AuthIntegrate: Toward Combating False Data on the Internet

Romila Pradhan
Department of Computer Science
Purdue University, IN, USA
rpradhan@purdue.edu

Sunil Prabhakar
Department of Computer Science
Purdue University, IN, USA
sunil@purdue.edu

ABSTRACT

The advent of the collaborative Web and the abundance of user-generated data has resulted in the problem of *information overload*; it is becoming increasingly difficult to discern relevant information and discard false data. Recently, a number of solutions for automated fact-checking have been proposed that view the problem from a largely linguistic perspective. We observe that the problem of false data detection has roots in several extensively studied research areas in data management and data mining such as data integration, data cleaning, crowdsourcing and machine learning. Specifically, detection of false data has significant overlap with data fusion, an active area of research in data integration that focuses on distinguishing correct from incorrect information in a structured data setting. In this vision paper, we propose the architecture of AuthIntegrate, an end-to-end system that ingests conflicting data from disparate information providers, curates and presents highly accurate data to end-users. We discuss the technical challenges in building this system and outline an agenda for future research.

ACM Reference Format:
Romila Pradhan and Sunil Prabhakar. 2018. AuthIntegrate: Toward Combating False Data on the Internet. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

With the abundance of information on the Internet from multiple providers, it has become essential for data curators to ingest, standardize and clean data before extracting any value from it. While the volume and variety of data has rocketed over the years, often there is little to no restraint over their quality; data sources often provide conflicting information for the same data item (a real-world entity or event) and as a result, information on the Internet are rife with inconsistencies.

Resolving conflicting information is important because inaccurate data may result in unfavorable consequences such as unexpected financial losses. A perfect example of the damage inconsistent, unverified information can inflict is the steady rise of "fake" news in the media and popular culture. Increasingly, it is becoming difficult for consumers to fathom whether or not a particular piece of information should be trusted unambiguously. The urgency of this matter has prompted concern from inter-governmental agencies [1, 2] that consider the dissemination of trustworthy information to be of paramount importance.

Previous work related to the detection of false information falls under three major themes – (1) leveraging linguistic cues in specific communities, (2) structured conflict resolution (or, data fusion) mechanisms, and (3) soliciting expert intervention. In the following, we discuss related work in each of these themes and examine how advances in these areas can benefit the cause of combating misinformation on the Internet.

Assessing Claims Individually. Fabricated information has been around on the Internet in different forms such as deception, fake reviews, vandalisms, controversies, rumors and hoaxes. There has been a surge of research in recent years on the credibility of claims in social media, specific communities and the Web, and can be broadly categorized as being either *language-based* or *structure-based*. The language-based false data detection approaches heavily rely on different aspects of language – tone, stance, objectivity, hedges, negation – to infer the correctness of claims [15, 25, 26, 29]. On the other hand, the structure-based models are specific to communities such as social networks [43]. For example, the problems of detecting vandals, controversies and hoaxes have primarily been studied in the context of Wikipedia [6, 18, 19, 30] whereas rumor identification has mostly been studied on microblogging websites and social media [17, 23, 33, 37, 41] and detecting false reviews has been an active area of research in the services business [9]. While there has been significant progress in detecting different forms of false data, there is a lack of consolidated efforts from these different, though related, research areas.

Structured Data Fusion. In light of the growing discord over structured data extracted from disparate data sources, recent years have witnessed significant research in the area of *data fusion*, or *truth discovery* [21] – a key step in the data integration pipeline. A growing list of data fusion systems over the years can be found in [4] and [21]. Data fusion combines multiple instances of the same real-world data item from heterogeneous data sources to produce a single consistent record. State-of-the-art data fusion systems propose conflict resolution, i.e., distinguishing correct from incorrect information, as a solution to integrating inconsistent data from multiple providers and presenting end-users with the most accurate data. From trusting data sources uniformly, fusion mechanisms over the years have moved toward identifying credible sources and trusting them over others. Sophisticated fusion systems characterize data sources through quality measures, such as accuracy, precision, recall and false positive rate, and use a variety of techniques,

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

[1] http://www.un.org/apps/news/story.asp?NewsID=56336
[2] http://reports.weforum.org/outlook-14/top-ten-trends-category-page/10-the-rapid-spread-of-misinformation-online/

such as Bayesian analysis [10], probabilistic graphical models [27], optimization [22] and probabilistic soft logic [36], to jointly infer correctness of claims and source credibilities. The idea that not all data sources are of equal value was further studied in the source dependence [10, 28, 38] and source selection [34] problems.

While data fusion has proved quite successful in resolving inconsistencies in structured data, it has not been fully explored for the resolution of unstructured data conflicts. Moreover, data fusion stands on the assumption that information providers are inherently honest; however, in the present era and in the context of "fake" news, this assumption no longer holds true. The existence of possibly malicious players has been an active area of research in data classification and machine learning [8, 16] but has not been studied in the context of data integration and, in particular, data fusion.

User Interaction. In the era of "alternative" facts, fact-checking websites, such as Snopes and PolitiFact, have emerged as vanguards having dedicated teams of employees who comb through speeches, news stories, press releases to verify rumors and political claims. Solicitation of human input has been studied in various data management problems [12, 32] and, in general, has been found to improve effectiveness of the concerned tasks. In particular, [32] studies the problem of efficiently involving users to validate claims for the data fusion problem. Ongoing research in collecting input from a crowd of workers, termed *crowdsourcing* [13], adds a new dimension to user interaction where instead of domain experts, the task can be outsourced to workers on crowdsourcing portals such as Amazon Mechanical Turk and CrowdFlower. The presence of noisy data sources and noisy users requires jointly estimating true labels, source credibilities and user qualities [39]. We contend that advances made in characterizing users and incorporating user feedback in data management tasks will help in leveraging expert teams on the fact-checking websites for effective and efficient detection of misinformation.

In this paper, we propose that the problem of detecting false data from the sea of conflicting information on the Internet would benefit from recent advances in the fields of data fusion, natural language processing techniques and effective integration of user input. We address key challenges in directly applying existing approaches to data provided by (possibly biased) data sources and briefly discuss the implementation of AuthIntegrate, a system that ingests (possibly) conflicting data from heterogeneous data providers, distinguishes correct from incorrect information and provisions strategies to limit the spread of false data on the Internet.

Organization: We present the architecture of our envisioned system and its various components in Section 2 where we discuss in detail specific research problems pertaining to each system component. In Section 2.1, we present information extraction and knowledge management strategies for processing the data. We present in Section 2.2, key challenges in the data fusion module, which is the next step in the proposed system pipeline. We address strategies to identify misinfluencers and limit the spread of misinformation in Section 2.3 and conclude in Section 3.

2 ARCHITECTURE OVERVIEW

In this section, we discuss architecture of the AuthIntegrate system (shown in Figure 1). Our system AuthIntegrate has the single agenda of tackling false data on the Internet. To this end, we focus on the (a) detection of false data (Sections 2.1 and 2.2) coupled with combating its spread through identifying mis-influencers and installing corrective measures (Section 2.3).

2.1 Knowledge Management Module

The foremost step toward combating misinformation is extracting structured information from the collection of unstructured and noisy textual data provided by disparate data sources such as news agencies and social media. This process, termed as *information extraction* [14], is an area of growing relevance in the present-day information overload and forms the basis of our proposed system. Broadly, information extraction approaches can be categorized as based on (a) knowledge engineering techniques that leverage expert intervention in the form of rules, examples and domain knowledge, and (b) machine learning techniques that learn concept-specific mapping from text and generate rules from training data.

We envision the knowledge management module to take a hybrid approach learning from training data and external resources, such as general-purpose knowledge bases, master data and human input, to extract data items and their relationships. Several data management problems form pivotal building blocks of this module, e.g., entity resolution [11], that determines alternate representations of the same data item or claim; extracting data relationships to learn how data items (and claims) are related [24, 35], and establishing source dependencies [10, 28, 38] to distinguish originators and copiers; and provenance [7], to determine the origin and information on history of the life cycle of data.

Knowledge and Provenance. Although source dependencies have been extensively studied in data over the Internet [10, 28, 38], colluding sources may behave in unexpected ways: providing false information on related claims, evolving collusive relationships over time (both in terms of amount and direction of collusion) and colluding over emerging claims (in the presence of few sources). A comprehensive knowledge of the relationships between different stakeholders (entities, claims and sources) will prove instrumental in explaining the plausability of claims in the truth discovery module (Section 2.2). Provenance information and metadata associated with the claims, such as its context and fragments that have been used as is or have been altered, will be important in designing algorithms that assess sources and claims in a principled manner.

Classifying Claims. Drawing upon the breadth of research on detecting false information, we envision labeling claims as being facts, opinions, rumors, hoaxes, urban legends, vandalisms, joke, advertisement etc., their sentiment and establishing their temporal existence (happened in the past or is a prediction) – a natural language processing task made feasible with the help of domain experts, crowdsourcing platforms and knowledge bases. We discuss how these classifications help build the reputation of sources (in Section 2.2) and curb the rise of false data (in Section 2.3).

2.2 Truth Discovery Module

In recent years, a number of data fusion models have been proposed to automatically distinguish correct from incorrect information *structured* data conflicts. These fusion models consider source

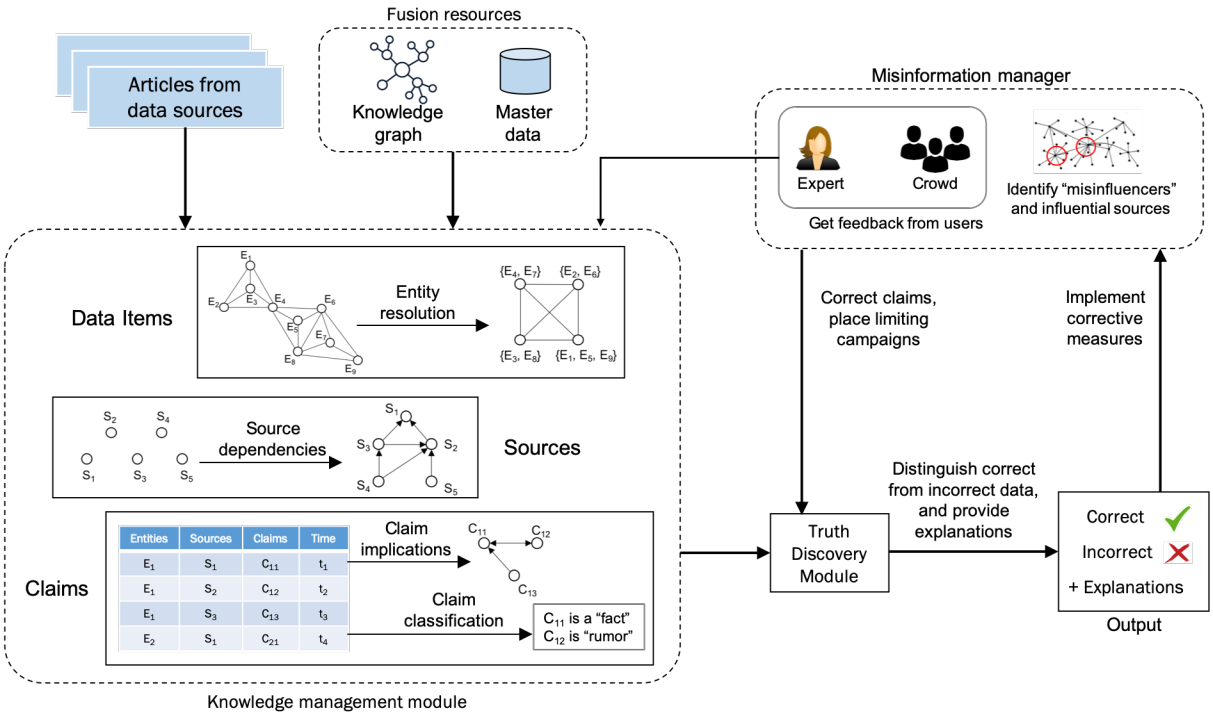

Figure 1: Figure depicts envisioned architecture of the AuthIntegrate system.

characteristics to play a pivotal role in estimating the correctness of claims – an approach that is in sharp contrast with most false data detection mechanisms that solely exercise natural language techniques to identify correct information. While the idea of *all important* sources has made tremendous advances in resolving conflicts, data fusion is designed on the single premise that sources are primarily *benevolent* – errors and inconsistencies creep into the data inadvertently because sources provide incomplete data, fail to update new values, lazily copy from other sources, or simply make errors and provide inaccurate data (*'none of the sources make errors on purpose'* [20]). Current times (of abundant false news), however, bear testimony to the fact that the "honest sources" assumption no longer holds true. Adversarial settings such as these are breeding grounds for false and biased data that have the potential to misguide fusion systems toward incorrect conclusions.

Modeling distrustful scenarios. We argue that it is imperative to revisit the problem of data fusion for identifying malicious sources and functioning successfully even in pessimistic settings. Researchers have only now begun to examine the economics of false data on the Internet [1, 44], and have identified the utility-maximizing *intent* of data providers as being one of the prime factors behind the generation and dissemination of false data.

The presence of correlated data sources has been studied before [28, 38]; however, the problem of adversarial and colluding data sources in fusion has not been addressed yet. Sources that provide falsified information may not behave consistent over time: driven by their interest, data sources may furnish data sporadically or continuously in large amounts. Modeling adversarial data sources and collusive relationships between sources can benefit a variety of applications such as information retrieval, news consolidation and web search, that have gained importance in the efforts to aggregate data from a multitude of sources. For example, malicious (or, biased) data sources may deliberately boost irrelevant documents during information retrieval tasks; knowledge of collusive relationships among sources can help retrieval systems make informed decisions on document relevance.

Source characterization. Data fusion models characterize data sources in terms of performance metrics, such as accuracy, precision, recall, false positive rate etc., that depend on the number of correct and incorrect claims provided by sources. Counting-based approaches fail to address the quality of sources where claims may span lengthy texts. We envision the data fusion module to re-invent source characteristics: (i) based on the kind of information a source provides (e.g., hoax, opinion, fact, prediction), and (ii) that effectively represent sources consistently through different tones and stances. Characterizing sources in this manner helps refine their reputation e.g., speculative facts and opinions make sources less credible than correct facts; in fact, speculations and opinion pieces may damage the credibility of a data source.

Integrating data relationships. Data sources often provide claims that may be related to each other through various entity-relationships: for example, claims *Hawaii* and *Honolulu* for the birthplace of *Barack Obama* can be abstracted to different granularities.

State-of-the-art fusion systems largely consider claims for data items to be unrelated to each other. [3] proposed using ontologies for the problem of truth discovery; however, their approach does not capture the full gamut of entity-relationships and does not guarantee a high recall. The approach in [31] proposes an *arbitrary directed graph* formalism to represent entity-relationships, such as subsumption, overlap, equivalence and mutual exclusion, among claims of data items and devise algorithms that integrate the data relationships with fusion models and improve their effectiveness.

2.3 Misinformation Manager Module

The objective of this module is two-pronged: one, identifying influential data sources that have the potential of inflicting maximum damage, and two, implementing corrective measures to minimize the damage. Toward this goal, we envision strategies to efficiently utilize human input and to limit the spread of false information.

Users as first-class citizens. Although automated fact-checking systems [42] enable deconstructing vague and countering questionable claims, the undeniable success of fact-checking websites (e.g., Snopes, PolitiFact) has made it clear that verification by experts is a stepping stone in the battle to counter false data. Corrective information published from an authoritative resource has the potential to diffuse enormously and prevent the rapid increase in false data [32, 40]. However, incorporating user input is challenging because there are a large number of claims and few experts with limited budgets to process the claims. This approach of vetting by experts is particularly important in the face of limited information on emerging claims. We intend to build upon strategies proposed in [32] to judiciously leverage user feedback by determining the most beneficial claims to be validated; these strategies can also be utilized for labeling different forms of claims (in Section 2.1) where the challenge is to prioritize labeling tasks for annotators.

Limiting the spread of false information. False data has the potential to be considered true by a large fraction of consumers; it is, therefore, of utmost importance to identify *misinfluencers* and prevent them from spreading misinformation. [2, 5] demonstrated that by placing limiting campaigns at influential nodes, it is possible to minimize the number of individuals that believe in a particular piece of misinformation and prevent the growth of false data. We propose to develop this idea of identifying misinfluencers to Bayesian networks of data items and sources, which is different from the influence maximization problem that examines the flow of a single propaganda (false data usually spans more than just one claim in a specific community (false data may extend to a multitude of communities such as social media, blogs and the Web).

3 CONCLUSION

We presented the design of AuthIntegrate, an end-to-end system aimed at combating false data on the Internet and identified key components of such a system. We envision AuthIntegrate as a system that (a) ingests conflicting data from multiple data sources, leverages authoritative resources of information, such as master data, knowledge bases, and domain experts, to maintain knowledge and provenance related to data items, claims and sources, (b) presents the problem of false data detection as the truth discovery of structured data that utilizes the extracted knowledge to distinguish correct from incorrect information, and (c) engages user feedback and corrective measures to recognize influential data providers and limit the dissemination of misinformation. Our proposed system has strong foundations housed in the principles of databases and data mining, and exploits research advances in the areas of information extraction, data fusion, adversarial machine learning and influence propagation.

REFERENCES

[1] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 2017.

[2] M. Amoruso, D. Anello, V. Auletta, and D. Ferraioli. Contrasting the spread of misinformation in online social networks. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 2017.

[3] V. Beretta, S. Harispe, S. Ranwez, and I. Mougenot. How can ontologies give you clue for truth-discovery? an exploratory study. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, 2016.

[4] J. Bleiholder and F. Naumann. Data fusion. *ACM Computing Surveys*, 2009.

[5] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web*, 2011.

[6] S. Bykau, F. Korn, D. Srivastava, and Y. Velegrakis. Fine-grained controversy detection in wikipedia. In *Proceedings of 2015 IEEE 31st International Conference on Data Engineering*, 2015.

[7] J. Cheney, L. Chiticariu, and W.-C. Tan. Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, 2009.

[8] N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma. Adversarial classification. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004.

[9] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, 2003.

[10] X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: The role of source dependence. *Proceedings of the VLDB Endowment*, 2009.

[11] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios. Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2007.

[12] D. Firmani, B. Saha, and D. Srivastava. Online entity resolution using an oracle. *Proceedings of the VLDB Endowment*, 2016.

[13] M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin. Crowddb: Answering queries with crowdsourcing. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, 2011.

[14] R. Grishman. Information extraction: Techniques and challenges. In *Intl. Summer School on Information Extraction*, 1997.

[15] N. Hassan, C. Li, and M. Tremayne. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th*

ACM International on Conference on Information and Knowledge Management, 2015.

[16] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar. Adversarial machine learning. In *roceedings of the 4th ACM workshop on Security and artificial intelligence*, 2011.

[17] F. Jin, E. R. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, pages 8:1–8:9, 2013.

[18] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: conflict and coordination in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–462, 2007.

[19] S. Kumar, R. West, and J. Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th International Conference on World Wide Web*, pages 591–602, 2016.

[20] Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han. A confidence-aware approach for truth discovery on long-tail data. *Proceedings of the VLDB Endowment*, pages 425–436, 2014.

[21] Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han. A survey on truth discovery. *SIGKDD Explorations Newsletter*, 17(2):1–16, 2016.

[22] Y. Li, Q. Li, J. Gao, L. Su, B. Zhao, W. Fan, and J. Han. On the discovery of evolving truth. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 675–684, 2015.

[23] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1751–1754. ACM, 2015.

[24] M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics*, 2009.

[25] S. Mukherjee and G. Weikum. Leveraging joint interactions for credibility analysis in news communities. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 353–362, 2015.

[26] N. Nakashole and T. M. Mitchell. Language-aware truth assessment of fact candidates. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014.

[27] J. Pasternack and D. Roth. Latent credibility analysis. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1009–1020, 2013.

[28] R. Pochampally, A. Das Sarma, X. L. Dong, A. Meliou, and D. Srivastava. Fusing data with correlations. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, pages 433–444, 2014.

[29] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum. Credibility assessment of textual claims on the web. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2173–2178, 2016.

[30] M. Potthast, B. Stein, and R. Gerling. Automatic vandalism detection in wikipedia. In *Proceedings of the IR research, 30th European conference on Advances in information retrieval*, pages 663–668. Springer-Verlag, 2008.

[31] R. Pradhan, W. G. Aref, and S. Prabhakar. Leveraging data relationships to resolve conflicts from disparate data sources. In *Database and Expert Systems Applications – 29th International Conference (DEXA)*, 2018.

[32] R. Pradhan, S. Bykau, and S. Prabhakar. Staging user feedback toward rapid conflict resolution in data fusion. In *SIGMOD*, pages 603–618, 2017.

[33] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011.

[34] T. Rekatsinas, A. Deshpande, X. L. Dong, L. Getoor, and D. Srivastava. Sourcesight: Enabling effective source selection. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2157–2160. ACM, 2016.

[35] S. Riedel, L. Yao, and A. McCallum. Modeling relations and their mentions without labeled text. In *Proceedings of the 2010 European conference on Machine learning and knowledge discovery in database*, 2010.

[36] M. Samadi, P. Talukdar, M. Veloso, and M. Blum. Claimeval: Integrated and flexible framework for claim evaluation using credibility of sources. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 222–228, 2016.

[37] J. Sampson, F. Morstatter, L. Wu, and H. Liu. Leveraging the implicit structure within social media for emergent rumor detection. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2377–2382. ACM, 2016.

[38] A. D. Sarma, X. L. Dong, and A. Halevy. Data integration with dependent sources. In *Proceedings of the 14th International Conference on Extending Database Technology*, 2011.

[39] V. S. Sheng, F. Provost, and P. G. Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 614–622, 2008.

[40] M. Takayasu, K. Sato, Y. Sano, K. Yamada, W. Miura, and H. Takayasu. Rumor diffusion and convergence during the 3.11 earthquake: a twitter case study. *Plos one, 10:e0121443-e0121443*, 2015.

[41] S. Wu, Q. Liu, Y. Liu, L. Wang, and T. Tan. Information credibility evaluation on social media. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 4403–4404. AAAI Press, 2016.

[42] Y. Wu, P. K. Agarwal, C. Li, J. Yang, and C. Yu. Toward computational fact-checking. *Proceedings of the VLDB Endowment*, 7(7):589–600, 2014.

[43] H. Zhang, M. A. Alim, X. Li, M. T. Thai, and H. T. Nguyen. Misinformation in online social networks: Detect them all with a limited budget. *Transactions on Information Systems*, 34(3):18:1–18:24, 2016.

[44] Y. M. Zhukov and M. A. Baum. Reporting bias and information warfare. *Intl. Studies Association Annual Convention*, 2016.