

Submitted by  
SOUHAM BISWAS  
CWID: A20365242



# Business Expansion Engine

---

# Problem Statement -

- ◉ When a new product is to be launched, it is crucial that the correct people are targeted for advertising.
- ◉ If those people having the highest affinity to the commodity you are publicizing are targeted, or in essence given a “free sample”, there is a high chance of the product gaining high popularity.
- ◉ The business expansion engine seeks to suggest a list of people within the user’s social network to whom the user should in essence give a “free sample” of the commodity being publicized so as to maximize the overall reach of the product.
- ◉ It is based on the concepts of Cascading.

# Approach

- Collect keywords pertaining to the commodity the user wishes to publicize.
- Get access to the user's social network and calculate the payoff of the commodity for each person in the network based on the pages/celebrities they follow.
- The payoff is calculated by computing a 'affinity coefficient' per input keyword per page followed by the person. Then, this quantity is normalized to get an approximate payoff.
- Get the edge weights, i.e. the influence each person has over his/her neighbors by analysing online interactions.
- Compute the set of initial adopters for maximum spread by analysing the payoffs, and the edge weights.

# Data

- List of friends of the user
- List of friends/pages followed by the user's friends.
- Mainly, the fields pertaining to the name, username, number of followers, 'verified' Boolean and list of friends were used.
- The data was collected using the Twitter API.
- Bing Search API used in the payoff computation process which involved programmatically searching word pairs to calculate their relevance respective of each other.

# Results

- Based on a small set of manually labelled test cases based on my friend list, the software performed with around 75% accuracy.
- For example, if the keywords are 'tech', 'electronics' etc., people with inclinations towards these domains tend to rank higher.
- Here, accuracy was measured by comparing the manually labelled set of 'initial adopters' to the predicted set.
- The payoffs for the manual labels comprised of labels assigned by me for my close friends, and some by directly asking people to 'rate' the importance of a given product on a scale from 1 to 10.
- The software seems to work well mainly on popular keywords like 'party', 'gossip' etc in which most people find interest.
- The performance was not up to the mark for general keywords such as 'books', 'sky', 'water' etc as these words are too general and are not good candidates to highlight specific groups.

# Conclusions

- From this project, we have learnt that the cascading algorithm can be modified to run inversely, to predict the set of initial adopters for maximum spread as opposed to computing the spread from the initial adopters.
- Also, a unique and fairly accurate method to compute the payoff of a given set of words to a person has been developed. This method could've been improved much more but the compute power available is a limitation.
- This project can be improved upon to serve as a commercial application to perform customer targeting with surgical precision by incorporating technologies like Deep Learning, TensorFlow by Google, Torch Framework by Facebook, CUDNN by NVIDIA etc.