Lecture Notes in Social Networks

Arash Shaban-Nejad
John S. Brownstein
David L. Buckeridge *Editors*

Public Health Intelligence and the Internet

Springer

Lecture Notes in Social Networks

More information about this series at http://www.springer.com/series/8768

Arash Shaban-Nejad · John S. Brownstein
David L. Buckeridge
Editors

Public Health Intelligence and the Internet

Springer

Editors
Arash Shaban-Nejad
Department of Pediatrics,
Center for Biomedical Informatics
The University of Tennessee Health Science Center—Oak-Ridge National Laboratory (UTHSC-ORNL)
Memphis, TN
USA

John S. Brownstein
Department of Pediatrics, Harvard Medical School, Boston Children's Hospital
Harvard University
Boston, MA
USA

David L. Buckeridge
Department of Epidemiology and Biostatistics and Occupational Health, McGill Clinical & Health Informatics
McGill University
Montreal, QC
Canada

ISSN 2190-5428 ISSN 2190-5436 (electronic)
Lecture Notes in Social Networks
ISBN 978-3-319-68602-8 ISBN 978-3-319-68604-2 (eBook)
https://doi.org/10.1007/978-3-319-68604-2

Library of Congress Control Number: 2017953826

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Severe pandemics due to highly-transmissible viruses continue to threaten the world in the twenty-first century. In a tightly interconnected world, infectious disease outbreaks can adversely affect economic growth, trade, tourism, business and industry, and social stability as well as public health. At the same time, noncommunicable diseases have become the main cause of global disability and death, imposing a crushing burden on societies and economies around the world. Public health authorities and researchers now collect data from many sources and analyze these data together to estimate the incidence and prevalence of different health conditions, as well as related risk factors. Modern surveillance systems employ tools and techniques to monitor direct and indirect signals and indicators of disease activities for early detection of outbreaks. Tracking of Internet-based health indicators complements other surveillance methods collecting data from clinical systems and registries. To provide proper alerts and timely response public health officials and researchers systematically gather news, and other reports about suspected disease outbreaks, bioterrorism, and other events of potential international public health concern, from a wide range of formal and informal sources. With the advent of modern communication technologies, many outbreak reports now originate in electronic media and electronic discussion groups. Given the ever-increasing role of the World Wide Web as a source of information in many domains including health care, accessing, managing, and analyzing its content have brought new opportunities and challenges. This is especially the case for nontraditional online resources such as social networks, blogs, news feed, twitter posts, and online communities with the sheer size and ever-increasing growth and change rate of their data. Web applications along with text processing programs are increasingly being used to harness online data and information to discover meaningful patterns identifying emerging health threats. The advances in web science and technology for data management, integration, mining, classification, filtering, and visualization have given rise to a variety of applications representing real time data on epidemics. Also, several public health surveillance tools have been recruited to use web data to detect health crises earlier than official monitoring systems. Some of the main technical and nontechnical challenges in these systems

include: reliability and representativeness of the online data; redundancy and inconsistency of data; generating predictive models; timely and early detection; issues related to verification and evaluation (of the sources (number, and qualities)); and ethics, security and privacy concerns.

This book aims to highlight the latest achievements in epidemiological surveillance and Internet interventions based on monitoring online communications and interactions on the web. It presents the state of the art and the advances in the field of online disease surveillance and intervention. The edited volume contains extended and revised versions of selected papers presented at the International World Wide Web and Population Health Intelligence (W3PHI) workshop series along with some invited chapters and presents an overview of the issues, challenges, and potentials in the field, along with the new research results. The book provides information for a wide range of scientists, researchers, graduate students, industry professionals, national and international public health agencies, and NGOs interested in the theory and practice of computational models of web-based public health intelligence.

Memphis, USA Arash Shaban-Nejad
Boston, USA John S. Brownstein
Montreal, Canada David L. Buckeridge

Contents

Contributors

David L. Buckeridge Department of Epidemiology & Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

Alan Chappell Pacific Northwest National Laboratory, Richland, WA, USA

Katia Charland Clinical and Health Informatics Research Group, McGill University, Montreal, QC, Canada

Lauren Charles-Smith Pacific Northwest National Laboratory, Richland, WA, USA

Soon Ae Chun City University of New York, New York, NY, USA

Courtney D. Corley Pacific Northwest National Laboratory, Richland, WA, USA

Vivek Datla Pacific Northwest National Laboratory, Richland, WA, USA

Rui Fang Research and Development, Thomson Reuters, New York City, NY, USA

James Geller New Jersey Institute of Technology, Newark, NJ, USA

Josh Harrison Pacific Northwest National Laboratory, Richland, WA, USA

Masoumeh Izadi Clinical and Health Informatics Research Group, McGill University, Montreal, QC, Canada

Xiang Ji The Bloomberg L.P., New York, NY, USA

Quanzhi Li Research and Development, Thomson Reuters, New York City, NY, USA

Xiaomo Liu Research and Development, Thomson Reuters, New York City, NY, USA

Hiroshi Mamiya Department of Epidemiology and Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

Anh Nguyen Saolasoft Inc., Centennial, CO, USA

Armineh Nourbakhsh Research and Development, Thomson Reuters, New York City, NY, USA

Umashanthi Pavalanathan Georgia Institute of Technology, Atlanta, GA, USA

Hoang Pham Piscataway, USA

Meg Pirrung Pacific Northwest National Laboratory, Richland, WA, USA

Arash Shaban-Nejad Department of Pediatrics, Center for Biomedical Informatics, University of Tennessee Health Science Center—Oak-Ridge National Lab (UTHSC-ORNL), Memphis, TN, USA

Sameena Shah Research and Development, Thomson Reuters, New York City, NY, USA

Eun Kyong Shin Department of Pediatrics, Center for Biomedical Informatics, University of Tennessee Health Science Center—Oak-Ridge National Lab (UTHSC-ORNL), Memphis, TN, USA

Tuan Tran Sullivan University, Louisville, KY, USA

Svitlana Volkova Pacific Northwest National Laboratory, Richland, WA, USA

Christopher C. Yang College of Computing and Informatics, Drexel University, Philadelphia, PA, USA

Mi Zhang College of Computing and Informatics, Drexel University, Philadelphia, PA, USA

Abbreviations

AHMM	Autoregressive Hidden Markov Models
ARMAX	Autoregressive Moving Average Exogenous Variables Model
BNs	Bayesian networks
CBOW	Continuous Bag of Words
CDC	Centers for Disease Control and Prevention
CF	Collaborative Filtering
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DBNs	Dynamic Bayesian Networks
ED	Emergency Department
EHR	Electronic Health Records
ESMOS	Epidemic Sentiment Monitoring System
FDA	Food and Drug Administration
GOARN	Global Outbreak Alert Response Network
GPHIN	Global Public Health Intelligence Network
H7N9	Avian Influenza Type A
HIV	Human Immunodeficiency Virus
HMMs	Hidden Markov models
ICD	International Classification of Diseases
ILI	Influenza-like Illness
LDA	Latent Dirichlet Allocation
LIWC	Linguistic Inquiry and Word Count
LOD	Linked Open Data
MeSH	Medical Subject Headings
NDOS	N-dimensional order statistics
NIH	National Institutes of Health
NLM	National Library of Medicine
NLP	Natural Language Processing
NSDUH	National Survey on Drug Use and Health
PROMed-mail	The Program for Monitoring Emerging Diseases

PTSD	Post Traumatic Stress Disorder
SARS	Severe acute respiratory syndrome
SBW	Star Bright World
SHR	Social Health Records
SNS	Social Networking Services
TB	Tuberculosis
UMLS	The Unified Medical Language System Metathesaurus
UTD	User Interest and Topic Detection Model
WCA	Wavelet Coherency Analysis
WHO	World Health Organization

Public Health Intelligence and the Internet: Current State of the Art

Eun Kyong Shin and Arash Shaban-Nejad

Abstract The increasing role of Internet and the World Wide Web as a widely accessible health information source creates opportunities, along with challenges, for researchers, healthcare workers and organizations across the globe, enabling them to collect and analyze data to improving patient care, disease surveillance, and delivering online preventive or therapeutic interventions. In this chapter, after reviewing some of the applications of Internet in public health, we analyze the use of online space in clinical trials, experiments, or observations performed in clinical research, in the United States. We provide visual analytics for health data and preliminary findings from a database that systematically comprises clinical trial records. Our study focuses on clinical trial data with an "online" component in its study design. We first parse out historical trajectories of online clinical trials since its first introduction to show the use of the Internet space in health studies.

Keywords Internet intervention · Clinical trials · Social media · Population health · Surveillance · Public health

1 Introduction

The Internet has expanded the public health research beyond its traditional realm. Not only the Internet serves as effective tools for health interventions and healthcare delivery methods but also it gathers health-related data in a way that was not available before. The Internet can provide [near] real-time data and knowledge necessary for fast and timely response to public health issues. Online data can be particularly helpful in epidemic and pandemic alert and response and emergencies

E.K. Shin (✉) · A. Shaban-Nejad
Department of Pediatrics, Center for Biomedical Informatics, University of Tennessee Health Science Center—Oak-Ridge National Lab (UTHSC-ORNL), Memphis, TN, USA
e-mail: eshin3@uthsc.edu

A. Shaban-Nejad
e-mail: ashabann@uthsc.edu

A. Shaban-Nejad et al. (eds.), *Public Health Intelligence and the Internet*,
Lecture Notes in Social Networks, https://doi.org/10.1007/978-3-319-68604-2_1

preparedness. In the United States, 72% of the Internet users have used the Internet for seeking health-related information whether it is directly related to themselves or for their cared ones [1]. Recent interest in social media and the use of online platforms for health research has been remarkably increased [2, 3]. There has been a wide array of application areas (e.g., epidemic detection, health intervention, discourse analysis, health education, surveillance and health promotion, etc.) and an exponentially growing use of social media analytics in health sciences [2]. Accordingly, public health practice and research made significant development using the Internet and online social media in health intervention and health surveillance.

We begin by summarizing recent studies on application of the Internet and online social media in public health and then examine the current use of the Internet in clinical trials (experiment studies or observations done in clinical research including biomedical or behavioral research studies on human participants), which are designed to answer specific questions about biomedical or behavioral interventions.

2 Digital Public Health

As with clinical applications, the digital space has been utilized in the public health field. Continuous monitoring and surveillance in public health are of critical importance and have long-established harnessing health data from healthcare facilities, which are typically resource-intensive endeavor. Digital public health has been actively explored, to partially addressing the cost and resource limitations in traditional health surveillance. In what follows, we look at some of the applications of the online public and population health intelligence.

2.1 Health Surveillance: Epidemic Detection

Digital health surveillance has been the most popular and successful area of adopting the new media. The Internet enables health professionals and the public to observe closely the most current health dynamic. Especially for detection of infectious disease dynamics, the social media serves as one of the prompt data sources that allow timely responses from health professionals and organizations at the scale beyond what traditional and formal methods have offered.

Researchers have shown that online social media data can provide a meaningful glimpse of the utmost state of infectious disease detection [4–13]. For example, influenza surveillance using online social media has attracted many attentions [4, 5]. Text mining of influenza mentions in the web environment and social media has been well correlated with actual influenza-like illness patient report data [6]. Broniatowski and his colleagues show that influenza tweet data move closely

together with outpatient surveillance data provided by Centers for Disease Control and Prevention (CDC) [7]. Trajectories from online health data sources match well and are in accordance with data available from the official and traditional health-related data sources.

Online discussions are highly pertinent with what is happening in real-life and the social media has become one of the most influential mediums for health information exchange. Bernardo et al., after reviewing 32 peer-reviewed journal articles published from 2010 and 2011 using Google program for influenza surveillance, show that 66% of the studies confirm that social media health surveillance can provide compatible data to existing traditional surveillance programs [8].

Digital surveillance systems provided important early detection of numerous infectious diseases such as the outbreak of severe acute respiratory syndrome (SARS) in 2003, the influenza A (H1N1) pandemic in 2009 [9, 10], the 2013 avian influenza Type A (H7N9) outbreak in China [11, 12], and the 2010 cholera epidemic in Haiti [13]. As demonstrated in these examples, prompt and timely analysis of outbreak information alerted the healthcare organizations and public faster that official traditional outlets and active online discussions disseminated pertinent information more effectively. Broniatowski and his colleagues show that their influenza prevalence estimates are strongly correlated with those of other conventional health data providers [7]. Their system can detect the directionality of increase and decrease in influenza prevalence with 85% accuracy.

2.2 Global Health Surveillance

Infectious diseases are among major challenges of public health globally. As results of the increase in geographical ranges, from local to global, burdens of infectious disease and health surveillance have been globalized [14, 15]. Digital health surveillance can overcome geopolitical boundaries efficiently. Due to increased traffics cross borders, most cases of epidemics are less likely to be constrained within a national boundary (as evidenced in SARS 2003 [16]). Recent internationally spreading infectious diseases arose urgent needs for global health surveillance [17–19]. Domestic disease control is neither sufficient nor invincible without deterring international disease transmission. In line with this concern, World Health Organization (WHO) has launched Global Outbreak Alert Response Network (GOARN) in 2000. WHO not only pay close attention to global disease control but also active support for its technical needs [20]. Global health collaboration usually requires collective efforts from many and often distinctive entities. Digital public health is expected to be among the most efficient and cost-effective responses to the ever-growing needs for a synthesized, comprehensive global health surveillance.

Digital health surveillance system allows large-scale health surveillance with a relatively low cost. For instance, HealthMap, which is a platform for bringing

together disparate data sources to achieve a unified and comprehensive view of the current global state of infectious disease, is utilizing Internet media reports [21, 22]. Flu Near You is a good example of a crowdsourced participatory epidemic monitoring system [23].

2.3 Behavior Monitoring and Sentiment Assessment

In addition, to detect disease pandemics, the new media enables the quantifying and monitoring of human behavior beyond what was possible in the past public opinion tracking at large-scale possible [24]. The Internet provides an effective platform to gauge and quantitatively measure public anxiety and concern through online social media. For example, by assessing Twitter posts or blogs, health professionals can collect and analyze quantitative and qualitative indicators to measure public anxiety about Ebola and other threatening infectious diseases [9]. Analyzing online blog posts or tweets also provides a way of measuring sentiment around any health-related event and behavior such as vaccine hesitancy [23, 25, 26].

Reviewing public data from Twitter, Salathé and Khandelwal show that sentiments about vaccination, which is a critical and effective measure in analyzing the incidence of several infectious diseases, can predicate vaccination rates in accordance with projected vaccination rates by the CDC (data came from phone survey) [27]. Running simulation models using twitter data, they also manifest how clusters of negative anti-vaccination sentiments can result in an increase in the number of unvaccinated individuals. Similarly, Mollema and colleagues scrutinized public opinion online displayed regarding measles in Netherlands [24]. Again, using Twitter data, they show that tweets are highly correlated with newly released paper articles and peaks in the online trend.

2.4 Public Health Communication

Online social media and the Internet are not only magnifying the scale but also improve the quality of public health communication. The Internet is proactively used in disseminating health-related information, both reliable and non-reliable, and delivering educational materials [2]. There is well-documented evidence showing social media and online communities are effective in disseminating health information [28–31]. Beneficial features of the digital communication are widely shared among health professionals. Physicians frequently seek and consult a wide array of social media venues and health professionals hear about the most up-to-date medical information, expose themselves to other experts, and discuss their patients