ABSTRACT
Social media sites such as Instagram have emerged as popular platforms for sharing ingestion and dining experiences. However, research on characterizing the nutritional information embedded in such content is limited. In this paper, we develop a computational method to extract nutritional information, specifically caloric content from Instagram food posts. Next, we explore how the community reacts specifically to healthy versus non-healthy food postings. Based on a crowdsourced approach, our method was found to detect caloric content in posts with 89% accuracy. We further show the use of Instagram as a platform where sharing of moderately healthy food content is common, and such content also receives the most support from the community.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous

Keywords
Instagram, nutrition, social media, food

1. INTRODUCTION
Social media platforms have emerged as popular tools allowing individuals to share the mundane goings on in their lives. An important element of our daily lives is food; it constitutes a central role in not only defining our identity, culture, and lifestyle, but also impacts our health and wellbeing. In this paper, we pursue the research question on how social media content may be leveraged to extract nutritional information of food people ingest, and the dynamics of how the broader community responds to sharing of food and dining experiences on social media.

Language of food and dietary patterns have been studied in the context of social media [2], however prior literature on extracting nutritional information from social media content is limited. Some works have focused on recipe websites and utilized food names in them as a proxy to derive consumption and dietary patterns of individuals [4, 3]. However the coverage of recipe websites can be limited, and the nutritional information may not always be normalized across websites. Close to our work is the work in [1], in which the website http://caloriecount.about.com/ was used to derive nutritional content in Twitter posts. However the method did not distinguish between actual reports of consumption versus cursory references to food items.

In this paper we address the limitations of prior literature by leveraging a novel dataset from a relatively underexplored social media, Instagram. Instagram has emerged as a highly popular photo sharing social platform, where food and ingestion related content are extensively shared. In 2013, the Business Insider reported that food photos are a phenomenon on Instagram1. A unique aspect of Instagram use is its ubiquitous nature. The platform can only be used via a smartphone, which has allowed individuals to share photos and videos of the food they are consuming anytime, anywhere. Consequently, the platform provides a reliable way to capture people's dining experiences, and thereby the nutritional information of the food they are consuming.

This paper makes the following contributions. First we leverage the textual content of tags associated with Instagram posts to extract nutritional information without actually analyzing the visual content in photos/videos. Second, via analysis of the patterns of “likes” and comments on posts we examine how the greater community responds to food and ingestion content on the platform.

2. DATA AND METHOD
Data Collection. We describe our data and method of extracting nutritional information from Instagram posts. First, we compiled a set of food-related words which could be used to query Instagram and collect a dataset of food posts. We obtained a list of 564 words (henceforth referred to as canonical food names) from an online food vocabulary word list2 [2]. Two researchers manually went through the list to remove words which were not indicative of any food item (e.g., “utensils”, “roast”, “hot”). Additionally we added 24 other words which were likely to be indicative of food posts, based on prior literature [1] and our qualitative inspection of food content sharing on Instagram. Some of them were: “breakfast”, “lunch”, “yummy”, “meal”, “foodporn”. Generic tags like “food” and “cooking” were considered but not finally included—qualitative inspections of searches of these tags on Instagram yielded noisy posts. The filtered canonical food names were then used as seed tags to get the posts.

Using the list of canonical names, next we collected 1,815,752 posts from Instagram between Nov 27-Dec 9, 2014. We used Instagram’s official API3 to get both image and video posts. For the purposes of this paper, we considered only English language public posts. We did not download the image or video themselves except for validation discussed later.

Extracting Nutritional Information. Following data collection,

2http://www.enchantedlearning.com/wordlist/food.shtml
3http://instagram.com/developer/
We observe a non-linear trend in the distribution—the maximum number of posts lie within the calorie range 150-300. Sharing content on high calorie food e.g., beyond 400 is less common among the Instagram sample of users we study. At the same time, while there are relatively more posts in the low calorie i.e., 1-100 range, still they are fewer than ones in the 150-300 range.

Next we investigate how the broader Instagram community responds to posts of food with different calorific content—we use the number of “likes” and comments on the posts for the purpose. We generated probability frequency distributions for both these attributes over calorific content of their corresponding posts. Then we compared each of these distributions with the distribution of posts over calories. Normalized mutual information was observed to be .88 and .84 respectively for the “likes” and comments distributions. This shows that “likes” and comments on the posts are higher for posts that are in the moderate calorie range, with less community endorsement and response on posts of very low or very high calorie food. Specifically we found that mean “likes” for posts in the 150-300 calorie range is 22.6 (5.2 for comments); the same is 12.7 for those over 400 calories (1.2 for comments), and it is 16 for posts between 1-100 calories (3.3 for comments). We observed these differences to be statistically significant based on Welch $t$-tests ($p < .001$).

### 4. CONCLUSION

Our results showed that reliable and high quality nutritional information (accuracy 89%) may be gleaned from tags associated with food-posts on Instagram. Further, the Instagram community tends to share food photos/videos of moderate calorific content, with high calorie food posts being less common. We further observed that the distribution of number of likes and comments on food posts show high correlated with the distribution of posts, showing that the community tends to like or comment more frequently on posts of moderate calorific content, compared to those of extremely low or extremely high calorie food items.

We note that we did not use any image content analysis in our nutritional information extraction method. Finally, as we relied on English language posts, cross-cultural generalizations cannot be derived. Addressing these limitations are promising directions for future research.

### 5. REFERENCES


---

**Table 1: Example posts and calorific content.**

<table>
<thead>
<tr>
<th>Post tags</th>
<th>Canonical name</th>
<th>Calorie</th>
</tr>
</thead>
<tbody>
<tr>
<td>miami, organic, garden, usa,</td>
<td>okra</td>
<td>28.33</td>
</tr>
<tr>
<td>food, fresh, creole, okra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bykaila, postre, luneslight,</td>
<td>flan</td>
<td>177.25</td>
</tr>
<tr>
<td>cuantocomer, dessert, flan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yam, instafood, hamburger,</td>
<td>yam, ham-</td>
<td>219.93</td>
</tr>
<tr>
<td>the-bird</td>
<td>burger</td>
<td></td>
</tr>
<tr>
<td>muesli, granola, easterngrano-</td>
<td>granola, muesli</td>
<td>330.87</td>
</tr>
<tr>
<td>nola, localislovely</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cheesecake, breakfast, cheese-</td>
<td>cheesecake</td>
<td>402</td>
</tr>
<tr>
<td>cakefactory, redvelvet</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 1: Distribution of number of posts over calorific values.**

![Figure 1: Distribution of number of posts over calorific values.](http://ndb.nal.usda.gov/)