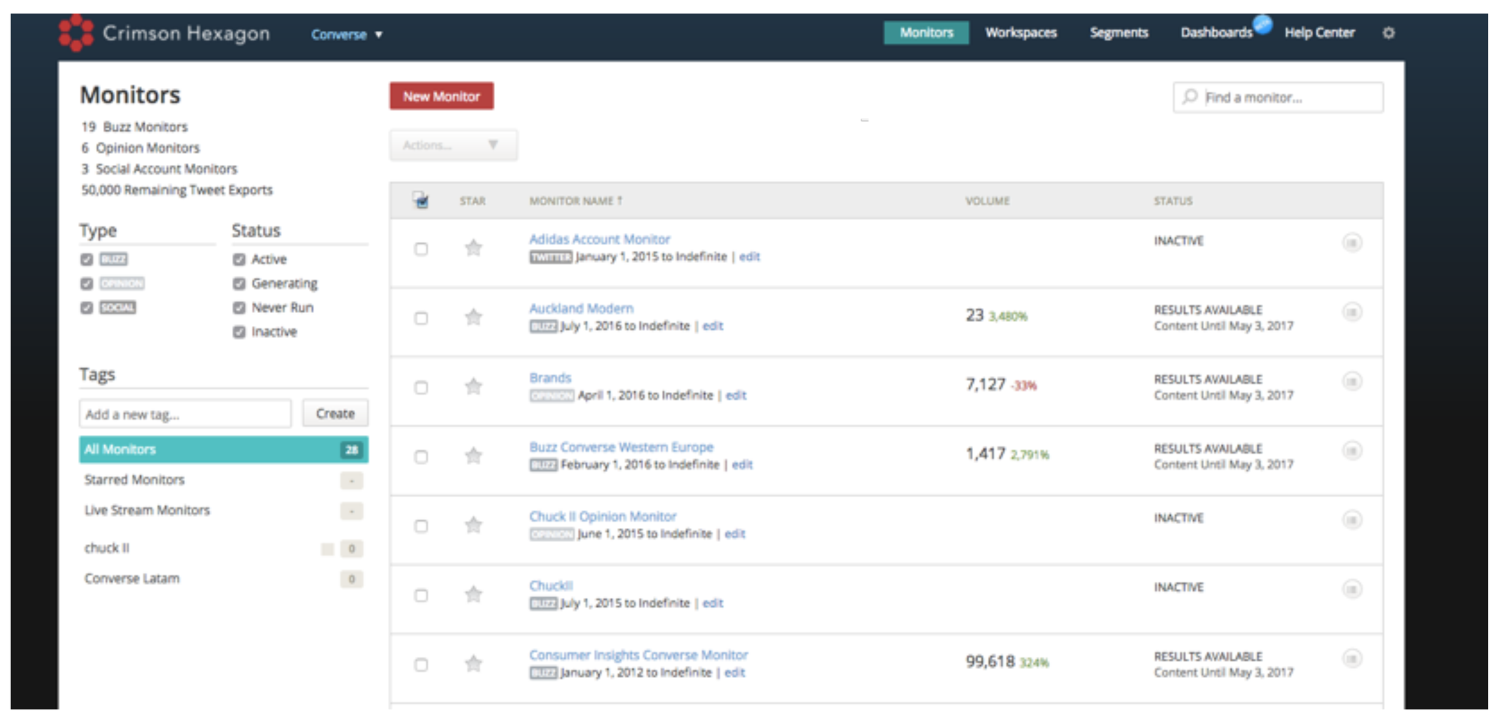
BU Department of Computer Science

bu.edu/cs

Understanding Information Propagation and Building Potential Influence Networks Using the Twitter API To Examine Converse Product Launch Effectiveness

Data Collection



Crimson Hexagon Monitor Page

Converse provided us with a fantastic tool called Crimson Hexagon, which bypasses Twitters 3-week data limitation. By setting up a monitor, Crimson is able to take in data marking a specific topic (ie: Converse, Chuck ii's, Stan Smith, Adidas) store the data over years of monitoring, and subsequently performs basic analysis such as sentiment analysis, geographic analysis, and word frequencies. We were able to get our preliminary data sets around the product launch times of Adidas' Stan Smith and Converse's Chuck ii which show the list of posts on Twitter around each of these launches. This got us as close to the source as possible.

ource	Klout Score	Gender	Posts	Followers	Following
vitter	71.0	NaN	53452	43263	2932
vitter	68.0	NaN	76618	10049	1224
vitter	69.0	M	31326	15785	6109

Example Dataframe Created from the pulled Crimson CSV's

We filtered these lists of posts by only looking at users with Klout scores above 65. For those unfamiliar to the index, Klout scores indicate a users influence across numerous social media platforms, and are indicative of a users ability to drive awareness. This served as the first layer to our network, and showed how many high influence users we had simply tweeting about the products. The question now was…what are these tweets doing?

Potential Influence Network Algorithm using Tweepy to Access Twitters API

1) Retrieve first level Users with high Klout from Crimson Hexagon Data and generate a list of these users
2) Search through each follower list of these users to find users with over 1000 followers. Return a list of these users
3) Recursively call the above steps for a max of 3 layers on user lists

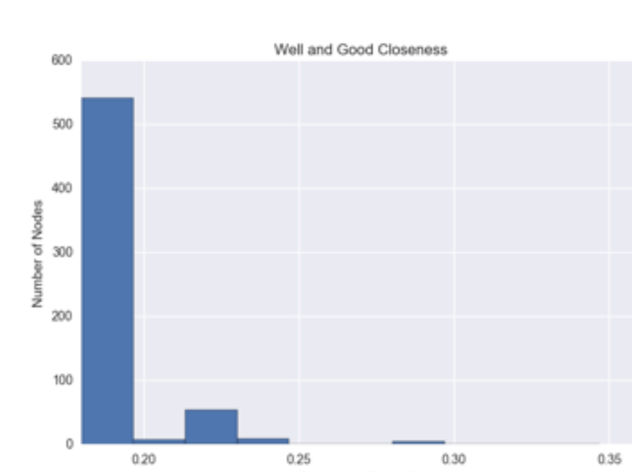
Abstract

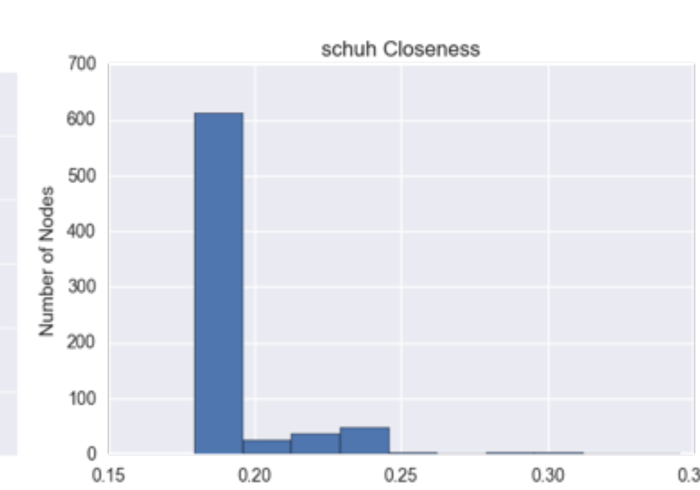
Social media analysis is a critical application of technology to market analysis. In this case, we employed Twitter analysis to analyze the social landscape around product launches. Mainly, we were interested in examining how dense networks are developed, and what users currently generate a lot of "buzz" around product launches. Our hypothesis rested on this central claim; successful product launches rely on people with **dense, influential follower networks** to tweet their product into new social realms. However, this might not always mean that getting high Klout, high status users to tweet will generate this dense propagation network. For instance, Uber recently launched a campaign, and found that 90% of traffic was driven by retweets from a tweet Justin Bieber made; not many of these users who tweeted were very influential. This leads us to believe that product launch effectiveness rests on the ability to jump into various social circles starting at one central node, which only occurs if, as the layers of propagation increase, influential users can influence influential users tweet the product into their respective realms of influence.

Thus, here we attempt to examine Converse vs Adidas launches on 2 hallmark products to see if we can see whether more successful launches have these dense, 'new market' networks we have described above.

Result Interpretation

It should be noted that the results serve as a baseline for understanding the various ways information can spread. Each of the networks were run on a maximum of 3 layers due to Twitters rate limit, and we were only limited to follower subsets from Twitters API rather than the whole lists. The system we have established, despite its data collection limits, still shows the various ways influence can be projected on social media, and how 'not all klouts are created equal' since some individuals may be able to reach different markets based on the people in their sphere of influence. Running this code using Firehose on powerful computers may prove to be a good way to target highly influential, cheaper spokespeople for product launches.





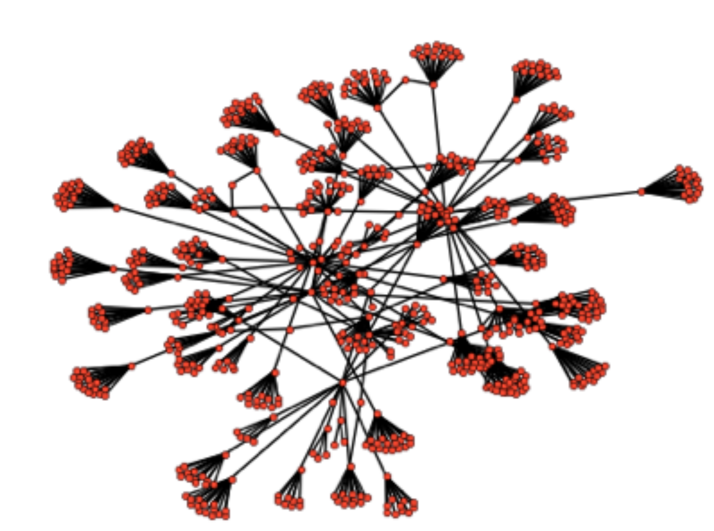
	Brand	User	Klout	Followers	Center	Nodes	Edges	Diameter	Average_Degree
0	StanSmith	@ochocinco	82	3593408	@ochocinco	147	146	6	1.986395
1	StanSmith	@WellandGoodNYC	81	66585	@WellandGoodNYC	617	637	6	2.064830
2	StanSmith	@HYPEBEAST	81	338942	@HYPEBEAST	207	219	6	2.115942
3	StanSmith	@HotNewHipHop	80	643159	EZZYSUMSERIOUS	141	143	4	2.028369

	Brand	User	Klout	Followers	Center	Nodes	Edges	Diameter	Average_Degree
0	Converse	champssports	80	368848	cameronjlogan	86	86	4	2.000000
1	Converse	Converse	74	1055504	Converse	308	309	6	2.006494
2	Converse	djtommytrash	81	375256	djtommytrash	270	272	6	2.014815
3	Converse	guestofaguest	82	29574	guestofaguest	246	247	6	2.008130
4	Converse	rikerR5	82	873141	rikerR5	350	360	6	2.057143
5	Converse	SavedYouAClick	77	199419	Codebryo	63	62	4	1.968254
6	Converse	schuh	79	142388	schuh	735	745	6	2.027211

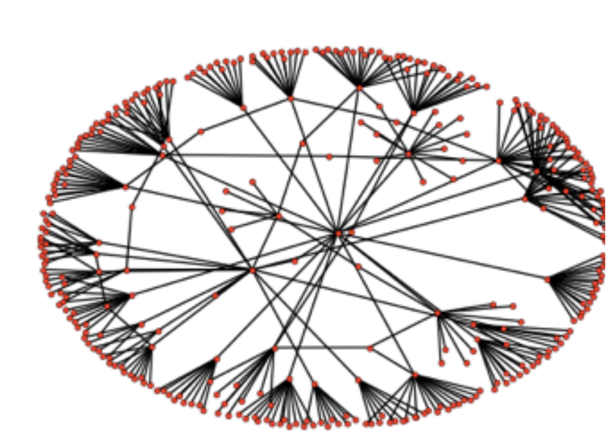
Data Analysis

Chuck 2; Top "Launch-Tweeters" Potential Influence Networks

schuh

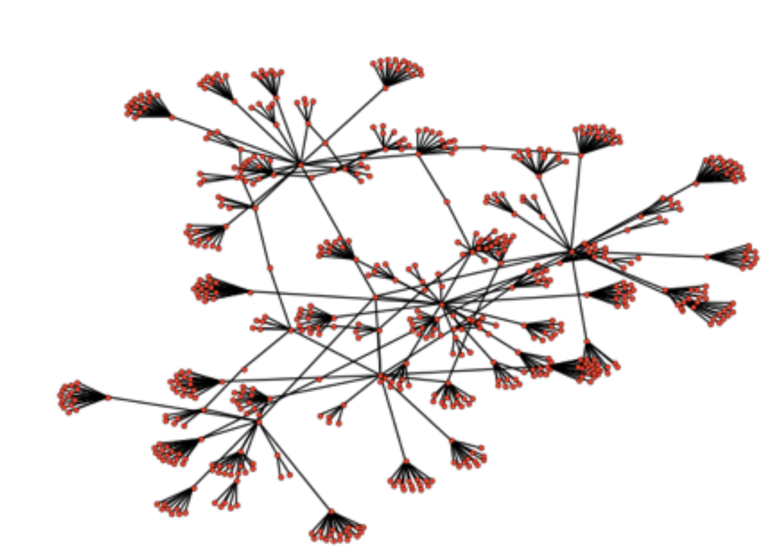


rikerR5

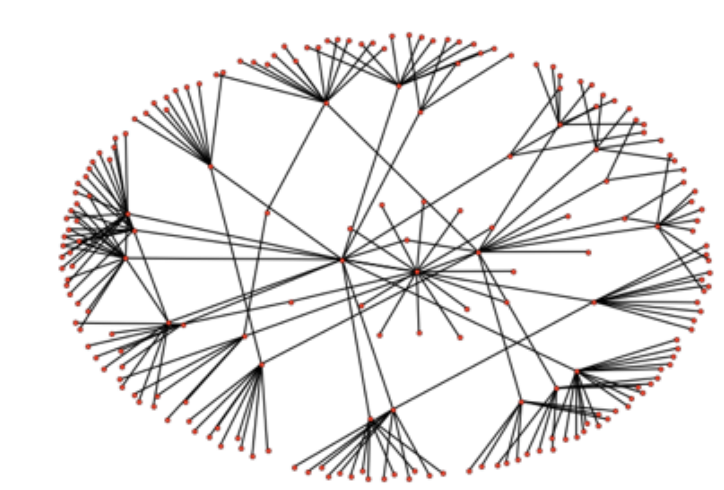


Stan Smith; Top "Launch-Tweeters" Potential Influence Networks

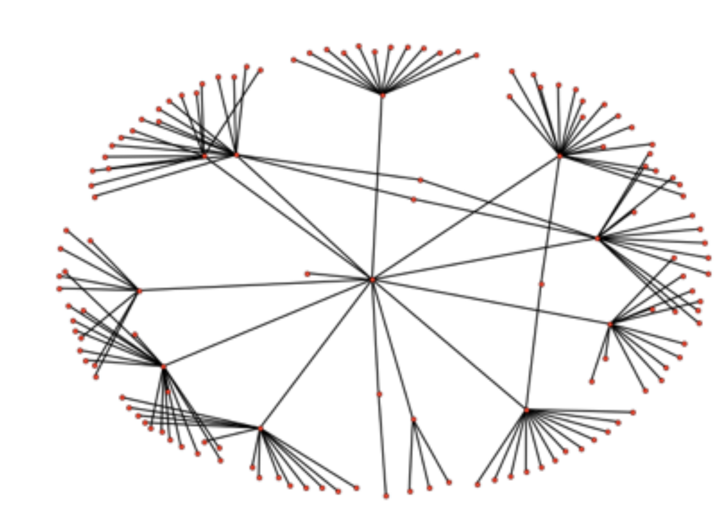
Well and Good

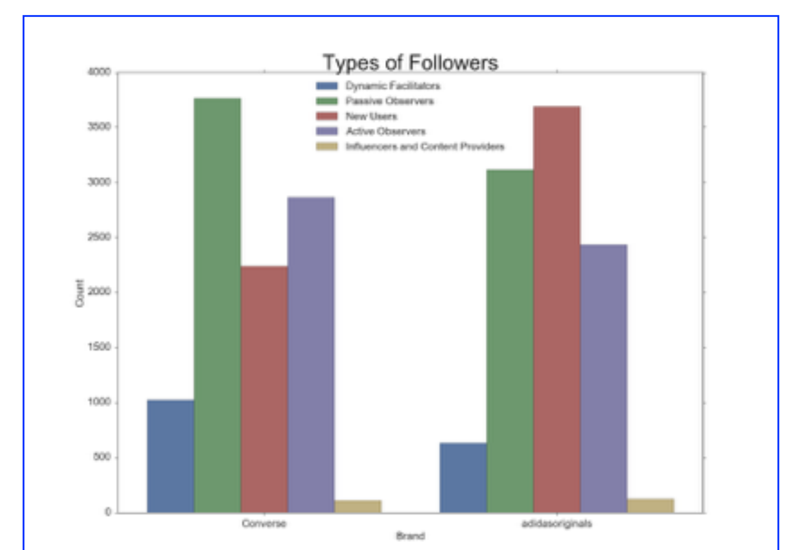


HYPEBEAST



HotNewHipHop





Clustering Results on Types of Users that Follow Adidas vs Converse on Twitter

Linear Regression Showing Strong Correlation Between Followers and Klout, which proves our algorithm has reliable influence determination

```
Linear Regression Output:

                            OLS Regression Results
==============================================================================
Dep. Variable:            Klout Score   R-squared:                       0.960
Model:                            OLS   Adj. R-squared:                  0.960
Method:                 Least Squares   F-statistic:                 3.711e+04
Date:                Thu, 04 May 2017   Prob (F-statistic):               0.00
Time:                        15:53:14   Log-Likelihood:                -15708.
No. Observations:                4589   AIC:                         3.142e+04
Df Residuals:                    4586   BIC:                         3.144e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Posts          1.4300      0.072     19.867      0.000         1.289     1.571
Followers      5.0026      0.117     42.700      0.000         4.773     5.232
Following     -1.1041      0.107    -10.354      0.000        -1.313    -0.895
==============================================================================
Omnibus:                       67.729   Durbin-Watson:                   1.949
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              118.981
Skew:                          -0.086   Prob(JB):                     1.46e-26
Kurtosis:                       3.770   Cond. No.                         15.8
==============================================================================
```

These following charts display info about each potential influence network, and shows how closeness in the network is a good indicator of how well these users are influencing various markets based on the followers they can reach

Note: "Well and Good" shows a high influence graph, "HYPEBEAST" shows how some followers spread out the graph while others don't, and HotNewHipHop isn't actually the center, one of his followers is, which shows the varied types of networks we can obtain.

@BUCompSci