

8-10-2020

## **A Structural Holistic Believability Metric for Influence in Linked Data**

Carolyn C. Matheus  
*Marist College*, [Carolyn.Matheus@marist.edu](mailto:Carolyn.Matheus@marist.edu)

Alan Labouseur  
*Marist College*, [alan.labouseur@marist.edu](mailto:alan.labouseur@marist.edu)

Follow this and additional works at: [https://aisel.aisnet.org/treos\\_amcis2020](https://aisel.aisnet.org/treos_amcis2020)

---

### **Recommended Citation**

Matheus, Carolyn C. and Labouseur, Alan, "A Structural Holistic Believability Metric for Influence in Linked Data" (2020). *AMCIS 2020 TREOs*. 35.  
[https://aisel.aisnet.org/treos\\_amcis2020/35](https://aisel.aisnet.org/treos_amcis2020/35)

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2020 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# A Structural Holistic Believability Metric for Influence in Linked Data

*TREO Talk Paper*

**Carolyn C. Matheus**  
Marist College  
Carolyn.Matheus@Marist.edu

**Alan G. Labouseur**  
Marist College  
Alan.Labouseur@Marist.edu

## Abstract

We live in an evolving world of rapidly accumulating and highly granular big data, constantly growing in volume and increasing in velocity. This TREO paper focuses on the data quality dimension *believability*, specifically as it applies to evaluating influence in linked data. Believability is highly relevant in social media and other forms of linked data. Referring to the extent to which data are regarded as credible, believability is closely related to the fourth “V” of big data, *veracity*, which describes accuracy and trustworthiness (Shankaranarayanan and Blake, 2017), and contributes to the fifth “V” *value*. Prior work has focused on *provenance*-based (Prat and Madnick, 2008), *context*-based (Serra and Marotta, 2016), and *reputation*-based (Cai and Zhu, 2016) approaches to believability. In each case, their efficacy is situational, depending on the specific data under analysis. We propose a **structural**-based approach, exploiting the fact that regardless of its dynamic content and meta-content (e.g., provenance, context, reputation), the structure of linked data remains the same (if not, it ceases to be linked data). We illustrate our structural approach to believability by analyzing influence in linked data using a network from Yelp, an online linked directory service and crowd-sourced review forum largely about food.

There are many ways to understand and measure influence in linked data. Consider one particular person in a graph containing people. We might be interested in determining their influence by looking at the number of their immediate friends (which can be structurally calculated by determining their vertex degree), along with how well connected and relevant those friends are (which can be structurally calculated through clustering coefficient and PageRank, respectively). Each of those metrics provide insight into influence (or lack thereof) within linked data, but none present the whole picture because there are ways to artificially inflate or otherwise “game” those individual measures. However, a **holistic** approach for evaluating the believability of influence measures in linked data seems attainable by combining individual measures. We will present phase one of this research-in-progress: combining graph analytics to develop a *Structural Holistic Believability Metric* for influence in linked data.

Intuitively, incorporating multiple believability measures seems like it could increase data quality by improving credibility and value judgements of influence in linked data. But how do we **know** that? How can we test and validate it? Generating a sense of believability is difficult, as it is an inherently human concept. Although this human ability seems like it might be helpful for validating believability metrics, it is not, because the volume, velocity, and variety inherent in big data is too great for humans to process. We will encourage discussion of this important topic, which addresses phase two of this research-in-progress.

## References

- Cai, Y., & Zhu, D. (2016). “Reputation in an open source software community: Antecedents and impacts” *Decision Support Systems*, vol. 91, pp. 103-112.
- Prat, N., & Madnick, S. (2008). “Measuring data believability: A provenance approach” *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, pp. 393.
- Serra, F., & Marotta, A. (2016). “Data Warehouse Quality Assessment Using Contexts” *Proceedings of the 17th International Conference on Web Information Systems Engineering*, pp. 436-448.
- Shankaranarayanan, G., & Blake, R. (2017). “From Content to Context: The Evolution and Growth of Data Quality Research” *Journal of Data and Information Quality* (8:2), pp. 9:1-9:28.