

# SpringerBriefs in Cybersecurity

## **Editor-in-chief**

Sandro Gaycken, Digital Society Institute, European School of Management and Technology (ESMT), Stuttgart, Baden-Württemberg, Germany

## **Series Editor**

Sylvia Kierkegaard, International Association of IT Lawyers, Highfield, Southampton, UK

John Mallery, Computer Science and Artificial Intelligence, Massachusetts Institute of Technology, Cambridge, MA, USA

Steven J. Murdoch, University College London, London, UK

Kenneth Geers, Taras Shevchenko University, Kyiv, Kiev's'ka, Ukraine

Michael Kasper, Department of Cyber-Physical Systems Security, Fraunhofer Institute SIT, Darmstadt, Hessen, Germany

Cybersecurity is a difficult and complex field. The technical, political and legal questions surrounding it are complicated, often stretching a spectrum of diverse technologies, varying legal bodies, different political ideas and responsibilities. Cybersecurity is intrinsically interdisciplinary, and most activities in one field immediately affect the others. Technologies and techniques, strategies and tactics, motives and ideologies, rules and laws, institutions and industries, power and money—all of these topics have a role to play in cybersecurity, and all of these are tightly interwoven.

The *SpringerBriefs in Cybersecurity* series is comprised of two types of briefs: topic- and country-specific briefs. Topic-specific briefs strive to provide a comprehensive coverage of the whole range of topics surrounding cybersecurity, combining whenever possible legal, ethical, social, political and technical issues. Authors with diverse backgrounds explain their motivation, their mindset, and their approach to the topic, to illuminate its theoretical foundations, the practical nuts and bolts and its past, present and future. Country-specific briefs cover national perceptions and strategies, with officials and national authorities explaining the background, the leading thoughts and interests behind the official statements, to foster a more informed international dialogue.

More information about this series at <http://www.springer.com/series/10634>

Samer Al-khateeb • Nitin Agarwal

# Deviance in Social Media and Social Cyber Forensics

Uncovering Hidden Relations Using Open  
Source Information (OSINF)



Springer

Samer Al-khateeb  
Department of Journalism,  
Media & Computing  
Creighton University  
Omaha, NE, USA

Nitin Agarwal  
Information Science Department  
University of Arkansas at Little Rock  
Little Rock, AR, USA

ISSN 2193-973X ISSN 2193-9748 (electronic)  
SpringerBriefs in Cybersecurity  
ISBN 978-3-030-13689-5 ISBN 978-3-030-13690-1 (eBook)  
<https://doi.org/10.1007/978-3-030-13690-1>

Library of Congress Control Number: 2019935524

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

*Samer would like to thank his parents, Ihsan Al-khateeb and Basima Al-Qaraghuli, and his siblings, Samar, Saja, and Hasanain Al-khateeb, for all the care and support provided to him throughout his life. He also would like to especially thank his wife Lina Al Azzawi for all the love, patience, and positive vibes given to him at the time of writing this manuscript. Nitin Agarwal would like to dedicate this book to his family and friends with much love and gratitude for everything. Without all their encouragements, this endeavor would not be possible.*

# Foreword

This book describes the methodologies and tools used to collect, analyze, and visualize social media data and conduct social cyber forensic analysis. By applying these methodologies and tools on various events observed in the case studies contained within, their effectiveness is highlighted. The methodologies blend computational social network analysis and cyber forensic concepts and tools in order to identify and study information competitors. Through cyber forensic analysis, metadata associated with propaganda-riddled websites is extracted. This metadata assists in extracting social network information such as friends and followers; communication network information, i.e., networks depicting flows of information among the actors such as tweets, replies, retweets, mentions, and hyperlinks; and social media footprint, i.e., users' or groups' activity on other social media sites. Through computational social network analysis, the authors identify influential actors and powerful groups coordinating the disinformation campaign. A blended social cyber forensic approach allows them to study cross-media affiliations of the information competitors. For instance, narratives are framed on blogs and YouTube videos, and then Twitter and Reddit, for instance, will be used to disseminate the message. Social cyber forensic methodologies enable researchers to study the role of modern information and communications technologies (ICTs) in the evolution of information campaign and coordination. In addition to the concepts and methodologies pertaining to social cyber forensics, this book also offers a collection of resources for readers including several datasets that were collected during case studies, up-to-date reference and literature surveys in the domain, and a suite of tools that students, researchers, and practitioners alike can utilize. Most importantly, the book demands a dialogue between information science researchers, public affairs officers, and policymakers to prepare our society to deal with the lawless “wild west” of modern social information systems triggering debates and studies on cyber diplomacy.

Springer Nature, Cham, Switzerland  
March 2018

Christopher T. Coughlin

# Preface

In order to address the problem of deviance in social media, we need to understand the power of social media analytics, the importance of open-source information (i.e., information from publicly available sources), and the impact of combining both to study deviant groups and tactics. We wrote this book to raise awareness of the deviant usage of social media by various groups, both virtual or physical. Hence, we introduce and blend a set of graph-theoretic concepts, social network analysis, and social cyber forensic tools and methodologies.

The readers will understand the aforementioned concepts, analysis, and tools to collect various social media data, and the metadata associated with various entities then analyze the collected data and derive insights. The methodologies and tools that we introduce throughout the book have been tested on many real-world events, e.g., the anti-NATO propaganda campaigns, ISIS/Daesh/ISILs propaganda, and the cyber campaigns of Blackhat hackers on social media. These methodologies have been developed with underpinnings in sociological theories of collective action, cyber forensic science, graph theory, and social network analysis. At the end of this text, we present a set of case studies that bring together all the concepts introduced in this book.

This book is suitable for undergraduate and graduate students as well as analysts who are interested in studying the problem of deviance in social media. After reading this book, readers will be enriched with a knowledge of various off-the-shelf tools. We hope this book can jump-start their skills on collecting, analyzing, and visualizing various social media datasets and use social cyber forensic analysis to unveil hidden relations. The book is written considering the wide range of disciplines that would benefit from it. For this reason, the concepts and tools covered in this book require absolutely zero technical background.

Omaha, NE, USA  
Little Rock, AR, USA  
December 2018

Samer Al-khateeb  
Nitin Agarwal

# Acknowledgments

We thank many colleagues who made substantial contributions in various ways to this project. The members at the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS) group at UA-Little Rock made this project much easier and enjoyable. We are grateful for their comments.

We really appreciate Springer and particularly Christopher Coughlin, Editor, for helping us throughout this project.

The research presented in this book is funded in part by the US National Science Foundation (IIS-1636933, IIS-1110868), US Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), US Air Force Research Laboratory, US Army Research Office (W911NF-16-1-0189), US Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden-Entergy Fund at the University of Arkansas at Little Rock, and Creighton University's College of Arts and Sciences. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

Last, and certainly not least, we thank our families, for supporting us through this fun but time-consuming project. We dedicate this book to them, with love.



# Contents

<b>1 Deviance in Social Media</b> .....	1
1.1 Introduction .....	1
1.2 Literature on Online Deviant Behaviors .....	3
1.2.1 Online Deviant Groups .....	4
1.2.2 Online Deviant Events .....	8
1.2.3 Online Deviant Tactics .....	11
1.3 Leveraging the Theory of Collective Action to Study DCFMs .....	13
1.3.1 The Case of DCFM-Success .....	16
1.3.2 The Case of DCFM-Failure .....	17
1.3.3 Conceptual Framework .....	17
1.3.4 A CFM Scenario .....	19
References .....	22
<b>2 Social Network Measures and Analysis</b> .....	27
2.1 Basics of Graph Theory .....	27
2.1.1 Graph Data Structures .....	31
2.2 Fundamentals of Social Network Analysis (SNA) .....	32
2.2.1 Centrality Measures .....	34
2.2.2 Triadic Closure and Clustering Coefficient .....	36
2.2.3 Modularity .....	38
2.2.4 Influential Blogs and Influential Bloggers .....	39
2.2.5 Focal Structures Analysis (FSA) .....	41
References .....	43
<b>3 Tools and Methodologies for Data Collection, Analysis, and Visualization</b> .....	45
3.1 TouchGraph SEO Browser .....	45
3.2 Twitter Archiving Google Sheet (TAGS) and TAGSExplorer .....	47
3.3 Network Overview, Discovery, and Exploration for Excel (NodeXL) .....	48
3.4 Gephi .....	48
3.5 CytoScape .....	51

- 3.6 Linguistic Inquiry and Word Count (LIWC) ..... 51
- 3.7 Organizational Risk Analyzer (ORA) NetScenes ..... 53
- 3.8 IBM Watson Analytics ..... 55
- 3.9 Web Content Extractor (WCE)..... 57
- 3.10 Blogtrackers ..... 58
- 3.11 YouTubeTracker ..... 60
- 3.12 Botometer ..... 61
- 3.13 Reaper: Social Media Scraping Tool ..... 62
- References ..... 64
- 4 Social Cyber Forensics (SCF): Uncovering Hidden Relationships ..... 67**
  - 4.1 Social Cyber Forensics Analysis (SCF) Using Maltego ..... 67
  - 4.2 Methodologies to Extract Open Source Information..... 71
    - 4.2.1 Finding Related Websites From Web Tracker Code (WTC) ..... 71
    - 4.2.2 Finding Blog Sites From Twitter Handles..... 72
    - 4.2.3 Inferring the Ownership or Hidden Connections Among Different Websites ..... 73
  - 4.3 Hands-On Exercises ..... 73
    - 4.3.1 Exercise A ..... 73
    - 4.3.2 Exercise B ..... 75
    - 4.3.3 Exercise C ..... 75
  - References ..... 76
- 5 Case Studies of Deviance in Social Media ..... 79**
  - 5.1 Introduction ..... 79
  - 5.2 Case Study 1: Propaganda During the 2014 Crimean Water Crisis ..... 82
  - 5.3 Case Study 2: Anti-NATO Propaganda During the 2015 Trident Juncture Exercise ..... 84
  - 5.4 Case Study 3: Anti-NATO Propaganda During the 2015 Dragoon Ride Exercise ..... 87
  - 5.5 Case Study 4: ISIS Beheading Propaganda in 2015 ..... 89
  - References ..... 92
- Glossary ..... 95**
- Index ..... 97**

## About the Authors

**Samer Al-khateeb** is an Assistant Professor in the Department of Journalism, Media and Computing, College of Arts and Sciences, at Creighton University and a former Postdoctorate Research Fellow at the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS) at the University of Arkansas at Little Rock (UA-Little Rock). He obtained his Ph.D. in Computer and Information Sciences, a master's degree in Applied Science, and a bachelor's degree in Computer Science from UA-Little Rock. He studies deviant acts (e.g., deviant cyber flash mobs and cyber propaganda campaigns) on social media that are conducted by deviant groups (e.g., Daesh, Black hat hackers, and Propagandist) which aim to influence individual's behaviors and provoke hysteria among citizens. He also studies the type of actors these deviant groups use to perform their acts, i.e., are they human (e.g., Internet trolls) or automated actors (e.g., social bots) by leveraging social science theories (e.g., the theory of collective action), social network analysis (e.g., centralities and community detection algorithms), and social cyber forensics (e.g., metadata collection to uncover the hidden relations among these actors across platforms). He has many publications including book chapters, journal papers (e.g., *Defence Strategic Communications*; *Journal of Digital Forensics, Security and Law*; *Journal on Baltic Security*; and the *IARIA (International Journal on Advances in Internet Technology)*), conference proceedings, and conference presentations. He won various awards such as the Staff Achievement Award for Educational Achievements, Excellence in Research Award, Outstanding Graduating Student Award (Master's Level), Who's Who Among Students in American Universities and Colleges, the Best Paper Award, 2nd Place Most Innovative Award, and 2nd Place Societal Impact Award, among others.

**Nitin Agarwal** is the Jerry L. Maulden-Entergy Endowed Chair and Distinguished Professor of Information Science at the University of Arkansas at Little Rock. He is the Founding Director of the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS) at UA-Little Rock. His research aims to push the boundaries of our understanding of cyber social behaviors that emerge and evolve constantly in the modern information and communication platforms with

applications in defense and security, health, business and marketing, finance, and education. At COSMOS, he is leading projects funded by over \$10 million from an array of federal agencies, including US National Science Foundation, Office of Naval Research, Army Research Office, Air Force Research Laboratory, Defense Advanced Research Projects Agency, and Department of State, and plays a significant role in the long-term partnership between UA-Little Rock and the Department of Homeland Security. He developed publicly available social media mining tools, viz., Blogtrackers, YouTube Tracker, and Focal Structure Analysis used by NATO Strategic Communications and public affairs, among others. Dr. Agarwal participates in the national Tech Innovation Hub launched by the US Department of State to defeat foreign-based propaganda.

His research contributions lie at the intersection of social computing, behavior-cultural modeling, collective action, social cyber forensics, AI, data mining, and machine learning. From Saudi Arabian women's right to drive cyber campaigns to autism awareness campaigns to ISIS and anti-West/anti-NATO disinformation campaigns, at COSMOS, he is directing several projects that have made foundational and applicational contributions to social and computational sciences. He has published 8 books and over 150 articles in top-tier peer-reviewed forums with several best paper awards and nominations. Dr. Agarwal obtained Ph.D. from Arizona State University with outstanding dissertation recognition in 2009. He was recognized as one of "The New Influentials: 20 In Their 20s" by Arkansas Business in 2012. He was recognized with the university-wide Faculty Excellence Award in Research and Creative Endeavors by UALR in 2015. Dr. Agarwal received the Social Media Educator of the Year Award at the 21st International Education and Technology Conference in 2015. In 2017, the *Arkansas Times* featured Dr. Agarwal in their special issue on "Visionary Arkansans: A Celebration of Arkansans with ideas and achievements of transformative power." Dr. Agarwal was nominated as International Academy, Research and Industry Association (IARIA) Fellow in 2017, Arkansas Academy of Computing (AAoC) Fellow in 2018, and Arkansas Research Alliance (ARA) Fellow in 2018.

Visit <http://ualr.edu/nxagarwal/> for more details.

# Acronyms

API	Application Programming Interface
ASAs	Automated social actors/agents
BRJP	Brilliant jump exercise
CA	Collective Action
CCA	Cyber collective action
CFM	Cyber flash mob
DCFM	Deviant Cyber Flash Mob
DDoS	Distributed denial-of-service
DHN	Deviant hacker networks
EEAS	European External Action Service
FM	Flash mob
FSA	Focal Structures Analysis
ICSR	International Centre for the Study of Radicalization and Political Violence
ICT	Information and communications technology
I2P	Invisible Internet Project
IRC	Internet Relay Chat
ISIL	Islamic State of Iraq and the Levant
ISIS	Islamic State of Iraq and Syria
ITAR	Information Telegraph Agency of Russia
LIWC	Linguistic Inquiry and Word Count
MMOGs	Massive Multiplayer Online Games
MUDs	Multi-User-Domains
NATO	The North Atlantic Treaty Organization
NIGMS	US National Institute of General Medical Sciences
NodeXL	Network Overview, Discovery and Exploration for Excel
NRNB	National Resource for Network Biology
NSMT	New Social Movement Theory
ODBC	Open Database Connectivity
ODGs	Online Deviant Groups
ORA	Organizational Risk Analyzer

OSINF	Open-Source Information
OSN	Online Social Network
PAO	Public Affairs Officer
SCF	Social cyber forensics
SNA	Social network analysis
TAGS	Twitter Archiving Google Sheet
TASS	Telegraph Agency of the Soviet Union
TCOs	Transnational Crime Organizations
TOR	The Onion Router
TRJE	Trident Juncture Exercise
WCE	Web Content Extractor
WoW	World of Warcraft
WTC	Web Tracker Code

# List of Figures

Fig. 1.1 The different forms of cyber collective action (CCA), e.g., parkour, social movements and campaigns, and flash mobs (FM) by their types, i.e., benign flash mobs (FM), cyber flash mobs (CFM), and deviant cyber flash mobs (DCFMs) ..... 3

Fig. 1.2 Some examples of online deviant groups (ODGs) ..... 5

Fig. 1.3 The banner Anonymous designed for their operation ..... 7

Fig. 1.4 A dancing flash mob at Toronto Eaton Centre ..... 9

Fig. 1.5 Bash mob in Long Beach California as an example of a deviant flash mob ..... 10

Fig. 2.1 (a) *Königsberg* bridge problem map, (b) representation of the *Königsberg* bridge problem as a graph..... 28

Fig. 2.2 Graph on left is an unweighted and undirected graph. Graph in the middle is a weighted and directed graph. Graph on the right is an unweighted and directed graph ..... 29

Fig. 2.3 Graph on the left represents a simple graph, graph in the middle represents a multigraph, and graph on the right represents a pseudograph ..... 30

Fig. 2.4 Graph on the left is a bipartite, graph in the middle is a tripartite, and graph on the right represents a heterogeneous graph ..... 30

Fig. 2.5 On left an incidence matrix, in the middle an adjacency matrix, and on right a distance matrix all for the graph on right in Fig. 2.2 ..... 32

Fig. 2.6 On left an undirected graph and on right a table of each vertex betweenness and closeness centralities ..... 35

Fig. 2.7 Two different triplets, same nodes but missing different edges ..... 37

Fig. 2.8 Examples of global and local clustering coefficient. Solid lines depict connected vertices, while dashed lines depict missing connections among vertices ..... 38

Fig. 2.9 A small world network with three communities on left. The same network presented on the right with block view of each community ..... 40

Fig. 3.1 TouchGraph results show how the three keywords “NATO,” “USA,” and “Russia” are connected through the respective domains as reported by Google search algorithm ..... 46

Fig. 3.2 TAGSExplorer results show @tim\_cook as the top conversationalist in the collected data ..... 47

Fig. 3.3 The social network of ISIL’s top 10 propaganda disseminators. Green edges are ISIL nodes following others. Red edges are other nodes following ISIL nodes. The nodes with different colors and bigger size represent ISIL top disseminators..... 49

Fig. 3.4 An undirected network of hashtags that co-occur with the hashtags #hurricaneharvey and #prayfortexas. The nodes are represented in gray color, the edges in light green, and the node’s labels in red color. The size of the node represents the degree centrality of the node ..... 50

Fig. 3.5 Eighteen bot accounts that are disseminating ISIL beheading of Ethiopian video on Twitter. Green ovals are Twitter bot accounts, while the red triangle is the tweet. Edges in black color represent tweet relationship ..... 52

Fig. 3.6 Sentiment trend in the blogosphere during the European Migrant Crisis between January 2015 and March 2016 ..... 53

Fig. 3.7 (a) “Refugees Welcome” banners in a major soccer game, (b) “Rapefugees Not Welcome” banners in a street protest..... 53

Fig. 3.8 The communication network of the deviant hackers on Twitter. Green edges represent “Retweets” relation. Black edges represent “Mentions” relation ..... 55

Fig. 3.9 The question along with its answer provided by Watson Analytics ..... 56

Fig. 3.10 Screenshot of web content extractor ..... 58

Fig. 3.11 The analytics page of Blogtrackers, showing all the current features ..... 59

Fig. 3.12 The analytics page of Blogtrackers, showing all the current features ..... 61

Fig. 3.13 The Botometer web interface showing a Twitter bot account has a bot score of 98% ..... 62

Fig. 3.14 Reaper GUI showing the welcome page ..... 62

Fig. 4.1 The available OSINT tools. Not many tools available ..... 70

Fig. 4.2 Maltego results after following the steps in Sect. 4.2.1 ..... 74

Fig. 4.3 Maltego results after following the steps in Sect. 4.2.2 ..... 75

Fig. 4.4 Maltego results after following the steps in Sect. 4.2.3 ..... 76



Fig. 5.1 A high-level view of the methodology followed to study various deviant acts and groups ..... 80

Fig. 5.2 Examples of offensive and biased memes observed on blogosphere during NATO’s exercises to delegitimize exercises’ objectives ..... 81

Fig. 5.3 Social bot network ..... 83

Fig. 5.4 A bridge blogger, Spanish and English blogs with anti-NATO narratives. The two blogs are kept unlinked. Connection between the blogs is discovered through Google Plus profile of the blogger using SCF methodologies ..... 85

Fig. 5.5 The additional blog sites that were identified using Maltego had three clusters labeled as 1, 2, and 3 based on their IP address geo-location ..... 86

Fig. 5.6 Two sub-networks, S1 and S2. S1 and S2 are un-collapsed. Edges in green denote mutually reciprocal relations (bidirectional edges) while edges in red color denote non-reciprocal relations (unidirectional edges). Nodes are sized based on their indegree centrality ..... 88

Fig. 5.7 Screen capture of (a) the beheading of Egyptians Copts in Libya by ISIL, (b) the beheading of the Arab-Israeli “Spy” in Syria by ISIL, and (c) the beheading of the Ethiopian Christians in Libya by ISIL ..... 89