DM STAT-1 CONSULTING

BRUCE RATNER, PhD

574 Flanders Drive   North Woodmere, NY 11581

516.791.3544 fax 516.791.5075 1 800 DM STAT-1

br@dmstat1.com

www.dmstat1.com

Social Network Analysis, Social Media Data, and Text Mining
to Boost Business Intelligence
Bruce Ratner, Ph.D.

The purpose of this article is to propose a methodology that has the potential of boosting business intelligence. I put forth the quantitative trinity – assembled for the business intelligence analyst – social network analysis, social media data, and text mining. I discuss the elements of the triad, and indicate how each fits together in the proposed methodology.

Social Network Analysis (SNA) is best described by an illustration. *Six Degree of Separation*, John Guare's play, in which everyone is linked to everybody else through a few highly connected intermediaries. SNA provides answers to four basis questions: 1) How highly connected is an individual within a network? 2) What is an individual's importance in a network? 3) How central is an individual within a network? 4) How does information flow within a network? [1, 2]

Using social media data (SMD) provides qualitative value for business intelligence applications, based on a 2008 survey. The top ten benefits are 1) Customer engagement, 2) Direct customer communications, 3) Speed of feedback/results, 4) Learning customer preference, 5) Low cost, 6) Brand building, 7) Market research, 8) Credibility of the "crowd," 9) Reach, and 10) Great lead generation source. [3]

The highly visible source of SMD is the swelling of online social networking such as MySpace, Facebook, Twitter, and YouTube videos, LinkedIn, marketing blogs, and social media message boards. Additionally, the rapidly increase in advice by freshly self-proclaimed social media experts, who exploit networking opportunities for profit, contributions to SMD. Not a well-known source of SMD, but surely rich in content, is the *Internet Public Library*: *Census Data & Demographics,* which includes five major categories: 1) Demographics, 2) Human cultural and artistic endeavors, 3) Hobbies, games, and activities for amusement, and pleasure, 4) Quick answers to factual questions and basic research tools, such as almanacs, dictionaries and encyclopedias, and 5) Social life and the behavior of individuals or groups within a society. [4] (For a list of *top* ten lists of social networking sites, see reference #5. For a list of social media message boards, see reference #6). [5, 6]

There is a new source of SMD that comes from a few start-up web-analytics companies that provide *timely* information about market, media, and market-/media-time shifts, as they have recruited a proprietary survey panel sample of 2,000,000+ U.S. Internet users that have given permission to analyze the web pages they visit and ask questions via surveys. These startup websites have useful statistics on social networking sites. For example, they compile a list of the ten top social networking sites along with site-statistics of the kind Rank-by (visits, page views, total time spent, and monthly attention), and Traffic-by (site-ranking, ranking by unique visitors, and ranking by all metrics). [7]

Websites, even those that are dwarfed by mega-sites like Amazon and Zappos, provide invaluable primary SMD. [8] On Amazon, current buyers' provide their reviews of books, and items of departments ranging from movies/music to tools/auto-industrial. On Zappos, active buyers' provide their reviews of shoes, and items of clothing and accessories. (Amazon and Zappos both originally only sold books, and shoes, respectively. Amazon from the get-go conceived and planned on expanding beyond books, and Zappos' business plan only focused on *consumer's feet*. [9] After all, everyone needs shoes.

> Consider U.S. President Abraham Lincoln, who starting walking 1.5 miles to school at the age of 8 years old. [10] Schoolboy Abe whose *large* shoe size and corresponding long steps made walking easy, but also required young Abe many a pair of shoes. (The size of young Abe's shoe size can be estimated from his adult shoe size 14.) [11, 12]

Website, mega-/mini-sites, engage their buyers in registering the presently purchased products. Buyers provide surrounding details about their purchases, and related almost-bought products. These SMD are prized primary information about products and services. Moreover, further primary data are obtained by websites crossing the line by gently asking their customers to rate, remark, and rant/rave *about an array of competitors'* products, services, and brands.

To briefly return to mega-sites Amazon and Zappos, one may recall their in-the-red *deaths were greatly exaggerated.* Their success stories are found *within their marketing space*. The founders created their success by unwittingly following the law of physics: Space Expands. Space Expands today is not a new concept. Over a century ago, it was an inconceivable thought, especially for one genius. This digression, which seemingly has no *relativity* to the current topic, is discussed later.

Text mining (virtually similar to data/*numerical* mining) of SMD uncovers unknown relationships within a single text document, and among aggregated documents. The findings are *input* for SNA to boost *business intelligence* – all that business analytics and database marketing can offer – for quantitative entities such as: 1) building scorecards for *consumer risk,* 2) developing models for *fraudulent transactions online,* 3) constructing segmentation models to *understand customer feelings or opinions* to prevent impair customer relationships, 4) identifying a set of segment-specific *key drivers of* acquisition, attrition, and/or retention within a segmented-population, and 5) cross-/up-selling to *target* the highest-influential segments. These quantitative entities are *some of my favorite things*, but there are many more. [13]

Digression:

I digress to illustrate an object lesson hidden in the success stories of Amazon and Zappos. As for buying books without seeing the cover, fanning the pages, and randomly reading a few passages, Amazon founder Jeff Bezos would have failed in 1994 if he believed the *thought of the day that it is impossible* to sell books <u>sight unseen</u>. As for buying shoes where it was the *thought of the day that it is impossible* to sell shoes without buyers bringing their feet along, Zappos founder Tony Hsieh also would have failed in 1999. But, Tony *opened our minds to a new possibility* – replace <u>bring your feet along</u> with *free shipping both ways if not satisfied. No reason required*. These two

mega-site founders created their own marketing space by *inserting* their space into the *overall* Space.

Amazon and Zappos' trek to success reminds me of Einstein's Theory of Relativity, where one stunning finding was: Space Expands. Unfortunately, but short-lived, the conventional wisdom of the today that *space doesn't expand* (actually terminology: space is static; neither contracts nor expands) forced Einstein to add a *fudge factor (!)* to his Theory of Relatively so the adjusted finding was: Space is Static. When Hubble showed that the universe indeed is expanding, Einstein declared that the inclusion of the fudge factor was his "biggest blunder."

> For me, I recall my elementary school teachers always told the students: Think twice about changing your first answer, as it is more likely to be correct then a changed-answer based upon a quick review of the exam before handing it in. What object lesson lies here? [13]

Conclusion:

I put forth the quantitative trinity – fitted to the business intelligence analyst – social network analysis, social media data, and text mining. I reviewed the elements of the triad, and discuss how each fits under the proposed methodology that has potential to boost business intelligence. I present my views about the each member of the trinity in order to arrive at a compelling argument for the proposed methodology to boost business intelligence. I digress to a faraway place, as my stream of consciousness writing is brought about how today's marketing and the effects of Einstein's genius are both found in cyberspace.

References and Footnotes:

1 - *Six Degrees of Separation* is the theory that anyone on the planet can be connected to any other person on the planet through a chain of acquaintances that has no more than five intermediaries. The theory was first proposed in 1929 by the Hungarian writer Frigyes Karinthy in a short story called "Chains." - http://whatis.techtarget.com/definition/0,,sid9_gci932596,00.html
2 - SAS© Social Network Analysis Software -
http://www.sas.com/news/sascom/2009q3/column_trends.html
3 - http://mashable.com/2008/12/29/benefits-of-social-media-marketing
4 - Internet Public Library*:* Census Data *&* Demographics - http://www.ipl.org/div/subject
5 - http://en.wikipedia.org/wiki/List_of_social_networking_websites,
http://traffikd.com/social-media-websites/,
http://mashable.com/2007/10/23/social-networking-god/,
http://social-networking-websites-review.toptenreviews.com/,
http://www.insidecrm.com/features/50-social-sites-012808/,
http://www.doshdosh.com/list-of-social-media-news-websites/,
http://www.ebizmba.com/articles/social-bookmarking-websites,
http://www.danah.org/researchBibs/sns.html,
http://bdnooz.com/lbsn-location-based-social-networking-links/, and

http://news.google.com/news?q=list+of+social+networking+sites&hl=en&sourceid=gd&rlz=1D1GGLD_enUS298US299&um=1&ie=UTF-8&ei=4as3S47RN8rClAfxx7WoBw&sa=X&oi=news_group&ct=title&resnum=11&ved=0CCsQsQQwCg

6- http://www.meetup.com/raygsmn/messages/boards/

7 - http://www.geniq.net/res/Top-25-Social-Network--Complete.com.pdf

8 - Primary data are collected directly from first-hand asking for specific question to a specific group of respondents. Once the primary data are collected, aggregated (summarized), and published, they are called secondary data.

9 - Amazon Buys Zappos - http://www.techcrunch.com/2009/07/22/amazon-buys-zappos

10 - http://www.atrochatro.com/quiz_personalities-5.html

11 - http://www.walktoschool.org.uk/content/fun_foot_facts.php

12 - Lincoln was a voracious reader. He had a habit of reading everything that came into his hands. He once walked six miles to borrow a book on English grammar, then read and understood it completely in six days. It was common in Lincoln's day to become a lawyer by reading and studying on one's own, and then passing the bar exams. That was the route Lincoln followed - http://wiki.answers.com/Q/Where_did_Abraham_Lincoln_go_to_law_school

13 - GenIQ Text Miner - http://www.geniq.net/res/GenIQ-Text-Miner.html

14 - When Einstein introduced the cosmological constant (a fudge factor!) in his theory of general relativity he did so because he was guided by the paradigm of the day that the universe was static (i.e. neither contracting nor expanding.) The cosmological constant provided a way of balancing the gravitational contraction caused by matter. It was later discovered by Edwin Hubble that other galaxies appear to be receding away from us, that the universe was actually expanding. When Einstein heard and fully appreciated these observations, he declared that the inclusion of the cosmological constant was his "biggest blunder."

Einstein's mistake was not a mathematical one but rather a philosophical one made many times over the history of thought. Einstein held too strongly to the belief that the universe was static and thus was unable to appreciate the power of his theory's predictions of a dynamic universe. His error serves as a lesson to all thinkers, that *we should never close our minds to new possibilities, even if the thought of the day is that they are impossible*. – http://super.colorado.edu/~michaele/Lambda/blund.html