WHAT IS YOUR OPINION WORTH?

A STUDY OF HOW TO MEASURE THE INFLUENCE OF DIGITAL PERSONAS

DIRECT MARKETING WITH COUPONS, PROMOTIONS, AND DISCOUNTS VIA SOCIAL MEDIA HAS OPENED A TWO-WAY CONVERSATION BETWEEN COMPANIES AND THEIR CUSTOMERS.  

USERS COUNT ON SOCIAL MEDIA WHEN MAKING A PURCHASE 

AS CONNECTED DEVICES BECOME A WAY OF LIFE FOR MANY, DIGITAL COMMUNITIES ARE MORE RELIABLE UPON. USERS LOOK TO SOCIAL MEDIA TO GATHER OPINIONS FROM TRUSTED SOURCES. UNDERSTANDING WHO HAS THE GREATEST PULL IN THESE COMMUNITIES BECOMES EXTREMELY VALUABLE.

TOTAL U.S. INTERNET TIME SPENT ON SOCIAL NETWORKING SITES 

WHEN OVER A QUARTER OF THE TIME SPENT ON THE INTERNET IS ON SOCIAL NETWORKING SITES, ADVERTISING DOLLARS COMPETE FOR SPACE AT A PREMIUM. TARGETING USERS WITH GREATER INFLUENCE AMONG COMMUNITIES OF INTEREST ALLOWS FOR A MUCH GREATER RETURN ON INVESTMENT.

TOPICS

THE FIRST STEP IN IDENTIFYING INFLUENCE IS TO FIND OUT WHAT AREA OF INTEREST AN INDIVIDUAL INTERACTS WITH. THIS TASK NARROWS THE FIELD TO A MORE MANAGEABLE SIZE FOR DETAILED ANALYSIS. GIVEN THAT 99% OF ALL DATA CREATED HAS BEEN OWNED IN THE PAST TWO YEARS ALONE, THIS PROBLEM IS A COMPARISON ONE. CHECKING DATA INTO INDIVIDUAL TOPICS DOESN’T MEAN THAT THERE Aren’T OVERLAPPING CORRELATIONS BUT IT PROVIDES THE FIRST FOCUS. IN THIS PROJECT WE CHOSE THE FOLLOWING THREE CATEGORIES FOR FURTHER ANALYSIS:

SPORTS, PHOTOGRAPHY, COOKING

IN THE CASE OF THESE THREE CATEGORIES, THERE ARE INSTANCES WHERE THEY COULD ALL OVERLAP WITH ONE ANOTHER. ON THEIR OWN, HOWEVER, THEY ARE QUITE DIFFERENT. THIS ALLOWS FOR AN INVESTIGATION TO UNCOVER THE MOST POPULAR SUBTOPICS AND WHETHER OR NOT THEIR OVERLAPS WILL SHOW UP IN A MEANINGFUL WAY.

RESULTS

USING A DEFINED METRIC FOR DOMAIN-SPECIFIC IMPACT FACTOR IN CONJUNCTION WITH PREDICTIVE MODELLING, WE ARE ABLE TO VISUALIZE SOME INTERESTING THEMES OF INFLUENCE AMONG TWITTER USERS. NOT ONLY CAN THESE FINDINGS ASSIST IN INTEGRATING SOCIAL MEDIA INTO AN EFFECTIVE MARKETING MODEL, THEY CAN POTENTIALLY OUTLINE A MORE DIRECT RETURN ON INVESTMENT.

THE FIRST STEPS ARE SHOWN TO THE RIGHT. FINDINGS OUTLINE SOME INTERESTING THEMES AND TOPICS FOR FURTHER ANALYSIS. WE FAB A NUMBER OF STATISTICS ON THE GRAPH BASED ON THE DATA COLLECTED, INCLUDING AVERAGE DEGREE, AVERAGE WEIGHT DEGREE, MODULARITY AND PAGE RANK MODULARITY IS AN INTERESTING MEASURE WHICH HELPS US IDENTIFY COMMUNITIES IN OUR GRAPH, EVEN WITHIN THE DATA BEING COLLECTED USING THE SAME KEYWORD.

TOP INFLUENCERS

IN THE RELATIONSHIP GRAPH, EACH COLOR REPRESENTS A DISTINCT MODULARITY CLASS. THE TWITTER ACCOUNTS DISPLAYED AS LARGER NODES HAS A HIGHER IMPACT FACTOR THAN OTHERS WITHIN THE COMMUNITY THEY BELONG TO. WHEN WE DIVIDE THE NODES WHICH REPRESENT EACH CLUSTER, WE FIND THATS CAN DIFFERENTIATE THOSE COMMUNITIES, SUCH AS CONTEXT, REGION, LANGUAGE AND SO ON. THE TOP THREE INFLUENCES ARE:

SCREEN NAME: CHEVY CAMARO (A COMMUNITY OF CAR RACING) https://twitter.com/ChevyCamaroPC
SCREEN NAME: BATTERER (A REAL RACING SPORT COMMUNITY) https://twitter.com/BATTERER
SCREEN NAME: SOCA_RACING (A SPANISH SPEAKING COMMUNITY) https://twitter.com/SOCA_RACING

ACCORDING TO OUR RESULTS, TOP TAB-RELATED INFLUENCES ARE NOT NECESSARILY THOSE WHO ARE PRESUMED AS THE MOST WELL KNOWN CELEBRITIES ESPECIALLY WHEN NARROWED DOWN TO A SPECIFIC TOPIC.

RESEARCHERS: YILI HUANG, FEI LANG, CHRIS MOUNT / UNIVERSITY OF WASHINGTON INFORMATION SCHOOL
SPONSOR: KUNPENG ZHANG, PHD / NORTHWESTERN UNIVERSITY

PROBLEM

OR PERHAPS BETTER PHRASED: OPPORTUNITY. FINDING A REPEATEDLY SUCCESSFUL MARKETING STRATEGY FOR SOCIAL MEDIA HAS BEEN ELUSIVE, DETERMINING A RELIABLE WAY TO GAUGE RETURN ON INVESTMENT HAS BEEN EVEN HARDER. THIS PROJECT PRESENTS FINDINGS TO UNDERSTAND SOCIAL INFLUENCE IN THE DIGITAL WORLD.

SOCIAL INFLUENCE IN THE PHYSICAL WORLD CAN BE CORRELATED WITH INTANGIBLE ATTRIBUTES LIKE CHARISMA, ATTRACTION, AND CHARM. IN TRANSLATION TO A DIGITAL CONTEXT, THE CONCEPT OF INFLUENCE IS STILL VERY MUCH CONNECTED TO THESE TYPES OF HUMAN ELEMENTS, BUT THEIR IMPACT BECOMES MEASURABLE. THE VALUE IN DETERMINING HOW MUCH IMPACT AN INDIVIDUAL HAS ON A SPECIFIC TOPIC OF INTEREST IS BECOMING CLEARER AS MORE COMPANIES DEPEND ON DIGITAL COMMUNITIES FOR PRODUCT AWARENESS.

ANALYSIS

AFTER ESTABLISHING A SET OF THREE SAMPLE TOPICS, THE NEXT STEPS IN THE PROCESS CAN BE OUTLINED AS FOLLOWS:

KEYWORD DEFINITIONS

WITHIN EACH TOPIC, THERE ARE A HUGE NUMBER OF KEYWORDS THAT NEED TO BE CORRELATED WITH THE HIGHER LEVEL CATEGORY.

ALGORITHM DEVELOPMENT

AN ALGORITHM WAS CREATED TO GATHER THE IMPACT RANKING OF TWITTER ACCOUNTS. AMONG OTHERS, PRIMARY WEIGHT CONTRIBUTORS INCLUDE THE NUMBER OF FOLLOWERS, RETWEETS, AND OUTBOUND LINKS.

DATABASE CONFIGURATION

TWO RELATIONAL TABLES IN MYSQL WERE CREATED, ONE FOR TWEETS AND ANOTHER FOR FOLLOWER METRICS.

DATA DOWN LOADING

A PYTHON SCRIPT WAS CONFIGURED TO DOWNLOAD DATA IN PARALLEL FROM TWITTER FOR 21 CONSECUTIVE DAYS.

METRIC COMPARISON

RELATIONSHIP VISUALIZATIONS WERE GENERATED USING GEPHI. INITIAL WORD EXPLORATIONS WERE CREATED USING WORDLE.

INSTANCE IN THE CLOUD

AMAZON’S EC2 SERVICE WAS USED TO HOST A VIRTUAL MACHINE DEDICATED FOR RESEARCH USE.

DATA PREPARATION

INVERSE AND UNIVERSAL SETTINGS WERE USED TO BRIDGE A RELATIONSHIP BETWEEN NODES AS A WAY OF TYPING THE NODES IN ON SOCIAL NETWORKING.