

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

## Twitter: A Digital Socioscope

How can Twitter data be used to study individual-level human behavior and social interaction on a global scale? This book introduces readers to the methods, opportunities, and challenges of using Twitter data to analyze phenomena ranging from the number of people infected by the flu, to national elections, to tomorrow's stock prices. Each chapter, written by leading domain experts in clear and accessible language, takes the reader to the forefront of the newly emerging field of computational social science.

An introductory chapter on Twitter data analysis provides an overview of key tools and skills and gives pointers on how to get started, while the case studies demonstrate shortcomings, limitations, and pitfalls of Twitter data as well as their advantages. The book will be an excellent resource for social science students and researchers wanting to explore the use of online data.

YELENA MEJOVA is a scientist in the Social Computing Group at Qatar Computing Research Institute (QCRI). Before joining QCRI, Yelena was a postdoc at Yahoo! Research in Barcelona. A part of the Web Mining Group, her work concerned building semantically enriched information retrieval systems, as well as examining user behavior through social media. Prior to that, her Ph.D. thesis at the University of Iowa concerned the design and application of sentiment analysis tools for mining a variety of social media discourses, including political speech.

INGMAR WEBER is a senior scientist in the Social Computing Group at QCRI. As an undergraduate, he studied mathematics at Cambridge University, UK, before moving to the Max Planck Institute for Computer Science, Germany, for his Ph.D. Before moving to Qatar, he spent two years working at the École Polytechnique Fédérale de Lausanne, Switzerland, and three years at Yahoo! Research in Barcelona.

MICHAEL W. MACY is Goldwin Smith Professor of Arts and Sciences and Director of the Social Dynamics Laboratory at Cornell University. His research team has used computational models, online experiments, and data from social media to explore enigmatic social patterns, the emergence and collapse of fads, the spread of self-destructive behaviors, cooperation in social dilemmas, the critical mass in collective action, the spread of contagions on small-world networks, the polarization of opinion, segregation of neighborhoods, and assimilation of minority cultures.

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

---

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

# Twitter: A Digital Socioscope

Edited by

YELENA MEJOVA

*Qatar Computing Research Institute*

INGMAR WEBER

*Qatar Computing Research Institute*

MICHAEL W. MACY

*Cornell University, Ithaca, New York*



**CAMBRIDGE**  
UNIVERSITY PRESS

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

**CAMBRIDGE**  
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781107500075](http://www.cambridge.org/9781107500075)

© Cambridge University Press 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Printed in the United States of America

*A catalog record for this publication is available from the British Library.*

*Library of Congress Cataloging in Publication Data*

Twitter : a digital socioscope / [edited by] Yelena Mejova, Ingmar Weber, Michael W. Macy.  
pages cm

Includes bibliographical references and index.

ISBN 978-1-107-10237-8 (hardback) – ISBN 978-1-107-50007-5 (paperback)

1. Twitter. 2. Dyadic data analysis (Social sciences) 3. Online social networks – Research. 4. Social sciences – Research – Methodology. 5. Webometrics. I. Mejova, Yelena, 1985– II. Weber, Ingmar, 1978– III. Macy, Michael W.

HM533.5.T95 2015

302.34072'7–dc23 2015003326

ISBN 978-1-107-10237-8 Hardback

ISBN 978-1-107-50007-5 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

## Contents

<i>List of Contributors</i>	<i>page</i> vii
<i>Preface</i>	ix
<b>Introduction</b>	1
SCOTT A. GOLDER AND MICHAEL W. MACY	
<b>1. Analyzing Twitter Data</b>	21
SHAMANTH KUMAR, FRED MORSTATTER, AND HUAN LIU	
<b>2. Political Opinion</b>	52
DANIEL GAYO-AVELLO	
<b>3. Socioeconomic Indicators</b>	75
HUINA MAO	
<b>4. Hyperlocal Happiness from Tweets</b>	96
DANIELE QUERCIA	
<b>5. Public Health</b>	111
PATTY KOSTKOVA	
<b>6. Disaster Monitoring</b>	131
BELLA ROBINSON, ROBERT POWER, AND MARK CAMERON	
<i>Index</i>	161

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

---

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

## Contributors

**Mark Cameron** *Senior Research Scientist in the Digital Productivity Flagship of the Commonwealth Scientific and Industrial Research Organisation*

**Daniel Gayo-Avello** *Associate Professor in the Department of Computer Science at the University of Oviedo*

**Scott A. Golder** *Ph.D. candidate in Sociology at Cornell University and Data Scientist and Staff Sociologist at the Context Relevant Department of Sociology, Cornell University*

**Patty Kostkova** *Principal Research Associate in e-Health at University College London*

**Shamanth Kumar** *Ph.D. candidate in Computer Science and Engineering at Arizona State University*

**Huan Liu** *Professor of Computer Science and Engineering at Arizona State University*

**Michael W. Macy** *Goldwin Smith Professor of Arts and Sciences and Director of the Social Dynamics Laboratory at Cornell University*

**Huina Mao** *Ph.D. candidate at the School of Informatics and Computing at Indiana University*

**Fred Morstatter** *Ph.D. student in Computer Science and Engineering at Arizona State University*

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

---

viii

List of Contributors

**Robert Power** *Team Leader in the Digital Productivity Flagship of  
the Commonwealth Scientific and Industrial Research Organisation*

**Daniele Quercia** *College Research Associate, University of Cambridge*

**Bella Robinson** *Senior Software Engineer in the Digital  
Productivity Flagship of the Commonwealth Scientific and  
Industrial Research Organisation*



Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

## Preface

More than 500 million tweets are shared on Twitter each day, each devoting 140 characters to anything from announcements of dinner plans to calls for revolutions. Given its sheer size and the fact that most of its content is public, it is not surprising that Twitter has been extensively used to study human behavior on a global scale.

This book uses thematically grouped case studies to show how Twitter data can be used to analyze phenomena ranging from the number of people infected by the flu, to national elections, to tomorrow's stock prices. The idea for the book grew out of a three-hour tutorial on "Twitter and the Real World" given in October 2013 at the Conference on Information and Knowledge Management. Expanding on the topics covered in that tutorial, the book provides a wider thematic scope of research and the most recent scientific work.

All chapters are written by leading domain experts and take the reader to the forefront of the emerging new field of computational social science, using topical applications to illustrate the possibilities, advantages, and limitations of large, semistructured social media data to gain insights into human behavior and social interaction. Although most of the authors are computer scientists, the book is intended for readers who have not been exposed to formal training on how to analyze "big data." To this end, unnecessary implementation details are avoided. Chapter 1, "Analyzing Twitter Data," introduces readers to the tools and skills used in these studies and gives pointers on how to take the first steps as a novice researcher.

The opening chapter lays the groundwork for the book by surveying the opportunities and challenges of a "Twitter socioscope." The chapters that follow are written to be stand-alone case studies that can be reordered according to the reader's needs and preferences.

Cambridge University Press

978-1-107-10237-8 - Twitter: A Digital Socioscope

Edited by Yelena Mejova, Ingmar Weber and Michael W. Macy

Frontmatter

[More information](#)

---

*x*

## Preface

Twitter will undoubtedly evolve and its usage will change, but even if it comes to be replaced by “the next big thing,” the detailed time-stamped records of hundreds of millions of global users will continue to be one of the most valuable sources of data on human behavior ever assembled, and these case studies will remain useful as an introduction to the methods, research opportunities, and challenges that these data present.

Finally, we would like to thank all the chapter authors for their hard work to ensure the highest quality and timeliness of this book.