Business Intelligence and Analytics: Research Directions

EE-PENG LIM, School of Information Systems, Singapore Management University
HSINCHUN CHEN, Eller College of Management, University of Arizona, Tucson
GUOQING CHEN, School of Economics and Management, Tsinghua University, China

Business intelligence and analytics (BIA) is about the development of technologies, systems, practices, and applications to analyze critical business data so as to gain new insights about business and markets. The new insights can be used for improving products and services, achieving better operational efficiency, and fostering customer relationships. In this article, we will categorize BIA research activities into three broad research directions: (a) big data analytics, (b) text analytics, and (c) network analytics. The article aims to review the state-of-the-art techniques and models and to summarize their use in BIA applications. For each research direction, we will also determine a few important questions to be addressed in future research.

Categories and Subject Descriptors: H 2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms: Algorithm, Design, Performance

Additional Key Words and Phrases: Business intelligence, business analytics

ACM Reference Format:
Lim, E.-P., Chen, H., Chen, G. 2012. Business intelligence and analytics: Research directions. *ACM Trans. Manage. Inf. Syst.* 3, 4, Article 17 (January 2013), 10 pages.
DOI:http://dx.doi.org/10.1145/2407740.2407741

1. BUSINESS INTELLIGENCE AND ANALYTICS (BIA)

1.1. Introduction

Business intelligence and analytics (BIA), a term coined in 1989, has gained much traction in the IT practitioner community and academia over the past two decades. BIA refers to: (1) the technologies, systems, practices, and applications that (2) analyze critical business data to (3) help an enterprise better understand its business and market.

Traditionally, business intelligence (BI) has been used as an umbrella term to describe concepts and methods to improve business decision making by using fact-based support systems. BI also includes the underlying architectures, tools, databases, applications, and methodologies. BI's major objectives are to enable interactive and easy access to diverse data, enable manipulation and transformation of these data, and provide business managers and analysts the ability to conduct appropriate analyses and perform actions [Turban et al. 2008; Wixom et al. 2011]. Successful BI initiatives have been reported for major industries, from healthcare and airlines, to major IT and

This work is supported by the National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.
Authors' addresses: E.-P. Lim, School of Information Systems, Singapore Management University, Singapore; email: eplim@smu.edu.sg; H. Chen, Eller College of Management, University of Arizona, Tucson; email: hchen@eller.arizona.edu; G. Chen, School of Economics and Management, Tsinghua University, Beijing, China; email: chengq@sem.tsinghua.edu.cn.

DOI:http://dx.doi.org/10.1145/2407740.2407741

telecommunication firms [Anderson-Lehman et al. 2004; Carte et al. 2005; Turban et al. 2008].

As a datacentric approach, BI heavily relies on various advanced data collection, extraction, and analysis technologies [Turban et al. 2008; Watson and Wixom 2007]. These technologies are collectively known as business analytics (BA). Data warehousing is often considered the foundation of BI. Design of data marts and tools for extraction, transformation, and load (ETL) are essential for converting and integrating enterprise-specific data. Database query, online analytical processing (OLAP), and advanced reporting tools are often adopted next, to explore important data characteristics. Business performance management (BPM) using scorecards and dashboards can be used to analyze and visualize various employee performance metrics. In addition to these well-established business analytics functions, advanced knowledge discovery using data and text mining can be adopted for association rule mining, database segmentation and clustering, anomaly detection, and predictive modeling in various information systems and human resources, accounting, finance, and marketing applications. Given that modern business intelligence has to heavily depend upon data analytics, it is timely to adopt business intelligence and analytics (BIA) as the preferred combined term.

Since about 2004, Web intelligence, Web analytics, Web 2.0, social networking, and microblogging sites have begun to usher in a new and exciting era of Business Intelligence 2.0 (BI 2.0) research [Nelson 2010]. In BI 2.0, an immense amount of company, industry, product, and consumer information can be gathered from both enterprise databases and the Web. These data are then organized and visualized through various knowledge mapping, Web portal, and multilingual retrieval techniques [Chung et al. 2005; Marshall et al. 2004]. By analyzing customer clickstream data logs, Web analytics tools such as Google Analytics provide a trail of the user's online activities and reveal the user's browsing and purchasing patterns. Web site design, product placement optimization, customer transaction analysis, and product recommendations can be easily accomplished through Web analytics.

More recently, the social media phenomena have created an abundance of user-generated contents from various online sites such as forums, product review and rating sites, Web blogs, social networking sites, media sharing sites (for photos and videos), and even virtual worlds. By amassing a large volume of timely feedback and opinions from diverse social media users and analyzing them using social media analytics, one can derive a wide range of social and business insights much needed for social policy formulation, customer relationship management, and product innovation. Many believe Web analytics and social media analytics present a unique opportunity for business researchers to treat the market as a conversation between businesses and customers instead of traditional business-to-customer marketing. Advanced information extraction, topic identification, opinion mining, and time-series analysis techniques can be applied to traditional business information and the new BI 2.0 contents for various accounting, finance, and marketing applications, such as enterprise risk assessment and management, credit rating and analysis, corporate event analysis, stock and portfolio performance prediction, viral marketing analysis, and so on.

1.2. Emerging Trends

Industry Trends. In a press release dated 2 April 2012, Gartner reported that BI is the highest priority technology item for CIOs in 2012. It also estimated that BI revenue will reach $7.8 billion in 2011, which is an increase of 16% over that in 2010 [Gartner 2012]. Through BI initiatives, businesses are gaining insights from the growing volumes of transaction, product, inventory, customer, competitor, and industry data generated by enterprise-wide applications such as enterprise resource planning (ERP), customer relationship management (CRM), supply-chain management (SCM), knowledge management, collaborative computing, Web analytics, and so on. According to a USA Today article, IBM spent $14B in BI in 24 acquisitions in five years and its BI revenue reached $9B in 2010 [Acohido 2010]. IBM expects to employ 10,000 BI software developers, 8000 BI consultants, and 200 BI mathematicians. The demand for BIA professionals in the US will also be 50 to 60 percent larger than its projected supply by 2018 [Manyika et al. 2011).

Data Trends. Over the past decade the “Big Data” era has quietly descended on many communities, from governments and e-commerce to health and sports organizations [Beyer 2011]. With the overwhelming amount of Web, social media, mobile, and sensor-generated data arriving at a terabyte and even petabyte scale [The Economists 2010], new science, discovery, and insights can be obtained from the highly detailed, contextualized, and rich contents of relevance to businesses and organizations. Enterprise database systems, search systems, advanced data, text and Web analytics are becoming important for turning data into actionable knowledge and intelligence. As the data volume is large, the analytics can only be possible if we have highly efficient algorithms and software.

Businesses have been collecting and processing traditional structured payroll, employee, supplier, and product information for years, often via relational database management systems (RDBMS). Some large corporations have also resorted to the transaction-friendly column-based DBMS and the more powerful parallel DBMS. More recently, businesses and organizations are facing a new tsunami of unstructured text contents and user log information collected from e-commerce sites and via many customer-facing social media platforms (e.g., forums, Twitter, Facebook). Text data are fast becoming a major part of enterprise data, especially for multinational corporations and e-commerce firms. Increasingly, multimedia contents such as images, photos (Flickr), and videos (YouTube) can also contain significant product- or customer-related information. Text data can be in different languages but there is so far no good translation technology that handles all of these languages. Understanding text and extracting knowledge from it remain challenging research tasks.

Platform Technology Trend. Several platform technology trends are relevant to business analytics [IBM 2011]. Among them, cloud computing and mobile computing are of critical important. Coincidentally, the former has a major impact on the design of business analytics servers while the latter changes the way business applications reach out to consumers.

A cloud computing platform is one that is built upon a large number of low cost computers to meet the needs of storing and computing big data in BIA applications. Instead of focusing on cloud computing platform and infrastructure development (such as the Google App Engine, the Amazon EC2, Microsoft Azure), there are great opportunities for cloud application development in various critical industry sectors including: government, defense, security, health, education, and entertainment. Recent advances include many applications rapidly spreading across a wide variety of other sectors such as banking, telecommunications, energy, retailing, and so on, witnessing technological and service development in large data centers and different clouds (not only public clouds but also private/enterprise ones).

Mobile computing is firmly established in the marketplace and offers a means for IT professional growth as more and more organizations build mobile business applications. Android, iOS (iPhone and iPad), and Windows 8 are the key mobile development platforms that compete with one another for users and application developers. Services built on mobile devices also provide many opportunities for application development and use, which may vary significantly depending upon the types of the devices. Examples include the browsing and e-reading functionalities of tablets and regular mobile phones, which are used by different customer groups (e.g., middle-class users vs. farmers/low-income users) and often involve different technical algorithms and features.

1.3. Objectives

BIA research has to cope with rapid changes in the industry, data, and technology landscapes. As database and data analytics companies, as well as IT consulting firms continue to introduce new BIA products and new features into their existing BIA products, there are also parallel multiple threads of data analytics research activities happening in research labs and universities. Most university research in BIA topics happens in the business schools and computer science departments.

In this article, we will categorize these research activities into three broad research directions: (a) big data analytics, (b) text analytics, and (c) network analytics. The objectives here are to review the state-of-the-art techniques and models and to give a quick summary of their use in BIA applications even though they may not have been widely adopted. For each research direction, we will also determine a few important questions to be addressed in future research. It is our hope that these research questions will help steering future work extending BIA capabilities as well as bridging the gaps between research techniques and industry solutions.

Due to the page limit, this article does not seek to cover all relevant BIA topics and case studies. Readers should refer to the cited references for more detailed information. For BIA works in the area of data mining and machine learning, one can refer to the conference proceedings of the ACM SIGKDD Conference, International Conference on Machine Learning (ICML), World Wide Web (WWW), ACM Conference on Web Search and Data Mining (WSDM), and International AAAI Conference on Weblogs and Social Media (ICWSM), as well as journals such as IEEE Transactions on Knowledge and Data Engineering (TKDE), ACM Transactions on Knowledge Discovery from Data (TKDD), ACM Transactions on Intelligent Systems and Technology (TIST), and ACM Transactions on Web. There are also a lot of great reading materials about BIA in the management science discipline. They include ACM Transactions on Management Information Systems (TMIS), Information Systems Research (ISR), Management Science, Marketing Science, and the Proceedings of the National Academy of Sciences (PNAS).

2. RESEARCH DIRECTION I: BIG DATA ANALYTICS

Data analytics using Hadoop. Inspired in part by a 2004 Google white paper about its use of the parallel MapReduce techniques, Hadoop is a Java-based software framework for distributed processing of data-intensive transformation and analytics. The top three commercial database suppliers–Oracle, IBM, and Microsoft–have all adopted Hadoop recently. The open source Apache Hadoop has also gained significant traction for business analytics, including Chukwa for data collection, HBase for distributed data storage, Hive for data summarization and ad hoc querying, and Mahout for data mining [Henschen 2011; Watson 2012].

Hadoop has been shown to be highly efficient for processing big data that are particularly structured in applications that involve computation of simple summary statistics. Hadoop however has not yet been widely used for complex data analysis that involves many record comparisons and massive data movement among servers. For BIA that involves unstructured data, new advanced text analytics, image indexing, and ad hoc one-time processing are also yet to be developed in the distributed Hadoop or MapReduce environments.

Big data can also be presented in the form of graphs with nodes representing users or product items, and edges representing social relationships, information flows, and product adoptions. To gain insights about consumer behavior in these graphs, we need to mine graphs or conduct graph-mining. To compute the diameter of big graph data, Kang et al. [2011] proposed HADI (HAdoop DIameter and radii estimator), which that runs efficiently on the Hadoop/MapReduce systems. Other than computing such simple graph statistics, very few graph mining works have been carried out using the Hadoop/MapReduce framework and these are clearly important future research topics to be investigated.

Research Questions. From the information systems research perspective, one also has seek to answer the following key questions in future big data analytics research.

(a) Given a diverse set of business analytics application requirements, how can one tell if a Hadoop/MapReduce framework should be used for a given business analytics application? To answer this question, one has to understand the strengths and limitations of the framework much better as more analytics models and algorithms are developed. Should the Hadoop/MapReduce framework be only applicable to a subset of analytics operations, can we still leverage its strengths by applying the framework partially, perhaps in data preprocessing or postprocessing?
(b) What is the cost of migrating legacy data and applications from existing servers to the Hadoop/MapReduce framework? Switching from traditional application/database servers to the Hadoop/MapReduce framework is considered a major infrastructure change as it involves moving legacy data from standard relational databases to possibly non-SQL ones, creating new indexing schemes, as well as reimplementation of existing applications. In addition, existing BIA developers and analysts also have to be trained to use this new framework. The time and financial costs associated with these changes should be well understood and accurately estimated. Unfortunately, to the best of our knowledge, there has not been any research that addresses these costs. It is therefore a challenge for businesses to embrace BIA using Hadoop/MapReduce even if the expected benefits can be determined.

3. RESEARCH DIRECTION II: TEXT ANALYTICS

From Search Engines to Enterprise Search Systems. Since its humble beginning in information retrieval (IR) systems in the 70s, search engines have evolved into a complex system that consists of fast, distributed crawling, efficient inverted indexing, inlink-based page ranking, and search logs analytics. Much of the text processing and indexing components have been deployed in text-based enterprise search and document management systems. More recent advances in this area include in-memory and real-time processing for large-scale or dynamic contents. Other efforts include semantic search either for a search engines' functionality or for an enterprise information service, which considers text analytics in meaning via semantic match technologies, as well as in relevance via semantic transfer measures [Agarwal et al. 2006; Guha et al. 2003].

From Information Extraction to Question Answering Systems. The field of natural language processing (NLP) has also advanced significantly over the past decade, leveraging the power of Big Data (for training) and statistical NLP (for building language models). In addition to the traditional text representations such as bag of words, phrases, entities, and relationships, NLP techniques have been successfully adopted for event and topic detection, machine translation, and more recently in question-answering (Q/A) systems. IBM Watson is a good example of an advanced Q/A system that adopts sophisticated analytics to understand the meaning and context of human language. Many promising Q/A system application areas have been identified, including education, health, and defense [IBM 2011].

Another related direction in information extraction, as well as in search engines, is the representation of search queries and outcomes in light of different user preferences. In addition to existing page-rank based practices [Page et al. 1999], demands for other representation schemes become more usual and meaningful. An example is the display of diversified query results for various online product reviews. Another example is representation of compact search outcomes (that are both less redundant and more information rich) for mobile Web queries and advertisements.

From Sentiment Analysis to Opinion Mining. Opinion mining, a subdiscipline within data mining and computational linguistics, refers to the computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated content. Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online text. The advent of Web 2.0 and social media contents has stirred much excitement and created abundant opportunities for understanding the opinions of the general public and consumers towards social events, political movements, company strategies, marketing campaigns, and product preferences [Abbasi and Chen 2008; Chen and Zimbra 2010].

Research Questions. As search engine, information extraction and sentiment analysis technologies become more mature, future text analytics research has to seek answers to the following more challenging research questions.

(a) How can one perform text analytics in noisy unstructured data? Most text analytics research has been carried out on well written text documents including news, company, and government reports. In Web and social media, the user generated text often contains grammar and spelling errors, emoticons, mixed languages, abbreviations, and other idiosyncrasies. The standard text analytics solutions therefore do not work perfectly on these data. The short message length imposed by popular microblogging sites (e.g., Twitter) further limits the amount of context available to understand the text content [Lin et al. 2011]. To address these challenging text analytics tasks, we need more out-of-the-box research approaches. For example, Duolingo is a new crowdsourcing site that recruits a large number of users to manually perform translation of text as part of their efforts to learn new languages [Savage 2012].[1]
(b) How can text analytics be performed for stream data? Stream data are continuously generated by online sensors or applications, so as to be received and processed by BIA applications in real time. While there have been many data analytics techniques developed for structured data streams, to perform quick numerical computations, research works on unstructured data streams are very scarce [Gaber et al. 2005]. With the increasing amount of text data gathered from web and social media, the future of data stream research will have to focus on processing text streams extracting and summarizing them for easy consumption as well as detecting their underlying events and topics.

4. RESEARCH DIRECTION III: NETWORK ANALYTICS

Traditionally, customer and transactional data are treated as independent records in company databases. This view, however, has to change, given that records are often

[1]Duolingo.com

connected in one way or another. In social media and social network sites, users are connected with one another by a variety of links. These links may represent friendships, trusts, message interactions, shared communities, and other forms of relationships. In an online auction business, a buyer is linked to sellers who sell her some items. A bidding transaction is linked to a purchase transaction when both involve the same bidding item. The links among records turn the latter into network data from which new insights can be discovered about the customers, their consumption patterns, and the relationships among them. Leveraging these new insights, one can develop new businesses and services to meet customers' preferences. For example, in a telecommunications company, phone calls among customers may allow us to infer relationships that may be used to keep the customers from churning.

Network analytics is a nascent research area. Some of the important topics in network analytics research are the following.

Link mining. In link mining, one seeks to discover links between nodes of a network.[2] Within a network, nodes may represent customers, end users, products, and/or services. The links between nodes may represent social relationships, collaboration, email exchanges, or product adoptions. Not all of these links can be observed, as the network data may be incomplete and some of the links may only appear in the future. To recover the missing links and to predict new links, we need a variety of link mining techniques.

One can conduct link mining using only topology information [Liben-Nowell and Kleinberg 2007]. Techniques such as *common neighbors*, *Jaccard's coefficient*, *Adamic Adar measure*, and *Katz measure* are popular for predicting missing or future links. The common assumption behind these techniques is that nodes having higher topological proximity between them are more likely to have links. These topology-based techniques clearly do not work well for new nodes joining the network. Link mining accuracy can be further improved when the node and link attributes are considered. This topology and attribute approach to link mining can also be used to predict links for new nodes.

Community Detection. User communities are formed in a network for a variety of reasons. Users may link with other users due to their family relationships, friendships, or similar product adoption patterns. These relationships bring users together to form dense clusters within a network. Due to the homophily effect, we see many similar users who tend to be linked to one another, however this does not happen among dissimilar users. Detecting the user communities in networks therefore helps to uncover the common preferences and foci shared by users in the same communities. For example, in banking applications, user communities may represent different customer segments. It is thus important to design different product and service packages for the targeted customer segments.

Community detection is a very active research area. Several good survey papers about the topic have been published [Fortunato 2010; Porter et al. 2009]. By representing networks as graphs, one can apply graph partitioning to find a minimal cut to obtain dense subgraphs representing user communities. The main idea here is to divide a network into multiple subgraphs such that links between the resultant subgraphs are minimal. In more recent works, some network level goodness measure such as *modularity* is introduced to determine how well the network is partitioned. A centrality measure such as betweenness is then computed for each network link, and

[2]Link mining may be defined to cover a wider range of research problems [Getoor and Diehl 2005]. In this article, we choose a more focused definition.

links with higher centrality values are removed iteratively until we obtain a good partitioned network that yields the optimal goodness value.

Social Recommendation. Retail industries in developed countries are becoming digital as consumers spend more time and money on the Web. In addition to large B2C and C2C e-businesses such as Amazon.com, Apple iTunes, Dell.com, and E-Bay.com, many other companies have created a presence on the Web publicizing and selling their products. While Geographic constraint does not apply to businesses operating on the Web, consumers are spoiled with many retailer and product options. For a business to succeed in this e-business environment, knowing the consumers well and recommending the right products to the right consumers are of utmost importance. To this end, collaborative filtering has been used to recommend a new product to a user based on the products purchased by other users sharing similar purchase patterns with the target user [Herlocker et al. 1999]. Collaborative filtering works well when the target users have previously purchased some products. It however fails to yield good recommendations when the target user is new. This is also known as the *cold-start user problem*.

Network data about consumers fortunately helps to address the cold-start user problem. The users who are linked to the target user may serve as very good proxies of the target user. One can therefore add the social aspect to collaborative filtering by recommending to the target user those products purchased by users directly linked to the target user. This is also known as *social recommendation*, an emerging research topic. There are several new social recommendation methods that introduce social factors into existing methods. For example, Ma et al. [2008] extended matrix factorization with social link weights for rating prediction. Chua et al. [2011] determine social correlations between users and use them to improve prediction of item adoptions, using the Latent Dirichlet Allocation (LDA) model.

Research Questions. As network data become important in profiling users, communities, and social recommendations, one has to address hard research questions in network analytics.

(a) How can the existing network analytics techniques be extended to cope with multidimensional networks? The vast majority of existing network analytics techniques have been developed for simple networks consisting of one type of nodes and one type of edges. Networks, nevertheless, are often *multidimensional*, as nodes may be of different types (e.g., buyers, sellers, products, etc.) and links may also be labeled differently (e.g., buyer-purchases-item, seller-sells-item, buyer-is-friend-of-buyer, buyer-rates-seller, etc.) [Contractor 2009]. Link mining in multidimensional networks will therefore have to consider the semantics of different node and link types when links are to be predicted. Similarly, community detection and social recommendation will have to perform differently in multidimensional networks, as user nodes of different types may interact differently based on the different types of links [Sun and Han 2012].

(b) How can one distinguish social influence from selection in networks? Social recommendation and marketing are important applications using network analytics, and both assume that users near to one another in a network share similar preferences. This homophily effect can be attributed to two mechanisms, namely *self selection* and *social influence*. The former says that people sharing similar attributes tend to form social links with one another. The latter says that linked users may influence each other to adopt similar preferences or attributes. Distinguishing the two mechanisms is currently a hard research problem as shown in Shalizi and Thomas [2011]. Nevertheless, as determining the cause of homophily will allow one to design appropriate incentives or recommendation approaches to achieve business goals, there are still much interest in this research topic for the next few years.

5. CONCLUSIONS

Business intelligence/business analytics have attracted attention from enterprises, the computing industry, and the research community due to the availability of big data and new business needs. While several off-the-shelf BIA tools and systems are already available in the market, there is much room for further research and development due to the emergence of new data genres, computing paradigms, and mobile technologies. This article walks through the research trends of BIA in three areas, namely big data analytics, text analytics, and network analytics. The article also identifies important research questions that will drive future BIA research in these areas. Meanwhile, we witness vibrant research activity both in industry labs and universities. Enterprises are also spearheading initiatives to strengthen their BIA capabilities through acquisition of technologies and collaboration with researchers. Notable examples of such BIA research collaboration include the Living Analytics Research Center, jointly established by the Singapore Management University and Carnegie Mellon University to discover consumer and social insights from company datasets through experiment-driven analytics, and the Center for Business Analytics at the NYU Stern School of Business. In this collaboration model, researchers have direct access to real datasets and develop both descriptive and predictive models about user preferences and behavior. With the increasing need for business innovation through BIA and the many intellectually challenging research problems waiting to be addressed, we believe more research collaboration models between industry and academia will emerge, accelerating the pace of research discoveries and technology transfers.

REFERENCES

Abbasi, A. and Chen, H. 2008. CyberGate: A system and design framework for text analysis of computer-mediated communication. *MIS Quart. 32*, 4, 811–837.

Acohido, B. 2010. Tech-savvy put business intelligence to work. *USA Today* 11/17/10.

Agarwal, A., Chakrabarti S., and Aggarwal, S. 2006. Learning to rank networked entities. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 14–23.

Anderson-Lehman, R., Watson, H. J., Wixom, B. H., and Hoffer, J. A. 2004. Continental Airlines flies high with real-time business intelligence. *MIS Quart. Exec. 3*, 4, 163–176.

Beyer, M. 2011. Gartner says solving 'big data' challenge involves more than just managing volumes of data. http://www.gartner.com/it/page.jsp?id=1731916.

Carte, T. A., Schwarzkopf, A. B., Shaft, T. M., and Zmud, R. W. 2005. Advanced business intelligence at Cardinal Health. *MIS Quart. Exec. 4*, 4, 413–424.

Chen, H. and Zimbra, D. 2010. AI and opinion mining. *IEEE Intell. Syst. 25*, 3, 74–76.

Chua, F. C. T., Lauw, H. W., Lim E.-P. 2011. Predicting item adoption using social correlation. In *Proceedings of the SIAM International Conference on Data Mining.* 367–378.

Chung, W., Chen, H., Nunamaker Jr., J. 2005. A visual framework for knowledge discovery on the Web: An empirical study of business intelligence exploration. *J. Manage. Inform. Syst. 21*, 4, 57–84.

Contractor, N. 2009. The emergence of multidimensional networks. *J. Comput.-Mediated Commun. 14*, 3.

Dean, J. and Ghemawat, S. 2010. MapReduce: A flexible data processing tool. *Comm. ACM 53*, 1.

The Economist. 2010. Data, data everywhere. 2/10.

Fortunato, S., 2010. Community detection in graphs. *Phys. Rep. 486*, 3–5, 75–174.

Gaber, M. M., Zaslavsky, A., and Krishnaswamy, S. 2005. Mining data streams: A review. *SIGMOD Record 34*, 2.

Gartner. 2012. Gartner says worldwide business intelligence, analytics and performance management software market surpassed the $12 billion mark in 2011. http://www.gartner.com/it/page.jsp?id=1971516.

Getoor, L. and Diehl, C. P. 2005. Link mining: A survey. *ACM SIGKDD Explor. Newsl. 7*, 2.

Guha, R., McCool, R., and Miller E. 2003. Semantic search. In *Proceedings of the International World Wide Web Conference.*

Henschen, D. 2011. Why all the Hadoopla? *Information Week* 11/14/11, 19–26.

Herlocker, J. L., Konstan, J. A., Borchers, A., and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* 230–237.

IBM. 2011. The 2011 IBM tech trends report, November 15. http://ibm.com/developerworks/techntrendsreport.

Kang, U., Tsourakakis, C. E., Appel, A. P., Faloutsos, C., and Leskovec, J. 2011. HADI: Mining radii of large graphs. *ACM Trans. Knowl. Discov. Data.*

Liben-Nowell, D. and Kleinberg, J. 2007. The link-prediction problem for social networks. *J. Amer. Soc. Inform. Sci. Technol. 58*, 7, 1019–1031.

Lin, J., Snow, R., and Morgan, W. 2011. Smoothing techniques for adaptive online language models: Topic tracking in tweet streams. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Ma, H., Yang, H., Lyu, M. R., and King, I. 2008. Sorec: Social recommendation using probabilistic matrix factorization. In *Proceedings of the ACM Conference on Information and Knowledge Management*. 931–940.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. 2011. Big data: The next frontier for innovation, competition, and productivity. *The McKinsey 2011 Big Data Report.*

Marshall, B., McDonald, D., Chen, H., and Chung, W. 2004. EBizPort: Collecting and analyzing business intelligence information. *J. Amer. Soc. Inform. Sci. Technol. 55*, 10, 873–891.

Nelson, G. 2010. Business Intelligence 2.0: Are we there yet? *SAS Global Forum.*

Page, L., Brin, S., Motwani, R., and Winograd, T. 1999. The PageRank citation ranking: Bringing order to the Web. http://dbpubs.stanford.edu/pub/1999-66.

Porter, M. A., Onnela, J.-P., and Mucha, P. J. 2009. Communities in Networks. *Notice of AMS 56*, 9, 1082–1097.

Savage, N. 2012. Gaining wisdom from crowds. *Comm. ACM 55*, 3, 13–15.

Shalizi, C. R. and Thomas, A. C. 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociol. Meth. Res. 40*, 211–239.

Sun, Y. and Han, J. 2012. *Mining Heterogeneous Information Networks: Principles and Methodologies.* Morgan & Claypool Publishers.

Turban, E., Sharda, R., Aronson, J. E., and King, D. 2008. *Business Intelligence: A Managerial Approach.* Pearson Prentice Hall.

Watson, H. J. 2012. This isn't your mother's BI architecture. *Bus. Intell. J. 17*, 1, 4–6.

Watson, H. J. and Wixom, B. H. 2007. The current state of business intelligence. *IEEE Comput. 40*, 9, 96–99.

Wixom, B. H., Watson, H. J., and Werner, T. 2011. Developing an enterprise business intelligence capability. *MIS Quart. Exec. 10*, 2.

Received March 2012; revised August 2012; accepted October 2012