From the Editors

Thick Data: A New Qualitative Analytics for Identifying Customer Insights

Jinan Fiaidhi
Lakehead University

Sabah Mohammed
Lakehead University

Editor: Jinan Fiaidhi [jinan.fiaidhi@lakeheadu.ca]

All sort of businesses and organizations are now online, and they leave a trail of data on social media sites, blogs and portals, messages of all types, and lots of traces on search engines. Enterprises can no longer escape the need to monitor and analyze social media outlets such as Facebook, Twitter, Pinterest, news sites, blogs, forums, video sites, and microblogs. To succeed and grow, a business needs to be able to acquire, retain, satisfy, and engage their customers effectively. Embracing social media analytics is vital for assessing how well the business does this. Social media analytics is the process of accessing data generated on social media such as ideas, sentiments, and customer feedback. This information can then be analyzed and fed into the decision-making process across all business activity, including campaign orchestration, product development, recruitment, customer advocacy and engagement processes, sales input, and much more. A business that knows its customers' insights from the social media analytics can make smarter decisions that maximize consumer loyalty without spending money on unnecessary exercises (e.g., general advertising, rebates, discounts, and special offers). In contrast, textual analytics form the basis of most of the real insights we draw from social media today. From simplistic bag of words counting algorithms through advanced deep learning approaches, it is text that forms the lens through which we see social media. We count hashtags, compute word and phrase histograms, measure textual sentiment, flag brand mentions, compute follower and retweet graphs, measure meme velocity, and so on. Any social analytics firm can tell you how many tweets per day over the past month have mentioned your company in the tweet text or hashtag. Nearly all of them will offer you a timeline of the average "tone" of those tweets by day as well, along with who

Digital Object Identifier 10.1109/MITP.2019.2910982
Date of current version 21 May 2019.

1520-9202 © 2019 IEEE

are the most positive and most negative tweeters about your brand. However, the general belief has always been to rely on big data and predictive analytics to make relevant, personalized, and precisely timed offers to customers. Most automated sentiment analysis, for example, is deemed useless or poor by insight clients and market researchers, which basically means that social media usage requires people with some expertise, and that tends to make the whole process prohibitively expensive and slower. Actually, the promise of social media analytics is on a decline because "social analytics is not panning out the way it was expected."[1] Gartner reported in 2017 that 60% of all big data projects failed. As bad as that sounds, the reality is actually worse. The big data projects failed because of the obsession with quantitative data at the expense of qualitative consumer centric data. Well, we are not talking about the demise of social media analytics, but the buzz has definitely died down. Although social media monitoring and mining reaps (collection and automated analysis of quantitative amounts of naturally occurring text from social media) provided plentiful silos of results, however, the success is still limited with at most a 20% success rate, as Gartner predicted for the social media analytics by 2020 (https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/).

> In today's data-driven and consumer-centric world, decision makers invest in trying to reach the best outcome and value-driven services. Much of the investment is directed toward analytics, CRMs, and loyalty programs—all of which provide plenty of quantitative data about customers. This type of data offers the when, where, what, and how of the customers' interactions with their experiences, services, or products. But critically, it does not provide the why, as numbers alone (the quantitative data) are not enough for understanding what makes their customers tick. To understand the whys behind your customer's actions, you need to have thick data.

PATCHING ANALYTICS WILL NOT WORK

Once the failure rates of big data and predictive analytics started to appear frequently on the media, a variety of suggested additional quantitative analytics patches started to appear for securing better success rates including

- customer segmentation analytics;
- customer engagement analytics;
- customer churn analytics;
- customer acquisition analytics;
- consumer analytics using maturity models.

Developing a patch for enhancing the accuracy of identifying consumer insights requires a thorough understanding of the nature of data analyzed, and it must go beyond the quantitative nature of the analytics used. It makes no sense to add more quantitative analytics to failed analytics. Adding more patching analytics proved to be daunting if it is built on prediction models that deal with only quantitative indicators and do not incorporate important consumer values. According to Wang,[2] "What is measurable isn't the same as what is valuable." A common error that many social media analytics make is diving into marketing without first understanding the target audience. In order to make wise social media marketing decisions, you need to understand the pain points shared by the target customer, as well as how the target customer prefers to make purchasing decisions. Wang, in a 2017 TED talk (https://www.ted.com/speakers/tricia_wang), described the danger of relying on social media quantitative data analytics as it was the reason behind the failures of important vendors like Nokia, although Nokia has collected a huge amount of data about their customers and the market but failed to alert the company about the consumer's values and the market trends for smartphone technology.[3] Actually, failing to capitalize from social media analytics is expected if the organization is looking at social media data on its own, rather than in conjunction with email data, call recordings, surveys, and countless other internal and external sources.[4] Moreover, all leading web analytics software used by several organizations have excellent quantitative analytics capabilities where they can easily provide information, such as number of users who visited the website during a given time period, geographic locations from which the website was accessed, time period for which user stayed on the website, etc. However, when it comes to the "true business value" questions (such as why did the customer not make the purchase, what factors influenced the customer's buying decision, how did the customer feel after the purchase?), these web analytics software at best provide vague answers. So, if an organization is narrowly focused on the clickstream statistics, they will never be able to comprehend the thoughts going through customer's mind as they move through the purchasing journey. What is most needed is to understand the nuances of consumer behavior with help of new analytical techniques. The new techniques need to answer several questions about the consumer behaviors, as follows.

1) How do consumers think and feel about different alternatives (brands, products, services, and retailers)?
2) How do consumers reason and select between different alternatives
3) What is the behavior of consumers while researching and shopping?
4) How is consumer behavior influenced by their environment (peers, culture, media)?
5) How do we navigate for insights when the consumer data are in doubt or associated with uncertainty due to data inconsistency and incompleteness, ambiguities, latency, deception, model approximations, low accuracy, low quality, unclear truthfulness or with lower trustworthiness?
6) How do we navigate for insights when consumer data continuously change through the use of the differing ways of data interpretation?
7) What is the relative importance of different complex consumer data from different perspectives and locations?

Answering these questions will enable us to understand the demographics, interests, preferences, and the satisfaction rate of the customers. However, understanding these insights can never be captured through quantitative analytics alone. No amount of quantitative guessing can match the business value of "voice of customer." Having said that, we need to differentiate between qualitative and quantitative analytics before going further as the differentiation is not always obvious. For example, during Text Analytics, measuring the frequency of certain words would be considered quantitative analytics, whereas exploring the contextual meaning of a conversation would be considered qualitative analytics. In other words, qualitative analytics includes the analysis of context, human behavior, emotions, and other factors that are hard to digitize without losing any meaning. Qualitative analytics is a great approach to bridge the gap between insights provided by quantitative research and providing in-depth understanding of the underlying reasons and motivations for a given phenomenon situation. Qualitative analytics is not an added patching analytics because it will not simply add more data points to adjust inaccurate prediction algorithm outputs. Actually, no output will be able to predict human behavior until inputs are as complex, unexpected, and sometimes as contradictory as humans themselves. This where the notion of Thick Data came to surface.

THICK DATA APPROACH

The thick data concept has been popularized by Tricia Wang to understand the quirks of human behavior and predict how an individual's relationship with the business service or product will evolve over time. Without this understanding, quantitative patterns that the business may uncover that suggest people will behave in a certain way could be based in a world that no longer exists. Thick data takes individual consumers' temperature more precisely and offers depth analytics to the consumer data story. The thick data differs from big data by its qualitative approach, obtaining ethnographic data that allow contexts and emotions of the analyzed subjects to be revealed, while big data requires an algorithmic process usually carried out by statesmen and data scientists. The problem is that while big data is big, it can also be thin in producing effective analytics. It has not been the silver bullet that some commentators expected. CEOs like Tricia have discovered that when it is combined with other methods, qualitative research in particular, or "thick" data, if you will, then you can really start building insights and make better-informed decisions. Big brands have recognized this. For example, Netflix combined their big data on customer viewing habits (who was watching, when, and what) with ethnographic studies to better understand the viewing habits of subscribers (why they were watching, what needs they had, and how they were delivered against), which gave direction to strategy.[5] In this way, they have been able to leverage the modern viewing phenomena of binge watching to great success. Paradoxically, then, the era of big data needs even more qualitative, granular knowledge of local contexts. For big data to be analyzable, it has to normalize, standardize, and define certain parameters and assumptions to sort, organize, and disseminate information. The workability of its models depends on certain conjectures. If the underlying assumptions go wrong, the entire results may go wrong. This risk can be minimized using thick data. While big data relies on machine learning, thick data relies on the social context of connections between data points. It goes beyond big data to explaining the everyday lives of consumers to understanding why they have a certain set of preferences. This way, it reduces the depreciation that the data go through in order to become usable for analysis. Using an anthropological approach to understand thick data means we expand the scope to include aspects such as attitudes, experiences, opinions, emotions, behavior, context, social dynamics, and sensory information—in short, all the complexities of human life that cannot be reduced to numbers. Taking these factors into account means that thick data can help us understand not just what people do and have done, but also why they do it. This allows us to enrich big data with insights into what drives people, not just, e.g., as consumers, but as human beings. For this reason, we will see an increasing use of thick data techniques for decision making as thick data makes remarkable business improvement by discovering consumer insights. Lego, the Danish Toy company, as another example, was near collapse in the early 2000s because they lost touch with their customers. After failed attempts to reposition the company with action figures and other concepts, Jorgen Vig Knudstorp, CEO of the Danish Lego firm, decided to engage in a major qualitative research project. Children in five major global cities were studied to help Lego better understand the emotional needs of children in relation to legos. After evaluating hours of video recordings of children playing with legos, a pattern emerged. Children were passionate about the "play experience" and the process of playing. Rather than the instant gratification of toys like action figures, children valued the experience of imagining and creating. The results were clear; Lego needed to go back to marketing its traditional building blocks and focus less on action figures and toys. Today, Lego is once again a successful company, and thick data proved to be its savior.[6] Thus, thick data can bring a top-notch differentiator to any organization by helping businesses to uncover the kinds of insights that most of the quantitative approaches do not give. Figure 1 illustrates some of the traditional assistive technologies for helping to provide qualitative data for thick health data analytics, including forming focus groups, conducting surveys, scrapping health blog comments, patient reviews and complaints, searching Q&A sites for relevant cases, as well as collecting helpful thin data from Google Analytics, popular health sites like Patients LikeMe, Drugs.com, and from governmental systems like the FDA FREAS.

However, finding the hidden values in customer feedback, for example, is exceptionally complicated, and it is still too early to leave it to machine learning as the sample size is quite small (on average only 4% of unsatisfied customers in any business provide feedback). Indeed, the lack of customer feedback does not necessarily mean your customers are completely satisfied with your business. If we rely solely on big data with some data that are fed by some of assistive technologies mentioned in Figure 1, we may end up with a warped sense of the world in which human beings are simply numbers to be fed into a machine learning algorithm. However, companies should invest more in gathering and analyzing thick data to uncover the deeper, more human meaning behind big data. Moreover, investment is also needed to learn consumer insights not only from structured platforms like review sites and forums but also throughout social media networks. Therefore, it is necessary to analyze online discussions and conversations on social media, as well as through data gathered from search engine keyword trends associated with your brand and

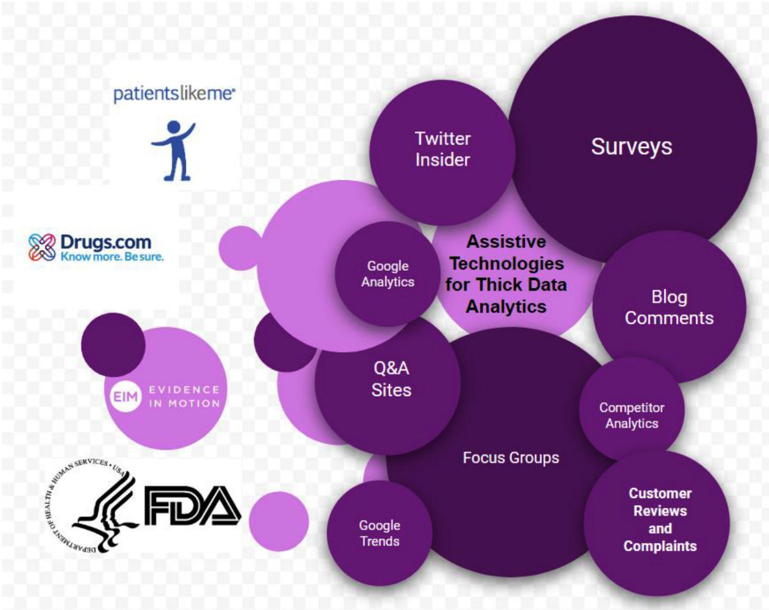

Figure 1. Thick health data analytics assistive technologies.

products to get a more well-rounded understanding of the customers' attitudes and perceptions. This article focuses on the later approach on developing thick data analytics from social media by identifying conversations and learning the outcome of the consumer satisfactions from these discussions on social media.

Discovering Thick Data From Social Media Conversations

Online social networks (e.g., Twitter) have dramatically grown in usage and popularity in recent years. In addition to keeping track of friends and acquaintances, these networks provide powerful means of exchanging and sharing information on many different topics of interest (e.g., sports, religion, politics, health concerns, etc.). The use of these networks has introduced a completely new way of collaboration and conversation among people, virtually creating spheres of friends who generally feel comfortable discussing a variety of subjects and even helping each other to grow in knowledge about certain subjects. The automatic detection of dialog structure is an important first step toward deep understanding of human conversations. Dialog provides an initial level of structure by annotating utterances with shallow discourse roles such as "statement," "question," and "answer." Actually, conversations or dialog over Twitter create networks with identifiable contours as people reply to and mention one another in their tweets. These conversational structures differ, depending on the subject and the people driving the conversation. Six structures are regularly observed: divided, unified, fragmented, clustered, and inward and outward hub, and spoke structures.[7] These are created as individuals choose whom to reply to or mention in their Twitter messages, and the structures tell a story about the nature of the conversation. However, locating conversations from Twitter starts by identifying conversation clusters in what we call socio graph (similar to ethnography[8] or netnography[9]). Socio graphs provide a wealth of qualitative information on twitter user's conversations, as well as implicit social circles. The twitter conversational structure (see Figure 2) creates variety of socio graphs that we can capture if proper graph retrieval algorithms are used.

However, it is important to note that the standard Twitter Search API or other similar APIs (e.g., Tweepy) that are used to retrieve tweets are based on keyword filtering which produce matching tweets to particular keywords. Using these APIs, retrieval can be customized according to the attributes of the Tweet Object. The Tweet object has a long list of "root-level" attributes, including fundamental attributes such as user profile, tweet id, created at, and tweet text. Tweet objects are also the "parent" object to several child objects. Tweet child objects include user, entities, and extended entities. Tweets that are geo-tagged will have a location child object. However, this retrieval method with whatever selected attributes cannot recreate complete individual conversations since tweets within a single conversation may not all match a single keyword.[10] Therefore, we need to collect a massive number of tweets (i.e., silo) related to specific attributes like the topic using tools like NodeXL Pro (https://www.smrfoundation.org/nodexl/), NetworkX (https://networkx.github.io/), Scraper (https://www.scraperapi.com/), Gephi (https://gephi.org/), Proxycrawl (https://proxycrawl.com/), Zignal Labs Discover (https://zignallabs.com/products/zignal-discover/), Neo4j (https://neo4j.com/), or Tags (https://tags.hawksey.info/). One can develop a crawler using the mentioned tools or the simple searching APIs provided by twitter to collect a silo that is more suitable for thick data analytics. The crawler then continuously adds to the silo whatever relevant tweets as tweets get published or from older twitter repositories. All these tweets in the silo will later be converted into a structured conversation with other associated information. The conversation structures can be constructed using a recursive searching algorithm to select a random tweet and then traces the chain of replies back to the root (i.e., the initial tweet) of the conversation. Once the root has been found, all of the tweets within the conversation can be tracked. Every recursive cycle will provide the capability of reconstructing complete conversation around initial Tweets, and we then can explore how to characterize or learn the outcome from these conversation graphs instead of looking into a silo of unrelated or partially related tweets. There are a number of ways in which a conversation graph can be retrieved or defined from the silo of tweets collected. Figure 3 illustrates our investigation on finding suitable algorithms for detecting socio graphs from social media in general.

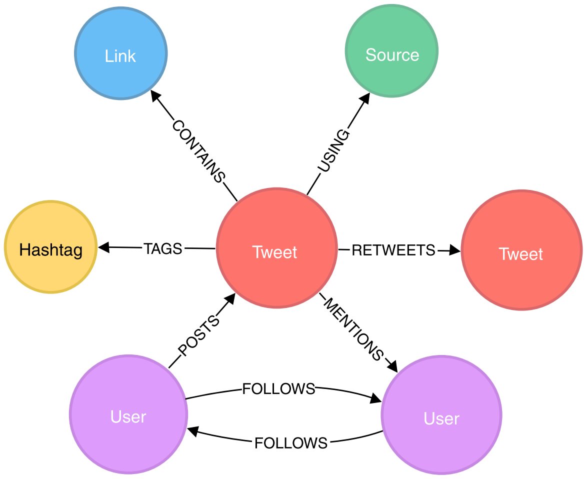

Figure 2. Twitter conventional conversation structure.

Thick Data Crawler

Social media crawling is widely applied in many areas today. It targets fetching new or updated data from any website or blog and then stores the collected data for easy access, processing, and content analysis. However, not every crawler is a focused one. A focused crawler is programmed to look for the pages/tweets classified as relevant to the specified attributes. For this purpose, the focused crawler needs to have a classifier that defines words, their classes, and rules of joining words into the phrases relevant to the topic or attributes selected. The classifier may use several machine learning and similarity algorithms (e.g., TextRank,[19] TF-IDF, or Cosine Similarity), as well as several databases and rule repositories for building the classification. Based on the weights of words and phrases, the page/tweet text weight is calculated, which indicates relevance to the topic/attributes selected. The crawler starts by reading the words stream and convert the word sequence into a matrix in which the rows are documents/tweets and the columns are words. The values matching a document with a word in the matrix are counted using the tf-idf (http://www.tfidf.com/) algorithm. This matrix is then used by machine learning algorithms like the TextRank to produced highly focused graph of documents/tweets that are related to the search attributes selected. Figure 4 illustrates the main components of a focused crawler where it converts the stream of text into a weighted graph and refines it through a classifier. Luckily, there are a variety of implementations of such a crawler that are already available on the web, like the PyTextRank (https://github.com/DerwenAI/pytextrank), which is a Python library that uses spaCy, datasketch, and NetworkX.

Socio Graph Detection Algorithm	Comment	Reference
Edge betweenness	Define densely connected regions of a socio graph via graph partitioning.	[11]
Walktrap	Define a measure of similarity between vertices based on random walks.	[12]
Leading Eigenvectors	Define groups of vertices with a higher-than-average density of edges connecting them. This maximization process can be written in terms of the Eigen spectrum of a matrix.	[13]
Fast Greedy	Define a hierarchical agglomeration algorithm for detecting community structure.	[14]
Label Propagation	Define a simple label propagation algorithm that uses the network structure alone as its guide.	[15]
Louvain	Define a greedy optimization method that attempts to optimize the "modularity" of a partition of a socio graph.	[16]
Spinglass	Define a semi-supervised spin-glass model that enables current community detection methods to incorporate background knowledge in the forms of individual labels and pairwise constraints	[17]
InfoMap	Defines a method that decomposes a network into modules by optimally compressing a description of information flows on the network. The result is a map that both simplifies and highlights the regularities in the structure and their relationships.	[18]

Figure 3. Candidate algorithms for extracting socio graphs from social media.

The crawler can store the graph data as JSON, CVS, or XML. To the best of our knowledge, this is the best crawling method that is able to reconstruct complete conversations around initial tweets. The thick data crawl provides an opportunity to build a focused graph around certain topic or attributes. Based on this graph, we can compute a variety of thick data indicators like the centrality of the conversations, the importance of individual nodes in a graph, and the influencers in each of these conversations.

Human Conversation Detector and Thick Data Indicators

There has been a lot of research carried out in this topic for social network analysis to answer the some of the questions that can contribute to breathing life into flat data and brings clarity to data analysis. However, identifying meaningful data is the biggest challenge here (i.e., kind of data that incorporates quantitative metrics with human insights and provides decision makers with an effective framework for evaluation). Meaningful data tells a more complete story—ultimately getting closer to the "why." With complex information, what indicators you can get depends on what you put in. If the inputs are not accurate, the results will not be either. This is why the initiative of "big data analytics" did not fly so high as it mostly depends on the bag of words input. Finding useful patterns and indicators that involve human factors can point the way to new efficiencies, new ways to fight crime and disease, and new trends in business. All that starts with a meaningful input from the information crawler in the form of weighted graph. Difficult questions can come up after that like "how many conversation communities are in a focused graph?" and "what are the topics addressed by each of these conversations?" Moreover, what are the individual metrics that we can extract from these conversation structures? For example, in a global trade, we would like to know which countries tend to trade only with a select subgroup of other countries. Which goods are traded in networks where one country dominates trade? These questions

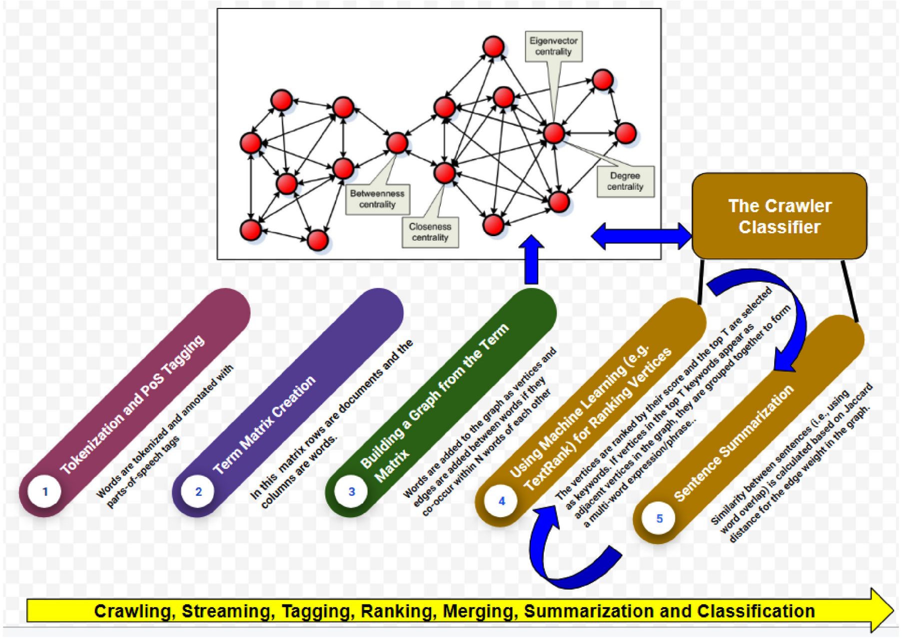

Figure 4. Main components of a thick data crawler.

relate theoretically to the respective graph theory concepts of clustering and centrality. The graph built by the crawler is most likely to end as adjacency matrix or what is generally known in the terminology of NetworkX as the ego-networks. If our focus is on conversations between people over the social network, then our ego network will be composed of rows where each row is the list of the friends of the first user in the line who is directly part of the ego's network. Analyzing such ego-graphs will provide answers to lots of questions we raised earlier. Processing these ego-graphs in Python based on the NewtworkX API is simple as this API enables us to construct an adjacency matrix and store it as CSV file. Pochetti[20] provided a clear way to convert ego-graphs into an adjacency matrix in his blog using the Python NetworkX API. Because egos or friends can be broken down into different groups, you may wonder if we could identify these different communities in the social network. The answer is yes. Using community detection algorithms, we can break down a social network into different potentially overlapping communities. The criterion for finding good communities is similar to that for finding good clusters. We want to maximize intracommunity edges while minimizing intercommunity edges. Formally, the algorithm tries to maximize the modularity of network or the fraction of edges that fall within the community minus the expected fraction of edges if the edges were distributed by random. Good communities should have a high number of intracommunity edges, so by maximizing the modularity, we detect dense communities that have a high fraction of intracommunity edges. While there is no community detection method in NetworkX, there is a helping library called community (https://python-louvain.readthedocs.io/en/latest/api.html) for detecting communities built on top of NetworkX. However, in R, the iGraph library can help to build adjacency matrices from ego-graphs and visualize the conversation communities.[21]

However, the NetworkX API provides a vast number of algorithms (https://networkx.github.io/documentation/latest/reference/algorithms/index.html) for extracting indicators from the generated graphs. Figure 5 summarizes the NetworkX algorithms suitable for extracting indicators from the conversation graphs.

Conversation community detection based on these algorithms can provide valuable indicators for decision makers. To such provide insightful information about the conversation communities, much research has been conducted in the form of surveys, systematic literature reviews, and visual studies. But only a few of them show how this field may contribute to thick data

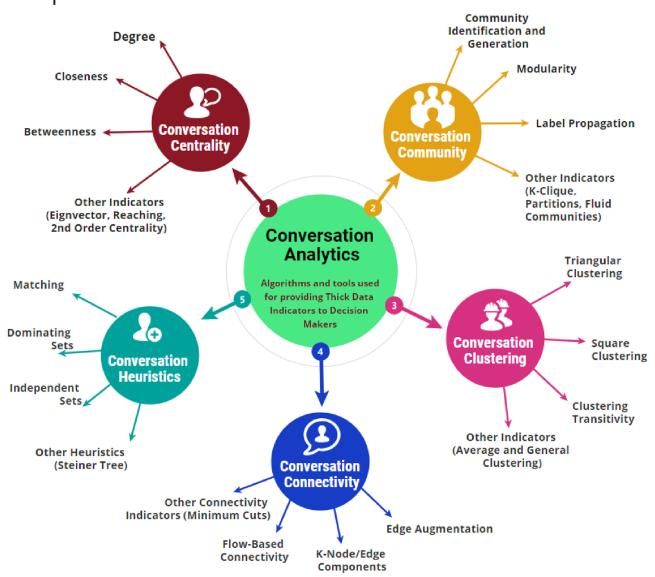

Figure 5. NetworkX conversation algorithms and techniques.

analytics.[22] In an attempt to demonstrate the usefulness of the thick data analytics framework and the benefits of using such qualitative data for physicians to learn about patient insights and associated indicators like adherence to medications or complaining about adverse events, we developed a sandbox that can be used by physicians to monitor useful indicators from the conversations with their patients or with other twitter users over a certain topic. Figure 6 illustrates the components of this sandbox.

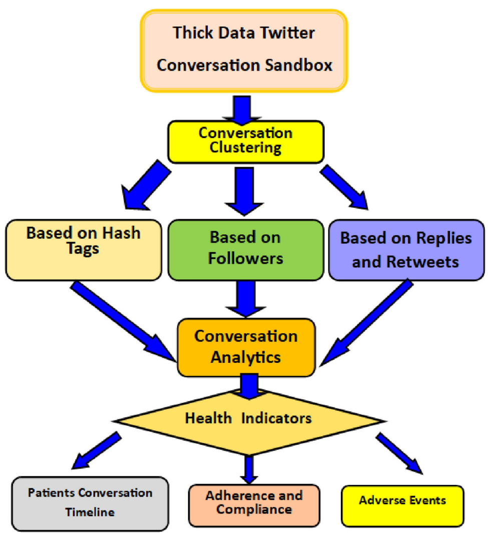

Figure 6. Twitter thick data analytics sandbox.

In this sandbox, we assumed that the physician needs to create a visible profile over twitter where patients can use it for building conversations. The conversation sandbox can help the physician to localize conversations and visualize them in a time line according to a selected adverse event like pain or a reaction to certain medication. The conversation timeline component uses three functions: get-patients (), lookup-conversations-according-ADE (), and scale-conversation ().

This type of analytics leaves the door wide open for massive research on how to learn from the consumer conversations and to identify the outcomes of these conversations and the effect of some of the conversation circles on others.

CONCLUSION

We introduce the growing paradigm of thick data analytics and focused on the importance of identifying the conversation communities in social media. Such conversation networks contain a set of strong dominant communities, which interfere with other weak or strong structures. Representing and analyzing these conversation structures have a massive impact on how to understand social network's conversational outcomes. However, we have only touched the surface of this exciting topic. We encourage you to contribute to this column by writing to the editor at jfiaidhi@lakeheadu.ca.

DISCLAIMER

The authors are completely responsible for the content in this article. The opinions expressed here are completely and personally their own.

REFERENCES

1. R. Bhatia, "The inconvenient truth—Broken promise of social media analytics," Analytics India, Apr. 3, 2017. [Online]. Available: https://www.analyticsindiamag.com/inconvenient-truth-broken-promise-social-media-analytics/
2. T. Wang, "Why big data needs thick data," Medium.com blog, Jan. 20, 2016. [Online]. Available: https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7

3. Brand Minds, "Why did Nokia fail and what can you learn from it?" Jul. 24, 2018. [Online]. Available: https://medium.com/multiplier-magazine/why-did-nokia-fail-81110d981787
4. C. Pettey, "5 reasons you're failing with social analytics," Gartner, Mar. 7, 2017. [Online]. Available: https://www.gartner.com/smarterwithgartner/5-reasons-youre-failing-with-social-analytics
5. F. Raben, "A fusion of thick and thin data," Nov. 1, 2018. [Online]. Available: https://greenbookblog.org/2018/11/01/a-fusion-of-thick-and-thin-data/
6. Buitelce Rep., "Competitive edge that thick data can bring to your business," Jul. 1, 2017. [Online]. Available: http://www.busitelce.com/big-data/45-competitive-edge-that-thick-data-can-bring-to-your-business
7. I. Himelboim, "Mapping twitter topic networks: From polarized crowds to community clusters," Blog Post, Pew Research Center's Internet Amer. Life Project, 2014. [Online]. Available: https://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/
8. A. E. Marwick, "Ethnographic and qualitative research on Twitter," *Twitter Soc.*, vol. 89, pp. 109–121, 2014. [Online]. Available: http://www.tiara.org/wp-content/uploads/2018/05/Marwick_Ethnographic-and-Qualitative-Research-on-Twitter_2013.pdf
9. R. V. Kozinets, P.-Y. Dolbec, and A. Earley, "Netnographic analysis: Understanding culture through social media data," in *Sage Handbook of Qualitative Data Analysis*, U. Flick, Ed. London, U.K.: Sage, 2014, pp. 262–275.
10. P. Cogan, M. Andrews, M. Bradonjic, W. S. Kennedy, A. Sala, and G. Tucci, "Reconstruction and analysis of twitter conversation graphs," in *Proc. 1st ACM Int. Workshop Hot Topics Interdisciplinary Social Netw. Res.*, 2012, pp. 25–31.
11. R. Kaptein, E. Boertjes, and D. Langley, "Analyzing discussions on Twitter: Case study on HPV vaccinations," in *Proc. 36th Eur. Conf. IR Res.*, Amsterdam, The Netherlands, Apr. 13–16, 2014, pp. 474–480.
12. N. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.
13. P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *Proc. Int. Symp. Comput. Inf. Sci.*, 2005, pp. 284–293.
14. M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, 2006, Art. no. 036104.
15. A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, 2004, Art. no. 066111.
16. U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E*, vol. 76, no. 3, 2007, Art. no. 036106.
17. V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Statist. Mech., Theory Experiment*, vol. 10, 2008, Art. no. P10008.
18. E. Eaton and R. Mansbach, "A spin-glass model for semi-supervised community detection," in *Proc. AAAI Conf. Artif. Intell.*, 2012, pp. 00–906.
19. R. Mihalcea and P. Tarau, "TextRank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 404–411.
20. F. Pochetti, "Community detection in social networks," Oct. 19, 2014. [Online]. Available: http://francescopochetti.com/community-detection-social-networks/
21. S. Whalen, "Creating beautiful Twitter graphs with python, towards data science," Dec. 14, 2018. [Online]. Available: https://towardsdatascience.com/creating-beautiful-twitter-graphs-with-python-c9b73bd6f887
22. B. S. Khan and M. A. Niazi, "Network community detection: A review and visual survey," Cornel Univ. Archive, Unpublished Preprint, arXiv:1708.00977, 2017, Available: https://arxiv.org/ftp/arxiv/papers/1708/1708.00977.pdf

Jinan Fiaidhi is a Full Professor of computer science and the Graduate Coordinator of the BioTech Ph.D. program at Lakehead University. She is also an Adjunct Research Professor with the University of Western Ontario and the Editor-in-Chief of IGI Global *International Journal of Extreme Automation and Connectivity in Healthcare*. She is also the chair of Big Data for eHealth with the IEEE ComSoc. Contact her at jfiaidhi@lakeheadu.ca.

Sabah Mohammed is a Full Professor with the Department of Computer Science and the Supervisor of the Smart Health FabLab at Lakehead University. He is also an Adjunct Professor with the University of Western Ontario. Moreover, he is the Chair of Smart and Connected Health with the IEEE ComSoc. Contact him at mohammed@lakeheadu.ca.