A Study on Social Data Analytics and Privacy Concern among Social Media Users

Yogeswari Suppiah
Faculty of Computer
Science and Information
Technology,
Universiti Selangor,
Malaysia

Raja Mohd Tariqi Raja
Lope Ahmad
Faculty of Computer
Science and Information
Technology,
Universiti Selangor,
Malaysia

Mohd Fahmi Mohamad
Amran
Computer Science
Department,
Faculty of Science and
Defense Technology,
National Defence University of
Malaysia,
Kuala Lumpur, Malaysia

ABSTRACT
Social Data Analytics represent a new era in the ability of organizations and businesses to tap the potential of the information economy. But this new opportunity comes with hazards and risks as well. Privacy is the major concern among social media users. Protecting their personal information privacy has become a controversial issue among online social network providers and users. This study aims to provide a literature study on social data analytics and privacy concern among social media users.

Keywords
Big Data, Big Data Analytics, Social Data Analytics, Privacy, Social Data Privacy

1. INTRODUCTION
We are living in Digital Era where everything is digital such as an e-library, e-mail, e-shopping, e-ticket, e-payment, e-governance and many more. People use social media sites for entertainment such as Facebook, Twitter, and You Tube for video, photos, twits, and data downloads as well as uploads in the internet. Internet has stored a massive amount of data or information that is in the zeta or in Exabyte's and it is nothing but the Big Data [1]. Dynamic nature of social media data is a significant challenge for continuously and speedily evolving social media sites.

There are many additional interesting questions related to human behavior can be studied using social media data. Social media can also help advertisers to find the influential people to maximize the reach of their products within an advertising budget. Social media can help sociologists to uncover the human behavior such as in-group and out-group behaviors of users [2].

The business world is undergoing a revolution driven by the use of data and analytics to guide decision-making. While many forces are at work, a major reason for the business analytics revolution is the rapid proliferation of the amount of data available to be analysed. Leading organizations increasingly recognize the importance of leveraging their data as a strategic asset. Some organizations undertake analytics initiatives to improve the quality of customer experience by measuring and acting on sentiments expressed by customers or by linking various divisions and operating units that customers tend to connect with. Others analyse data to predict a customer's propensity to buy new products or services in order to proactively make recommendations for future purchases or offer discounts to encourage a longer-term relationship [3].This study aims to understand big data and social data analytics. It also investigates the privacy concern among social media users on their personal information on these sites.

2. LITERATURE REVIEW
Social media and networks provide powerful systems for businesses to learn and use productively. Based on the pervasive use and prevailing impact social media have on society, our position is that it is essential for leaders to integrate these technologies and seek the best way to use social media and networks to the advantage of the business and for him/herself, personally. They also explain some of the reason using social media which are viral growth, connectivity, ease of use, anonymity, global reach community, relevance, and a "smart" trend. The viral growth of usage requires everyone who wants to compete in the business environment of today to understand and utilize social technologies [5].

Most people have already shared parts of their personal information with several organizations, and have accepted the loss of partial privacy. For example, one may have accepted terms to share personal health information with healthcare networks and insurance agencies, and purchasing power and trends with sales and marketing agencies. However, such data may be shared between the data holders, either due to mergers or data sharing agreements between the organizations. Combination of such data may result in a more complete view of the individual, which may be unacceptable to the concerned individuals, especially if they perceive that it may be used against their interests. Extending this from a practical perspective, it would be interesting to explore economic models that allow incentive or compensation mechanisms for individuals to trade off their privacy [17].

2.1. Social Media
As can be seen in Figure 1, there are many types of social media. They include the following:

   a. SNSs: Social networking sites are networks enable and help people to connect and interact with each other through a website and to expand their personal networks (e.g., Facebook, MySpace).

   b. Blogs: These are websites function as online personal journals. They enable writers to post their opinions online and allow readers to comment (e.g., Blogger, WordPress).

   c. Micro-blogs: These are similar to blogs. They offer views of various updates in the world and the

happening in people's lives (e.g., Twitter).

d. Virtual worlds: These are online communities in the form of computer-based simulated environments where users can interact with each other (e.g., Second Life, Zwinky).

e. Really simple syndication (RSS): These are sites which enable frequently feed and used to publish full or summarised text, updated, news headlines, audio, and video. The RSS feed includes metadata such as publishing dates and authorship (e.g., Bahrain Gate Forum, iLounge Forum).

f. Wikis: Wikis are web pages with informational content that visitors can edit (e.g., Wikipedia).

g. Multimedia-sharing sites: These sites allow individuals to upload and share various types of multimedia contents, such as photos and videos. There are three types of media contents which can be shared: video-sharing (e.g., YouTube), photo-sharing (e.g., Flickr), and audio-sharing (e.g., Podcasts).
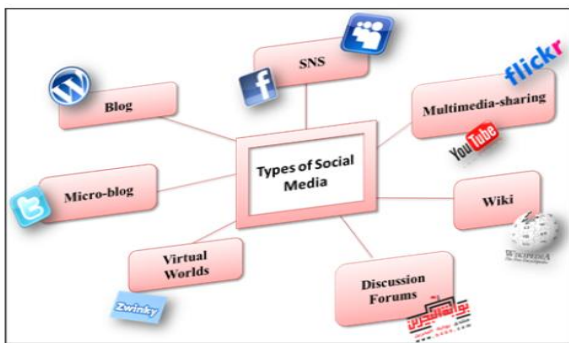

Figure 1: Types of Social Media

2.2 Social Media User

Contrary to some assertions that social media isolates people since you can communicate without ever coming face-to-face with another human being, the authors' perspective is social media actually aids the formation of communities. Without the discomfort of personal interactions and bias or filters we administer in face-to-face interactions, social media provides the opportunity for people to form connections intellectually and based on common interests and points of view. It also permits access to others who share a perspective around the globe, not just "down the street." The potential to leverage these communities for all kinds of reasons has been discovered by some businesses and leaders [5].

Leaders are aggressively addressing the linkages that make sense for their business. With the capabilities that are available with enterprise social networking, organizations can devise a social strategy to reach out to four key stakeholder groups: internal (employees), external (customers and suppliers), governing (board, regulators, and shareholders); public (influencers, industry, and public). These stakeholders are creating demand for use of social media for the following:

• Marketing: promote, sell, campaign, viral release

• Communications: information, knowledge, news, announcements

• Public Relations: quick release, public influence, offset, response

• Disaster Control: rapid, comprehensive reach (Japanese Nuclear Disaster)

• Operations: trouble shooting, radiological reviews, diagnoses (Antarctica)

• Knowledge and research: ideas, market research, open source

Social media was a very effective tool for businesses to actively involve and engage target market customers and influence customer behaviour. The ability of social media to create interactivity and dialogue represents a powerful way to engage customers and develop long-term relationships with the business and the brand. Increasingly, perhaps the most important reason for businesses to emphasize social media use is that promotion on social media is more affordable compared to traditional media. It also has the potential to target consumers more effectively. Overall, social media is the most cost-effective form of information dissemination and customer interaction [6]

2.3 Big Data

Big data as an abstract concept. Apart from masses of data, it also has some other features, which determine the difference between itself and "massive data" or "very big data." At present, although the importance of big data has been generally recognised, people still have different opinions on its definition. In general, big data mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. Because of different concerns, scientific and technological enterprises, research scholars, data analysts, and technical practitioners have different definitions of big data [7].

Big data is notably different from traditional information management and analytics in this regard. Instead of just creating reports or presentations that advise senior executives on internal decisions, big data scientists commonly work on customer-facing products and services. This is particularly true in big data start-ups, but it's also the case in larger, more established companies.

Big Data' is recently been arrived and applied in to dataset, big data growth is very large and even they don't work with traditional database management system. According to the IDC Big Data is a new generation technology and architecture designed to efficiently extract value from huge volume of wide variety of data. Regarding to Jason Bloom Big Data is a huge amount of both structured and unstructured types of data that is so large that it's difficult to process using traditional database and other software. Big data is a term for massive data sets having large, varied & complex type of structure with the challenges of analyzing, storing and visualizing for extra processes or results. Big data is a kind of datasets that are not only big but also is in form of high variety with fast velocity, which is not easy to handle through traditional tools [1].

Computers, smart phones, GPS devices, embedded microprocessors, sensors all connected by the mobile Internet are forming a 'societal nervous system' that is generating a cloud of data about people that is growing at an exponential rate. Every time we perform a search, tweet, send an email, post a blog, comment on one, use a cell phone, shop online, update our profile on a social networking site, use a credit card, or even go to the gym, we leave behind a mountain of data, a digital footprint, that provides a treasure trove of

information about our lifestyles, financial activities, health habits, social interactions, and much more [8].

BD has three defining features. The first is the availability of data at a massive scale collected not only online but through the use of mobile devices with location tracking capabilities and thousands of 'apps' that share data with multiple parties, interactions with smart environments (sometimes referred to as ambient intelligence or the internet of things), monitoring systems in the physical environment, and the human body itself, which is being used not only to harness data for genetic testing but also for authentication via biometric data. Additionally, Web 2.0 services enable users to create and voluntarily share vast amounts of personal data about themselves and their friends and family. Although individuals mostly volunteer these data for social purposes, organizations are happy to collect and profit from their analysis. The second defining feature is the use of high speed, high-transfer rate computers, coupled with petabytes (i.e millions of gigabytes) of storage capacity, resulting in cheap and efficient data processing. This increasingly means reliance on the cloud-computing model. The third and final feature is the use of new computational frameworks (such as Apache Hadoop) for storing and analysing this huge volume of data [9].

2.4. Social Data Analytics

Social data analytics comprises two main constituent parts: 1) data generated from social networking sites (or through social applications), and 2) sophisticated analysis of that data, in many cases requiring real-time (or near real-time) data analytics, measurements which understand and appropriately weigh factors such as influence, reach, and relevancy, an understanding of the context of the data being analyzed, and the inclusion of time horizon considerations. In short, social data analytics involves the analysis of social media in order to understand and surface insights which is embedded within the data. Online platform providers process and analyse big data to find meaningful correlations in order to specifically target products and services to individual consumers. Platforms are poised to become an increasingly important point of aggregation in the advertising market with unique reach, customer data and measurability. By combining their traditional expertise with new techniques, including data analysis and viral marketing, publishers/content aggregators can retain their valuable position as curators in the digital space and play a key role in ensuring quality content finds its audience. Big data will be used online and with mobile to power the content recommendation engine [10].

The highest and best use of big data – mining it for discovery and experimentation – is still in the learning phase in most companies. To date, the primary focus of business and technology organizations has been to automate data analysis processes such as marketing, sales and service. Analytics has been used to understand and tune such business processes, keeping management informed and alerting them to anomalies – "exception reporting" has been a key aspect of business intelligence [11]. Big Data and analytics represent a new era in the ability of organizations to tap the potential of the information economy. But this new capability is not without hazards. The ability to ingest massive volume, variety, and velocity of data for analytics does not eliminate risk—if not properly addressed such ability can often compound it, creating the risk of data use outside the bounds of law, regulation, and ethical practice. To derive the greatest benefit from Big Data and analytics, institutions will need to understand and address the implications of choices about data and analytic tools. They will need to carefully assess the integrity of their analytic processes and the accuracy of their findings, and to consider the legal consequences for privacy and data protection of applying the outcomes of analytic models to information about individuals. Organizations concerned about brand, reputation, and the privacy interests of their customers must bring to bear broad expertise to evaluate their processes and decisions and to foster trusted, ethical outcomes [12].

Some organizations undertake analytics initiatives to improve the quality of customer experience by measuring and acting on sentiments expressed by customers or by linking various divisions and operating units that customers tend to connect with. Others analyse data to predict a customer's propensity to buy new products or services in order to proactively make recommendations for future purchases or offer discounts to encourage a longer-term relationship [3].

Social media analytics refer to the analysis of structured and unstructured data from social media channels. Social media is a broad term encompassing a variety of online platforms that allow users to create and exchange content. The key characteristic of the modern social media analytics is its data-centric nature. The research on social media analytics spans across several disciplines, including psychology, sociology, anthropology, computer science, mathematics, physics, and economics. Marketing has been the primary application of social media analytics in recent years [13].

2.5 Privacy of Personal Information

Protecting the privacy of personal information is one of the biggest challenges facing website developers, especially social network providers. Several researchers have discussed the issue of privacy [14].

The focus is on the concerns that individuals have about who have access to their private information and the information that is being used. Certainly, with the increased use of the Internet, social networking, and other forms of information sharing, concerns over privacy continue to be a source of much research and discussion. Some of the factors that affect privacy concern include information usage, awareness, information sensitivity, familiarity with the firm/organization, and compensation [15].

2.6 Social Data Privacy

BD raises a number of policy considerations, with privacy scholars tending to highlight two main issues which are privacy and discrimination. BD also raises a related issue concerning automated decision making, which relegates decisions about an individual's life—such as credit ratings, job prospects, and eligibility for insurance coverage or welfare benefits—to automated processes based on algorithms and artificial intelligence. Not surprisingly, BD intensifies the use of automated decision making by substantially improving its accuracy and scope. Because decisions based on data mining are largely invisible to their subjects, significant issues arise around access to, and the accuracy and reliability of, the underlying data [9]. Privacy principles pose particular compliance challenges for Big Data analytics processes that use data about individuals. This is because Big Data analytics often draws disparate datasets from multiple sources and looks experimentally for relationships or correlations that may generate useful insights, over- collecting of information (to sift for possible correlations) is the norm. Potential uses often cannot be identified and therefore the subject of meaningful notice to the individual until potentially useful insights are identified [16].The topic of privacy comes up in nearly every

client conversation we have around Big Data and analytics. Individuals that are the source and subject of data have an increasing say in what is done with that data; in some cases, as a matter of law. Many privacy preserving data mining standards (such as k-anonymity) once thought to be secure have been shown to be vulnerable. It's not currently easy for individuals to get visibility into what is known about them, get alerts to possible inaccuracies and to what other organizations might do with that information in conjunction with other data [3].

Perhaps, a first step to enabling individual privacy could be to create a specification framework allowing individuals to express their privacy constraints. Robust search and correlation mechanisms could then be built informing individuals of the impact on their online privacy that any potential new online sharing might entail. Network data are actually the most reliable sources of real-life big data, thanks to well-known Web social networks like Facebook and Twitter. Here, mining such data is of primary interest, but the need for privacy and security very often limits the real impact of these tasks [17].Ensuring individual control over personal data is not only an increasingly unattainable objective of data protection, but in many settings it is an undesirable one as well. The discussions also addressed the advent of Big Data,

the increasingly ubiquitous nature of data collection and use, and the technological developments that expand our capacity to interconnect, analyse, identify, and extract new and unanticipated value from even old or seemingly worthless data as factors that require new approaches to data protection [18].

3. CONCLUSION

Social media plays a significant role in today's networked society. It has affected the online interaction between users, whom shares a lot of personal details and information online. Big data analytics is another technology that expanding drastically and provides great opportunity for businesses to increase strategic competitiveness. Social media analytics refer to the analysis of data from social media channels. Privacy is a major issue that need to be addressed in this era. There is a need for creating awareness among the social media users about the usage of their personal information for analysis purposes.

The summary of the study as shown on Table 1:

Table 1: Social data analytic developments and issues

Author	Critical point	Year
Billington, M. G.,&Billington, P. J.	Social media and networks provide powerful systems for businesses to learn and use productively. The reason using social media which are viral growth, connectivity, ease of use, anonymity, global reach community, relevance, and a "smart" trend. The viral growth of usage requires everyone who wants to compete in the business environment of today to understand and utilize social technologies.	2012
Jones, N., Borgman, R., &Ulusoy, E.	Social media was a very effective tool for businesses to actively involve and engage target market customers and influence customer behaviour. The ability of social media to create interactivity and dialogue represents a powerful way to engage customers and develop long-term relationships with the business and the brand. The most important reason for businesses to emphasize social media use is that promotion on social media is more affordable compared to traditional media.	2015
De Prato, G., & Simon, J. P.	Online platform providers process and analyse big data to find meaningful correlations in order to specifically target products and services to individual consumers. Platforms are poised to become an increasingly important point of aggregation in the advertising market with unique reach, customer data and measurability. By combining their traditional expertise with new techniques, including data analysis and viral marketing, publishers or content aggregators can retain their valuable position as curators in the digital space and play a key role in ensuring quality content finds its audience.	2015
Waterman, K. K., Bruening, P. J., Krasnow	Big Data and analytics represent a new era in the ability of organizations to tap the potential of the information economy. The ability to ingest massive volume, variety, and velocity of data for analytics does not eliminate risk—if not properly addressed such ability can often compound it, creating the risk of data use outside the bounds of law, regulation, and ethical practice. To derive the greatest benefit from Big Data and analytics, institutions will need to understand and address the implications of choices about data and analytic tools.	2014
Gopalkrishnan, V., &Steier, D.	Privacy comes up in nearly every client conversation we have around Big Data and analytics. Individuals that are the source and subject of data have an increasing say in what is done with that data; in some cases, as a matter of law. Many privacy- preserving data mining standards (such as k-anonymity) once thought to be secure have been shown to be vulnerable. It's not currently easy for individuals to get visibility into what is known about them, get alerts to possible inaccuracies and to what other organizations might do with that information in conjunction with other data.	2012
Tan, X., Qin, L., Kim, Y., & Hsu, J.	The concerns that individuals have about who have access to their private information and how such information will be used. Certainly, with the increased use of the Internet, social networking, and other forms of information sharing, concerns over privacy continue to be a source of much research and discussion. Some of the factors that affect privacy concern include information usage, awareness, information sensitivity, familiarity with the firm/organization, and compensation.	2012
Rubinstein, I. S.	BD raises a number of policy considerations, with privacy scholars tending to highlight two main issues: privacy and discrimination. BD also raises a related issue concerning automated decision making, which relegates decisions about an individual's life—such as credit ratings, job prospects, and eligibility for	2013

	insurance coverage or welfare benefits—to automated processes based on algorithms and artificial intelligence.	
Leonard, P., Crompton, M., & Barrett-glasgow, J.	Privacy principles pose particular compliance challenges for Big Data analytics processes that use data about individuals. Big Data analytics often draws disparate datasets from multiple sources and looks experimentally for relationships or correlations that may generate useful insights, over- collecting of information (to sift for possible correlations) is the norm.	2014
Cate, F. H., & Mayer-Schönberger, V.	Ensuring individual control over personal data is not only an increasingly unattainable objective of data protection, but in many settings it is an undesirable one as well. The discussions also addressed the advent of Big Data, the increasingly ubiquitous nature of data collection and use, and the technological developments that expand our capacity to interconnect, analyse, identify, and extract new and unanticipated value from even old or seemingly worthless data as factors that require new approaches to data protection.	2013

4. FUTURE WORK

This study provides a brief understanding on social data analytics and privacy concern among social media users. Future research could explore more on privacy perspective

of social data analytics. The future findings would be beneficial for social network providers and policy makers

to provide more concern on privacy of personal information of users on social media sites. There is immense scope in Social Data Analytics and a huge scope for research and Development.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] Gondaliya, T. P. (2015). New Big Things in Era of Digital Data : " Big Data " & Big Data Challenges with its, 496–507.

[2] Gundecha, P., & Liu, H. (2012). Mining Social Media : A Brief Introduction, (Dmml), 1–17.

[3] Gopalkrishnan, V., & Steier, D. (2012). Big data, big business: bridging the gap. *BigMine '12: Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining*, 7–11. http://doi.org/10.1145/2351316.2351318

[4] Saleh Zolait, A. H., Al-Anizi, R. R., Ababneh, S., BuAsalli, F., & Butaiba, N. (2014). User awareness of social media security: The public sector framework. *International Journal of Business Information Systems*, *17*(3), 261–282. http://doi.org/10.1504/IJBIS.2014.064973

[5] Billington, M. G., & Billington, P. J. (2012). Social media tools for leaders and managers. *Journal of Leadership, Accountability & Ethics*, *9*(6), 11–19.

[6] Jones, N., Borgman, R., &Ulusoy, E. (2015). Impact of social media on small businesses.*JrnlOf Small Bus EnteDev*, *22*(4), 611-632. http://dx.doi.org/10.1108/jsbed-09-2013-0133

[7] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209. http://doi.org/10.1007/s11036-013-0489-0

[8] Schaeffer, D. M., & Olson, P. C. (2014). Big Data Options For Small And Medium Enterprises. *Review of Business Information Systems*, *18*(1), 41–46.

[9] Rubinstein, I. S. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, *3*(2), 74–87. http://doi.org/http://dx.doi.org/10.1093/idpl/ips036

[10] De Prato, G., & Simon, J. P. (2015). The next wave: "big data"? (*). *Communications & Strategies*, (97), 15–39,153,155. Retrieved from http://search.proquest.com.idpproxy.reading.ac.uk/docview/1678885914?accountid=13460

[11] H. Davenport, T. (2014). How Strategists use "big data" to Support Internal Business Decisions, Discovery and Production. *Strategy & Leadership*, *42*(4), 45–50. http://doi.org/10.1108/SL-05-2014-0034

[12] Waterman, K. K., Bruening, P. J., Krasnow Waterman, K., & Bruening, P. J. (2014). Big Data analytics: risks and responsibilities. *International Data Privacy Law*, *4*(2), 89–95. http://doi.org/10.1093/idpl/ipu002

[13] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. http://doi.org/10.1016/j.ijinfomgt.2014.10.007

[14] Aldhafferi, N., Watson, C., & Sajeev, A. S. M. (2013). P ERSONAL I NFORMATION PRIVACY S ETTINGS OF O NLINE S OCIAL N ETWORKS AND T HEIR, *2*(2), 1–17.

[15] Tan, X., Qin, L., Kim, Y., & Hsu, J. (2012). Impact of privacy concern in social networking web sites. *Internet Research*, *22*(2), 211–233. http://doi.org/10.1108/10662241211214575

[16] Leonard, P., Crompton, M., & Barrett-glasgow, J. (2014). Customer data analytics : privacy settings for " Big Data " business, *4*(1), 53–69.

[17] Cuzzocrea, A. (2014). Privacy and Security of Big Data : Current Challenges and Future Research Perspectives, 45–47.

[18] Cate, F. H., & Mayer-Schönberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, *3*(2), 67–73. http://doi.org/http://dx.doi.org/10.1093/idpl/ipt005