Overview of the Special Issue on Trust and Veracity of Information in Social Media

SYMEON PAPADOPOULOS, Centre for Research and Technology Hellas
KALINA BONTCHEVA, University of Sheffield
EVA JAHO, Athens Technology Center
MIHAI LUPU, Vienna University of Technology
CARLOS CASTILLO, Sapienza University of Rome

CCS Concepts: • **Information systems** → **Information retrieval;** *Web searching and information discovery*; • **Human-centered computing** → *Social networking sites*

Additional Key Words and Phrases: Information veracity, social media, news verification, fake content, rumour propagation

ACM Reference Format:
Symeon Papadopoulos, Kalina Bontcheva, Eva Jaho, Mihai Lupu, and Carlos Castillo. 2016. Overview of the special issue on trust and veracity of information in social media. ACM Trans. Inf. Syst. 34, 3, Article 14 (April 2016), 5 pages.
DOI: http://dx.doi.org/10.1145/2870630

From a business and government point of view, there is an increasing need to interpret and act upon information from large-volume media, such as Twitter, Facebook and Web news. However, knowledge gathered from such online sources comes with a major caveat—it cannot always be trusted, nor is it always factual or of high quality. Rumors tend to spread rapidly through social networks, and their veracity is hard to establish in a timely fashion. For instance, during an earthquake in Chile, rumors spread through Twitter that a volcano became active and there was a tsunami warning in Valparaiso [Castillo et al. 2013]. Later, these reports were found to be false. Another example concerns "astroturf campaigns"—a malicious use of Twitter and other social media during election campaigns to provide fake support of a message or project by grassroots participants, while at the same time hiding the original campaign sponsors (usually elite groups or their lobbies). Researchers have identified numerous sources of untrusted content, and through them they found that several online communities interact with narratives stemming from conspiracy theories [Bessi et al. 2015]. A 2012

This work is supported by the REVEAL and PHEME projects under grant numbers 610928 and 611233 respectively.

Authors' addresses: S. Papadopoulos, Information Technologies Institute, Centre for Research and Technology Hellas; Thessaloniki, Greece; email: papadop@iti.gr; K. Bontcheva, Department of Computer Science, University of Sheffield, Sheffield, UK; email: k.bontcheva@sheffield.ac.uk; E. Jaho, Innovation Lab, Athens Technology Center, Athens, Greece; email: e.jaho@atc.gr; M. Lupu, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria; email: mihai@mihailupu.net; C. Castillo, Department of Computer, Control, and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy; email: chato@acm.org.

DOI: http://dx.doi.org/10.1145/2870630

report from Pew Internet Research on the future of big data [Anderson and Rainie 2012] argues that, even though by 2020 big data is likely to have a transformational effect on our knowledge and understanding of the world, there is also high risk of "distribution of harms" due to the abundance of inaccurate and false information.

There has been increasing research interest in computational models of assessing properties of content and information, such as veracity and quality, and automatic methods of uncovering deception. So far, information quality and deception has been studied in relation to the manipulation of content, comments and tags, as well as search engine results. For instance, this has been the focus of a series of relevant workshops, namely AIRWeb (Adversarial Information Retrieval on the Web), with the latest edition taking place in 2009 [Fetterly and Gyöngyi 2009]. More focus on information credibility on the Web was given by the WICOW (Workshop on Information Credibility on the Web) workshop series, organized until 2010 [Tanaka et al. 2010]. Later, the aforementioned two workshops merged into the WebQuality conferences [Nielek et al. 2015]. However, the previously listed workshops have not sufficiently addressed the social computing and multimedia analysis aspects of the problem, such as characteristics and impact of short messages, their real-time nature, and interweaving of social media users and content. These are increasingly important in view of the ubiquitous use of social networking and media sharing applications, and the growing reliance of information seekers on user-generated content in a number of domains (e.g., news, shopping, travelling).

Emphasis on social media aspects was given by another set of workshops, including the workshop series on Making Sense of Microposts, with the latest edition being organized for the fifth time in 2015 [Rowe et al. 2015], the RAMSS 2013 workshop on Real-time Analysis and Mining of Social Streams [Zubiaga et al. 2013], and the SNOW workshops on Social News on the Web, with the latest edition including a data challenge on real-time topic detection [Papadopoulos et al. 2014]. Such workshops have placed emphasis on the real-time nature and particular characteristics of microblogging platforms (with Twitter being one of the most popular platforms of study), and have addressed a number of challenging problems, such as trend detection and tracking, event monitoring, influencer mining, and information spread.

Given the real-time nature of social media and their wide usage in cases of breaking events, such as natural disasters or terrorist attacks, there have been several efforts to study the quality and reliability of content that is circulated in such contexts [Castillo et al. 2013; Gupta et al. 2013], and to compare the features of trustworthy content with those of content carrying malicious or erroneous information [Canini et al. 2011; Castillo et al. 2011; Gupta and Kumaraguru 2012; O'Donovan et al. 2012]. Most of the aforementioned studies used Twitter as the source of social media content, and examined a number of Twitter-specific features, such as number of hashtags, number of retweets, number of followers, etc., as well as some more generic features, primarily related to the text of tweets. An interesting question when comparing these studies is whether their findings are transferrable to other social networks or whether they are tightly related to the particular datasets/events that were used in these studies. A recent benchmarking task [Boididou et al. 2015] in the context of the MediaEval workshop series attempted to explore such questions. In parallel, another MediaEval evaluation task focused on the credibility of image annotations in social media [Ionescu et al. 2015].

Related efforts focused on devising automated methods of assessing the credibility of shared content [Gupta et al. 2014], spotting fake content [Martinez-Romo and Araujo 2013; Boididou et al. 2014], and detecting rumors spreading in social media [Qazvinian et al. 2011; Ratkiewicz et al. 2011; Seo et al. 2012; Liu et al. 2015; Ma et al. 2015; Zubiaga et al. 2015]. Methods focusing on the credibility-veracity of individual items (tweets) typically adopt a supervised learning approach: models are built based on past cases of true/fake items and then applied to new/unseen data. A key question related to these approaches pertains to the transferability of their results, i.e., how applicable are the pre-trained models in new domains/events. For instance, the study of Boididou et al. [2014] demonstrated that previous methods overestimated the classification accuracy that could be achieved by a trained model due to the fact that they did not account for the change in the test domain/event. Other recent methods do not only solely rely on features of individual content items and of their authors (users), but also take into account the event to which these items refer, as well as the reactions of other users in the network to these items. Furthermore, several of the aforementioned works are not agnostic to the online context of a test content item, but study how false information spreads over a social network and examine how feasible it is to detect the rumor at an early stage. A noteworthy such study was carried out by Seo et al. [2012], which deals with the problem of identifying the source (provenance) of rumors by injecting a number of monitor nodes in the social network.

For this special issue, we solicited articles reflecting the state of the art and emerging trends in Trust and Veracity of Information in Social Media. Although manuscripts focusing on all areas of trust and veracity of information were considered, we especially encouraged submissions that focused on social networking environments and online context; in particular, topics such as estimating trust on a user graph, detecting misinformation given observations of social network connections and flows, and detecting false content by leveraging geosemantic information. From 10 submissions, we selected four high-quality articles that represent current themes of research on rumor and fake content detection, and trust in social networking systems. In the following, we briefly summarize each of the accepted articles.

Digital wildfires are rumors that spread uncontrollably on social media. The contribution, *Digital Wildfires? Propagation, Verification, Regulation and Responsible Innovation,* by Webb et al., is a multi-disciplinary survey on the area of false/malicious information spread in online settings that results in the creation of digital wildfires. Their survey brings together work from the areas of computational and social science, and identifies research gaps, while proposing an agenda and a research methodology to systematically explore the problem of unverified content propagation and to address the issues of and governance by state agencies, news media, or even social media users themselves.

A more specific problem is studied in the article, *Geoparsing and Geosemantics for Social Media: Spatio-Temporal Grounding of Content Propagating Rumours to Support Trust and Veracity Analysis during Breaking News,* by Middleton and Krivcovs. In particular, the authors present a social media analytics system to support journalists in their verification processes, along with a novel approach to perform geosemantic feature extraction and classification of evidence sourced from social media content in terms of situatedness, timeliness, confirmation, and validity. Although the article focuses primarily on information that can be automatically extracted through processing social streams, it also discusses human aspects of verification, namely the assistance of the journalistic workflow through appropriate visualizations and insights that can be supported by the proposed framework.

The article *TISON: Trust Inference in Trust-Oriented Social Networks,* by Hamdi et al., focuses on the problem of computing trust on top of a social network expressing trust ratings among its members. This is of special interest in online communities, where interactions and explicit relations among members express a form of trust directed from one member to another. The authors investigate the properties of trust propagation within social networks, based on the notion of transitivity, and introduce a new model, termed TISON, for trust inference within the network. To this end, they develop a Trust Path Search (TPS) algorithm and different Trust Inference Measure (TIM) algorithms.

The fourth article of the special issue, *Misinformation in Online Social Networks: Catch Them All with Limited Budget,* by Zhang et al., focuses similarly to the work by Seo et al. [2012] on the general problem of misinformation detection in graphs, in cases where the knowledge about misinformation sources is lacking. The article demonstrates the equivalence of this problem to the one of influence maximization in the reverse graph. Moreover, considering node vulnerability, the authors attempt to detect the misinformation reaching to a specific user by placing a limited number of monitors on the network. They then propose a minimum monitor set construction algorithm to solve the problem in an optimal way and demonstrate its superiority in a number of real-world networks.

Overall, the increasing interest in work on information trust and veracity combined with the challenges and special nature of social media content make this a timely special issue that we hope will contribute to the further development of this area. Given the recent advances and breadth of the topic, as well as the level of interest in pertinent events, such as workshops, we aspire for the articles of the special issue to be both informative and thought provoking for readers. In addition, the variety of approaches proposed and used in the articles reveals that there are many possible aspects and nuances to this problem, and that a multi-disciplinary approach that combines different computer science fields with concepts and methods from social science can be beneficial. In this frame, we anticipate in the future a number of diverse and complementary approaches to be combined in order to tackle different aspects of trust and veracity of information in social media.

REFERENCES

Janna Anderson and Lee Rainie. 2012. The future of big data. Retrieved December 17, 2015 from http://www.pewinternet.org/2012/07/20/the-future-of-big-data/.

Alessandro Bessi, Mauro Coletto, George A. Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Science vs conspiracy: Collective narratives in the age of misinformation. *PLOS ONE* 10, 2 (23 Feb. 2015), e0118093+. DOI:http://dx.doi.org/10.1371/journal.pone.0118093

Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, and Yiannis Kompatsiaris. 2015. Verifying multimedia use at MediaEval 2015. In *Working Notes Proceedings of the MediaEval 2015 Workshop.*

Christina Boididou, Symeon Papadopoulos, Yiannis Kompatsiaris, Steve Schifferes, and Nic Newman. 2014. Challenges of computational verification in social multimedia. In *Proceedings of the 23rd International Conference on World Wide Web (WWW'14 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 743–748. DOI:http://dx.doi.org/10.1145/2567948.2579323

Kevin Robert Canini, Bongwon Suh, and Peter Pirolli. 2011. Finding credible information sources in social networks based on content and social structure. In *Proceedings of the 2011 IEEE 3rd International Conference on Privacy, Security, Risk and Trust and 2011 IEEE 3rd International Conference on Social Computing.* IEEE, 1–8. DOI:http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.91

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web (WWW'11)*. ACM, New York, NY, 675–684. DOI:http://dx.doi.org/10.1145/1963405.1963500

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2013. Predicting information credibility in time-sensitive social media. *Internet Research* 23, 5 (2013), 560–588. DOI:http://dx.doi.org/10.1108/IntR-05-2012-0095

Dennis Fetterly and Zoltán Gyöngyi (Eds.). 2009. *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web.*

Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media (PSOSM'12)*. ACM, New York, NY, Article 2, 7 pages. DOI:http://dx.doi.org/10.1145/2185354.2185356

Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. TweetCred: Real-time credibility assessment of content on twitter. In *Proceedings of the 6th International Conference, SocInfo 2014*. Luca Maria Aiello and Daniel A. McFarland (Eds.), Vol. 8851. Springer, 228–243. DOI:http://dx.doi.org/10.1007/978-3-319-13734-6_16

Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking Sandy: Characterizing and identifying fake images on Twitter during hurricane Sandy. In *Proceedings of the 22nd International Conference on World Wide Web (WWW'13 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 729–736.

Bogdan Ionescu, Alexandru-Lucian Gînsca, Bogdan Boteanu, Adrian Popescu, Mihai Lupu, and Henning Müller. 2015. Retrieving diverse social images at MediaEval 2015: Challenge, dataset and evaluation. In *Working Notes Proceedings of the MediaEval 2015 Workshop*. Martha A. Larson, Bogdan Ionescu, Mats Sjöberg, Xavier Anguera, Johann Poignant, Michael Riegler, Maria Eskevich, Claudia Hauff, Richard F. E. Sutcliffe, Gareth J. F. Jones, Yi-Hsuan Yang, Mohammad Soleymani, and Symeon Papadopoulos (Eds.). 2015. *Working Notes Proceedings of the MediaEval 2015 Workshop.* CEUR Workshop Proceedings, Vol. 1436

Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-time rumor debunking on Twitter. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM'15)*. ACM, New York, NY, 1867–1870. DOI:http://dx.doi.org/10.1145/2806416.2806651

Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM'15)*. ACM, New York, NY, 1751–1754. DOI:http://dx.doi.org/10.1145/2806416.2806607

Juan Martinez-Romo and Lourdes Araujo. 2013. Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Syst. Appl.* 40, 8 (June 2013), 2992–3000. DOI:http://dx.doi.org/10.1016/j.eswa.2012.12.015

Radoslaw Nielek, Adam Wierzbicki, Adam Jatowt, and Katsumi Tanaka (Eds.). 2015. *WebQuality 2015, 5th International Workshop on Web Quality, Co-Located with the 24th International World Wide Web Conference (WWW'15)*.

John O'Donovan, Byungkyu Kang, Greg Meyer, Tobias Hollerer, and Sibel Adalii. 2012. Credibility in context: An analysis of feature distributions in twitter. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust (SOCIALCOM-PASSAT'12)*. IEEE Computer Society, Washington, DC, 293–301. DOI:http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.128

Symeon Papadopoulos, David Corney, and Luca Maria Aiello (Eds.). 2014. *Proceedings of the SNOW 2014 Data Challenge co-located with the 23rd International World Wide Web Conference (WWW'14).* CEUR Workshop Proceedings, Vol. 1150.

Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'11)*. Association for Computational Linguistics, Stroudsburg, PA, 1589–1599.

Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. 2011. Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference Companion on World Wide Web (WWW'11)*. ACM, New York, NY, 249–252. DOI:http://dx.doi.org/10.1145/1963192.1963301

Matthew Rowe, Milan Stankovic, and Aba-Sah Dadzie (Eds.). 2015. *Proceedings of the the 5th Workshop on Making Sense of Microposts Co-Located with the 24th International World Wide Web Conference (WWW'15).* CEUR Workshop Proceedings, Vol. 1395.

Eunsoo Seo, Prasant Mohapatra, and Tarek Abdelzaher. 2012. Identifying rumors and their sources in social networks. *Proc. SPIE* 8389 (2012), 83891I–83891I–13. DOI:http://dx.doi.org/10.1117/12.919823

Katsumi Tanaka, Xiaofang Zhou, Min Zhang, and Adam Jatowt (Eds.). 2010. *Proceedings of the 4th ACM Workshop on Information Credibility on the Web (WICOW'10).*

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Kalina Bontcheva, and Peter Tolmie. 2015. Towards detecting rumours in social media. *CoRR* abs/1504.04712 (2015). http://arxiv.org/abs/1504.04712

Arkaitz Zubiaga, Damiano Spina, Maarten de Rijke, and Markus Strohmaier (Eds.). 2013. *RAMSS 2013: Proceedings of the Second Workshop on Real-time Analysis and Mining of Social Streams, Co-Located with the 22nd International World Wide Web Conference (WWW'13)*.

Received December 2015; accepted December 2015