	0.005114	133	Chittagong	0.003629
134	Saint Petersburg	0.007166	134	Jeddah	0.005111	134	San Jose CA	0.003611
135	Addis Ababa	0.007112	135	Basel	0.005081	135	Xiamen	0.003569
136	Montreal	0.007102	136	Moscow	0.005054	136	Sofia	0.003569
137	Edinburgh	0.007000	137	Bonn	0.005049	137	Tianjin	0.003566
138	Newcastle	0.006990	138	Salvador	0.005046	138	Santiago	0.003560
139	Las Vegas	0.006970	139	Khartoum	0.005000	139	Melbourne	0.003559
140	Sacramento	0.006930	140	Surat	0.004981	140	Alexandria	0.003547
141	Rio de Janeiro	0.006919	141	Monrovia	0.004959	141	Glasgow	0.003540
142	Monterrey	0.006886	142	Brisbane	0.004928	142	Nanjing	0.003537
143	Minsk	0.006869	143	Rio de Janeiro	0.004923	143	Paris	0.003533
144	Birmingham	0.006854	144	Quito	0.004912	144	Bangkok	0.003532
145	Amman	0.006842	145	Harare	0.004894	145	Wuxi	0.003529
146	Edmonton	0.006823	146	Lusaka	0.004893	146	Seattle	0.003509
147	Osaka	0.006820	147	Stuttgart	0.004879	147	Mombasa	0.003502
148	Glasgow	0.006819	148	Qingdao	0.004858	148	Sacramento	0.003477
149	Dar es Salaam	0.006810	149	Jinan	0.004858	149	Lima	0.003460
150	Budapest	0.006807	150	Belgrade	0.004855	150	Edinburgh	0.003453
151	Helsinki	0.006790	151	Rome	0.004799	151	Douala	0.003449
152	Nanjing	0.006766	152	Beirut	0.004772	152	Budapest	0.003445
153	Aleppo	0.006765	153	Alexandria	0.004728	153	Tampa	0.003421
154	Accra	0.006738	154	Tbilisi	0.004727	154	Kobe	0.003421

APPENDIX D: (continued)

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
155	Bogota	0.006734	155	Vienna	0.004711	155	New Orleans	0.003418
156	Paris	0.006692	156	Saint Petersburg	0.004705	156	Bucharest	0.003407
157	Caracas	0.006656	157	Copenhagen	0.004688	157	Cape Town	0.003397
158	Lisbon	0.006618	158	Mexico City	0.004684	158	Manchester	0.003380
159	Sofia	0.006589	159	Stockholm	0.004654	159	Jeddah	0.003364
160	Wuhan	0.006582	160	Liverpool	0.004605	160	Wuhan	0.003362
161	Shenyang	0.006533	161	Zurich	0.004589	161	Turin	0.003359
162	Seoul	0.006531	162	San Jose CR	0.004575	162	Kuala Lumpur	0.003357
163	Milan	0.006525	163	Istanbul	0.004568	163	Vancouver	0.003354
164	Harare	0.006525	164	Suzhou	0.004564	164	Salvador	0.003343
165	Baghdad	0.006514	165	Kuala Lumpur	0.004548	165	Buffalo	0.003339
166	Bonn	0.006487	166	Sofia	0.004546	166	Dubai	0.003331
167	Ankara	0.006467	167	Havana	0.004488	167	Calgary	0.003318
168	Lagos	0.006439	168	Changchun	0.004475	168	Lyons	0.003301
169	Tel Aviv	0.006436	169	Port-au-Prince	0.004428	169	Stuttgart	0.003292
170	Cairo	0.006393	170	Addis Ababa	0.004406	170	Vienna	0.003286
171	Stockholm	0.006389	171	Hamburg	0.004385	171	Bangalore	0.003284
172	Amsterdam	0.006375	172	Tel Aviv	0.004383	172	Tashkent	0.003284
173	Kiev	0.006361	173	Milan	0.004375	173	Qingdao	0.003279
174	Tashkent	0.006333	174	Conakry	0.004364	174	Shenyang	0.003271
175	Mexico City	0.006327	175	Lucknow	0.004356	175	Montreal	0.003269
176	Kuwait City	0.006312	176	Newcastle	0.004305	176	Ahmedabad	0.003260
177	Madrid	0.006300	177	Minsk	0.004302	177	Rome	0.003258
178	Changsha	0.006277	178	Yerevan	0.004298	178	Edmonton	0.003256
179	Luanda	0.006276	179	Abu Dhabi	0.004289	179	Taiyuan	0.003240
180	Khartoum	0.006275	180	Ciudad Juarez	0.004273	180	Durban	0.003234
181	Tunis	0.006222	181	Caracas	0.004272	181	Prague	0.003231
182	Guadalajara	0.006187	182	Guadalajara	0.004264	182	Changchun	0.003212
183	Jaipur	0.006175	183	Nagoya	0.004242	183	Chengdu	0.003212
184	Stuttgart	0.006140	184	Colombo	0.004241	184	Rio de Janeiro	0.003203
185	Tehran	0.006100	185	Monterrey	0.004234	185	Sydney	0.003202
186	Durban	0.006091	186	Cologne	0.004228	186	Munich	0.003199
187	Buffalo	0.006085	187	Shenyang	0.004227	187	Birmingham	0.003189
188	Kunming	0.006041	188	Mombasa	0.004210	188	Auckland	0.003174
189	Yangon	0.006036	189	Ho Chi Minh City	0.004191	189	Kunming	0.003157
190	Bucharest	0.006025	190	Algiers	0.004169	190	Abu Dhabi	0.003129
191	Hanoi	0.006009	191	Honolulu	0.004149	191	Kiev	0.003128
192	Copenhagen	0.005991	192	Kunming	0.004130	192	Ho Chi Minh City	0.003113
193	Berlin	0.005983	193	Guangzhou	0.004124	193	Shantou	0.003111

APPENDIX D: (continued)

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
194	New Orleans	0.005977	194	Durban	0.004100	194	Saint Petersburg	0.003107
195	Istanbul	0.005965	195	Harbin	0.004067	195	Zurich	0.003098
196	Baku	0.005956	196	Frankfurt	0.004066	196	Nagoya	0.003092
197	Liverpool	0.005956	197	Kobe	0.004059	197	Surat	0.003085
198	Lille	0.005828	198	Luxembourg	0.004057	198	Madrid	0.003041
199	Algiers	0.005810	199	Osaka	0.004012	199	Pune	0.003025
200	Tbilisi	0.005809	200	Accra	0.004001	200	Bandung	0.003022
201	Yerevan	0.005792	201	Pune	0.003994	201	Hamburg	0.003022
202	Managua	0.005772	202	Kiev	0.003986	202	Shenzhen	0.003016
203	Barcelona	0.005768	203	Chengdu	0.003967	203	Osaka	0.003012
204	Panama City	0.005763	204	Jaipur	0.003960	204	Ulan Bator	0.003007
205	Kabul	0.005722	205	Naples	0.003959	205	Changsha	0.002989
206	Belgrade	0.005704	206	Nanjing	0.003946	206	Amsterdam	0.002967
207	Tripoli	0.005701	207	Douala	0.003915	207	Guadalajara	0.002955
208	Tokyo	0.005688	208	Wuhan	0.003904	208	Las Vegas	0.002955
209	Salvador	0.005684	209	Bucharest	0.003897	209	Perth	0.002955
210	Yaoundé	0.005677	210	East Rand	0.003881	210	Xi'an	0.002947
211	Curitiba	0.005652	211	Guayaquil	0.003864	211	Frankfurt	0.002946
212	Kobe	0.005626	212	Bogota	0.003848	212	Copenhagen	0.002908
213	Hamburg	0.005607	213	Tianjin	0.003824	213	Medan	0.002907
214	Nagoya	0.005602	214	Dusseldorf	0.003820	214	Guangzhou	0.002883
215	Port-au-Prince	0.005580	215	Tunis	0.003816	215	Marseille	0.002873
216	Maputo	0.005532	216	Panama City	0.003808	216	Rotterdam	0.002870
217	Batam	0.005509	217	Xi'an	0.003799	217	Liverpool	0.002863
218	Marseille	0.005495	218	Zhengzhou	0.003750	218	Barcelona	0.002862
219	Rotterdam	0.005483	219	Casablanca	0.003709	219	Buenos Aires	0.002858
220	Xi'an	0.005468	220	Yangon	0.003694	220	Newcastle	0.002840
221	Colombo	0.005465	221	Bordeaux	0.003665	221	Naples	0.002836
222	Brasilia	0.005440	222	Hangzhou	0.003607	222	Casablanca	0.002819
223	Quito	0.005381	223	Valencia	0.003597	223	Hangzhou	0.002812
224	Lima	0.005376	224	Antwerp	0.003574	224	Montevideo	0.002804
225	Damascus	0.005364	225	Manaus	0.003547	225	Recife	0.002794
226	Turin	0.005361	226	Yaoundé	0.003541	226	Guayaquil	0.002780
227	Belo Horizonte	0.005328	227	Oslo	0.003535	227	San Salvador	0.002766
228	Alexandria	0.005310	228	Ulan Bator	0.003519	228	Antwerp	0.002762
229	Vienna	0.005251	229	Taiyuan	0.003496	229	San Jose CR	0.002730
230	Oslo	0.005244	230	Xiamen	0.003496	230	Brisbane	0.002720
231	Mombasa	0.005216	231	Changsha	0.003493	231	Quito	0.002714
232	Freetown	0.005130	232	Curitiba	0.003489	232	Lille	0.002701

APPENDIX D: (continued)

Ranking	City	Economic / Total Words	Ranking	City	Cultural / Total Words	Ranking	City	Political / Total Words
233	Naples	0.005122	233	Tokyo	0.003486	233	Foshan	0.002690
234	Palermo	0.005074	234	Recife	0.003431	234	Kuwait City	0.002687
235	Ciudad Juarez	0.005063	235	Kyoto	0.003429	235	Belo Horizonte	0.002673
236	Cologne	0.005014	236	Lille	0.003422	236	Zhengzhou	0.002664
237	Manaus	0.005011	237	Shenzhen	0.003419	237	Dongguan	0.002648
238	Fortaleza	0.005002	238	Helsinki	0.003415	238	Hanoi	0.002610
239	Havana	0.004994	239	Tegucigalpa	0.003410	239	Manaus	0.002602
240	Guatemala City	0.004977	240	Hanoi	0.003399	240	Cologne	0.002601
241	Tegucigalpa	0.004958	241	Tashkent	0.003362	241	Guatemala City	0.002594
242	Bordeaux	0.004943	242	Luanda	0.003318	242	Dusseldorf	0.002585
243	Rome	0.004940	243	Santo Domingo	0.003179	243	Honolulu	0.002583
244	Santo Domingo	0.004938	244	Belo Horizonte	0.003179	244	East Rand	0.002575
245	Suzhou	0.004920	245	Krakow	0.003144	245	Suzhou	0.002559
246	Krakow	0.004917	246	Dongguan	0.003100	246	Milan	0.002555
247	Medellin	0.004654	247	Guatemala City	0.003098	247	Jaipur	0.002549
248	Valencia	0.004624	248	Shantou	0.003018	248	Krakow	0.002536
249	Montevideo	0.004575	249	Marseille	0.002995	249	Santo Domingo	0.002529
250	Medan	0.004546	250	Fortaleza	0.002898	250	Medellin	0.002480
251	Ulan Bator	0.004523	251	Lagos	0.002876	251	Valencia	0.002395
252	Rabat	0.004489	252	Maputo	0.002828	252	Panama City	0.002395
253	Guayaquil	0.004478	253	San Salvador	0.002799	253	Helsinki	0.002340
254	San Salvador	0.004467	254	Montevideo	0.002776	254	Tijuana	0.002335
255	Porto Alegre	0.004443	255	Palermo	0.002712	255	Lagos	0.002300
256	Jerusalem	0.004367	256	Medellin	0.002663	256	Bordeaux	0.002287
257	Tijuana	0.004337	257	Kuwait City	0.002613	257	Tokyo	0.002202
258	Recife	0.004332	258	Medan	0.002573	258	Fortaleza	0.002182
259	Antwerp	0.004292	259	Tijuana	0.002487	259	Palermo	0.002168
260	Lyons	0.003994	260	Bandung	0.002288	260	Batam	0.002051
261	Bandung	0.003953	261	Foshan	0.002254	261	Harbin	0.001963
262	La Paz	0.003601	262	Batam	0.002244	262	Lahore	0.001717
263	Kyoto	0.003017	263	Lahore	0.001829	263	Kyoto	0.001621
264	Lahore	0.001725	264	La Paz	0.001715	264	La Paz	0.001100

APPENDIX E: WORD FREQUENCY AND PROPORTION FOR QCA

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Abidjan	9037832	74936	47715	63285	185936	0.403	0.257	0.340	2.057
Abu Dhabi	2716491	27400	11651	8500	47551	0.576	0.245	0.179	1.750
Accra	16921772	114024	67708	73938	255670	0.446	0.265	0.289	1.511
Addis Ababa	3148406	22391	13872	24594	60857	0.368	0.228	0.404	1.933
Adelaide	5336315	47344	40936	22252	110532	0.428	0.370	0.201	2.071
Ahmedabad	3856268	45217	19840	12571	77628	0.582	0.256	0.162	2.013
Aleppo	14006859	94759	78492	129462	302713	0.313	0.259	0.428	2.161
Alexandria	17806323	94554	84180	63167	241901	0.391	0.348	0.261	1.359
Algiers	12147571	70574	50647	49552	170773	0.413	0.297	0.290	1.406
Almaty	17241181	151764	93483	84801	330048	0.460	0.283	0.257	1.914
Amman	23411673	160188	134831	90477	385496	0.416	0.350	0.235	1.647
Amsterdam	30282418	193062	160884	89844	443790	0.435	0.363	0.202	1.466
Ankara	18886099	122130	98248	100882	321260	0.380	0.306	0.314	1.701
Antwerp	17195632	73801	61456	47499	182756	0.404	0.336	0.260	1.063
Asuncion	2940669	24752	19264	15269	59285	0.418	0.325	0.258	2.016
Athens	11375863	100215	74500	45279	219994	0.456	0.339	0.206	1.934
Atlanta	30384884	285236	287520	123538	696294	0.410	0.413	0.177	2.292
Auckland	3495533	29490	21945	11095	62530	0.472	0.351	0.177	1.789
Baghdad	18995874	123739	120931	208920	453590	0.273	0.267	0.461	2.388
Baku	10416642	62042	63471	54369	179882	0.345	0.353	0.302	1.727
Baltimore	15182477	113969	109716	65185	288870	0.395	0.380	0.226	1.903
Bandung	6027544	23825	13794	18215	55834	0.427	0.247	0.326	0.926
Bangalore	5715029	58759	34279	18769	111807	0.526	0.307	0.168	1.956
Bangkok	2521122	25497	13458	8905	47860	0.533	0.281	0.186	1.898
Barcelona	6290613	36284	40307	18005	94596	0.384	0.426	0.190	1.504
Basel	2017224	20955	10249	9701	40905	0.512	0.251	0.237	2.028
Batam	5595675	30825	12557	11477	54859	0.562	0.229	0.209	0.980
Beijing	19306975	196887	129302	78003	404192	0.487	0.320	0.193	2.094
Beirut	14016796	106752	66886	75933	249571	0.428	0.268	0.304	1.781
Belgrade	19420095	110763	94292	92128	297183	0.373	0.317	0.310	1.530
Belo Horizonte	12553571	66890	39902	33557	140349	0.477	0.284	0.239	1.118
Berlin	15499995	92736	97588	59689	250013	0.371	0.390	0.239	1.613
Birmingham	6013217	41215	34370	19178	94763	0.435	0.363	0.202	1.576
Bogota	2638250	17767	10151	12022	39940	0.445	0.254	0.301	1.514
Bonn	9355749	60690	47238	46067	153995	0.394	0.307	0.299	1.646
Bordeaux	851307	4208	3120	1947	9275	0.454	0.336	0.210	1.090
Boston	13386808	128845	119074	63490	311409	0.414	0.382	0.204	2.326
Brasilia	6949738	37809	47076	37602	122487	0.309	0.384	0.307	1.762

APPENDIX E: (continued)

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Brazzaville	13540425	181020	74504	124294	379818	0.477	0.196	0.327	2.805
Brisbane	20884164	244803	102922	56804	404529	0.605	0.254	0.140	1.937
Brussels	9261065	88332	75400	58990	222722	0.397	0.339	0.265	2.405
Bucharest	21884542	131851	85284	74558	291693	0.452	0.292	0.256	1.333
Budapest	27746105	188863	186479	95575	470917	0.401	0.396	0.203	1.697
Buenos Aires	25570105	201544	149267	73074	423885	0.475	0.352	0.172	1.658
Buffalo	4262737	25939	21885	14233	62057	0.418	0.353	0.229	1.456
Busan	5907677	44785	37595	37824	120204	0.373	0.313	0.315	2.035
Cairo	19639818	125557	115784	120641	361982	0.347	0.320	0.333	1.843
Calgary	28192149	202734	159969	93540	456243	0.444	0.351	0.205	1.618
Cape Town	25342351	202756	154474	86077	443307	0.457	0.348	0.194	1.749
Caracas	14288306	95103	61036	72188	228327	0.417	0.267	0.316	1.598
Casablanca	17998317	161735	66758	50735	279228	0.579	0.239	0.182	1.551
Changchun	6782254	51323	30348	21786	103457	0.496	0.293	0.211	1.525
Changsha	1758915	11040	6144	5257	22441	0.492	0.274	0.234	1.276
Charlotte	24383140	249443	220299	110412	580154	0.430	0.380	0.190	2.379
Chengdu	19760388	143820	78398	63473	285691	0.503	0.274	0.222	1.446
Chennai	563908	7520	4362	2440	14322	0.525	0.305	0.170	2.540
Chicago	25331489	226406	234558	112138	573102	0.395	0.409	0.196	2.262
Chittagong	12088351	169661	63920	43865	277446	0.612	0.230	0.158	2.295
Chongqing	14001948	101134	93747	54999	249880	0.405	0.375	0.220	1.785
Cincinnati	8984130	77346	79687	38217	195250	0.396	0.408	0.196	2.173
Ciudad Juarez	5366090	27168	22927	35384	85479	0.318	0.268	0.414	1.593
Cleveland	7298117	64945	61720	31972	158637	0.409	0.389	0.202	2.174
Cologne	5156937	25856	21801	13412	61069	0.423	0.357	0.220	1.184
Colombo	21609806	118104	91648	101212	310964	0.380	0.295	0.325	1.439
Columbus	25733682	205264	187477	117100	509841	0.403	0.368	0.230	1.981
Conakry	465829	3817	2033	3025	8875	0.430	0.229	0.341	1.905
Copenhagen	22751452	136314	106656	66152	309122	0.441	0.345	0.214	1.359
Curitiba	12002643	67833	41872	45975	155680	0.436	0.269	0.295	1.297
Dakar	5996689	49421	34489	38227	122137	0.405	0.282	0.313	2.037
Dalian	3015993	71375	16788	15464	103627	0.689	0.162	0.149	3.436
Dallas	17273240	154871	156452	72026	383349	0.404	0.408	0.188	2.219
Damascus	15752678	84504	86316	164943	335763	0.252	0.257	0.491	2.131
Dar es Salaam	12392041	84389	77880	58979	221248	0.381	0.352	0.267	1.785
Denver	8661278	76476	60458	32984	169918	0.450	0.356	0.194	1.962
Detroit	38820742	335583	340590	182082	858255	0.391	0.397	0.212	2.211
Dhaka	2802399	22601	16239	16075	54915	0.412	0.296	0.293	1.960
Doha	21275737	235945	126014	91053	453012	0.521	0.278	0.201	2.129

APPENDIX E: (continued)

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Dongguan	11432236	131477	35443	30267	197187	0.667	0.180	0.153	1.725
Douala	10567146	76530	41372	36441	154343	0.496	0.268	0.236	1.461
Dubai	10568231	128047	62474	35202	225723	0.567	0.277	0.156	2.136
Dublin	29037009	260223	192487	107921	560631	0.464	0.343	0.192	1.931
Durban	22159243	134972	90842	71652	297466	0.454	0.305	0.241	1.342
Dusseldorf	21500862	171642	82126	55590	309358	0.555	0.265	0.180	1.439
East Rand	15776870	250479	61232	40621	352332	0.711	0.174	0.115	2.233
Edinburgh	5235397	36646	31004	18080	85730	0.427	0.362	0.211	1.638
Edmonton	30300842	206735	167782	98652	473169	0.437	0.355	0.208	1.562
Fortaleza	15451092	77290	44771	33717	155778	0.496	0.287	0.216	1.008
Foshan	19825245	259993	44683	53332	358008	0.726	0.125	0.149	1.806
Frankfurt	13896682	143113	56500	40939	240552	0.595	0.235	0.170	1.731
Freetown	8964883	45988	50320	57559	153867	0.299	0.327	0.374	1.716
Geneva	1310990	10603	7732	7389	25724	0.412	0.301	0.287	1.962
Glasgow	21468890	146390	126510	76005	348905	0.420	0.363	0.218	1.625
Guadalajara	3372067	20863	14378	9966	45207	0.461	0.318	0.220	1.341
Guangzhou	22231155	217486	91687	64099	373272	0.583	0.246	0.172	1.679
Guatemala City	18661998	92882	57807	48404	199093	0.467	0.290	0.243	1.067
Guayaquil	2123564	9509	8206	5903	23618	0.403	0.347	0.250	1.112
Harbin	3062921	27012	12458	6012	45482	0.594	0.274	0.132	1.485
Hamburg	5935561	33282	26027	17935	77244	0.431	0.337	0.232	1.301
Hangzhou	10811942	96123	38997	30405	165525	0.581	0.236	0.184	1.531
Hanoi	4778722	28713	16243	12474	57430	0.500	0.283	0.217	1.202
Harare	13746287	89695	67278	76955	233928	0.383	0.288	0.329	1.702
Hartford	25783629	225053	223654	113850	562557	0.400	0.398	0.202	2.182
Havana	18861722	94189	84644	73797	252630	0.373	0.335	0.292	1.339
Helsinki	966752	6564	3301	2262	12127	0.541	0.272	0.187	1.254
Ho Chi Minh City	25366781	257478	106320	78970	442768	0.582	0.240	0.178	1.745
Hong Kong	177955	1949	1206	1045	4200	0.464	0.287	0.249	2.360
Honolulu	12261932	91722	50876	31677	174275	0.526	0.292	0.182	1.421
Houston	6660831	68602	70454	25092	164148	0.418	0.429	0.153	2.464
Hyderabad	30053531	256449	173117	114599	544165	0.471	0.318	0.211	1.811
Indianapolis	7852073	61867	71485	31396	164748	0.376	0.434	0.191	2.098
Islamabad	3919519	35065	22368	44318	101751	0.345	0.220	0.436	2.596
Istanbul	22053090	131553	100730	89897	322180	0.408	0.313	0.279	1.461
Jaipur	13808061	85268	54680	35192	175140	0.487	0.312	0.201	1.268
Jakarta	13432991	108976	72628	52827	234431	0.465	0.310	0.225	1.745
Jeddah	960179	8020	4907	3230	16157	0.496	0.304	0.200	1.683
Jerusalem	30183079	131821	158654	167629	458104	0.288	0.346	0.366	1.518

APPENDIX E: (continued)

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Jinan	12476445	109568	60612	47262	217442	0.504	0.279	0.217	1.743
Johannesburg	1026195	8876	5443	3729	18048	0.492	0.302	0.207	1.759
Kabul	9143233	52322	58032	97521	207875	0.252	0.279	0.469	2.274
Kampala	2692323	24166	16063	15349	55578	0.435	0.289	0.276	2.064
Kano	6839757	54086	38762	62966	155814	0.347	0.249	0.404	2.278
Kansas City	26464690	233281	162400	99458	495139	0.471	0.328	0.201	1.871
Karachi	28261978	315231	165173	188534	668938	0.471	0.247	0.282	2.367
Khartoum	5472964	34344	27363	55998	117705	0.292	0.232	0.476	2.151
Kiev	7240963	46059	28863	22652	97574	0.472	0.296	0.232	1.348
Kinshasa	466545	4155	3188	6018	13361	0.311	0.239	0.450	2.864
Kobe	16676991	93817	67684	57052	218553	0.429	0.310	0.261	1.311
Kolkata	418828	4496	2932	1774	9202	0.489	0.319	0.193	2.197
Krakow	5248786	25809	16502	13310	55621	0.464	0.297	0.239	1.060
Kuala Lumpur	23795150	286881	108220	79876	474977	0.604	0.228	0.168	1.996
Kunming	11482905	69364	47422	36253	153039	0.453	0.310	0.237	1.333
Kuwait City	18620306	117523	48650	50040	216213	0.544	0.225	0.231	1.161
Kyoto	165377	499	567	268	1334	0.374	0.425	0.201	0.807
La Paz	1195030	4303	2049	1314	7666	0.561	0.267	0.171	0.641
Lagos	817831	5266	2352	1881	9499	0.554	0.248	0.198	1.161
Lahore	1060262	1829	1939	1820	5588	0.327	0.347	0.326	0.527
Las Vegas	21274198	148291	110984	62870	322145	0.460	0.345	0.195	1.514
Lille	21664908	126256	74136	58510	258902	0.488	0.286	0.226	1.195
Lima	10186855	54760	52222	35245	142227	0.385	0.367	0.248	1.396
Lisbon	2709428	17931	19615	10893	48439	0.370	0.405	0.225	1.788
Liverpool	4101979	24430	18890	11744	55064	0.444	0.343	0.213	1.342
London	3434145	38387	25740	12544	76671	0.501	0.336	0.164	2.233
Los Angeles	2292603	20067	18137	8626	46830	0.429	0.387	0.184	2.043
Luanda	13567289	85146	45019	51765	181930	0.468	0.247	0.285	1.341
Lucknow	23282760	188459	101427	98948	388834	0.485	0.261	0.254	1.670
Lusaka	617392	5291	3021	3280	11592	0.456	0.261	0.283	1.878
Luxembourg	6605872	92743	26798	37139	156680	0.592	0.171	0.237	2.372
Lyons	5669049	22644	32348	18716	73708	0.307	0.439	0.254	1.300
Madrid	24852097	156576	151186	75572	383334	0.408	0.394	0.197	1.542
Managua	9229699	53271	47717	35033	136021	0.392	0.351	0.258	1.474
Manaus	13924934	69779	49396	36226	155401	0.449	0.318	0.233	1.116
Manchester	1898618	17339	12295	6417	36051	0.481	0.341	0.178	1.899
Manila	25288910	187181	150277	114018	451476	0.415	0.333	0.253	1.785
Maputo	7411060	40998	20958	32120	94076	0.436	0.223	0.341	1.269
Marseille	3298472	18124	9880	9477	37481	0.484	0.264	0.253	1.136

APPENDIX E: (continued)

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Medan	14095501	64083	36268	40973	141324	0.453	0.257	0.290	1.003
Medellin	12205583	56805	32508	30268	119581	0.475	0.272	0.253	0.980
Melbourne	4714621	45792	39434	16778	102004	0.449	0.387	0.164	2.164
Mexico City	12258349	77556	57423	46581	181560	0.427	0.316	0.257	1.481
Miami	2546088	19462	18731	9260	47453	0.410	0.395	0.195	1.864
Milan	2043805	13336	8941	5222	27499	0.485	0.325	0.190	1.345
Minneapolis	1821861	15559	18181	8051	41791	0.372	0.435	0.193	2.294
Minsk	10291898	70690	44278	46933	161901	0.437	0.273	0.290	1.573
Mombasa	549443	2866	2313	1924	7103	0.403	0.326	0.271	1.293
Monrovia	383775	3026	1903	2938	7867	0.385	0.242	0.373	2.050
Monterrey	17143984	118061	72595	79665	270321	0.437	0.269	0.295	1.577
Montevideo	8479363	38797	23541	23772	86110	0.451	0.273	0.276	1.016
Montreal	24995526	177520	142738	81700	401958	0.442	0.355	0.203	1.608
Moscow	13356074	103713	67502	58171	229386	0.452	0.294	0.254	1.717
Mumbai	3060726	33441	22457	11181	67079	0.499	0.335	0.167	2.192
Munich	21969153	168364	113797	70279	352440	0.478	0.323	0.199	1.604
Nagoya	4748084	26597	20143	14681	61421	0.433	0.328	0.239	1.294
Nairobi	18070625	188385	98706	108083	395174	0.477	0.250	0.274	2.187
Nanjing	3788089	25632	14948	13399	53979	0.475	0.277	0.248	1.425
Naples	7029885	36008	27833	19940	83781	0.430	0.332	0.238	1.192
New Delhi Delhi	6326262	54795	41035	29204	125034	0.438	0.328	0.234	1.976
New Orleans	19418466	116065	113800	66374	296239	0.392	0.384	0.224	1.526
New York	28164926	239779	203661	123219	566659	0.423	0.359	0.217	2.012
Newcastle	30945827	216326	133221	87885	437432	0.495	0.305	0.201	1.414
Omaha	9633589	78797	72710	38275	189782	0.415	0.383	0.202	1.970
Osaka	11356722	77458	45564	34206	157228	0.493	0.290	0.218	1.384
Oslo	1680460	8812	5941	6308	21061	0.418	0.282	0.300	1.253
Ottawa	8493582	80390	74825	40354	195569	0.411	0.383	0.206	2.303
Palermo	18744579	95108	50836	40638	186582	0.510	0.272	0.218	0.995
Panama City	19113030	110154	72791	45777	228722	0.482	0.318	0.200	1.197
Paris	14437515	96611	86549	51014	234174	0.413	0.370	0.218	1.622
Perth	5782104	60621	31328	17084	109033	0.556	0.287	0.157	1.886
Philadelphia	15595960	124943	137325	61032	323300	0.386	0.425	0.189	2.073
Phoenix	6805862	58786	46294	28098	133178	0.441	0.348	0.211	1.957
Pittsburgh	23731366	184720	157479	90462	432661	0.427	0.364	0.209	1.823
Port-au-Prince	10718783	59808	47467	57086	164361	0.364	0.289	0.347	1.533
Portland	15020925	124294	161502	70499	356295	0.349	0.453	0.198	2.372
Porto Alegre	11881635	52793	94353	44374	191520	0.276	0.493	0.232	1.612
Prague	20720602	153044	111050	66952	331046	0.462	0.335	0.202	1.598

APPENDIX E: (continued)

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Pretoria	1047671	8276	5969	4931	19176	0.432	0.311	0.257	1.830
Pune	6648810	65372	26556	20113	112041	0.583	0.237	0.180	1.685
Pyongyang	5091040	39098	32383	56685	128166	0.305	0.253	0.442	2.517
Qingdao	2154616	22307	10468	7066	39841	0.560	0.263	0.177	1.849
Quito	12779781	68772	62780	34688	166240	0.414	0.378	0.209	1.301
Rabat	16384697	73558	93802	61595	228955	0.321	0.410	0.269	1.397
Rawalpindi	4819368	45930	29449	36225	111604	0.412	0.264	0.325	2.316
Recife	10977502	47558	37664	30673	115895	0.410	0.325	0.265	1.056
Richmond	10268420	74909	71408	43634	189951	0.394	0.376	0.230	1.850
Rio de Janeiro	8200775	56741	40373	26265	123379	0.460	0.327	0.213	1.504
Riyadh	13488351	160470	87054	57648	305172	0.526	0.285	0.189	2.262
Rochester	1255963	11118	8528	4916	24562	0.453	0.347	0.200	1.956
Rome	11514864	56880	55260	37511	149651	0.380	0.369	0.251	1.300
Rotterdam	15539597	85207	81989	44598	211794	0.402	0.387	0.211	1.363
Sacramento	16185900	112165	104775	56278	273218	0.411	0.383	0.206	1.688
Saint Louis	27885820	281072	245955	135481	662508	0.424	0.371	0.204	2.376
Saint Petersburg	5134115	36789	24156	15953	76898	0.478	0.314	0.207	1.498
Salvador	329351	1872	1662	1101	4635	0.404	0.359	0.238	1.407
San Diego	12203965	114150	86717	47654	248521	0.459	0.349	0.192	2.036
San Francisco	36031143	309792	286982	153245	750019	0.413	0.383	0.204	2.082
San Jose CA	18619413	161974	128444	67226	357644	0.453	0.359	0.188	1.921
San Jose CR	10586237	85957	48429	28896	163282	0.526	0.297	0.177	1.542
San Salvador	5208990	23269	14579	14409	52257	0.445	0.279	0.276	1.003
Santiago	16935477	123632	104665	60293	288590	0.428	0.363	0.209	1.704
Santo Domingo	8404817	41506	26721	21258	89485	0.464	0.299	0.238	1.065
Sao Paulo	15963058	222961	92319	68116	383396	0.582	0.241	0.178	2.402
Seattle	569102	4778	4141	1997	10916	0.438	0.379	0.183	1.918
Seoul	21585313	140970	157192	90427	388589	0.363	0.405	0.233	1.800
Shanghai	11530468	230594	75147	46496	352237	0.655	0.213	0.132	3.055
Shantou	13978901	143242	42191	43495	228928	0.626	0.184	0.190	1.638
Shenyang	11739072	76687	49616	38403	164706	0.466	0.301	0.233	1.403
Shenzhen	15641824	192930	53485	47169	293584	0.657	0.182	0.161	1.877
Singapore	25906399	292493	175571	114542	582606	0.502	0.301	0.197	2.249
Sofia	14760818	97258	67108	52682	217048	0.448	0.309	0.243	1.470
Stockholm	14875673	95040	69231	58918	223189	0.426	0.310	0.264	1.500
Stuttgart	16120270	98976	78643	53066	230685	0.429	0.341	0.230	1.431
Surat	5549236	53889	27640	17119	98648	0.546	0.280	0.174	1.778
Suzhou	19957339	98193	91089	51068	240350	0.409	0.379	0.212	1.204
Sydney	25675395	253049	176612	82204	511865	0.494	0.345	0.161	1.994

APPENDIX E: (continued)

City	**T**otal **W**ords	**E**conomic	**C**ultural	**P**olitical	**T=E+C+P**	**E / T**	**C / T**	**P / T**	**T / TW** (%)
Taipei	1104119	10215	5881	4871	20967	0.487	0.280	0.232	1.899
Taiyuan	2183154	15845	7633	7074	30552	0.519	0.250	0.232	1.399
Tampa	9556802	75285	54888	32698	162871	0.462	0.337	0.201	1.704
Tashkent	8403718	53220	28252	27595	109067	0.488	0.259	0.253	1.298
Tbilisi	14642569	85060	69210	81096	235366	0.361	0.294	0.345	1.607
Tegucigalpa	10374593	51441	35381	40023	126845	0.406	0.279	0.316	1.223
Tehran	14493863	88414	92919	121818	303151	0.292	0.307	0.402	2.092
Tel Aviv	3674305	23647	16106	14329	54082	0.437	0.298	0.265	1.472
The Hague	16310140	128649	89713	94877	313239	0.411	0.286	0.303	1.921
Tianjin	19712519	162990	75387	70288	308665	0.528	0.244	0.228	1.566
Tijuana	19793356	85847	49235	46210	181292	0.474	0.272	0.255	0.916
Tokyo	5450	31	19	12	62	0.500	0.306	0.194	1.138
Toronto	19791296	197430	153962	83361	434753	0.454	0.354	0.192	2.197
Tripoli	14398005	82081	77050	128841	287972	0.285	0.268	0.447	2.000
Tunis	9134005	56829	34856	41918	133603	0.425	0.261	0.314	1.463
Turin	27119048	145376	138698	91100	375174	0.387	0.370	0.243	1.383
Ulan Bator	4259607	19265	14988	12810	47063	0.409	0.318	0.272	1.105
Valencia	18599590	86004	66907	44548	197459	0.436	0.339	0.226	1.062
Vancouver	2478525	18912	18093	8314	45319	0.417	0.399	0.183	1.828
Vienna	16472703	86505	77599	54130	218234	0.396	0.356	0.248	1.325
Warsaw	2373437	17711	15688	10976	44375	0.399	0.354	0.247	1.870
Washington	7046126	60847	58611	45640	165098	0.369	0.355	0.276	2.343
Wuhan	7284049	47942	28439	24486	100867	0.475	0.282	0.243	1.385
Wuxi	280571	2610	1522	990	5122	0.510	0.297	0.193	1.826
Xiamen	18312188	215969	64011	65359	345339	0.625	0.185	0.189	1.886
Xi'an	17037827	93157	64734	50207	208098	0.448	0.311	0.241	1.221
Yangon	18521379	111796	68418	71134	251348	0.445	0.272	0.283	1.357
Yaound é	8079448	45866	28607	31040	105513	0.435	0.271	0.294	1.306
Yerevan	15170940	87865	65205	76458	229528	0.383	0.284	0.333	1.513
Zhengzhou	14788533	232430	55454	39398	327282	0.710	0.169	0.120	2.213
Zurich	2842408	27859	13043	8806	49708	0.560	0.262	0.177	1.749

APPENDIX F: CORRELATION COEFFICIENT FOR EACH CITY

Rank	City	ρ	Rank	City	ρ	Rank	City	ρ	Rank	City	ρ
1	Chicago	0.4284	67	Vancouver	0.2633	133	Perth	0.1801	199	Mombasa	0.0866
2	Atlanta	0.4149	68	Oslo	0.2630	134	Stockholm	0.1793	200	Guatemala City	0.0856
3	Boston	0.3866	69	Bonn	0.2623	135	Freetown	0.1782	201	Honolulu	0.0824
4	Houston	0.3854	70	Dakar	0.2615	136	Hyderabad	0.1741	202	Shenyang	0.0819
5	Kano	0.3849	71	Karachi	0.2593	137	Frankfurt	0.1724	203	Barcelona	0.0816
6	Brazzaville	0.3805	72	Las Vegas	0.2593	138	Surat	0.1720	204	Dar es Salaam	0.0788
7	Conakry	0.3805	73	Seattle	0.2575	139	Harbin	0.1710	205	Auckland	0.0787
8	Monrovia	0.3704	74	Busan	0.2569	140	Berlin	0.1702	206	Jeddah	0.0764
9	Ciudad Juarez	0.3687	75	Luanda	0.2563	141	Johannesburg	0.1654	207	Helsinki	0.0762
10	Tripoli	0.3668	76	Caracas	0.2561	142	Zurich	0.1653	208	Alexandria	0.0719
11	Charlotte	0.3501	77	Dalian	0.2548	143	Manila	0.1619	209	Birmingham	0.0706
12	Portland	0.3478	78	Phoenix	0.2526	144	San Francisco	0.1617	210	Kiev	0.0615
13	Tehran	0.3425	79	Lusaka	0.2524	145	Dubai	0.1583	211	Manaus	0.0587
14	Pune	0.3398	80	Yerevan	0.2520	146	Tianjin	0.1577	212	Rio de Janeiro	0.0587
15	Port-au-Prince	0.3317	81	Edmonton	0.2507	147	Nanjing	0.1575	213	Managua	0.0569
16	Khartoum	0.3294	82	Yaoundé	0.2503	148	Fortaleza	0.1570	214	Durban	0.0541
17	Bogotá	0.3267	83	Ankara	0.2448	149	Krakow	0.1563	215	Asunción	0.0519
18	Denver	0.3252	84	Bandung	0.2436	150	Hong Kong	0.1560	216	Qingdao	0.0423
19	Detroit	0.3219	85	Rome	0.2434	151	Ulan Bator	0.1550	217	Tokyo	0.0419
20	Rawalpindi	0.3197	86	Athens	0.2385	152	Douala	0.1501	218	Taipei	0.0416
21	Baghdad	0.3194	87	Batam	0.2367	153	Tel Aviv	0.1493	219	La Paz	0.0370
22	The Hague	0.3180	88	Nairobi	0.2367	154	Recife	0.1472	220	Osaka	0.0365
23	Tunis	0.3171	89	Lucknow	0.2358	155	Sofia	0.1444	221	Newcastle	0.0359
24	Islamabad	0.3166	90	San Diego	0.2356	156	Istanbul	0.1442	222	Amsterdam	0.0357
25	Edinburgh	0.3153	91	Algiers	0.2341	157	Mexico City	0.1390	223	Prague	0.0325
26	Mumbai	0.3137	92	Omaha	0.2340	158	Lahore	0.1388	224	Lima	0.0299
27	Calgary	0.3125	93	Taiyuan	0.2326	159	Xi'an	0.1363	225	Hartford	0.0286
28	Shantou	0.3116	94	San Salvador	0.2311	160	Belo Horizonte	0.1350	226	Brasília	0.0280
29	Philadelphia	0.3107	95	Minsk	0.2304	161	Bucharest	0.1347	227	Cape Town	0.0277
30	Toronto	0.3086	96	Montevideo	0.2275	162	Jaipur	0.1316	228	Valencia	0.0254
31	Kabul	0.3072	97	Kolkata (Calcutta)	0.2268	163	Düsseldorf	0.1308	229	São Paulo	0.0214
32	Ottawa	0.3071	98	Aleppo	0.2264	164	Changsha	0.1300	230	Guadalajara	0.0210
33	Damascus	0.3066	99	Rochester	0.2258	165	Lagos	0.1288	231	Doha	0.0205
34	Chennai	0.3058	100	Monterrey	0.2258	166	Vienna	0.1288	232	Copenhagen	0.0178
35	Xiamen	0.3055	101	Kampala	0.2243	167	Zhengzhou	0.1263	233	Naples	0.0166
36	Harare	0.3038	102	Singapore	0.2237	168	Lille	0.1240	234	Hamburg	0.0084
37	Ahmedabad	0.2994	103	Baku	0.2230	169	Colombo	0.1233	235	Stuttgart	0.0080
38	Indianapolis	0.2993	104	East Rand	0.2180	170	Paris	0.1230	236	Addis Ababa	0.0075
39	Maputo	0.2991	105	Marseille	0.2168	171	Buffalo	0.1228	237	Madrid	0.0057
40	Minneapolis	0.2975	106	Beirut	0.2166	172	Wuhan	0.1202	238	Lisbon	-0.0010
41	Glasgow	0.2947	107	Bangkok	0.2149	173	Lyons	0.1200	239	Milan	-0.0041
42	New York	0.2935	108	Hangzhou	0.2147	174	Tbilisi	0.1179	240	Antwerp	-0.0170
43	Columbus	0.2923	109	Kuala Lumpur	0.2120	175	Guangzhou	0.1177	241	Kansas City	-0.0174
44	Miami	0.2899	110	Cairo	0.2118	176	Chengdu	0.1150	242	San José	-0.0184
45	Cincinnati	0.2891	111	Accra	0.2101	177	Shenzhen	0.1109	243	Budapest	-0.0217
46	Foshan	0.2871	112	Ho Chi Minh City	0.2080	178	Kobe	0.1095	244	Almaty	-0.0351
47	Tegucigalpa	0.2865	113	Tampa	0.2047	179	Medellin	0.1078	245	Turin	-0.0377
48	Sydney	0.2859	114	Moscow	0.2043	180	Chittagong	0.1055	246	Rotterdam	-0.0386
49	Medan	0.2851	115	Shanghai	0.2042	181	Manchester	0.1046	247	Bordeaux	-0.0459
50	Dublin	0.2835	116	Tijuana	0.2039	182	Montreal	0.1037	248	Buenos Aires	-0.0560
51	Saint Louis	0.2833	117	Geneva	0.2033	183	Jerusalem	0.1017	249	Riyadh	-0.0627
52	Kinshasa	0.2813	118	Cleveland	0.2008	184	Abu Dhabi	0.1007	250	Adelaide	-0.0718
53	Pittsburgh	0.2811	119	Bangalore	0.2007	185	Pretoria	0.0994	251	Rabat	-0.0760
54	Curitiba	0.2805	120	Jakarta	0.1984	186	Saint Petersburg	0.0991	252	Chongqing	-0.0891
55	Dhaka	0.2798	121	Jinan	0.1980	187	Seoul	0.0974	253	Sacramento	-0.0907
56	Kuwait City	0.2794	122	Palermo	0.1967	188	Nagoya	0.0970	254	Guayaquil	-0.0924
57	Abidjan	0.2790	123	Belgrade	0.1966	189	Beijing	0.0955	255	Panama City	-0.1167
58	Liverpool	0.2766	124	Basel	0.1960	190	New Delhi (Delhi)	0.0954	256	Amman	-0.1222
59	Baltimore	0.2742	125	Brisbane	0.1950	191	Casablanca	0.0951	257	Cologne	-0.1237
60	New Orleans	0.2732	126	Wuxi	0.1950	192	Kunming	0.0930	258	Suzhou	-0.1256
61	Washington	0.2727	127	Richmond	0.1924	193	Salvador	0.0908	259	Los Angeles	-0.1334
62	Pyongyang	0.2726	128	Melbourne	0.1907	194	Warsaw	0.0907	260	San Jose	-0.1545
63	London	0.2724	129	Hanoi	0.1900	195	Santo Domingo	0.0901	261	Porto Alegre	-0.1941
64	Dongguan	0.2697	130	Dallas	0.1876	196	Changchun	0.0892	262	Santiago	-0.1945
65	Tashkent	0.2690	131	Havana	0.1860	197	Luxembourg	0.0882	263	Quito	-0.2190
66	Yangon	0.2653	132	Brussels	0.1839	198	Munich	0.0872	264	Kyoto	-0.2376