



Using NodeXL to decipher big data

| By Michael Lieberman

snapshot

Michael Lieberman tests a free, open-source social media data analysis tool.

In a recent piece in *Forbes*, Mark Fidelman asked, “What if instead of a score, you could visualize the impact a person, business or topic has in a social network? What if instead of using complicated listening tools, you could see in an instant who is talking about your company or its products and how you’re connected to them?” This, of course, is a central theme in the ongoing campaign to tame big data.

One of the most onerous challenges facing the marketing research industry today is to dam and direct the raging flow of social network data being generated each second. As researchers, we need to make things relevant. We need to tell the story.

Marc Smith is a sociologist specializing in the social organization of online communities and computer-mediated interaction; he and I met at a recent predictive analytics conference. Smith leads the Connected Action Consulting Group in Silicon Valley and cofounded the Social Media Research Foundation. Smith has pioneered a free, open-source graphics program, NodeXL, which synthesizes and clusters social network data. This analysis is called social network analysis (SNA). Instead of a complicated listening platform, NodeXL – an Excel add-on – is able to synthesize, for example, various Twitter feeds and produce a relevant graphic and report. NodeXL creates maps that make sense of social media – and that is just the beginning.

Thanks to Smith’s mentoring, I have functionally mastered NodeXL and learned how to make its output relevant to the marketing research industry. This article will present

the fundamentals of SNA and NodeXL. I will provide sample Twitter and Facebook maps and show how they may be used for understanding brand conversations taking place in the realms of social media.

Network theory

Social network analysis views social relationships in terms of network theory, consisting of nodes (representing individual actors within the network) and ties (which represent relationships between the individuals, such as Facebook friendships, e-mail correspondence, hyperlinks or Twitter responses). These networks are often depicted in a social network diagram, where nodes are represented as points and ties are represented as lines.

Figure 1 is an example of a Twitter NodeXL social network graph. Instead of using listening tools, these Twitter maps organize and visualize content in a way that makes them very easy to interpret.

- We can see who is talking about the brand (in this case, BMW).
- We can tell who are major influencers or connectors and what they are saying. (A connector is someone whose tweets are heard by a large number of people, oftentimes in different “clouds” within the graphic.) If we look at Figure 1, we see an oval that shows several connectors. These are people that the brand might want to contact directly through Twitter.

At first it may be hard to decode the map. In short, given that BMW is a major brand,



quirks.com/articles

ID 20140105

Figure 1

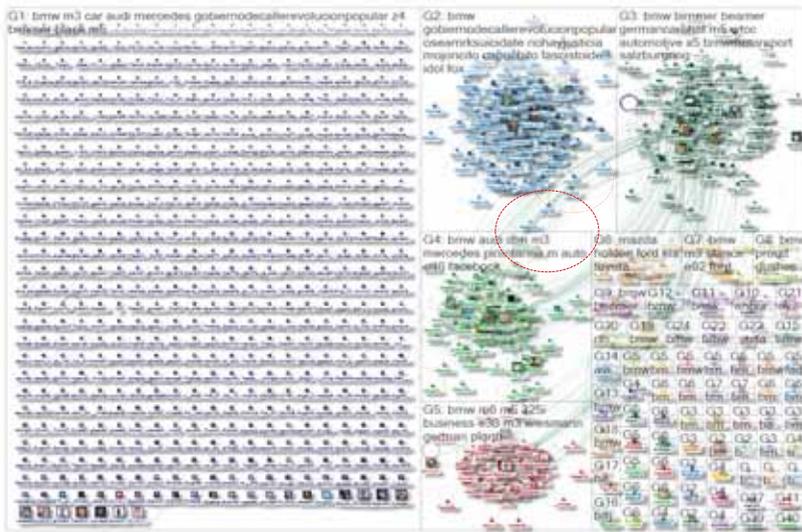


Figure 2



Twitter map. This is called a broadcast map, where an individual account creates a large buzz. Examples might be Lady Gaga, the GOP, a news channel or the Chicago Bulls.

In our example, CNNMoney is the center of the broadcast network. Its hashtags cluster around the day's major stories. Interestingly enough, in our example Twitter is a major story for CNNMoney, probably because Twitter announced its IPO the day before I ran the map. Also trending for CNNMoney is Obamacare. Not surprising, CNNMoney is connected to other news outlets, the most visible being CNN.

Again, where to begin? Looking closely at Figure 2, one can see that CNNMoney has many direct followers and broadcasts to a large number of small clusters (G3 thru G50) and islands (G2). These kinds of broadcast maps are useful not only for content but also for reaction. For example, if a football team bombs on Sunday, what are the responses by its fans? (Though not included in the map, NodeXL does record all tweets within the search for keyword and sentiment analysis.)

In addition to Twitter, NodeXL analyzes e-mail networks, hyperlinks, Flickr and Facebook friends, Likes and groups. Figure 3 is an example of a Facebook-focused map. Unlike Twitter, Facebook networks are not publicly available and a password is required to collect data from a user, so I used my own network of Facebook friends. I do not use Facebook for my firm Multivariate Solutions, so my account reflects life clusters. It is instructive to see how NodeXL clusters my Facebook friends.

NodeXL surveys all my friends and clusters those in my network who are connected to one another. It then maps these people around these shared connections.

Examining Figure 3, we see three dominant clusters. One is my high school class – the original reason I got on Facebook. One is my current social circle. The third is family. I have several other small clusters of friends but to show how robust NodeXL is, I put a small blue circle around a tiny cluster at the bottom, right-most corner of the

there are a lot of people who are talking about it who are not connected to other Twitter users. These nodes are referred to as islands. Most brands will have a large cluster of islands. In Figure 1, section G1 is the large group of dots to the left. These are people not directly connected to the main BMW Twitter account but are discussing topics related to BMW. These are people with whom BMW might want to talk in the future.

In sections G3 and G4, clustered people are discussing different aspects of BMW. The lines that connect these mushroom-shaped clusters are “connectors,” the people whose voice is heard in both groups.

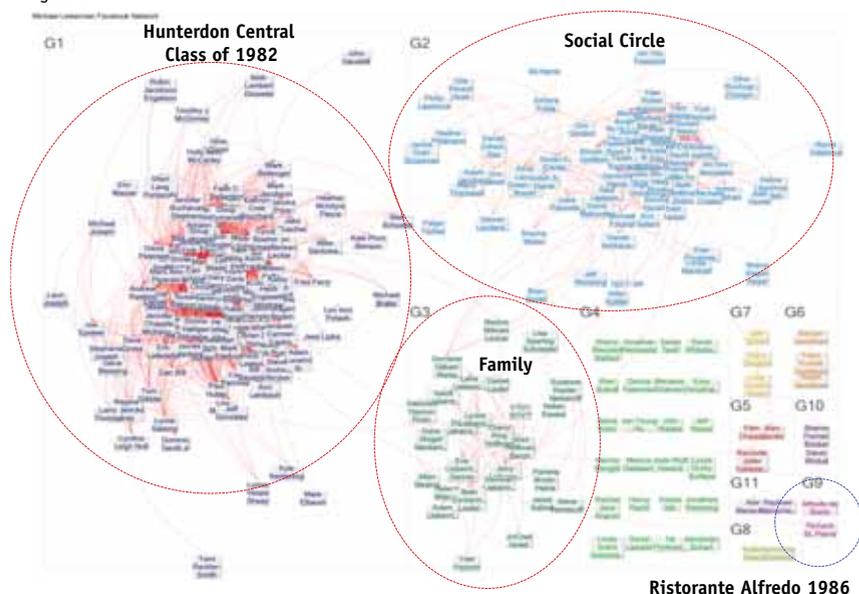
The clusters in the BMW map, the

multicolored clouds of names, are people who are in contact with each other either by retweeting or responding. If we look to see which Twitter hashtags cluster in each group we get a pretty good idea of what they are saying about BMW.

What's exciting about this is that NodeXL allows companies to understand not only what is being said in the social media sphere but also to identify their most efficient messengers. This allows firms to ferret out prospective customers and identify influencers and allows brands to test social media campaigns by monitoring NodeXL Twitter maps over specified periods of time.

Figure 2 is another type of common

Figure 3



map. These are a few people with whom I worked at Ristorante Alfredo in New Brunswick, N.J., in 1986 while I was an undergraduate at Rutgers University.

The Facebook Likes page of a company like Starbucks is a treasure trove of social network information. NodeXL pinpoints the influencers

and their opinions. This would allow Starbucks to directly contact a well-connected Facebook customer and offer him a free latté. NodeXL could also cluster people who are saying negative things about Starbucks and allow the company to reach out to them in a positive way.

Harness the power

We are living in an increasingly saturated world. Facebook, Twitter, Google, smartphones – more data is being produced daily than was created in the first 4,000 years of human existence. There is so much marketing noise that even big TV networks and news giants no longer have the ability to create overnight product success. The result is a growing effort by smart organizations to augment their campaigns through proven media channels with social network advertisements. That is, these companies have learned to harness the power of thought leaders, experts and influencers to promote their products. As we move forward, social network analysis, using tools such as NodeXL, will no doubt play a larger and larger part in this emerging field. ①

Michael Lieberman is founder and president of Multivariate Solutions, a New York statistical consulting firm. He can be reached at 646-257-3794 or at michael@mvsolution.com.