

Do you really know what lies under the hood of your social media analytics tool?



Between Twitter, Facebook, Instagram, Snapchat, Tumblr, LinkedIn, YouTube, VK, Weibo, Pinterest, Flickr, Medium and the multitude of other such platforms, social media seems to be “driving” almost everything today - political campaigns, news broadcasting services, sports and entertainment events, brand marketing, customer service, public relations and brand reputation. Enterprises are investing millions of dollars into improving their social media presence, and are using social media analysis to drive their decisions, managing their brand’s reputation and assess return on investments.

If you are planning to use, or are currently using, an enterprise level social media analytics tools to inform enterprise level decision process regarding marketing campaigns, investment strategies or public relations, then some of the questions you should be asking your analytics tool vendors or your analytics teams are

1. What is the accuracy and completeness of the data driving the analytics?
2. If needed, can you perform data filtering or cleansing of the input data?
3. How are the metrics/analytics computed or is it being sold as dark art under the guise of AI and machine learning?
4. Does the analytics tool allow you to use analytics as a filter to derive actionable intelligence?

Without first having answers to these questions, tread carefully and don’t get enamored by the pretty graphics, ROI metrics and sales pitches. While this may seem as “*data science 101*” to many, in the world of social media analytics tools, interestingly enough⁴, these attributes are difficult to find in many of the enterprise level social media analytics tools currently available in the market. The heart of the problem lies in the fact that, in the world of social media data and analytics are delivered as an

integrated package. This packaging is done, in part to remain compliant terms of service agreements of the social media data providers. This is further amplified by the fact that, this still an evolving area and standards for metrics have not been defined. Unfortunately, the lack of separation of data from analytics, leaves the burden on the acquirer to now have a much better understanding of what is being offered and how the two are connected. This is quite different from the more traditional model where businesses intelligence (BI) intelligence or analytics tools such as Tableau, SAP, SAS, Pentaho or Micro strategy, are bought to bear on financial, inventory or sales data generated within the enterprise. Data analysts have full access to the data and the freedom to configure the BI tools as needed

To help decision makers better understand their choices while selecting a social media analytics tool, in this two part blog series I will further elaborate on the questions identified in this blog and as to why - you the CxO should care!!!. Questions 1 and 2 are discussed below, while questions 3 and 4 will be addressed in part 2 of this blog.

Accuracy and completeness of the data

Full access to social media data is not easy to procure, nor is it cheap. While most social media platforms offer public API's for their tools, these API's are mostly designed to enable integration of data into vendor apps. In some tools, only a small percentage of the total traffic is exposed via public API's, while in others rate limits, and limitations imposed by terms of service agreements limit the amount of data that can be downloaded or exported. Others provide limitations on how the data can be used. To get full access to data, analytics vendors need to purchase what is commonly referred to "firehose data" from the social media platform vendors. GNIP and Datasift are two of the dominant data providers in today's market. Prior to making any decisions that are driven by the analytics provided by your tool of choice, make sure you have an in-depth understanding of some of the following aspects:



- What data sources does tools have access to? For each data source what kind of access does your enterprise need? Do you need access to the firehose, or just a sampling? If access is managed, then what are limitations? Are you losing any data because of rate limits? Is this loss of data relevant to your analysis?
- For monitoring campaigns in real-time, does the tool provide access to real-time streaming data? For historic data how far back can the data be accessed, last 30 days, since inception etc.
- When searching for data are you limited to use of keywords, or can you use advanced search grammars to refine your searches. What is the complexity of the searches and are there limitations to the number of rules?
- Can you conduct searches in multiple languages?
- When looking at geo-spatial data, is the data being provided only geo-coded or geo-profiled or geo-inferred or all of the above. Are the analytics being compensated for this skewness in data.

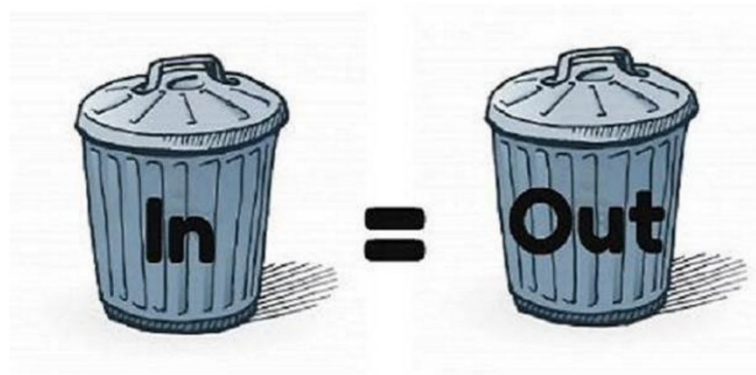
An understanding of this is especially important, since in some platforms only 2- 3% of the data is geo-coded and any analytics based on such data may result in incorrect or biased results.

- Are the enrichments associated with gender, language, and other demographics a part of the raw data from the social media platform are they being enriched/inferred? If inferred what is the accuracy?
- Can your tool of choice provide validation that the data collected meets the substantiated claims?

Data filtering and cleansing

Once you have a good understanding of the data sources, the next step is understanding, whether your tool provides you access to the posts that have been collected based on your search criteria. This is important because the

underlying data used to compute metrics is typically large, in the big-data sense, and is inherently prone to noise. As such, the well-known adage “garbage-in garbage out” applies really well here. 7-10% of users on social media platforms are bots and they trend to drive a significant part of the traffic on social



media. Social media trolls hijack hashtags and user handles to discussing unrelated content, or use it as platform to start unrelated conversations or propagate pornographic material. In many situations 95% of the traffic is generated by 5% of the user-base. Follower count in social media is easily manipulated. As discussed in Forbes <http://www.forbes.com/sites/johngreathouse/2014/01/07/the-cost-social-media-fame-in-2014-about-6800/#518dbbc69ec6>, followers can be purchased for an average cost of .01 per follower (or \$10 for 1K followers). Location data is constantly spoofed by numerous apps. Given these uncertainties in the input data, depending on the use case, some questions to ask of your analyst and tool vendors are

- What level of access do you have to the raw data (actual posts) being used to drive the analytics? Any access, only recent posts, or full access?
- What types of filters can be applied to the raw data for purposes of cleansing, removing outliers, normalizing etc, prior to running analytics?
- Does the tool detect bots and trolls, and if so can this data be filtered/removed?
- Can the data be exported?

Without the ability to filter out the noise and outliers, aggregate analytics don't quite provide the accurate picture needed to derive actionable intelligence.

While understanding the quality and completeness of the data, and the ability to access and filter data is important, the range and fidelity of the analytics offered and the easy with which the analytics can be used to derive actionable intelligence are of even more importance. Stay tuned for part 2 of this blog where I discuss these features.

Understanding the metrics and analytics provided by your tools

Social media metrics and analytics provided by most tools today can broadly be categorized into the following categories (i) basic statistics - “count” related metrics associated with top users, mentions posts, demographics, devices, timelines etc; (ii) social engagement metrics - those related to reach, engagement, impressions etc. (iii) content analytics – those related to named entity recognition, topic modeling, term frequency counts (word clouds) and sentiment analysis (iv) graph analytics – influence detection community detection, link analysis, social graphs, friendship graphs (v) geo-spatial analytics (vi) media analytics and (vii) predictive analytics

While few tools provide all the above capabilities, for those that provide some subset of these, it is important to understand how their metrics or analytics are calculated and what is their value towards providing actionable insights. This is

especially important since there has been no standardization of metrics in this field of social network analytics, nor is there a standardized data-set against which performance of vendor analytics can be compared. As an example, even a simple metric such as Twitter engagement has different definitions. Some tools define it as the sum of mentions and retweets, some include likes in the definition and others include clicks and follows. For other more complicated analytics such as sentiment, and influence detection for proprietary reasons many vendors will not divulge the details of the underlying algorithms.



In the context of metrics and analytics some questions that should be asked are

- What is coverage of the advanced analytics available in the tool and how many of these are relevant and needed to meet your objectives.
- What are definitions of social engagement metrics being provided and how are they different from the native definitions of the platforms
- What is the size of the dataset on which these advanced analytics are being performed? Unfortunately, due to the computation complexity of several of these advanced analytics, several tools only run these analytics on a small subset of the data or a random sampling. The details are hidden in a footnote. Unfortunately depending on the analytic, results run on small data sets are not representative of the what the results might look like when run on the entire data.
- What level of sentiment analysis is needed for your analysis? Is sentiment analysis being performed using a dictionary-based approach, or is machine learning being used. If machine learning is being used what is the training data? How many languages can the analysis be performed in? Are only aggregate statistics available or are post-by post?
- Is topic analysis being performed or is only term frequency count being performed. How does topic modeling work across multiple languages?

- How is name entity recognition being performed? is it gazetteer based or NLP –based? How many languages can NER be performed in? Can entity resolution be performed?
- Can analytics be performed across multiple data sources?

Analytics as a filter to find actionable insights

Finally, what is probably the most important of all is to evaluate whether the tool allows for the use analytics as a filter to find actionable insights or are the results “static”. Once you look past the pretty graphics, a true analyst or a senior executive who has to make decisions based on the analytics, should want to evaluate alternate scenarios.

For example in the context of understanding your brands reach and engagement, can you drill down and understand who the primary contributors of the reach and engagement metrics are? If the large reach is primarily due to a few individuals or influencers amplifying your brand, what would your reach and engagement metric look like if you filtered out these influencers or outliers and re-computed the reach? Similarly, are your engagement metrics primarily being driven by bots. Can you eliminate the bots and re-compute your engagement?



Similarly, when looking at brand sentiment analytics, brand needs to understand what is driving the positive or negative sentiment. Is the sentiment being dominated by a few posts and their re-posts, or by a particular topic or a particular community? What are the sentiment trends outside these communities? For brand reputation management, brands need to be able to execute workflows where-in they can drill down into particular subset of posts, topics or communities, filter by these communities, re-compute all the basic and advanced analytics. For example, if a brand can identify the influencers in the communities that are driving the negative (or positive) sentiment, then they can they can engage with these influencers and execute the appropriate courses of action needed to maintain brand reputation

While the above examples illustrate some simple use cases, as one can imagine the opportunities to explore alternatives, should your tool allow you to do it, are unlimited, and are needed to get the actionable insights from analytics.

Conclusion

If you are planning to use, or are currently using, an enterprise level social media analytics tools to inform the enterprise level decision process regarding marketing campaigns, investment strategies or public relations, it is really important to be able to understand, and drill down into any analytics being presented to you by various tools. Without the ability to do so, analytics are nothing but a set of pretty graphics that can drive you into making the wrong choices with respect to your brand. Unfortunately,

with the flooding of the market with social media analytics tools, and all the buzz about engagement metrics and ROI, choosing the right tool can be quite a challenge. If your enterprise does not have a team to help you navigate this complex maze of social media tools, engage with an un-biased set of subject matter experts with in depth understanding of this space, and those that know how to use a proven analytic process to help brands quantify performance, drive strategic decision making, and improve brand awareness.

About Scraawl



Scraawl is a product of BlueHalo, a defense industry leader in artificial intelligence and machine learning technologies including text, video, and sensor data analytics. Our mission is to develop a comprehensive solution for brands to gain actionable insights by listening, searching, extracting, and analyzing social media conversations and large text corpuses. We have developed a set of advanced analytics that leverages the state-of-the-art in big-data, machine learning, natural language processing, and graph theory. Our business case centers on helping brands and agencies understand their global audience, manage customer experience, improve public relations, drive strategic growth, and develop new clients. With analytic capabilities across multiple languages, and its ability to execute on desktops, tablets or mobile devices, our goal is to provide all the analytics a user will ever need – at their fingertips, anytime, and anywhere. For more information on Scraawl, visit www.scraawl.com to request a demo to learn more about the wide range of advanced analytics offered in Scraawl’s professional, premium, or enterprise packages. You can also sign up for a free personal account and start exploring some of Scraawl’s basic search and analytics capabilities.