SOCIAL NETWORK ANALYSIS OF AN AMAZON PRODUCT BASED ON CUSTOMER COMPLAINTS

Group Members:

Mr. Sandeep Ramesh

Mr. Mohamed Naumani

Mr. Venkatesh Singanallur

Table of Contents

Project Description:

The success and failure of a product rests with its user. User based reviews are more common to find hence the project is solely based on the customer complaints. The product in focus here is Asus C201 11.6inch Chromebook laptop with a decent specification. It has a 4star product rating with a nominal price of $180. The plan of attack for this project was to collect the 1star and 2star customer reviews through Web Scraping. The collected reviews were then tokenized and then subjected to pre-processing steps to get unique and lemmatized words. The pre-processed words were then given as an input to a Word2Vec Model. Google's Word2Vec is a deep-learning inspired method that focuses on the meaning of words. Word2Vec attempts to understand meaning and semantic relationships among words. It works in a way that is similar to deep approaches, such as recurrent neural nets or deep neural nets, but is computationally more efficient.

The result is a list of lists with similar meaning words of a given word. The words are then visualized as a network in NetDraw software and further analysis like Degree Centralization, Eigen Vector are done to check out the main complaint of the product. Through this project, we have eliminated the step of reading all the bad reviews and focused on the main complaint of the product which may not have been possible by reading all reviews.

Dataset:

The dataset is a list of customer reviews of an Asus Chromebook Laptop from Amazon.com. Web Scraping of 1star and 2star customer complaints were obtained using Beautiful Soup library in Python. The resultant data was stored in a text file for further analysis. Link of the product is given below.

https://www.amazon.com/Chromebook-Rockchip-Pearl-White-Light/dp/B01EGBAQXY/ref=cm_cr_arp_d_product_top?ie=UTF8

Data Cleaning:

The reviews were first tokenized to words. The tokenized words were Lemmatized and stop words along with punctuation were removed. Textacy Library was used to remove Punctuation, Numbers, Accents and Currency Symbols. The final set of words were then stored as a list which was then fed as an input to Word2Vec Model.

Model:

Google's Word2Vec model was used to generate our list of words with similar meaning. The result set were the most common words occurring in the dataset along with their list of similar words for each word. Words with a similarity measure greater than 0.5 were considered. The words were then paired and converted to a Ucinet compatible data to be displayed as a network.

Analysis of Network Data:

- N Cliques:

A n-clique of an undirected graph is the maximal subgraph in which every pair of nodes is connected by a path of length n or less. This analysis gives information on the number of times each pair of nodes are in the same n-clique and gives a hierarchical clustering based upon this information.

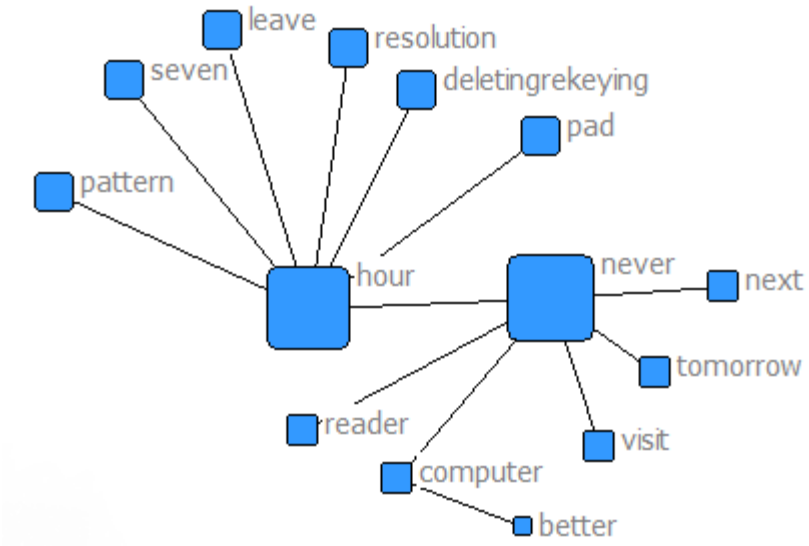
The Value of N is set as 2 which means that all members of an n-clique are connected by a path of length 2 or less. The minimum set size of the clique is 6, this will give a list of issues with the product. The result of this analysis is shown below, the cliques are ranked based on their size with the largest clique being of Size 8.

While components give the linkage of words with any distance of n, usage of cliques will give a closely-knit group of words. The below cliques give an idea of the most closely related issues that occur with the product.

```
N-CLIQUES
------------------------------------------------------------------------------

Max Distance (n-):                  2
Minimum Set Size:                   6
Input dataset:                      word_ucinet (C:\Users\lamps08\Documents\UCINET data\word_ucinet)

272
10 2-cliques found.

   1:  never hour pad resolution leave deletingrekeying pattern seven
   2:  computer never reader visit hour next tomorrow
   3:  screen request biggest crashing air thing
   4:  support said keyboard hdmi priced ill
   5:  like wifi t100ha displaying extremely suggest help
   6:  like point best loved wifi login
   7:  warranty said keyboard reappeared reimburse quickly
   8:  coincide still removed caused interesting replaces
   9:  quick new statusi replaced company crappy
  10:  go bookmark video comment actual microsoft
```

- Eigen Vector Centrality:

The centrality measure in this method is calculated based on the sum of the centralities of the adjacent nodes. The approach is to find the most central nodes of the network.

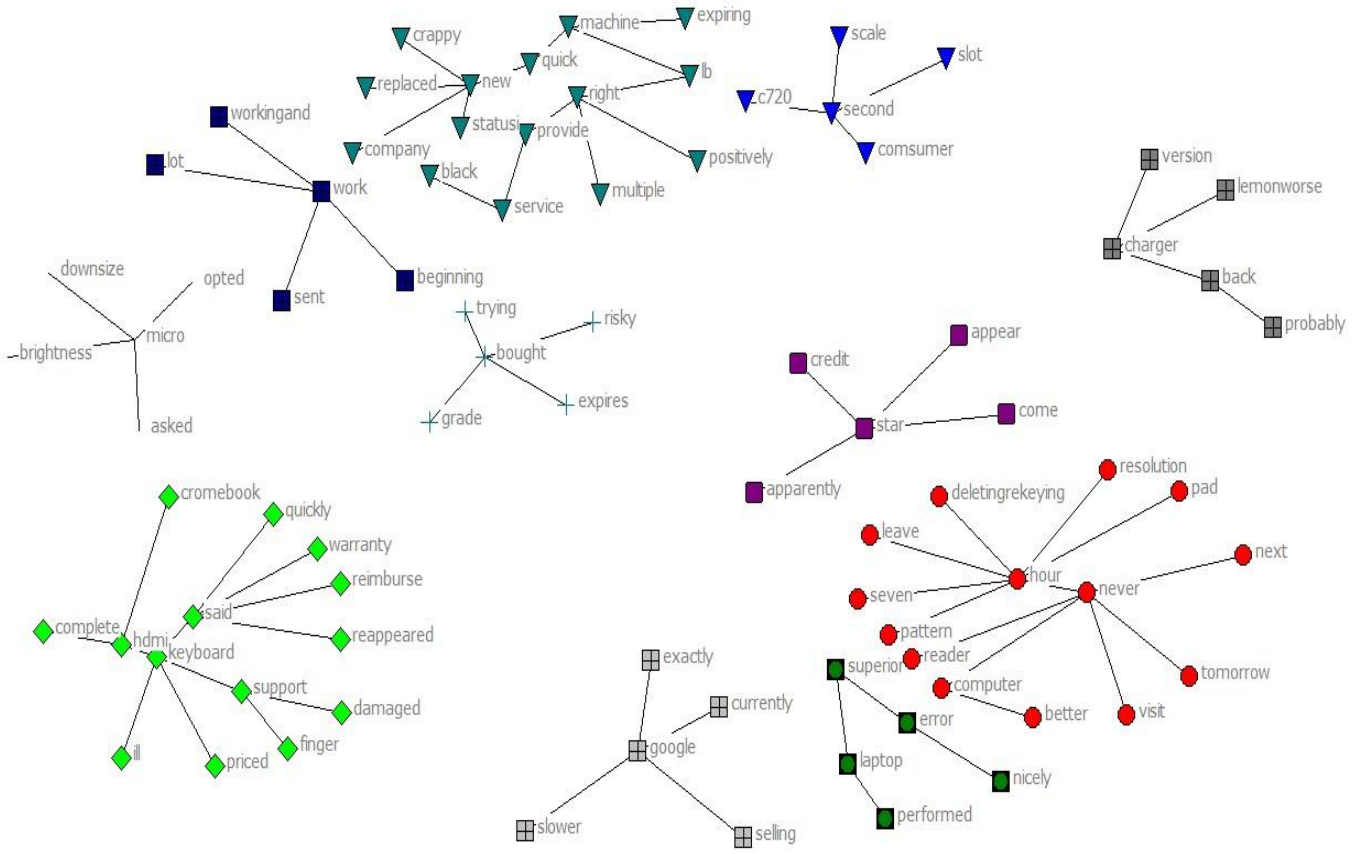


The above component is the only one with high Eigen Vector centrality. This component gives an insight to a range of issues that are most central to the product and needs to be addressed immediately. The example for such issues are screen resolution, delete key, card reader, customer support and Key pad. Below are the values for the eigen vector centrality of the network. High values indicate the node is central.

Positive eigenvalues of word_ucinet

	1 Eigenvalue	2 Pct Variance	3 Cum Pct
1	2.918	2.323	2.323
2	2.690	2.141	4.464
3	2.662	2.118	6.582
4	2.358	1.877	8.459
5	2.358	1.877	10.336
6	2.303	1.833	12.169
7	2.236	1.780	13.949
8	2.216	1.764	15.713
9	2.194	1.746	17.459
10	2.175	1.731	19.191
11	2.149	1.710	20.901
12	2.101	1.672	22.573
13	2.074	1.651	24.224
14	2.074	1.651	25.875
15	2.061	1.640	27.515
16	2	1.592	29.107
17	2	1.592	30.699

- Component Analysis:



A total of 57 components with component sizes ranging from 2 – 15 were obtained from the analysis. Components of larger sizes such as 14,15 gave a generalized information on the product reviews whereas components of smaller sizes like 5 gave a specific information on the complaints faced by the users in the product. Component analysis mainly helped in interpreting the core issues of the product from the superficial ones.

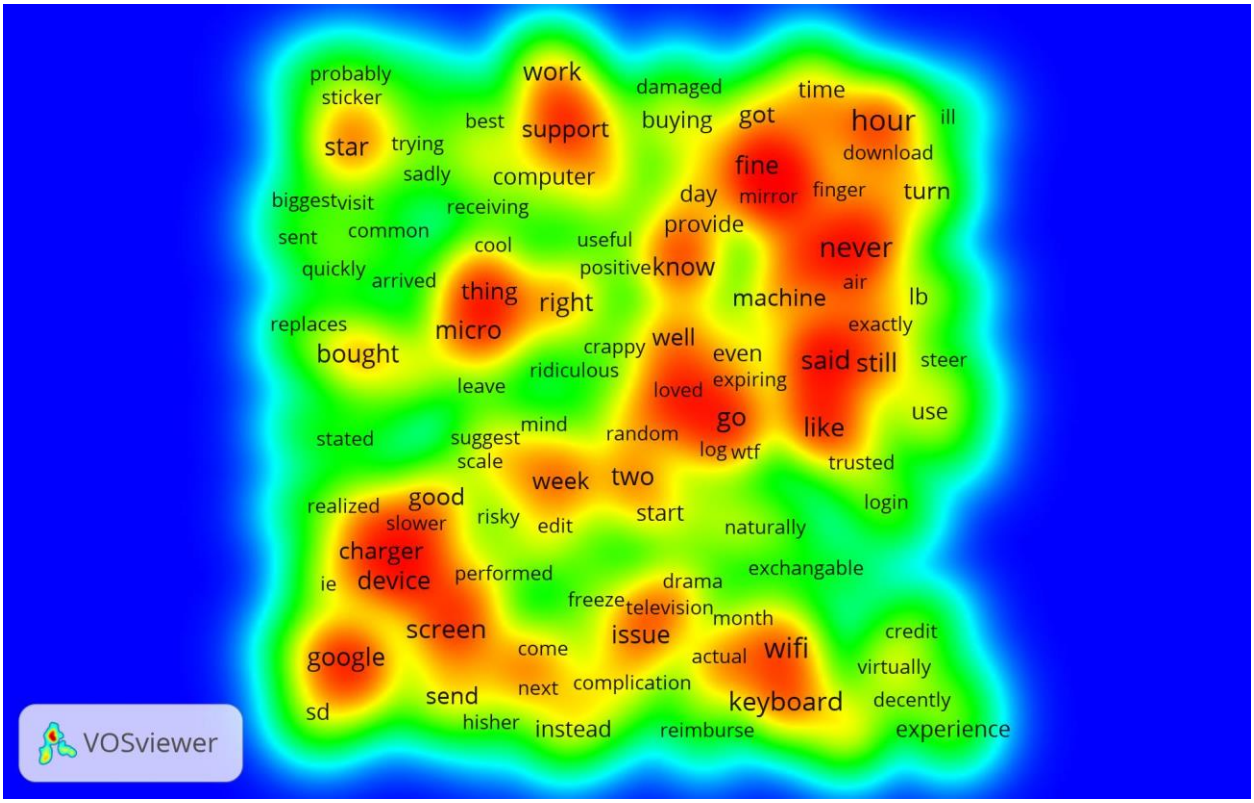
For example, consider the components of size 5(node – second, node - charger). These nodes pertain to a certain problem with the laptop. Lemon worse is a way of saying that the charger is bad. Comparatively components of size 15 (node – Keyboard) is facing a range of issues like HDMI, Support issues, Warranty, Reimbursement, Damage which covers a broad topic of complaints.

- Betweenness Centrality:

The Betweenness Centrality gives a measure of the most frequent nodes which participated in most of the complaints from the users. Through this analysis, we can once again infer that the node with the highest Betweenness centrality is the most occurring word in most of the reviews. The Betweenness and nBetweenness values of the most important nodes are given below.

Word	Betweenness	nBetweenness
keyboard	62	0.169
never	56	0.153
lb	48	0.131
wifi	35	0.096
support	23	0.063
hdmi	23	0.063
screen	18	0.049
replacement	13	0.036
service	13	0.036

Density Visualization:



The Density Visualization gives a perfect picture of the complaints addressed by the users. The areas in red signifies the critical issues of the product or the components of the product itself and the areas in green represent more general words used in the reviews. This correlates with our results obtained using Word2Vec Model and more significantly correlates with the analysis of our results obtained using Ucinet.

Conclusion:

Through these analysis, we have been able to achieve the objective of our project which is to identify the key issues pointed out by the users. This idea can be implemented on any product to correct the issues and focus more on what the people need. A more up-scale idea based on this project can be used to compare a successful product versus an unsuccessful product.

References:

Dr. Sridhar Nerur, Professor, University of Texas At Arlington

https://rare-technologies.com/word2vec-

http://textacy.readthedocs.io/en/latest/api_reference.html#module-textacy.extract