

# Social Set Visualizer (SoSeVi) II: Interactive Social Set Analysis of Big Data

Benjamin Flesch<sup>1</sup>, Ravi Vatraapu<sup>1,2</sup>, Raghava Rao Mukkamala<sup>1</sup>, and Abid Hussain<sup>1</sup>

<sup>1</sup> Centre for Business Data Analytics (<http://bda.cbs.dk>),  
Department of IT Management, Copenhagen Business School, Denmark

<sup>2</sup> Westerdals Oslo School of Arts, Comm & Tech, Norway  
{bf.itm, rv.itm, rrm.itm, ah.itm}@cbs.dk

**Abstract** Current state-of-the-art in big social data analytics is largely limited to graph theoretical approaches such as social network analysis (SNA) informed by the social philosophical approach of relational sociology. This paper proposes and illustrates an alternate holistic approach to big social data analytics, social set analysis (SSA), which is based on the sociology of associations, mathematics of set theory, and advanced visual analytics of event studies. We first presented the SSA approach to a wider audience at IEEE Big Data 2015 [6], IEEE EDOCW 2015 [7], and IEEE EDOCW 2016 [8]. Since then we worked on improving SoSeVi in order to further demonstrate its usefulness in large-scale visual analytics tasks of individual and collective behavior of actors in social networks. The current iteration of the Social Set Visualizer (SoSeVi) builds upon some of the concepts laid out by the *UpSet* project [14] and aims to further improve the capabilities of researchers and practitioners in big social data analytics alike. We then illustrate our new approach by reporting on the design, development, and evaluation results of a state-of-the-art visual analytics dashboard, the Social Set Visualizer (SoSeVi). The development of the dashboard involved cutting-edge open source visual analytics libraries (D3.js) and creation of new visualizations such as visualizations of actor mobility across time and space, conversational comets, and more. Evaluation of the dashboard consisted of technical testing, usability testing, and domain-specific testing with CSR students and yielded positive results. In conclusion, we discuss the new analytical approach of social set analysis and conclude with a discussion of the benefits of set theoretical approaches based on the social philosophical approach of associational sociology.

**Keywords:** Big Social Data, Social Set Analysis, Computational Set Analysis, Big Data Visual Analytics, Event Studies

## 1 Introduction

This paper introduces a new research approach situated in the domains of Data Science [4,15,22] and Computational Social Science [13] with practical applications to Big Social Data Analytics in organizations [26,24,23]. It addresses one of the important theoretical and methodological limitations in the emerging paradigm of Big Data Analytics of social media data [25]. In particular, it address the major limitation in existing research on Big Social Data analytics that computational methods, formal models and software tools are largely limited to graph theoretical approaches [9] (such as SNA [2]), and are informed by the social philosophical approach of relational sociology [5]. There are no other unified modeling approaches to social data that integrate the conceptual, formal, software, analytical and empirical realms [20]. This results in a research problem when analyzing Big Social Data from platforms like Facebook and Twitter, as such data consists of not only dyadic relations but also individual associations [21]. For Big Social Data analytics of Facebook or Twitter data, the fundamental assumption of SNA, that social reality is constituted by dyadic relations and interactions that are determined by structural positions of individuals in social networks [19], is neither necessary nor sufficient [27].

For example, consider a Facebook post made on the official Facebook wall of Lionel Messi, the soccer prodigy who plays for FC Barcelona and Argentina's national football team. Each official post by Messi to his Facebook page typically receives more than 100,000 likes, 25,000 comments and 18,000 shares. Such association-based and content-driven social media interactions involving large number of social actors are unlike the other social interactions such as face-to-face, email, phone and instant messaging in the sense that what binds the interacting social actors together in the first instance is not so much the relational ties (strong vs. weak ties) but associations ranging from the player himself, the teams that he plays for, to the cultural, ethnic, national and linguistic attributes. Modeling such Facebook interactions using affiliation networks creates the problem of an extremely low number of nodes with an extremely high number of nodes as spokes. Further, such SNA assumes the central social psychological concept of "homophily" that social actors with similar interests (that is, associations) prefer to interact with each other. To overcome this limitation and address the research problem, this paper proposes an alternative holistic approach to Big Social Data analytics that is based on the sociology of associations as well as the mathematics of set theory and offers to develop fundamentally new methods and tools for Big Social Data analytics, Social Set Analysis (SSA). Our overarching research question is stated as: *How, and in what way, can methods and tools for Social Set Analysis, derived from the alternative holistic approach to Big Social Data analytics based on the sociology of associations and the mathematics of set theory, result in meaningful facts, actionable insights and valuable outcomes?*

The rest of the paper is organized as follows. First, we present a philosophical template for holistic approaches to computational social sciences, compare and contrast the dominant approach of social network analysis with the proposed novel approach of social set analysis and discuss the benefits of set theoretical

approaches based on the social philosophical approach of associational sociology in section 2. Then, we illustrate our new analytical approach by reporting on the design and development of our state-of-the-art visual analytics dashboard, the Social Set Visualizer (SoSeVi), that builds on and extends the UpSet visualizations of set intersections.

## 2 Theoretical Framework

The theoretical concepts behind our proposed approach of Social Set Analysis are discussed here.

### 2.1 Set Theoretical Big Social Data Analytics

Social Set Analysis (SSA) as employed in this paper is concerned with the mobility of social actors across time and space. For mobility across time, we conduct SSA of big social data from the Facebook walls of eleven companies from the same industry with an analytical focus on the set of actors that interacted with the company before, during and after the real-world events, and set theoretical intersections of the three time periods. Similarly, for mobility across space, we conduct set inclusions and exclusion of actors who interacted with different Facebook walls. This will allow us to uncover not only the interactional dynamics over time and space but also identify actor sets that correspond to marketing segmentations such as brand loyalists, brand advocates, brand critics and social activists.

### 2.2 Event Study Methodology

Event studies is a finance methodology to assess an impact on corporate wealth (e.g. stock prices) caused by events such as restructuring of companies, leadership change, mergers & acquisitions [3,17,16]. It has been a powerful tool since the late 1960s to assess financial impact of changes in corporate policies and used exclusively in the area of investments and accounting to examine stock price performance and the dissemination of new information [1].

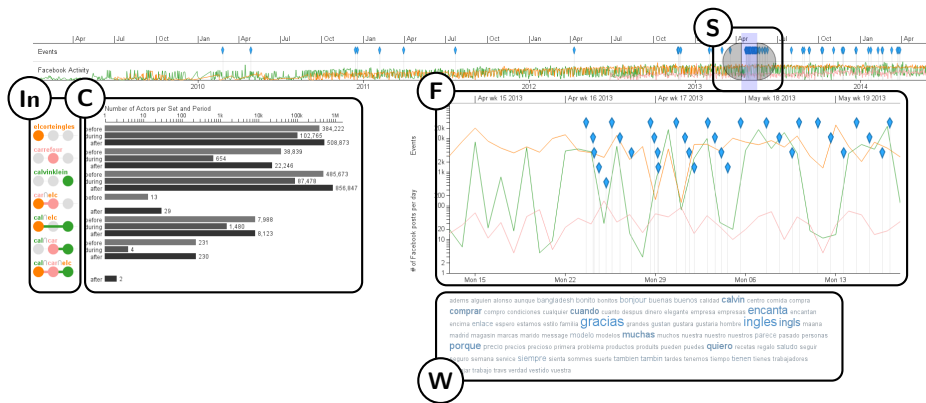
While there is no unique structure for event study methodology, at a higher level of abstraction, it contains identifying three important time periods or windows. First, defining an event of interest and identifying the period over which it is active (event window), the second involves identifying the estimation period for the event (pre-event or estimation window) and the final one being identifying the post-event window [16]. In social set analysis of a real-world event, we have applied event study methodology to identify the three important time periods of user interactions on social media platforms: *before* (pre-event window), *during* (event window) and *after* (post-event window).

### 3 Related Work

The improved version of the Social Set Visualizer (SoSeVi) has been influenced by the previous work of the *UpSet* project [14], and the visualizations of set intersection based on innovative approaches to combination matrices. In contrast to UpSet, SoSeVi uses server-side calculations on its underlying big social data corpus, and therefore is able to handle much larger volumes of data with 100,000s to millions of actors moving between set intersections. Both projects strive to provide a real-time visual analytics tool.

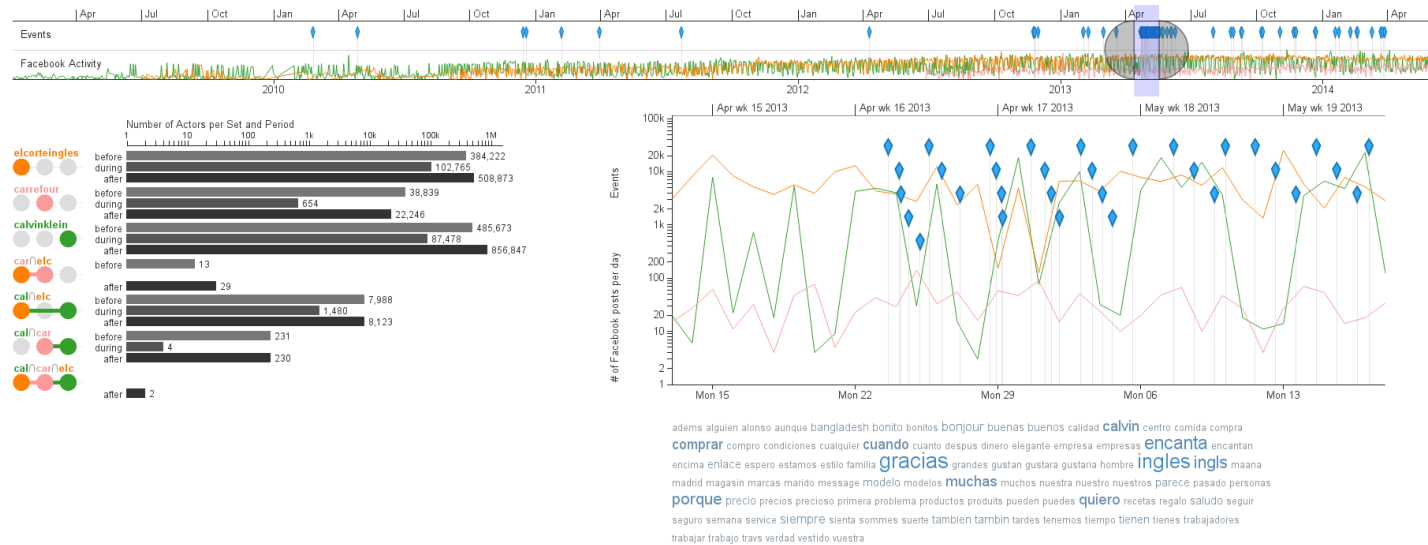
The previous version of SoSeVi used Venn diagrams to showcase actor migration between time periods and different Facebook walls, with focus shifting towards the use of Euler visualizations as discussed in [18] for version 2 of SoSeVi.

### 4 The Visual Analytics Tool



**Figure 1.** Social Set Visualizer showing 8M Facebook entries from the *Carrefour*, *CalvinKlein* and *El Corte Ingles* Facebook pages adapted to the computational set analysis approach: [F] main activity chart zoomed in on the user-selected time period using the [S] selection tool. Underneath the dynamically calculated word cloud [W] is located. On the left side we see all set intersections [In] displayed as a combination matrix, where [C] displays the cardinality of each individual set intersection over the *Before*, *During* and *After* periods. On clicking or hovering, we see a visualization of period-over-period actor migration between set intersections.

We showcase version II of the Social Set Visualizer (SoSeVi) tool which is adapted to the challenges of visualizing user-selected and dynamically calculated large-scale set intersections depicting aggregated actor behavior in social networks such as Facebook. Figure 1 illustrates the major design shift from the SoSeVi version which was presented in [6].



**Figure 2.** Social Set Visualizer II with interactive selection of time period for analysis and calculation of intersections for *Before*, *During* and *After* periods over the dataset fetched from Facebook. Set intersections are dynamically encoded through a combination matrix. The migration of actors between time periods and set intersections is showcased through cardinality changes on the left hand side. User-provided event markers signify important real-world events with relevance for the underlying research domain of the investigating analyst.

Figure 1 depicts the *DashboardView* which is the central interface of the web application. It contains all visualizations and is initially shown to the user. It consists of both small and large overall social media activity visualizations [F]. The researcher can use the time period selection tool [S] to navigate the data, and to toggle data sets from different Facebook walls depending on the analysis tasks at hand. Based on the user-selected time period, which we label as the *During* period, the tool is able to deduct *Before* and *After* time periods by looking at the beginning (earliest event) and end (most recent event) of the underlying data. An alphabetical word cloud [W] underneath the main activity chart [F] illustrates the most important conversation topics in the during period. This showcases the pluggable architecture of SoSeVi, which facilitates diverse real-time content analysis tasks on the underlying data, all based on a user-selected time frame for analysis. To the left of the main activity visualization [F], set intersections [In] are dynamically visualized based on the user selection of the time period. Set intersections are encoded in a combination matrix. Each data source uses a distinct color.

For each set intersection, we render up to three bar graphs in [C]. One bar each is drawn for every single *Before*, *During* and *After* period, when the underlying set has a cardinality of at least one actor. The bars are horizontally stacked, with the topmost bar signifying the *Before* period, the center bar *During*, and the lowest *After*. More information on the visualization of actor migration through time and space sets is shown in figure 3.

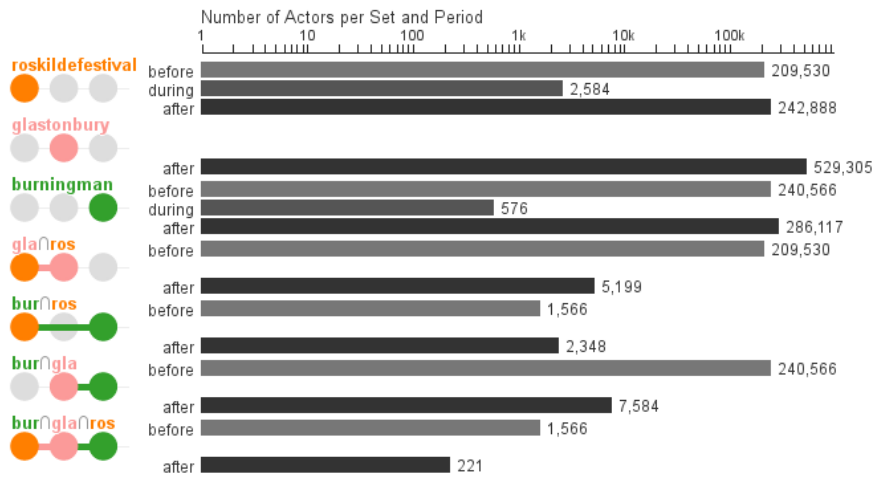
When clicking on or hovering over the set cardinality visualizations in [C], a visualization of actor migration between periods and set intersections is displayed as shown in figure 4. SoSeVi calculates all possible set intersection for each set with all sets of the following period (migrations from *Before* to *During*, and migrations from *During* to *After*) based on the user-selected time period using the selection tool [S]. Cardinality numbers illustrate the actual migration volume, whereas the right-hand side bar labels indicate the destination of each migration.

Augmenting the extensive visual analytics features of SoSeVi, *RawdataView* presents a detailed search interface for the underlying Facebook activity data. It is accessible to the user through various means by interacting with the visualizations of the *DashboardView*. *ActorsView* presents a dedicated interface for analysis tasks related to Actor Mobility across time and space of companies' Facebook walls. The visualizations of actor mobility in *DashboardView* refer to *ActorsView* in order to provide the user with further details when requested. *ActorsView* presents a handy set of tools for analysis of actor mobility and cross-postings between different time periods and Facebook walls.

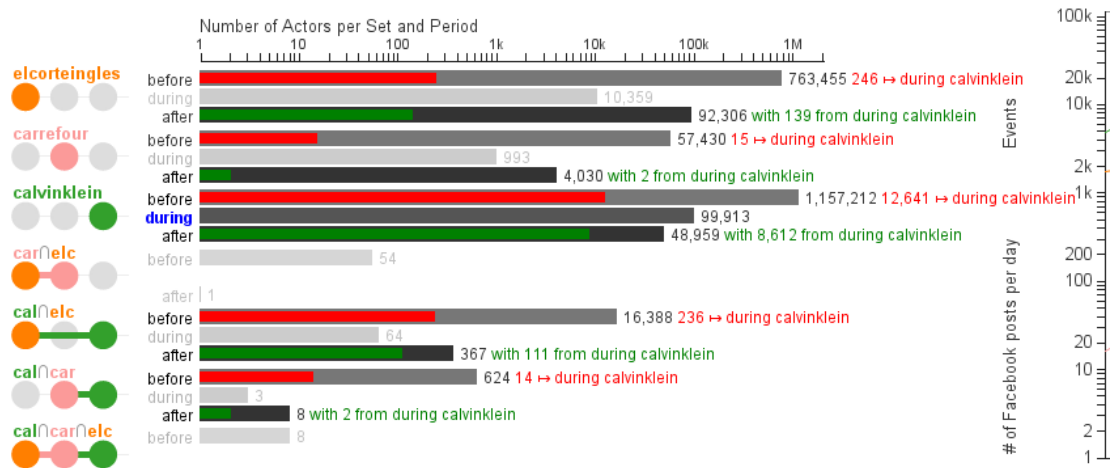
The Social Set Visualizer can be accessed at <http://bigdata:bigdata@5.9.5.20/> with user name and password *bigdata*. It has been tested by using Webkit and Gecko-based web browsers.

#### 4.1 Data Acquisition

Facebook data was collected through the Social Data Analytics Tool (SODATO) [11,10,12]. SODATO-provided Facebook activity datasets are generated as independent files



**Figure 3.** Visualization of set intersections and set intersection cardinality *Before*, *During*, and *After* the user-selected time period, illustrating the distribution of social media actors over time and space.



**Figure 4.** Visualization of actor migration concerning the *Calvin Klein* Facebook wall *During* the user-selected time period, showcasing strength and destinations of migrations. Migrations towards the Calvin Klein Facebook wall from the *Before* to the *During* Period are displayed in red color. Migrations originating from the Calvin Klein *During* period are received by to other intersections' *After* periods, and displayed in green color. The cardinality of each migration movement concerning the user-selected period and set is clearly visible in red (incoming migration) or green color (outgoing migration).

for each company's Facebook wall, and were combined into one for using them as a whole data set that can be filtered or expanded on demand.

## 5 Software Development

The SoSeVi dashboard is implemented as a client-side web application, and uses the D3.js Javascript SVG visualization framework. D3.js constitutes a lightweight and very extendable Javascript visualization framework which can display visualizations for a multitude of browser-based clients. The flexibility provided by D3.js enables the creation of new kinds of interactive visualizations which are able to run on any device with decent processing resources including Windows, MacOS and Linux based systems with screen sizes up to 4K, which gives SoSeVi the flexibility needed for various purposes in different research areas.

After thorough (re-)evaluation of Apache Spark and other NoSQL-based storage solutions, the decision to use PostgreSQL for data storage was not overthrown due to lack of empirically measured benefits in execution time with our test data. In version 2 of SoSeVi, all set intersection calculations have been outsourced from PostgreSQL to Redis. A dedicated Redis instance now performs memory-intensive set intersection calculations with a significantly better execution speed and pipes the calculation results back to the user-facing dashboard in real time.

## 6 Conclusion

Version 2 of the Social Set Visualizer (SoSeVi) project presented in this paper provides a significantly better interface for Social Set Analysis (SSA) tasks of big social data originating from Facebook. We showcase our interactive tool for large-scale real-time set intersection calculations to the research community. SoSeVi 2 depicts the first visual analytics tool to visualize migration flows between set intersections in big social data.

## 7 Future Work

We strive to add more customization features in order to provide the user with more viable investigation strategies, such as sorting and filtering of sets. This was also demonstrated in the *UpSet* project, but real-time implementation of those features was out of scope for version 2 of SoSeVi. Add extension points to perform statistical calculations over the set intersections and improve the overall measurement of migration flows. The research tool will be extended for non-Facebook data social media data.

## 8 Acknowledgments

The authors were partially supported by the project Big Social Data Analytics: Branding Algorithms, Predictive Models, and Dashboards funded by Industriens



Fond (The Danish Industry Foundation). Any opinions, findings, interpretations, conclusions or recommendations expressed in this paper are those of its authors and do not represent the views of the Industriens Fond (The Danish Industry Foundation).

## References

1. Binder, J.: The event study methodology since 1969. *Review of quantitative Finance and Accounting* 11(2), 111–137 (1998)
2. Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G.: Network analysis in the social sciences. *Science* 323(5916), 892–895 (2009)
3. Bromiley, P., Govekar, M., Marcus, A.: On using event-study methodology in strategic management research. *Technovation* 8(1), 25–42 (1988)
4. Cleveland, W.S.: Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review* 69(1), 21–26 (2001), <http://dx.doi.org/10.1111/j.1751-5823.2001.tb00477.x>
5. Emirbayer, M.: Manifesto for a relational sociology. *The American Journal of Sociology* 103(2), 281–317 (1997)
6. Flesch, B., Vatrapsu, R., Mukkamala, R.R., Hussain, A.: Social set visualizer: A set theoretical approach to big social data analytics of real-world events. In: *Big Data (Big Data)*, 2015 IEEE International Conference on. pp. 2418–2427 (Oct 2015)
7. Flesch, B., Hussain, A., Vatrapsu, R.: Social set visualizer: Demonstration of methodology and software. In: *2015 IEEE 19th International Enterprise Distributed Object Computing Workshops and Demonstrations (EDOCW)*. pp. 148–151 (Sept 2015)
8. Flesch, B., Vatrapsu, R.: Social set visualizer (sosevi) ii: Interactive computational set analysis of big social data. In: *2016 IEEE 20th International Enterprise Distributed Object Computing Workshops and Demonstrations (EDOCW)* (in press/2016)
9. Gross, J.L., Yellen, J.: *Graph theory and its applications*. CRC press (2005)
10. Hussain, A., Vatrapsu, R.: Social data analytics tool: Design, development, and demonstrative case studies. In: *Enterprise Distributed Object Computing Conference Workshops and Demonstrations (EDOCW)*, 2014 IEEE 18th International. pp. 414–417 (Sept 2014)
11. Hussain, A., Vatrapsu, R.: Social data analytics tool (sodato). In: *DESRIST 2014. Lecture Notes in Computer Science (LNCS)*. Springer, vol. 8463, pp. 368–372 (2014)
12. Hussain, A., Vatrapsu, R., Hardt, D., Jaffari, Z.: Social data analytics tool: A demonstrative case study of methodology and software. In: *Analysing Social Media Data and Web Networks*. Palgrave Macmillan (2014)
13. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. *Science* 323(5915), 721–723 (2009)
14. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., Pfister, H.: Upset: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics (IEEE InfoVis '14)* (2014), live Demo: <http://vcg.github.io/upset>
15. Loukides, M.: *What Is Data Science?* O'Reilly Media (2012)
16. MacKinlay, A.C.: Event studies in economics and finance. *Journal of economic literature* pp. 13–39 (1997)

17. McWilliams, A., Siegel, D.: Event studies in management research: Theoretical and empirical issues. *Academy of management journal* 40(3), 626–657 (1997)
18. Micallef, L., Rodgers, P.: euler ape: Drawing area-proportional 3-venn diagrams using ellipses. *PloS one* 9(7), e101717 (2014)
19. Mizruchi, M.S.: Social network analysis: Recent achievements and current controversies. *Acta sociologica* 37(4), 329–343 (1994)
20. Mukkamala, R.R., Hussain, A., Vatrappu, R.: Towards a formal model of social data. IT University Technical Report Series TR-2013-169, IT University of Copenhagen, Denmark (November 2013)
21. Mukkamala, R.R., Hussain, A., Vatrappu, R.: Towards a set theoretical approach to big data analytics. In: 3rd International Congress on Big Data (IEEE BigData 2014) (June 2014)
22. Ohsumi, N.: From data analysis to data science. In: *Data Analysis, Classification, and Related Methods*, pp. 329–334. Springer Berlin Heidelberg (2000), [http://dx.doi.org/10.1007/978-3-642-59789-3\\_52](http://dx.doi.org/10.1007/978-3-642-59789-3_52)
23. Sponder, M.: *Social media analytics: effective tools for building, interpreting, and using metrics*. McGraw-Hill (2012)
24. Sterne, J.: *Social media metrics: How to measure and optimize your marketing investment*. John Wiley & Sons (2010)
25. Tufekci, Z.: Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *arXiv preprint arXiv:1403.7400* (2014)
26. Vatrappu, R.: Understanding social business. In: *Emerging Dimensions of Technology Management*, pp. 147–158. Springer (2013)
27. Vatrappu, R., Mukkamala, R.R., Hussain, A., Flesch, B.: Social set analysis: A set theoretical approach to big data analytics. In: *IEEE Access*, 4. pp. 2542–2571 (2016)

All links were last followed on July 14th, 2016.