

Online Appendix For “Advertising Content and Consumer
Engagement on Social Media: Evidence from Facebook”,
Management Science

Appendix 1: Message Tag Examples for Different Page-Types and Visualizations

Sample Messages	Content Mostly	PageType (Page)	Content Tags
Welcome to the unveiling of the Pro Staff RF97 that I co-designed with Wilson Tennis. Learn more at http://bit.ly/29JXLdA . #FromFederer	Informative	Celebrity (Federer)	BRANDMENTION, PRODMENTION, PRODLOCATION, HTTP
Coach Seve and me. Excited to be back in Brisbane! Happy we got the 1st practice of year out of the way!	Brand Personality	Celebrity (Federer)	SMALLTALK, EMOTION
Hello fans from Colombia! I am very happy to see you at the exo I am playing vs Tsonga on Saturday, December 15th! Buy your tickets starting September 12th on www.tuboleta.com . I hope to see you all there!	Mixed	Celebrity (Federer)	EMOTION, SMALLTALK, TARGET, PRODAVAIL, PRODLOCATION, PRODMENTION, HTTP
The Walking Dead Season 1 DVD/Blu-ray is now available, purchase it now!!! http://blogs.amctv.com/the-walking-dead/2011/03/season-1-dvd-blu-ray.php	Informative	Entertainment (Walking Dead)	BRANDMENTION, PRODMENTION, PRODAVAIL, PRODLOCATION, HTTP
Daryl makes a funny. What are some of your favorite #TheWalkingDead quotes? The highest rated quote will be turned into a graphic! #tbt	Brand Personality	Entertainment (Walking Dead)	SMALLTALK, EMOTION, QUESTION, BRANDMENTION, ASKCOMMENT
Need some healing? This weekend, head over to The Walking Dead Social Game store and pick up a new Team Med-Kit to rejuvenate your ENTIRE team. It's "Buy 1 Get 1 Free" on med-kits through 12/9. http://bit.ly/18dTwVw	Mixed	Entertainment (Walking Dead)	SMALLTALK, BRANDMENTION, DEAL, PRODAVAIL, PRODLOCATION, PRODMENTION, QUESTION, HTTP
Open enrollment is here! You can sign up for health insurance or renew your plan from November 1, 2016 through January 31, 2017. Learn more and get covered: http://p.ppfa.org/2eRaZem	Informative	Organizations (Planned Parenthood)	PRODAVAIL, PRODLOCATION, PRODMENTION, HTTP
Studies continue to prove what we already know — the HPV vaccine helps prevent cancer.	Brand Personality	Organizations (Planned Parenthood)	REMFAC, PHILANTHROPIC
Everyone knows it's legal to buy alcohol when you turn 21. But did you know that women should also have their first Pap test when they turn 21? We've got all the details. Cheers. http://bit.ly/xzJRZC	Mixed	Organizations (Planned Parenthood)	REMFAC, SMALLTALK, EMOTION, TARGET, PRODMENTION, HTTP, QUESTION
Score an iPad 3 for an iPad2 price! Now at your local store, \$50 off the iPad 3. Plus, get a \$30 iTunes Gift Card. Offer good through 12/31 or while supplies last.	Informative	PlaceBusiness (Walmart)	PRODMENTION, DEAL, PRODLOCATION, PRODAVAIL, PRICE
Maria's mission is helping veterans and their families find employment. Like this and watch Maria's story. http://walmarturl.com/VzWF1h	Brand Personality	PlaceBusiness (Walmart)	PHILANTHROPIC, SMALLTALK, ASKLIKE, HTTP
Cheers! Let Welch's help ring in the New Year.	Mixed	PlaceBusiness (Walmart)	BRANDMENTION, SMALLTALK, HOLIDAYMENTION, EMOTION
Welcome home, #GoogleHome. Ask it questions. Tell it to do things. Available today, your own Google, ready to help. https://goo.gl/Ugb9kl#madebygoogle	Informative	Website (Google)	BRANDMENTION, PRODMENTION, PRODAVAIL, PRODLOCATION, HTTP
#ItTakesATeacher to prepare the leaders of tomorrow. Celebrate them on #WorldTeachersDay by sharing a story about your favorite teacher or what teachers mean to you. g.co/ItTakesATeacher	Brand Personality	Website (Google)	EMOTION, HOLIDAYMENTION, SMALLTALK, ASKCOMMENT, HTTP
Sure it can translate languages, but did you know Google Translate could beat box? Watch it throw down at www.demoslam.com	Mixed	Website (Google)	BRANDMENTION, PRODMENTION, REMFAC, SMALLTALK, QUESTION, HTTP
It's time to #ChooseTheRockyRoad with the NEW Rocky Road Quest Nutrition Bar - now available at GNC: http://spr.ly/61828EuPi	Informative	Consumer Product (GNC)	BRANDMENTION, PRODMENTION, PRODAVAIL, PRODLOCATION, HTTP
Today's the day! Exercise your muscles and your right to vote. #election2016	Brand Personality	Consumer Product (GNC)	SMALLTALK, EMOTION, PHILANTHROPIC
Heroes don't take a day off. GNC honors the sacrifices veterans and their families make, today and every day. Have a vet in your life? Tag your own hero to say thanks. #VeteransDay	Mixed	Consumer Product (GNC)	BRANDMENTION, SMALLTALK, EMOTION, HOLIDAYMENTION, PHILANTHROPIC, QUESTION, ASKCOMMENT

Table 1: **Different Content Examples by Page-Types:** We give real examples (not necessarily from our data) of each different content (mostly informative, mostly brand personality-related, mix of both) across different page-types tagged by AMT Workers and RAs based on survey instrument included in Appendix 2. Note that Facebook has updated their categorization since the collection of this data in 2011, however page-types stayed mostly the same. We find all types of posts in each page-types. Additionally, we confirmed manually (via hiring two RAs) that all pages in our dataset has at least one post with informative content and at least one post with brand personality-related content.

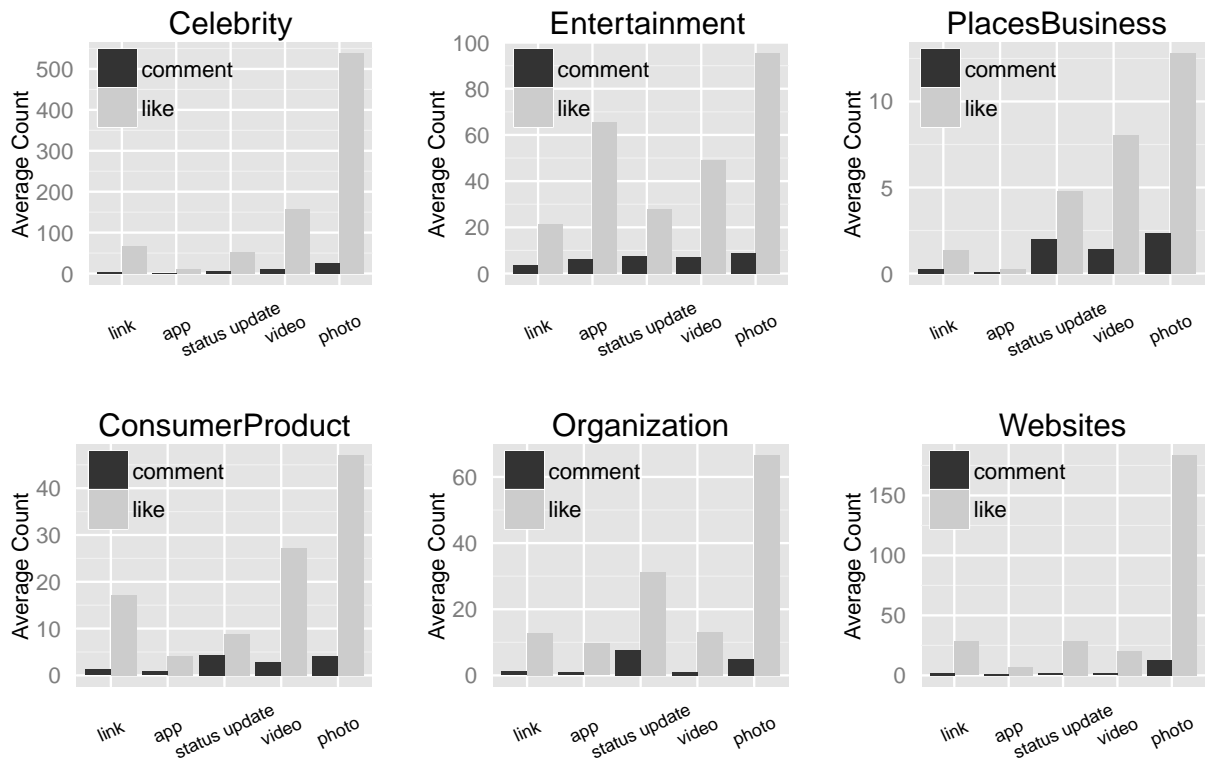


Figure 1: **Average Likes and Comments by Message Type by Page-Type:** This figure shows the average number of Likes and comments obtained by messages over their lifetime split by message type for each page-type.

Appendix 2: Survey Instrument

We are interested in identifying the content of social media messages. Please read the following short message posted on a social media site and answer the following questions. Thank you.

{Message Content}

Indicate the message characteristics by (Yes/No).

The Message,

1: contains an interesting or remarkable fact.

2: includes emotional content.

3: uses humor.

4: includes **philanthropic, awareness or activist content** (promotes social welfare, social change, or informs the reader about certain facts to improve people's lives).

5: **mentions a company (organization) name or brand name** (e.g. Nike, Red Cross, VitaminWater or musicians such as The Beatles).

If yes, please type the company or brand name as it appears in the message in the following box (only one per box)

Most relevant company/brand mentioned _____

Second company/brand mentioned (if any) _____

6: provides information about any type of **discounts or freebies** (e.g., deals, coupons, promotional offers, sweepstakes, rewards, free items).

7: makes a **price comparison** or a **price-match guarantee** against competitors' product(s) or service(s).

8: contains content **other than** about a product or company business (e.g. small talk, social pleasantries, casual banter, "Happy Halloween", "what was your day like?").

9: **is targeted towards an audience segment**. E.g., a particular demographic (race, gender, age, location) or people with certain qualifications or characteristics. (e.g. "all moms, do you like Gerber?", "All chocolate lovers, get this deal").

10: provides information about the **availability of a product or service** (e.g. Only 5 deals remain, available only until tomorrow, new product coming tomorrow).

11: provides information about **where or how to obtain a product or service** (e.g. link, physical location to buy, general location)?

12: does the message have a dollar sign (\$)?

13: Are your friends on social media likely to post messages such as the above?

14: The message is specific to a **particular product or service** (e.g. electronics, drinks, music, etc.).

If yes, please type the mentioned product or service:

Most relevant product mentioned _____

Figure 2: Survey Form Used in Amazon Mechanical Turk

Appendix 3: Amazon Mechanical Turk – Robust Content Extraction

Following best-practices in the literature, we employ the following strategies to improve the quality of classification by the Turkers in our study.

1. For each message, at least 9 different Turkers’ inputs are recorded. We obtain the final classification by a majority-voting rule.
2. We restrict the quality of Turkers included in our study to comprise only those with at least 100 reported completed tasks and 97% or better reported task-approval rates.
3. We use only Turkers from the US so as to filter out those potentially not proficient in English, and to closely match the user-base from our data (recall, our data has been filtered to only include pages located in the US).
4. We refined our survey instrument through an iterative series of about 10 pilot studies, in which we asked Turkers to identify confusing or unclear questions. In each iteration, we asked 10-30 Turkers to identify confusing questions and the reasons they found those questions confusing. We refined the survey in this manner till almost all queried Turkers stated no questions were confusing.
5. To filter out participants who were not paying attention, we included an easily verifiable test question “does the message have a dollar sign (\$)?”. Responses from Turkers that failed the verification test are dropped from the data.
6. In order to incentivize workers, we awarded additional bonuses of \$2-\$5 to the top 20 workers with exceptional accuracy and throughput.
7. On average, we found that message tagging took a little over 3 minutes and it typically took at least 20 seconds or more to completely read the tagging questions. We defined less than 30 seconds to be too short, and discarded any message tags with completion times shorter than that duration to filter out inattentive Turkers and automated programs (“bots”).
8. Once a Turker tags more than 20 messages, a couple of tagged samples are randomly picked and manually examined for quality and performance. This process identified about 20 high-volume Turkers who completed all surveys in less than 10 seconds and tagged several thousands of messages (there were also Turkers who took time to complete the surveys but chose seemingly random answers). We concluded these were automated programs. These results were dropped, and the Turkers “hard blocked” from the survey, via the blocking option provided in AMT.

We believe our methodology for content-classification has good external validity. The binary classification task that we serve to the AMT Turkers in our study is relatively simpler than the more complex tasks for which AMT-based data have been employed successfully in the literature. The existing AMT literature

has documented evidence that several of the strategies implemented above improves the quality of the data generated (Mason and Suri (2012); Ipeirotis et al. (2010); Paolacci et al. (2010)). Snow et al. (2008) show that combining results from a few Turkers can produce data equivalent in quality to that of expert labelers for a variety of text tagging tasks. Similarly, Sheng et al. (2007) document that repeated labeling of the type we implement wherein each message is tagged by multiple Turkers, is preferable to single labeling in which one person tags one sentence. Finally, evaluating AMT based studies, Buhrmester et al. (2011) concludes that (1) Turkers are demographically more diverse than regular psychometric studies samples, and (2) the data obtained are at least as reliable as those obtained via traditional methods as measured by psychometric standards such as Cronbach’s Alpha, a commonly used inter-rater reliability measure.

Figure 3 presents the histogram of Cronbach’s alpha of all 5000 messages. The average Cronbach’s Alpha for our 5,000 tagged messages is 0.82 (median 0.84), above typically acceptable thresholds of 0.7. About 87.5% of the messages obtained an alpha higher than 0.7, and 95.4% higher than 0.6.

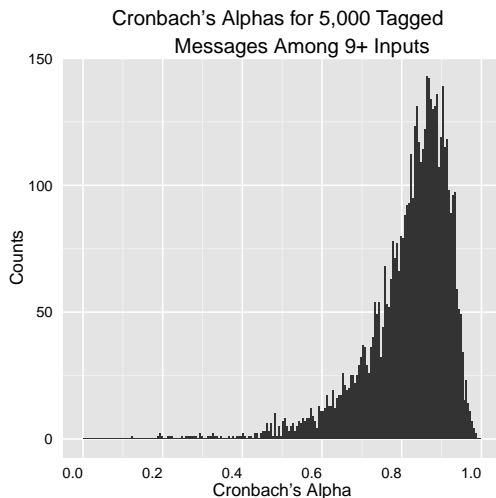


Figure 3: Cronbach’s Alphas for 5,000 Messages: This bar graph shows the inter-rater reliability measure of Cronbach’s Alpha among at least 9 distinct Turkers’ inputs for each 5,000 messages. The mean is 0.82 and the median is 0.84. We replicated the study with only those above 0.7 and found the result to be robust.

Appendix 4: NLP Algorithm

This section provides detailed outline of the algorithm used in the paper. Figure 4 shows the process visually. One can think of the NLP algorithm as emulating the Turkers’ collective opinion in content-coding.

Training The Algorithm

1. The raw textual data of 5,000 messages in the training sample are broken down into basic building blocks of sentences using *stop-words removal* (removing punctuation and words with low information such as the definite article “the”), *tokenization* (the process of breaking a sentence into words, phrases, and symbols or “tokens”), *stemming* (the process of reducing inflected words to their root form, e.g., “playing” to “play”), and *part-of-speech tagging* (determining part-of-speech such as nouns). For reference see Jurafsky and Martin (2008). In this process, the input to the algorithm is a regular sentence and the output is an ordered set of fundamental linguistic entities with semantic values. We use the well-regarded python-based NLP framework “NLTK” (Bird et al., 2009) to implement this step.
2. Once the messages are broken down as above, an algorithm extracts sentence-level attributes and sentence-structure rules that help identify the included content. Some examples of sentence-level attributes and rules include: frequent noun words (bag-of-words approach), bigrams, the ratio of part-of-speech used, tf-idf (term-frequency and inverse document frequency) weighted informative word weights, and whether “a specific key-word is present” rule. For completeness, we describe each of these in Table 2. The key to designing a successful NLP algorithm is to figure out what we (humans) do when identifying certain information. For example, what do we notice about the sentences we have identified as having emotional content? We may notice the use of certain types of words, use of exclamation marks, the use of capital letters, etc. The algorithms in this step try to reverse-engineer these patterns. At the end of this step, the dataset consists of sentence-level attributes generated as above (the x -variables), corresponding to a series of binary (content present/not-present) content labels generated from AMT (the y -variables).
3. For each binary content label, we then train a classification model by combining statistical and rule-based classifiers. In this step, the NLP algorithm fits the binary content label (the y -variable) using the sentence-level attributes as the x -variables. For example, the algorithm would fit whether or not a message has emotional content as tagged by AMT using the sentence attributes extracted from the message via step 2. We use a variety of different classifiers in this step including logistic regression with L1 regularization (which penalizes the number of attributes and is commonly used for attribute selection for problems with many attributes; see (Hastie et al., 2009)), Naive Bayes (a probabilistic classifier that applies Bayes theorem based on presence or absence of features), and *support vector machines* (a classification algorithm in machine learning that works well for high dimensional problems) with different

flavors of regularization and kernels.¹ To account for imbalance in positive and negative class labels in some content, we utilized combination of class-weighted classifiers and resampling methods.

4. To train the ultimate predictive classifier, we use ensemble methods to combine results from the multiple statistical classifiers we fit in step 3. The motivation for ensemble learning is that different classifiers perform differently based on underlying characteristics of data or have varying precision or recall in different locations of the feature vector space. Thus, combining them will achieve better classification output either by reducing variance (e.g. Bagging (Brieman, 1996)) or reducing bias (e.g. Boosting (Freund and Schapire, 1995)). Please see Xu and Krzyzak (1992); Bennett (2006) for further reading on ensemble methods. This step involves combining the prediction from individual classifiers by weighted-majority voting, unweighted-majority voting, or a more elaborate isotonic regression (Zadrozny and Elkan, 2002) and choosing the best performing method in terms of accuracy, precision and recall for each content profiles. In our case, we found that support vector machine based classifiers delivered high precision and low recall, while Naive Bayes based classifiers delivered high recall but low precision. By combining these, we develop an improved classifier that delivers higher precision and recall and in effect, higher accuracy. Table 3 shows the improvement of the final ensemble learning method relative to using only one support vector machine. As shown, the gains from combining classifiers are substantial. We obtain similar results for negative class labels.
5. Finally, we assess the performance of the overall NLP algorithm on three measures, viz., accuracy, precision, and recall (as defined in Footnote 4) using the “10-fold cross-validation” method. Under this strategy, we split the data randomly into 10 equal subsets before step 2. One of the subsets is used as the validation sample, and the algorithm trained on the remaining 9 sets. This is repeated 10 times, each time using a different subset as the validation sample, and the performance measures averaged across the 10 runs. The use of 10-fold cross-validation reduces the risk of overfitting and increases the external validity of the NLP algorithm we develop. Table 3 shows these metrics for different content profiles. The performance is extremely good and comparable to performance achieved by the leading financial information text mining systems (Hassan et al., 2011).
6. We repeat steps 2-5 until desired performance measures are achieved.

Tagging New Messages

1. For each new messages repeat steps 1-2 described above.
2. Use the ultimate classifier developed above to predict whether a particular type of content is present or not.

¹We tried support vector machines with L1 and L2 regularization and various kernels including linear, radial basis function, and polynomial kernels. For more details on these methods, please see Hastie et al. (2009).

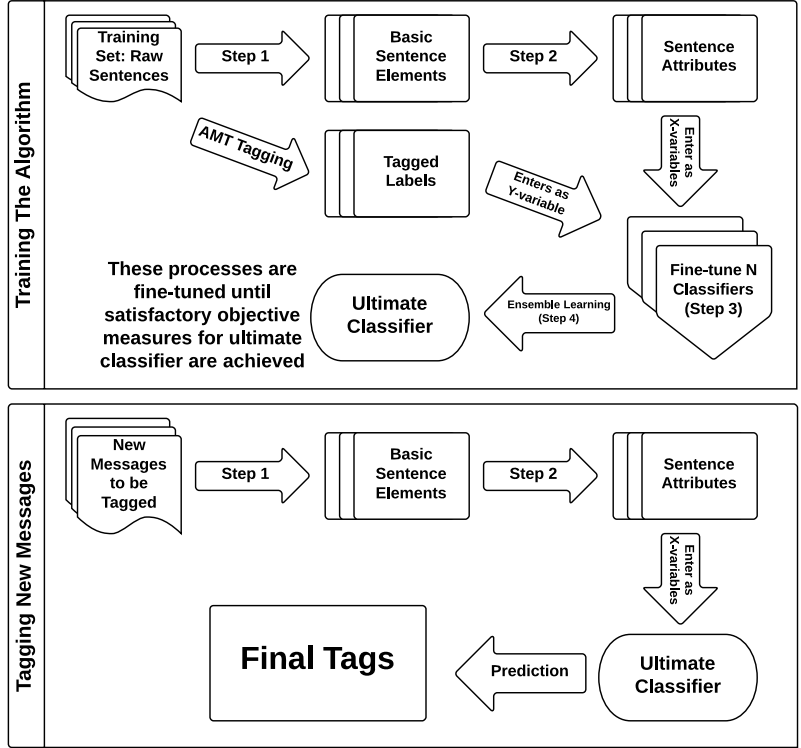


Figure 4: **Diagram of NLP Training and Tagging Procedure:** This diagram shows the steps of training the NLP algorithm and using the algorithm to tag the remaining messages. These steps are described in Appendix 3.

Rules and Attributes	Description
Bag of Words	Collects all the words and frequency for a message. Different variations include collecting top N most occurring words.
Bigram	A bigram is formed by two adjacent words (e.g. “Bigram is”, “is formed” are bigrams).
Ratio of part-of-speech	Part-of-speech (noun, verb, etc) ratio in each message.
TF-IDF weighted informative word	Term-Frequency and Inverse Document Frequency weighs each word based on their occurrence in the entire data and in a single message.
Specific Keywords	Specific keywords for different content can be collected and searched. e.g., Philanthropic messages have high change of containing the words “donate” and “help”. For brand and product identification, large online lists were scraped and converted into dictionaries for checking.
Frequency of different punctuation marks	Counts the number of different punctuations such as exclamation mark and question mark. This helps to identify emotion, questions, appearance of deals etc.
Count of non-alphanumerics	Counts the number of characters that are not A-Z and 0-9.

Table 2: **A Few Examples of Message Attributes Used in Natural Language Processing Algorithm**

	With Ensemble Learning (The Best Performing Algorithm)			Without Ensemble Learning (Support Vector Machine version 1 + Rule-based)		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
REMACT	0.94	0.99	0.68	0.88	0.99	0.33
EMOTION	0.97	0.99	0.87	0.94	0.98	0.65
HUMOR	0.98	1	0.90	0.97	1	0.14
PHILANTHROPIC	0.97	0.99	0.85	0.93	0.99	0.62
FRIENDLIKELY	0.94	0.99	0.68	0.90	0.99	0.41
SMALTALK	0.85	0.88	0.80	0.78	0.34	0.28
DEAL	0.94	0.99	0.65	0.90	1	0.43
PRICECOMPARE	0.99	0.99	1	0.99	1	0.85
TARGETING	0.98	0.99	0.89	0.95	0.99	0.71
PRODAVAILABILITY	0.96	0.99	0.76	0.91	1	0.10
PRODLOCATION	0.97	0.99	0.90	0.87	1	0.11

Table 3: **Performance of Text Mining Algorithm on 5000 Messages Using 10-fold Cross-validation:** This table presents metrics for performance of the classification algorithms used. The left 3 columns show the metrics for the final algorithm which combines classifiers via ensemble learning method while the right 3 columns show the metric for a support vector machine algorithm. Notice that the support vector machine classifier tends to have low recall and high precision. Naive Bayes tends to have high recall but low precision. Classifiers on their own cannot successfully overcome precision-recall tradeoff (if one is higher, one is lower). But combining many different classifiers with ensemble learning can increase both precision and recall. We obtain similar results for negative class labels.

Appendix 5: Obtaining the Number of Impressions

We discuss our procedure for constructing $n_{kjt}^{(d)}$, the number of impressions for each message k of firm j in day t split by demographic bin d for use in the EdgeRank correction model. As mentioned above, $n_{kjt}^{(d)}$ is not directly reported by Company X (or made available to page-owners by Facebook *Insights*). Instead, Company X reports $n_{jt}^{(d)}$, the number of impressions for all messages associated with firm j in demographic bucket d on day t , which is essentially $n_{kjt}^{(d)}$ summed across all k associated with j . In addition, we observe n_{kjt} , the total number of impression obtained by a specific post k by firm j on day t . To assess how we may split this across the various demographic buckets, we checked the extent to which pages release different types of messages over time. The bulk of impressions for a message occur within the first week of its release. Hence, the total impressions for a page on a given day out of a specific demographic bucket, $n_{jt}^{(d)}$, reflects the aggregate impressions to users in that bucket of all messages released by that firm over the past one week. Since EdgeRank allocates messages to users by message-type, if the firm releases the same *type of messages* (i.e., photos, videos, status updates, apps or links) over a week’s duration, then the the distribution of $n_{jt}^{(d)}$ across the various k messages released by firm j within the past week may be roughly the same. In other words, the distribution of demographics of the impressions of all messages released by a firm in the past week should be the same if all those released messages are similar.

To check informally if this is the case, we picked a random sample of 10,000 page-7-day combinations from our data. For each combination, we collated all the messages released by that page during that 7-day window and tabulated the type of these messages (i.e., photos, videos, status updates, apps or links). We then construct two concentration metrics, \mathbb{C}_1 , the proportion of messages released by that page during that 7-day window that belong to the highest type bucket, and \mathbb{C}_2 , the proportion of messages released by that page during that 7-day window that belong to the highest and second-highest type bucket. \mathbb{C}_1 and \mathbb{C}_2 are analogous to top-firm and top-two-firm concentration ratios used in industry-concentration studies, and measure the extent to which the messages released by a page in a given 7-day period are spread across types. If all messages released by a page during that 7-day window are of the same type, \mathbb{C}_1 and \mathbb{C}_2 will be both 1. The spread away from 1 thus indicates higher variation in message-types released by an average Facebook page over a week’s duration. Table 4 reports on the distribution of \mathbb{C}_1 and \mathbb{C}_2 we computed in this manner. Looking at Table 4, we find that the median \mathbb{C}_1 is .71 (mean .72) and the median \mathbb{C}_2 is 1.0 (mean .94). Most pages seem to be releasing at-most 2-types of messages within a week window, and more than $2/3^{rd}$ of messages released by an average page in an average week are of the same type. Given this, we assume that n_{kjt} is split into $n_{kjt}^{(d)}$ with the same distribution as given by $n_{jt}^{(d)}$. We construct the variable $n_{kjt}^{(d)}$ in the left hand-side of the EdgeRank correction equation 2 in this manner. We replicated the study with different split distributions such as 7-day aggregated distribution of $n_{jt}^{(d)}$, kernel smoothed distribution, and the results were similar.

The method is not without its limitations. We view it as a practical way to deal with the lack of data-reporting by Facebook, while exploiting the variation embedded in the observed impressions and to correlate it with the observed variation in the the potential market for each message in each demographic bucket

	<i>Min.</i>	<i>1st Qu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rd Qu.</i>	<i>Max.</i>
\mathbb{C}_1	0.250	0.535	0.706	0.719	0.915	1.000
\mathbb{C}_2	0.500	0.889	1.000	0.942	1.000	1.000

Table 4: **Distribution of the Top (\mathbb{C}_1) and Top-two (\mathbb{C}_2) Concentration Ratios of the type of messages Served by a Facebook Page over a Randomly picked 7-day period**

$\left(\mathbb{N}_{jt}^{(d)}\right)$. The method produces potential measurement error in the dependent variable, $n_{kjt}^{(d)}$ in the EdgeRank correction stage. Measurement error in the dependent variable is absorbed into the RHS unobservables and is usually less of a concern unless it is correlated with the unobservables. The fact that we include page-fixed effects separately for each demographic ($\theta_{1j}^{(d)}$ in Equation 3) mitigates concerns to a large extent that these unobservables may systematically be correlated with included characteristics.

Appendix 6: Result for AMT-Tagged & Different Models

	NO ER COMMENT	OR	ER COMMENT	OR	NO ER LIKE	OR	ER LIKE	OR
Constant	-5.431*** (0.047)		-5.682*** (0.041)		-3.326*** (0.014)		-4.199*** (0.012)	
SMOG	-0.098*** (0.001)	0.907	-0.125*** (0.001)	0.882	-0.045*** (0.000)	0.956	-0.066*** (0.000)	0.936
MSGLEN	-0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000
HTTP	-0.253*** (0.005)	0.776	-0.143*** (0.005)	0.867	-0.097*** (0.002)	0.908	-0.025*** (0.002)	0.975
QUESTION	0.164*** (0.004)	1.178	0.191*** (0.004)	1.210	-0.325*** (0.001)	0.723	-0.220*** (0.001)	0.803
BLANK	0.739*** (0.012)	2.094	0.755*** (0.012)	2.128	-1.009*** (0.010)	0.365	-1.098*** (0.010)	0.334
ASKLIKE	-0.376*** (0.033)	0.687	-0.336*** (0.033)	0.715	-0.100*** (0.009)	0.905	-0.072*** (0.009)	0.931
ASKCOMMENT	0.927*** (0.049)	2.527	0.272*** (0.049)	1.313	0.111*** (0.020)	1.117	-0.406*** (0.020)	0.666
Brand personality-related								
REMACT	0.105*** (0.006)	1.111	0.145*** (0.005)	1.156	-0.063*** (0.002)	0.939	-0.032*** (0.002)	0.969
EMOTION	0.159*** (0.006)	1.172	0.135*** (0.006)	1.145	0.084*** (0.002)	1.088	0.106*** (0.002)	1.112
EMOTICON	-0.383*** (0.026)	0.682	-0.361*** (0.026)	0.697	-0.416*** (0.008)	0.660	-0.192*** (0.008)	0.825
HOLIDAYMENTION	-1.465*** (0.033)	0.231	-1.293*** (0.033)	0.274	-0.433*** (0.006)	0.649	-0.266*** (0.006)	0.766
HUMOR	-0.086*** (0.012)	0.918	-0.163*** (0.012)	1.177	-0.275*** (0.004)	0.760	-0.019*** (0.004)	1.019
PHILANTHROPIC	0.142*** (0.007)	1.153	0.138*** (0.007)	1.148	0.182*** (0.002)	1.200	0.168*** (0.002)	1.183
FRIENDLY	-0.159*** (0.005)	0.853	-0.130*** (0.005)	0.878	0.206*** (0.002)	1.229	0.159*** (0.002)	1.172
SMALLTALK	-0.005 (0.004)	0.995	-0.051*** (0.004)	0.950	0.022*** (0.001)	1.022	-0.028*** (0.001)	0.972
Directly informative								
BRANDMENTION	-0.179*** (0.004)	0.836	-0.228*** (0.004)	0.796	-0.141*** (0.001)	0.868	-0.224*** (0.001)	0.799
DEAL	-0.151*** (0.008)	0.860	0.011 (0.007)	1.011	-0.478*** (0.002)	0.620	-0.302*** (0.002)	0.739
PRICECOMPARE	0.688*** (0.193)	1.990	0.302 (0.193)	1.353	-0.498*** (0.103)	0.608	-0.857*** (0.103)	0.424
PRICE	-0.051*** (0.014)	0.950	-0.500*** (0.014)	0.607	-0.316*** (0.005)	0.729	-0.712*** (0.005)	0.491
TARGET	0.188*** (0.010)	1.207	0.076*** (0.010)	1.079	0.130*** (0.003)	1.139	0.059*** (0.003)	1.061
PRODAVAIL	-0.281*** (0.007)	0.755	-0.360*** (0.007)	0.698	0.073*** (0.002)	1.076	-0.022*** (0.002)	0.978
PRODLOCATION	-0.151*** (0.006)	0.860	-0.015** (0.005)	0.985	-0.239*** (0.002)	0.787	-0.080*** (0.002)	0.923
PRODMENTION	-0.170*** (0.005)	0.844	-0.349*** (0.005)	0.705	0.291*** (0.001)	1.338	0.095*** (0.001)	1.100
Message Type - App is the base								
-Link	0.226*** (0.008)	1.254	-0.051*** (0.008)	0.950	0.004 (0.003)	1.004	-0.310*** (0.003)	0.733
-Photo	0.633*** (0.008)	1.883	0.379*** (0.008)	1.461	0.718*** (0.002)	2.050	0.418*** (0.002)	1.519
-Status Update	1.325*** (0.009)	3.762	1.148*** (0.009)	3.152	0.671*** (0.003)	1.956	0.553*** (0.003)	1.738
-Video	-0.191*** (0.027)	0.826	-0.447*** (0.027)	0.640	-0.636*** (0.011)	0.529	-0.928*** (0.011)	0.395
Page-type Category - Celebrity is the base								
-ConsumerProduct	-0.024*** (0.007)	0.976	-0.368*** (0.007)	0.692	-0.541*** (0.002)	0.582	-0.761*** (0.002)	0.467
-Entertainment	0.014* (0.006)	1.014	0.221*** (0.006)	1.247	-0.472*** (0.002)	0.624	-0.169*** (0.002)	0.845
-Organization	0.222*** (0.007)	1.249	0.072*** (0.007)	1.075	-0.163*** (0.002)	0.850	-0.161*** (0.002)	0.851
-PlaceBusiness	0.600*** (0.014)	1.822	0.339*** (0.014)	1.404	-0.698*** (0.007)	0.498	-0.869*** (0.007)	0.419
-Websites	-0.076*** (0.009)	0.927	0.152*** (0.008)	1.164	0.099*** (0.002)	1.104	0.398*** (0.002)	1.489
McFadden R-sq.	0.288		0.239		0.295		0.214	
Nagelkerke R-sq.	0.288		0.239		0.296		0.216	
Log-likelihood	-299060.434		-428096.493		-1728807.768		-3,119,513.948	
Deviance	552293.46		810390.763		3365047.493		6,146,513.932	
AIC	598244.869		856316.985		3457739.535		6,239,151.895	
N	38706		38706		38706		38706	
Significance **** 0.001 *** 0.01 ** 0.05 * 0.1								

Table 5: **Aggregate Logistic Regression Results For Comments and Likes (5000 Messages):** This table presents the aggregate logistic regression on comments and *Likes* for both EdgeRank-corrected (ER) and uncorrected (NO ER) for 5000 messages data tagged by Turkers. OR means Odds ratio and shows the odds ratio for the estimates left of the column.

	NO ER COMMENT	OR	ER COMMENT	OR	NO ER LIKE	OR	ER LIKE	OR
Constant	-8.122*** (0.004)		-7.129*** (0.004)		-5.168*** (0.001)		-4.205*** (0.001)	
SMOG	-0.047*** (0.000)	0.954	-0.059*** (0.000)	0.943	-0.031*** (0.000)	0.969	-0.049*** (0.000)	0.952
MSGLEN	0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000
HTTP	-0.545*** (0.002)	0.580	-0.301*** (0.002)	0.740	-0.388*** (0.000)	0.678	-0.085*** (0.000)	0.919
QUESTION	0.488*** (0.001)	1.629	0.581*** (0.001)	1.788	-0.284*** (0.000)	0.753	-0.186*** (0.000)	0.830
BLANK	1.002*** (0.003)	2.724	1.186*** (0.003)	3.274	-0.709*** (0.002)	0.492	-0.538*** (0.002)	0.584
ASKLIKE	0.080*** (0.010)	1.083	0.259*** (0.010)	1.296	0.494*** (0.003)	1.639	0.598*** (0.003)	1.818
ASKCOMMENT	0.514*** (0.021)	1.672	0.413*** (0.021)	1.511	-0.244*** (0.011)	0.783	-0.557*** (0.011)	0.573
Brand personality-related								
REMFAC	-0.026*** (0.002)	0.974	0.017*** (0.002)	1.017	-0.055*** (0.001)	0.946	-0.015*** (0.001)	0.985
EMOTION	0.213*** (0.002)	1.237	0.217*** (0.002)	1.242	0.218*** (0.001)	1.244	0.191*** (0.001)	1.210
EMOTICON	0.171*** (0.004)	1.186	0.101*** (0.004)	1.106	-0.056*** (0.001)	0.946	-0.087*** (0.001)	0.917
HOLIDAYMENTION	-0.485*** (0.014)	0.616	-0.378*** (0.014)	0.685	-0.329*** (0.004)	0.720	-0.162*** (0.004)	0.850
HUMOR	0.030*** (0.001)	1.030	0.094*** (0.002)	1.099	-0.049*** (0.000)	0.952	0.047*** (0.000)	1.048
PHILANTHROPIC	0.182*** (0.002)	1.200	0.261*** (0.002)	1.298	0.019*** (0.001)	1.019	0.157*** (0.001)	1.170
FRIENDLIKELY	0.008*** (0.002)	1.008	-0.053*** (0.002)	0.948	0.076*** (0.001)	1.079	0.005*** (0.001)	1.005
SMALLTALK	0.059*** (0.002)	1.061	-0.111*** (0.002)	0.895	-0.061*** (0.001)	0.941	-0.172*** (0.001)	0.842
Directly informative								
BRANDMENTION	0.013*** (0.002)	1.013	0.084*** (0.002)	1.088	-0.015*** (0.000)	0.985	0.043*** (0.000)	1.044
DEAL	-0.134*** (0.002)	0.875	-0.245*** (0.002)	0.783	-0.183*** (0.001)	0.833	-0.357*** (0.001)	0.700
PRICECOMPARE	-0.038*** (0.001)	0.963	0.008*** (0.001)	1.008	-0.033*** (0.000)	0.968	-0.020*** (0.000)	0.980
PRICE	-0.031*** (0.005)	0.969	-0.303*** (0.005)	0.739	-0.185*** (0.001)	0.831	-0.459*** (0.001)	0.632
TARGET	-0.021*** (0.002)	0.979	-0.135*** (0.002)	0.874	0.036*** (0.001)	1.037	-0.120*** (0.001)	0.887
PRODAVAIL	-0.090*** (0.002)	0.914	-0.003 (0.002)	0.997	-0.119*** (0.001)	0.888	0.038*** (0.001)	1.039
PRODLOCATION	-0.056*** (0.002)	0.946	-0.043*** (0.002)	0.958	0.072*** (0.001)	1.075	0.034*** (0.001)	1.035
PRODMENTION	-0.087*** (0.002)	0.917	-0.128*** (0.002)	0.880	0.076*** (0.001)	1.079	0.039*** (0.001)	1.040
Message Type - App is the base								
-Link	0.201*** (0.003)	1.223	-0.301*** (0.003)	0.740	0.116*** (0.001)	1.123	-0.389*** (0.001)	0.678
-Photo	0.876*** (0.003)	2.401	0.318*** (0.003)	1.374	1.036*** (0.001)	2.818	0.475*** (0.001)	1.608
-Status Update	1.118*** (0.003)	3.059	0.666*** (0.003)	1.946	0.458*** (0.001)	1.581	-0.015*** (0.001)	0.985
-Video	0.641*** (0.009)	1.898	0.108*** (0.009)	1.114	0.218*** (0.003)	1.244	0.203*** (0.003)	1.225
Page-type Category - Celebrity is the base								
-ConsumerProduct	0.068*** (0.002)	1.070	-0.062*** (0.002)	0.940	-0.446*** (0.001)	0.640	-0.544*** (0.001)	0.580
-Entertainment	0.420*** (0.002)	1.522	0.660*** (0.002)	1.935	-0.260*** (0.001)	0.771	-0.067*** (0.001)	0.935
-Organization	0.497*** (0.002)	1.644	0.631*** (0.002)	1.879	-0.005*** (0.001)	0.995	0.110*** (0.001)	1.116
-PlaceBusiness	0.381*** (0.005)	1.464	0.234*** (0.005)	1.264	-0.726*** (0.002)	0.484	-0.933*** (0.002)	0.393
-Websites	0.218*** (0.003)	1.244	0.380*** (0.003)	1.462	0.183*** (0.001)	1.201	0.314*** (0.001)	1.369
1st Stage Residual	-0.000*** (0.000)	1.000	0.000*** (0.000)	1.000	-0.000*** (0.000)	1.000	0.000*** (0.000)	0.000
McFadden R-sq.	0.229		0.199		0.298		0.270	
Nagelkerke R-sq.	0.230		0.199		0.298		0.272	
Log-likelihood	-2575187.42		-3460249.739		-14563950.18		-24911467.22	
Deviance	4745694.741		6516530.605		28179842.5		48875491.61	
AIC	5150474.839		6920599.479		29128000.36		49823034.45	
N	665916		665916		665916		665916	
Significance '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1								

Table 6: **Aggregate Logistic Regression Results For Comments and Likes (All Messages with the 1st Stage Residuals as a Control Function):** This table presents the aggregate logistic regression on comments and *Likes* for both EdgeRank-corrected (ER) and uncorrected (NO ER) for all messages data. OR means Odds ratio and shows the odds ratio for the estimates left of the column. This second-stage model includes first-stage residuals as a control function.

	NO ER COMMENT	ER COMMENT	NO ER LIKE	ER LIKE
Constant	-8.152*** (0.103)	-9.067*** (0.124)	-4.595*** (0.074)	-5.386*** (0.066)
SMOGE	-0.035*** (0.000)	-0.046*** (0.000)	-0.029*** (0.000)	-0.043*** (0.000)
MSGLEN	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)
HTTP	-0.329*** (0.002)	-0.382*** (0.002)	-0.214*** (0.001)	-0.282*** (0.001)
QUESTION	0.448*** (0.001)	0.474*** (0.001)	-0.179*** (0.000)	-0.165*** (0.000)
BLANK	1.009*** (0.003)	1.059*** (0.003)	-0.657*** (0.002)	-0.623*** (0.002)
ASKLIKE	-0.081*** (0.010)	0.030** (0.010)	0.388*** (0.003)	0.451*** (0.003)
ASKCOMMENT	0.745*** (0.022)	0.828*** (0.021)	0.415*** (0.011)	0.476*** (0.011)
Brand personality-related				
REMACT	0.053*** (0.002)	0.078*** (0.002)	-0.026*** (0.001)	-0.016*** (0.001)
EMOTION	0.046*** (0.002)	0.068*** (0.002)	0.110*** (0.001)	0.150*** (0.001)
EMOTICON	-0.028*** (0.005)	-0.066*** (0.005)	-0.050*** (0.001)	-0.080*** (0.001)
HOLIDAYMENTION	-0.620*** (0.014)	-0.604*** (0.014)	-0.236*** (0.004)	-0.195*** (0.004)
HUMOR	0.024*** (0.002)	0.067*** (0.002)	0.020*** (0.000)	0.047*** (0.000)
PHILANTHROPIC	-0.008*** (0.002)	-0.011*** (0.002)	-0.023*** (0.001)	-0.044*** (0.001)
FRIENDLY	0.018*** (0.002)	0.041*** (0.002)	0.046*** (0.001)	0.085*** (0.001)
SMALLTALK	0.019*** (0.002)	-0.015*** (0.002)	0.020*** (0.001)	-0.021*** (0.001)
Directly informative				
BRANDMENTION	0.043*** (0.002)	-0.014*** (0.002)	-0.050*** (0.001)	-0.093*** (0.001)
DEAL	-0.069*** (0.002)	-0.092*** (0.002)	-0.091*** (0.001)	-0.115*** (0.001)
PRICECOMPARE	-0.022*** (0.002)	-0.014*** (0.002)	-0.015*** (0.000)	-0.002*** (0.000)
PRICE	-0.065*** (0.005)	-0.158*** (0.005)	-0.096*** (0.002)	-0.226*** (0.002)
TARGET	-0.038*** (0.002)	-0.050*** (0.002)	-0.026*** (0.001)	-0.011*** (0.001)
PRODAVAIL	0.025*** (0.002)	0.011*** (0.002)	-0.007*** (0.001)	-0.028*** (0.001)
PRODLOCATION	-0.032*** (0.002)	-0.020*** (0.002)	0.009*** (0.001)	0.031*** (0.001)
PRODMENTION	-0.053*** (0.002)	-0.095*** (0.002)	0.047*** (0.001)	-0.025*** (0.001)
Message Type - App is the base				
-Link	0.022*** (0.003)	0.061*** (0.003)	-0.306*** (0.001)	-0.277*** (0.001)
-Photo	0.692*** (0.003)	0.959*** (0.003)	0.497*** (0.001)	0.839*** (0.001)
-Status Update	0.866*** (0.004)	1.083*** (0.004)	0.031*** (0.001)	0.297*** (0.001)
-Video	0.502*** (0.009)	0.758*** (0.009)	0.340*** (0.003)	0.695*** (0.003)
Page-type Category - Celebrity is the base				
-ConsumerProduct	-0.024 (0.114)	-0.407** (0.134)	-0.346*** (0.085)	-0.701*** (0.086)
-Entertainment	0.077 (0.134)	0.216 (0.141)	-0.387*** (0.106)	-0.017935
-Organization	0.098 (0.122)	-0.171 (0.148)	0.012 (0.095)	-0.219 (0.112)
-PlaceBusiness	0.468** (0.143)	-0.247 (0.171)	-0.197 (0.126)	-0.887*** (0.156)
-Websites	0.019 (0.171)	-0.070 (0.193)	0.097 (0.139)	0.023 (0.146)
Page random effect	0.88	0.95	0.78	1.233
Log-likelihood	-1770990.4	-2465463.7	-8801518.3	-17735437.4
Deviance	3134106.6	4522983.5	16653792.7	34521575.1
AIC	3542106.9	4931053.3	17603162.6	35471000.8
N	665916	665916	665916	665916
Significance '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1				

Table 7: **Aggregate Logistic Regression Results For Comments and Likes with Page-Level Random Effects:** This table presents the aggregate logistic regression on comments and *Likes* for both EdgeRank-corrected (ER) and uncorrected (NO ER) for all messages data. This second-stage model includes page-level random effects.

Variable	Intercept only	Controls	Friendly	Brand personality-related	Directly informative	All (No demo controls)
Constant	-7.052***(0.001)	-6.906***(0.003)	-6.952***(0.004)	-6.975***(0.004)	-6.861***(0.004)	-6.889***(0.004)
SMOG		-0.065***(0.000)	-0.065***(0.000)	-0.065***(0.000)	-0.067***(0.000)	-0.068***(0.000)
MSGLEN		-0.000***(0.000)	-0.000***(0.000)	-0.000***(0.000)	-0.000***(0.000)	-0.000***(0.000)
HTTP		-0.406***(0.001)	-0.400***(0.001)	-0.393***(0.001)	-0.390***(0.002)	-0.355***(0.002)
QUESTION		0.543***(0.001)	0.546***(0.001)	0.557***(0.001)	0.541***(0.001)	0.564***(0.001)
BLANK		1.172***(0.003)	1.168***(0.003)	1.187***(0.003)	1.144***(0.003)	1.146***(0.003)
ASKLIKE		0.222***(0.010)	0.229***(0.010)	0.220***(0.010)	0.223***(0.010)	0.221***(0.010)
ASKCOMMENT		0.508***(0.021)	0.529***(0.021)	0.465***(0.021)	0.506***(0.021)	0.469***(0.021)
Message Type - App is the base						
-Link		-0.388***(0.003)	-0.389***(0.003)	-0.394***(0.003)	-0.376***(0.003)	-0.370***(0.003)
-Photo		0.376***(0.003)	0.371***(0.003)	0.365***(0.003)	0.372***(0.003)	0.373***(0.003)
-Status Update		0.656***(0.003)	0.654***(0.003)	0.644***(0.003)	0.645***(0.003)	0.641***(0.003)
-Video		0.406***(0.009)	0.406***(0.009)	0.410***(0.009)	0.395***(0.009)	0.398***(0.009)
Page-type Category - Celebrity is the base						
-ConsumerProduct		-0.402***(0.002)	-0.392***(0.002)	-0.371***(0.002)	-0.374***(0.002)	-0.347***(0.002)
-Entertainment		0.514***(0.002)	0.519***(0.002)	0.537***(0.002)	0.514***(0.002)	0.529***(0.002)
-Organization		0.391***(0.002)	0.393***(0.002)	0.409***(0.002)	0.400***(0.002)	0.408***(0.002)
-PlaceBusiness		-0.087***(0.005)	-0.082***(0.005)	-0.052***(0.005)	-0.065***(0.005)	-0.021***(0.005)
-Websites		0.123***(0.002)	0.134***(0.002)	0.150***(0.003)	0.136***(0.003)	0.182***(0.003)
FRIENDLIKELY			0.064***(0.001)	-0.051***(0.002)		-0.006***(0.002)
REMFAC				-0.065***(0.002)		0.014***(0.002)
EMOTION				0.217***(0.002)		0.256***(0.002)
EMOTICON				0.109***(0.004)		0.121***(0.004)
HOLIDAYMENTION				-0.391***(0.014)		-0.388***(0.014)
HUMOR				0.022***(0.001)		0.072***(0.002)
PHILANTHROPIC				0.054***(0.002)		0.174***(0.002)
SMALLTALK				-0.055***(0.002)		-0.086***(0.002)
BRANDMENTION					0.071***(0.002)	0.081***(0.002)
DEAL					-0.124***(0.002)	-0.172***(0.002)
PRICECOMPARE					0.043***(0.001)	-0.006***(0.001)
PRICE					-0.368***(0.005)	-0.317***(0.005)
TARGET					0.049***(0.002)	-0.071***(0.002)
PRODAVAIL					0.028***(0.002)	-0.064***(0.002)
PRODLOCATION					0.023***(0.002)	0.011***(0.002)
PRODMENTION					-0.186***(0.002)	-0.151***(0.002)
Demographic proportion variables excluded						
McFadden R-sq.		0.161	0.161	0.165	0.164	0.171
Nagelkerke R-sq.		0.161	0.162	0.165	0.165	0.172
Log-likelihood	-4267283.759	-3612205.117	-3610817.72	-3597891.814	-3598824.088	-3570184.37
Deviance	8130598.643	6820441.36	6817666.566	6791814.753	6793679.301	6736399.867
AIC	8534569.517	7224474.234	7221701.44	7195863.627	7197730.175	7140466.74
N	665916	665916	665916	665916	665916	665916
Significance '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1						

Table 8: Logistic Regression EdgeRank-Corrected Estimates Model Comparison (Comments)

Variable	Intercept only	Controls	Friendly	Brand personality-related	Directly informative	All (No demo controls)
Constant	-4.662***(0.000)	-3.929***(0.001)	-3.979***(0.001)	-3.931***(0.001)	-3.921***(0.001)	-3.892***(0.001)
SMOG		-0.057***(0.000)	-0.057***(0.000)	-0.057***(0.000)	-0.060***(0.000)	-0.061***(0.000)
MSGLEN		-0.001***(0.000)	-0.000***(0.000)	-0.001***(0.000)	-0.000***(0.000)	-0.000***(0.000)
HTTP		-0.214***(0.000)	-0.207***(0.000)	-0.205***(0.000)	-0.205***(0.000)	-0.189***(0.000)
QUESTION		-0.198***(0.000)	-0.195***(0.000)	-0.181***(0.000)	-0.198***(0.000)	-0.175***(0.000)
BLANK		-0.603***(0.002)	-0.612***(0.002)	-0.592***(0.002)	-0.602***(0.002)	-0.596***(0.002)
ASKLIKE		0.576***(0.003)	0.583***(0.003)	0.576***(0.003)	0.564***(0.003)	0.565***(0.003)
ASKCOMMENT		-0.379***(0.011)	-0.359***(0.011)	-0.444***(0.011)	-0.404***(0.011)	-0.465***(0.011)
Message Type - App is the base						
-Link		-0.517***(0.001)	-0.519***(0.001)	-0.518***(0.001)	-0.511***(0.001)	-0.505***(0.001)
-Photo		0.572***(0.001)	0.567***(0.001)	0.561***(0.001)	0.566***(0.001)	0.561***(0.001)
-Status Update		-0.062***(0.001)	-0.067***(0.001)	-0.083***(0.001)	-0.067***(0.001)	-0.083***(0.001)
-Video		0.244***(0.003)	0.245***(0.003)	0.252***(0.003)	0.225***(0.003)	0.231***(0.003)
Page-type Category - Celebrity is the base						
-ConsumerProduct		-0.948***(0.001)	-0.937***(0.001)	-0.941***(0.001)	-0.939***(0.001)	-0.932***(0.001)
-Entertainment		-0.172***(0.001)	-0.168***(0.001)	-0.174***(0.001)	-0.189***(0.001)	-0.193***(0.001)
-Organization		-0.165***(0.001)	-0.164***(0.001)	-0.160***(0.001)	-0.171***(0.001)	-0.171***(0.001)
-PlaceBusiness		-1.305***(0.002)	-1.301***(0.002)	-1.292***(0.002)	-1.294***(0.002)	-1.275***(0.002)
-Websites		0.028***(0.001)	0.038***(0.001)	0.019***(0.001)	0.042***(0.001)	0.041***(0.001)
FRIENDLIKELY			0.072***(0.000)	0.051***(0.001)		0.080***(0.001)
REMACT				-0.065***(0.001)		-0.021***(0.001)
EMOTION				0.250***(0.001)		0.260***(0.001)
EMOTICON				-0.049***(0.001)		-0.020***(0.001)
HOLIDAYMENTION				-0.187***(0.004)		-0.183***(0.004)
HUMOR				-0.009***(0.000)		0.009***(0.000)
PHILANTHROPIC				-0.072***(0.001)		0.002*(0.001)
SMALLTALK				-0.112***(0.001)		-0.146***(0.001)
BRANDMENTION					0.012***(0.000)	0.021***(0.000)
DEAL					-0.166***(0.001)	-0.207***(0.001)
PRICECOMPARE					-0.036***(0.000)	-0.047***(0.000)
PRICE					-0.504***(0.001)	-0.471***(0.001)
TARGET					0.050***(0.001)	-0.019***(0.001)
PRODAVAIL					-0.027***(0.001)	-0.073***(0.001)
PRODLOCATION					0.134***(0.001)	0.138***(0.001)
PRODMENTION					-0.004***(0.001)	0.012***(0.001)
Demographic proportion variables excluded						
McFadden R-sq.		0.191	0.191	0.195	0.196	0.201
Nagelkerke R-sq.		0.192	0.193	0.197	0.197	0.203
Log-likelihood	-33968732.078	-27584573.527	-27566219.315	-27424516.699	-27419099.696	-27221231.997
Deviance	66990021.316	54221704.214	54184995.789	53901590.557	53890756.552	53495021.154
AIC	67937466.156	55169211.054	55132504.629	54849113.398	54838281.393	54442561.994
N	665916	665916	665916	665916	665916	665916
Significance **** 0.001 *** 0.01 * 0.05 ! 0.1						

Table 9: Logistic Regression EdgeRank-Corrected Estimates Model Comparison (*Likes*)

Appendix 7: Results Across Different Page-Types

While messages posted by different page-types do not seem to vary much from one another as shown in Appendix 1, the reaction from users on Facebook might differ by page-type source. In order to investigate how page-types might change the impact of content on engagement, we repeated the main analysis with the same specification on subsetted data on each page type as described and categorized in Table 2.

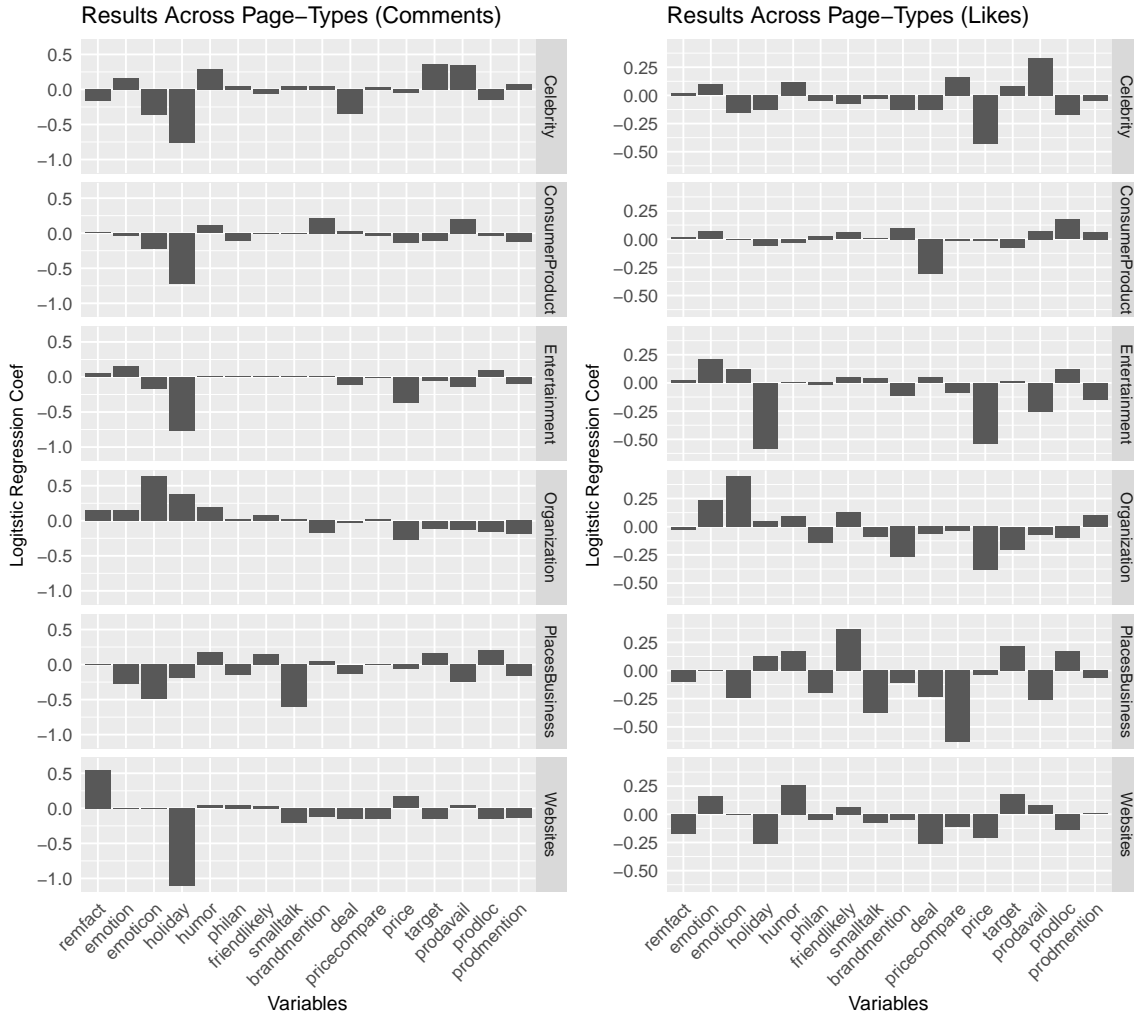


Figure 5: **Logistic Regression by Page Type (Comments and Likes)**: This bar graphs show the coefficients of logistic regression for EdgeRank-corrected model. Only the significant coefficients ($p < 0.05$) are graphed.

Figure 5 shows the results on content effects by page type. Only the statistically significant results are graphed and all results are EdgeRank-corrected. For clarity of visualization, we only present the coefficients for 16 content variables. For most of the content attributes, the direction of effects remain similar across page-types. Consistent with the main results, most informative content attributes have negative coefficients across all page-types. The brand personality-related content attributes show more variation across the page-types. Apart from two instances Celebrity (*Likes*) and PlacesBusiness (Comment), most of brand personality-related

content attribute effects remain positive while most of informative content effects are negative.

References

- Bennett, P. N.: 2006, 'Building Reliable Metaclassifiers for Text Learning'. Ph.D. thesis, CMU.
- Bird, S., E. Klein, and E. Loper: 2009, *Natural Language Processing with Python*. O'Reilly Media.
- Brieman, L.: 1996, 'Bagging predictors'. *Machine Learning* **24**, 123–140.
- Buhrmester, M., T. Kwang, and S. D. Gosling: 2011, 'Amazon's Mechanical Turk: a New Source of Inexpensive, Yet High-Quality, Data?'. *Psychological Science* **6**(1).
- Freund, Y. and R. E. Schapire: 1995, 'A decision-theoretic Generalization of on-line learning and an application to boosting'. *Computational Learning Theory: EUROCOLT* pp. 23–37.
- Hassan, M. H., B. S. Vikas, and H. Fiorletta: 2011, 'Accurate Information Extraction for Quantitative Financial Events'. *CIKM*.
- Hastie, T., R. Tibshirani, and J. Friedman: 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Ipeirotis, P. G., F. Provost, and J. Wang: 2010, 'Quality Management on Amazon Mechanical Turk'. *HCOMP*.
- Jurafsky, D. and J. H. Martin: 2008, *Speech and Language Processing*. Prentice Hall.
- Mason, W. A. and S. Suri: 2012, 'Conducting behavioral research on Amazon's Mechanical Turk'. *Behavior Research Methods* **44**(1), 1–23.
- Paolacci, G., J. Chandler, and P. G. Ipeirotis: 2010, 'Running Experiments on Amazon Mechanical Turk'. *Judgement and Decision Making* **5**(5).
- Sheng, V., F. Provost, and P. G. Ipeirotis: 2007, 'Get another label? Improving data quality and data mining using multiple, noisy labelers'. *14th ACM SIGKDD* pp. 614–622.
- Snow, R., B. O'Connor, D. Jurafsky, and A. Y. Ng: 2008, 'Cheap and fast - but is it good? Evaluating non-expert annotations for natural language tasks'. *Empirical Methods in Natural Language Processing*.
- Xu, L. and A. Krzyzak: 1992, 'Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition'. *IEEE Transactions on systems*.
- Zadrozny, B. and C. Elkan: 2002, 'Transforming Classifier Scores into Accurate Multiclass Probability Estimates'. *KDD*.