Discovering Communities in Social Networks

Depending on the social network, a search for posts based on keywords or tags can return hundreds of thousands, even millions of posts. For example, a Twitter search on the keyword Rio2016 resulted in over 30 million tweets. As one can imagine, even though all of the posts in this large corpus contain the keyword, there is in fact a wide variety of topics being discussed. There were communities of users discussing specific events such as the opening or closing ceremonies, while others were talking about medalists in one or more of the events such as aquatics, track and field, or gymnastics. Similarly, others were re-tweeting or posting about individual athletes such as Michael Phelps, Katie Ledecky, or Usain Bolt. Some users were even posting about or retweeting celebrities such as Katie Perry or Zac Efron (interestingly they were among the top retweeted and mentioned users in the corpus of tweets related to Rio2016), while others were talking about the achievement and medal counts of their countries, or about brands and their sponsorships. All this happened in over 50+ languages around the world.

Social media analytics tools help brands and agencies sift through this large volume of data to drill down into a subset of relevant posts, and then derive actionable insights that are relevant to their brands. Most analytics offered by tools are focused on analyzing the content of the post, and present analytics related to demographics, term frequency counts, sentiment, etc. But a natural question that comes to mind is – can one derive any actionable insights based on the rich structure of the underlying social network that is implicit in these conversations? In particular, does the underlying social graph reveal a community structure in which particular sets of nodes in the network are more densely connected with each other than with the rest of the network, and does this natural division in the network have any correlation with the topics of conversation of these communities?

Interestingly enough, it turns out that, depending on the representation used to define the underlying structure of the social network, graph theory reveals several interesting insights about the nature of the conversations. Conversations between members in a given community (as defined by graph partitioning) typically revolve around the same set of topics. This happens more naturally in social media platforms such as Twitter, Instagram, and Tumblr. In these platforms, users who have never interacted before can retweet or quote posts, or reply to or mention users who are not in their network. The behavior is an emergent one. This insight can provide a new way to identify topics in a corpus of social network data based on graph analysis - which is inherently language- and content–independent, and leverages the underlying structural dynamics of the social graph. The challenge lies in developing the appropriate graph representation and then being able to identify the underlying communities correctly. While several techniques such as hierarchical clustering, clique-based methods, statistical inference modularity maximization, label propagation and others exist in literature, the realization of this theory in the context of social media networks has been limited in-part by the complexity and scale of social networks and the need to identify meaningful communities in real-time or close to real-time.

Scraawl®, a social media analytics tool developed by BlueHalo, provides its users with the capability to detect communities in a social conversation. Scraawl's community detection analytic uses a graph-theoretic approach to partition the social interaction graph into communities. Scraawl provides community detection statistics including number of communities, and the average and maximum community sizes. As shown in the figure below, details of the top ten communities and the most connected users/hashtags in these communities are shown. A user can then filter the dataset by one or more communities of interest. Topic modeling or influence discovery analytics can then be performed on specific communities. Scraawl's community and social graph visualizations allow a user to further drill down into the communities to see how the members are connected.

Communities

Rank	Community Label	Community Size
1	#giveaway #samsung @accounts4cheap	1739
2	@bestbuy #ad @cmarienico	1716
3	@stacksocial @makeuseof @geekygadgets	1345
4	@youtube @androidauth @proandroid	1267
5	#browsefaster @operaindia #contest	1141
6	#tech #love #news	575
7	@tnwdeals @androidpitcom @thenextweb	550
8	#deals #ebay #usa	360
9	@kir1ca @gadget_shot @nokiadou_jp	331
10	@boingboing @mrjeremylcoope1 @geekdads	303

Top Mentions – '#giveaway #samsun...

Username	Interacting neighbors in community	Mentions in report	Posts in report	Users in report
@accounts4cheap 🔗	243	248	248	241
@giveawayfazz 🔗	85	N/A	N/A	N/A
@androids 🔗	69	N/A	N/A	N/A

Top Hashtags – '#giveaway #samsun...

Hashtag	Interacting neighbors in community	Hashtags in report	Posts in report	Users in report
#giveaway	345	452	448	345
#samsung	323	775	769	323
#android	214	384	383	214
#win	79	167	167	79
#galaxys7	78	198	198	78
#rt	76	263	263	76
#galaxys7edge	71	192	192	71

The figure above shows the results of a Scraawl community analysis on a set of tweets that included the keyword "Samsung S7 edge." Scraawl has partitioned the conversations into different communities which are labeled by the popular tags and users in the community. These labels provide insights into the topics of conversations and users participating in these conversations. A drill down into the communities shows that members of Community 1 were using hashtags #samsung, #giveaways, and #win and mentioning handles that were discussing deals, giveaways, exchanges, etc. The community graph belowshows the interactions between the users and tags in Community 1. The users and tags in Community 2 indicate that conversations in this community were focused on a deal from BestBuy for a

Scraawl
Actionable Insights from Analytics   www.scraawl.com   BLUEHALO
A product of BlueHalo

$150 gift card and a Galaxy S7 Edge Pink Gold. Similarly, Community 5 was focused on posts related to #BrowseFaster #Contest with @Gadgets360 in India. As discussed earlier, brands can filter by the community of interest to further analyze the data. What's interesting to note here is that all this was done based on graph analytics and no analysis of the post content was done!!
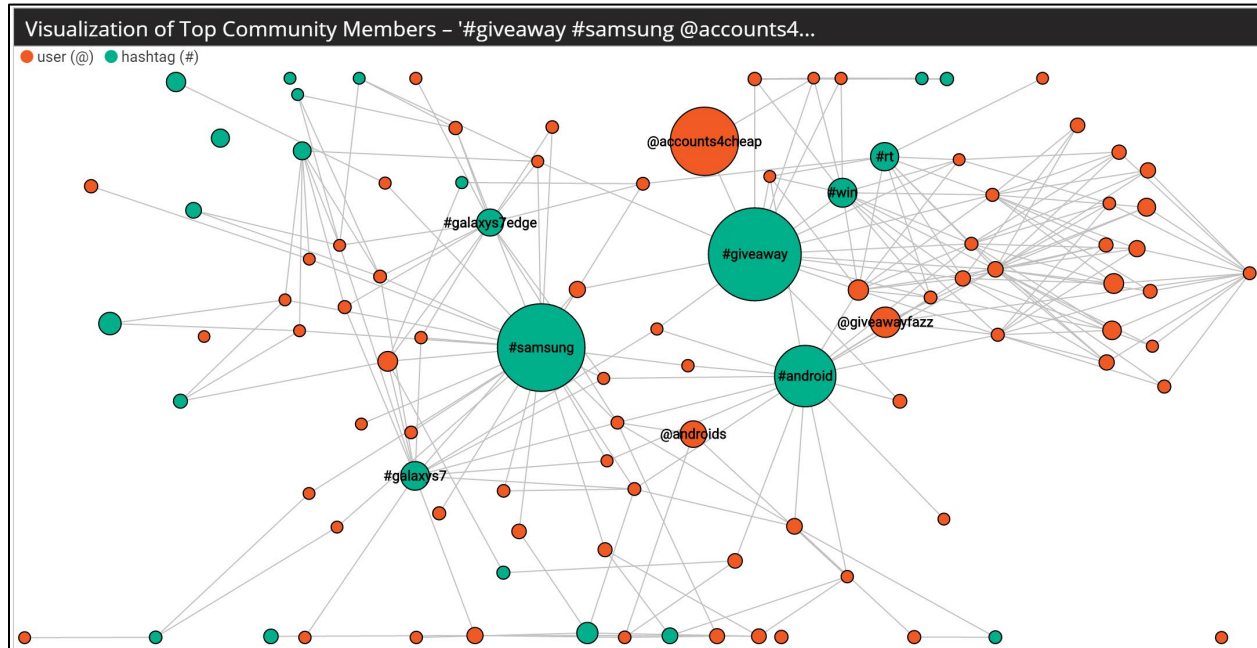
Visualization of Top Community Members – '#giveaway #samsung @accounts4...'

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 to start exploring some of Scraawl's basic search and analytics capabilities.

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.