On the Relations Between Cooking Interests, Hobbies and Nutritional Values of Online Recipes

Implications for Health-Aware Recipe Recommender Systems

Christoph Trattner
MODUL University Vienna, Austria
christoph.trattner@modul.ac.at

Markus Rokicki
L3S Research Center, Germany
rokicki@L3S.de

Eelco Herder
L3S Research Center, Germany & Radboud Universiteit Nijmegen, the Netherlands
eelcoherder@acm.org

ABSTRACT

In this paper, we investigate differences between recipes uploaded by users and recipes bookmarked by users. The results indicate that uploaded recipes outperform bookmarked recipes in terms of healthiness. Further, health scores and nutritional values of these recipes are highly related to the stated cooking interests: for example, Southern Food lovers do not eat as healthy as those who prefer the Mediterranean or Middle-Eastern cuisine. A disturbing finding is that interest in the category 'Kids' is associated with bad values for all nutritional measures. We also found some interactions between hobbies such as biking, hunting or knitting and nutritional values. These insights pave way to the design of health-aware recipe recommender systems that take a user's food preferences into account; in addition, taking a user's lifestyle and hobbies into account would provide valuable input to persuasive systems.

KEYWORDS

Online recipes; nutrition; food preferences; hobbies; cooking interests

1 INTRODUCTION

Several investigations[1] have shown that the recipes that users publish and read on online recipe sites are quite representative for their actual eating habits and preferences. Unfortunately, as we have shown in two recent earlier studies [22, 23], online recipes are typically not very healthy. Users also tend to interact most with recipes with lower scores in terms of nutritional value.

From social media research, it is known that users tend to present themselves in a positive light, by carefully selecting the content for their user profiles and timelines. Similarly, one would expect that users would only post their best creations on recipe websites such as Allrecipes[2]. It is not clear, though, whether these 'best' recipes are, on average, better in terms of nutrition, or rather in terms of taste or festiveness [23]. In this paper, we investigate the differences between the recipes that individual users bookmarked and the recipes that they have uploaded. We do so by comparing individual users' actual preferences (as reflected by their bookmarked recipes) with their self-presentation (their uploaded recipes) on Allrecipes. Particularly, we focus on nutritional value and accordance with health criteria as defined by the WHO and the FSA [12].

Further, it is likely that uploading and bookmarking behavior – and the differences between the two – are influenced by the user's preferred cuisine. For instance, users who mainly interact with recipes in the category 'Vegetarian' are probably more health-conscious (or are more likely to present them as such) than users who mainly interact with 'Grilling & Barbecue'. We investigate this by comparing upload and bookmark behavior for the different culinary interest categories on Allrecipes.

Finally, we explore the relation between individual users' hobbies and eating patterns in terms of nutritional values and types of recipes. It is likely that athletic users who state 'Biking' in their profiles have different preferences than users with 'WineTasting' as a hobby; 'Fishing' and 'Hunting' are expected to have some obvious impact on food preferences as well.

Contributions. We will show that, on average, the collection of users' uploaded recipes scores better in terms of nutritional values than the bookmarked recipes; this indicates that health-consciousness is part of users' self-presentation. We will show that some cuisines (e.g. Southern, Mexican) consistently perform worse in terms of healthiness than others (e.g. Middle-Eastern, Indian). These insights can be used for health-aware recommender systems that better take the user's culinary preferences into account. We also discuss patterns between a user's hobbies and eating habits, which can create a basis for persuasive systems that provide targeted motivating hints and relevant goals.

2 BACKGROUND

The way people interact with recipes online in the form of bookmarks, ratings or food posts on Twitter or other online food forums can give clues about their real-world food preferences, eating habits or health status [23]. Recently, there have been several notable studies on the online food context that analyzed the former mentioned aspects in more detail. This section reviews the most relevant and recent works in this area.

One of the most recent works in this context are the studies of Kusmierczyk et al. and Trattner et al. who analyzed upload and

[1] http://www.slate.com/articles/life/food/2016/05/allrecipes_reveals_the_enormous_gap_between_foodie_culture_and_what_americans.html
[2] http://www.allrecipes.com

UMAP'17 Adjunct, July 09-12, 2017, Bratislava, Slovakia

DOI: 10.1145/3099023.3099072

Table 1: Basic statistics of the Allrecipes.com dataset containing recipes, nutritional information and bookmarks of users between the years 2000 and 2015.

Total published recipes	60,983
Recipes containing nutrition information	58,263
Users with published recipes	25,037
Recipes bookmarked	58,194
Bookmarks	17,190,534
Users who provided bookmarks	155,769

rating interaction data from the German community platform Kochbar.de [13]. The findings of these studies is that there are clearly observable temporal trends in terms of nutrition (fat, proteins, carbohydrates, and energy) in the recipes. Similar patterns were also observed by Wagner et al. [25] and West et al. [26]. West and colleagues furthermore found correlations between recipes accessed via search engines and incidence of diet-related illness, which resembles findings reported recently by Said and Bellogin [19], De Coudhury et al. [6] and Abbar et al. [4] or Ofli et al. [14] in the context of Allrecipes.com, Instagram and Twitter respectively.

Rokicki et al. [16] investigated differences in nutritional values between user recipes created by different user groups finding, for example, that recipes uploaded by females are, on average, richer in carbohydrates. The carbohydrate content of recipes seems to decrease with the age of the user mirroring the advice given by most nutrition advice centers. Ahn et al. [5] mined and analyzed three different large-scale online food community platforms from Europe, the US and China to unveil patterns on how recipes vary between regions and to find out which flavor components make, for instance, Indian food different from the rest of the world. Yet another recent work are the studies of Trattner and Elsweiler [22, 23], who found that online recipes are not only significantly more unhealthy than supermarket ready meals and TV chef recipes but also that users of Allrecipes.com tend to interact with unhealthy recipes more likely than with healthy ones. Finally, Wagner and Aiello [24] and Rokicki et al. [17] studied gender differences in eating preferences in the context of the online platform Flickr and Kochbar.de.

Although all of these works provide useful insights into real-world eating habits, trends or health issues, not much work has been done yet to understand how online food preferences relate to nutritional properties of online recipes [23]. To contribute to knowledge in this area, we present in this paper a study, showing that there are significant observable differences in terms of nutritional values (health) of online recipes and people's uploading and bookmarking behavior. Furthermore, we study these in the context of stated cooking interests and hobbies and provide insights what that would mean for the implementation of health-aware food recommender systems [8, 11] that rely on the users bookmarks and upload history as training instances.

3 MATERIALS

In this work, we make use of a web crawl of the online platform Allrecipes.com that has been performed in July 2015. The crawler collected 60,983 recipes, 25,037 user profiles and 17,190,534 bookmarks generated between the years 2000 and 2015 on the Allrecipes.com website. We focus only on recipes that have been published on the main site and ignore personal recipes, which are often incomplete and do not provide nutrition information. Recipes and according user profiles were collected through the Allrecipes.com sitemap available through the robots.txt file.

Allrecipes.com was chosen as it claims to be the world's largest food-focused social network. The site has a community of 40 million users who annually access 3 billion recipes and who originate from about 24 countries [2].

In addition to the bookmarks and user profiles (stating among other things the users' hobbies and cooking interests), the following information was collected for each recipe: total energy (kCal), protein (g), carbohydrate (g), sugar (g), sodium (g), fat (g) and saturated fat (g). The nutritional meta-data was available via Allrecipes.com and collected during the main crawl. Allrecipes.com estimates the nutritional content for an uploaded recipe by matching the contained ingredients with those in the ESHA research database [3]. Table 1 provides an overview of the basic statistics of the dataset.

4 METHODS

In the following section, we describe briefly the methods employed in our research.

Data Preprocessing. Our analysis contrasts the nutritional properties and the healthiness of uploaded and bookmarked recipes. Furthermore, we are interested in studying this in the context of the users' stated cooking interests and hobbies. To this end, we restrict our analysis to users who have uploaded and bookmarked at least one recipe at the same time. In total, this reduces the original dataset with 25,037 users uploading recipes to 7644 users. In order to obtain the users' hobbies and cooking interests, the user profiles of all online cooks were parsed. In total, Allrecipes allows the users to state 20 unique hobbies and 20 unique interests which are pre-defined by the platform and can be added to the users profiles. On average, we could observe that in our data sample users in Allrecipes have five hobbies stated and seven cooking interests. An overview of the two distributions is given in Figure 1.

Measuring the Healthiness of Recipes. To determine the healthiness of an online recipe, we make use of two recognized international standards: The World Health Organization (WHO) guidelines [1] and the UK FSA "traffic light" system for labeling food [27]. The *FSA score* is based on traffic light labels – green for healthy, amber and red for unhealthy – for four macro-nutrients (sugar, sodium, fat and saturated fats). In order to derive a single metric, we follow the procedure of Sacks et al. [18] who first assign an integer value to each color (green=1, amber=2 and red=3) then sum the scores for each macro-nutrient, resulting in a final range from 4 (very healthy recipe) to 12 (very unhealthy recipe). The *WHO score* is based on compliance to recommended ranges for a larger number of macro-nutrients. We follow the approach of Howard et al. [12], who chose the 7 most important nutritional measures (i.e. proteins, carbohydrates, sugars, sodium, fats, saturated fats, and fibers) and their corresponding ranges to determine a so-called WHO health score, ranging from 0 to 7. A recipe with a WHO score of 7 (all WHO ranges are fulfilled) is interpreted as being very healthy, whereas a score of 0 (none of the ranges are met) is seen as very unhealthy.

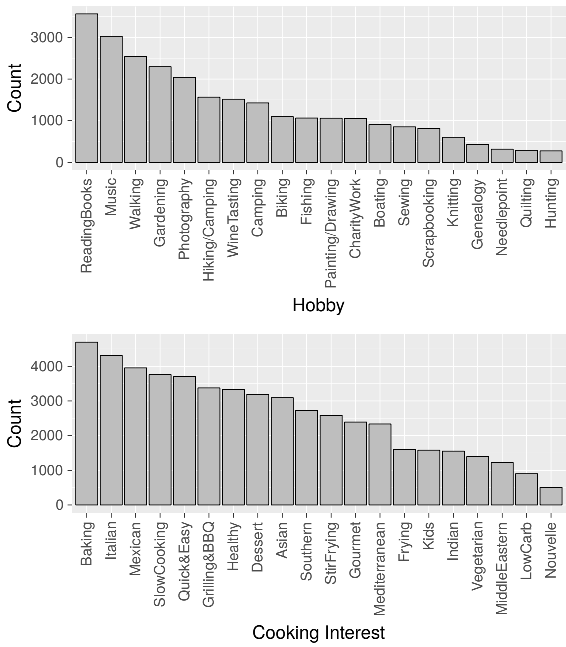

Figure 1: Distributions of interests and hobbies in user profiles for users with at least one upload and one bookmark.

This framework has already been successfully applied in our previous work that investigated the healthiness of internet-sourced recipes [22, 23].

Statistical Analysis. We test for significance of differences in the distributions of per user nutrition values for uploads and bookmarks by means of Wilcoxon Rank-Sum test. Further, we distinguish users based on the hobbies and cooking interests reported in their profiles on the platform. To test for differences in healthiness between individual hobbies or individual interests, we perform Kruskal-Wallis tests and employ Dunn's test (with Bonferroni correction) as a post-hoc test method to identify specific pairwise differences within the two groups (uploads and bookmarks). To test for significant differences between groups Wilcoxon Rank-Sum tests were performed. For space reasons, the specific p-values for these statistical comparisons could not be included into this paper, as for every sub-plot (macro-nutrient) as shown a Figure 3 and 4 one large p-value matrix was produced. However, the two plots provide error bars, which help to estimate whether or not there is a significant difference in terms of observable means. In fact, where the error bars of the upload and bookmark group overlap (and also between the two groups), no statistically significant differences are observable.

5 RESULTS

We start this section with a discussion on differences in nutritional values between uploaded and bookmarked recipes. After this general comparison, we separate users first with respect to their cooking interests and then with respect to their hobbies. Several interesting as well as disturbing patterns have been found.

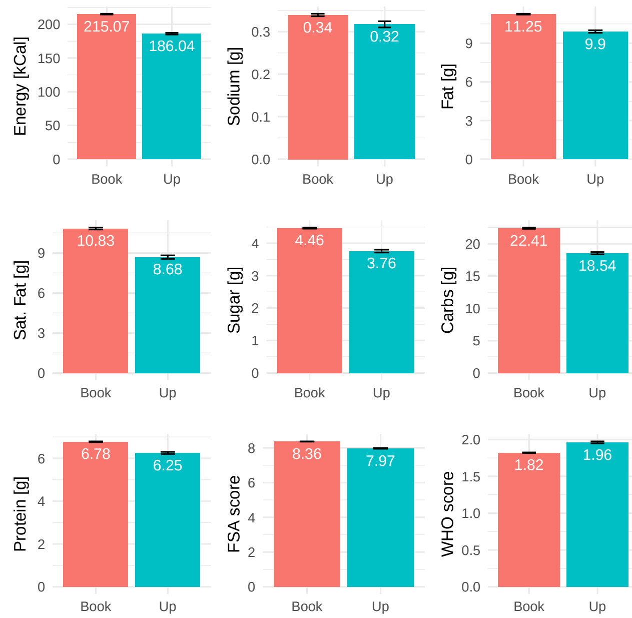

Figure 2: Means and std. errors in terms of macro nutrients (normalized to 100g per recipe) and health scores for recipes bookmarked (book) and uploaded (up). As shown there are sign. observable differences between what people actually upload and what they like (bookmark). All pairwise comparisons show statistically significant differences employing a Wilcox Ranksum test ($p < .001$).

5.1 Uploads vs Bookmarks

In Figure 2, the average values for several relevant nutritional aspects are given. Bottom-right are the calculated health scores according to the FSA and WHO (see Section 4 for more details). For FSA, lower values indicate healthier meals; for WHO, higher values are healthier. For both measures, the uploaded recipes outperform the bookmarked recipes in terms of healthiness significantly – with FSA scores of 8.36 (bookmarks) vs 7.97 (uploads, $p < .001$) and WHO scores of 1.82 (bookmarks) vs 1.96 (uploads, $p < .001$).

When looking at the individual nutritional aspects, we can see the largest differences for energy (215.07 vs 186.04, $p < .001$), saturated fat (10.83 vs 8.68, $p < .001$), sugar (4.46 vs 3.76, $p < .001$) and carbs (22.41 vs 18.54, $p < .001$) – aspects that are closely associated with illnesses such as heart diseases and diabetes [27]. Interestingly, the differences for sodium are relatively small.

In sum, on all aspects uploaded recipes have better nutrition scores than bookmarked recipes. Without a qualitative study, it is hard to prove that this is a result by users' desire for positive self-representation in terms of health. We suspect, however, that this is not the case: nutritional values are relatively invisible in the user interface compared to a recipe's title, ingredients, preparation steps and pictures [9].

5.2 Cooking Interests and Hobbies

Thus far, we have looked at differences between uploading and bookmarking behavior for recipes in general. However, it is likely that different types of users have different behavior. In this section,

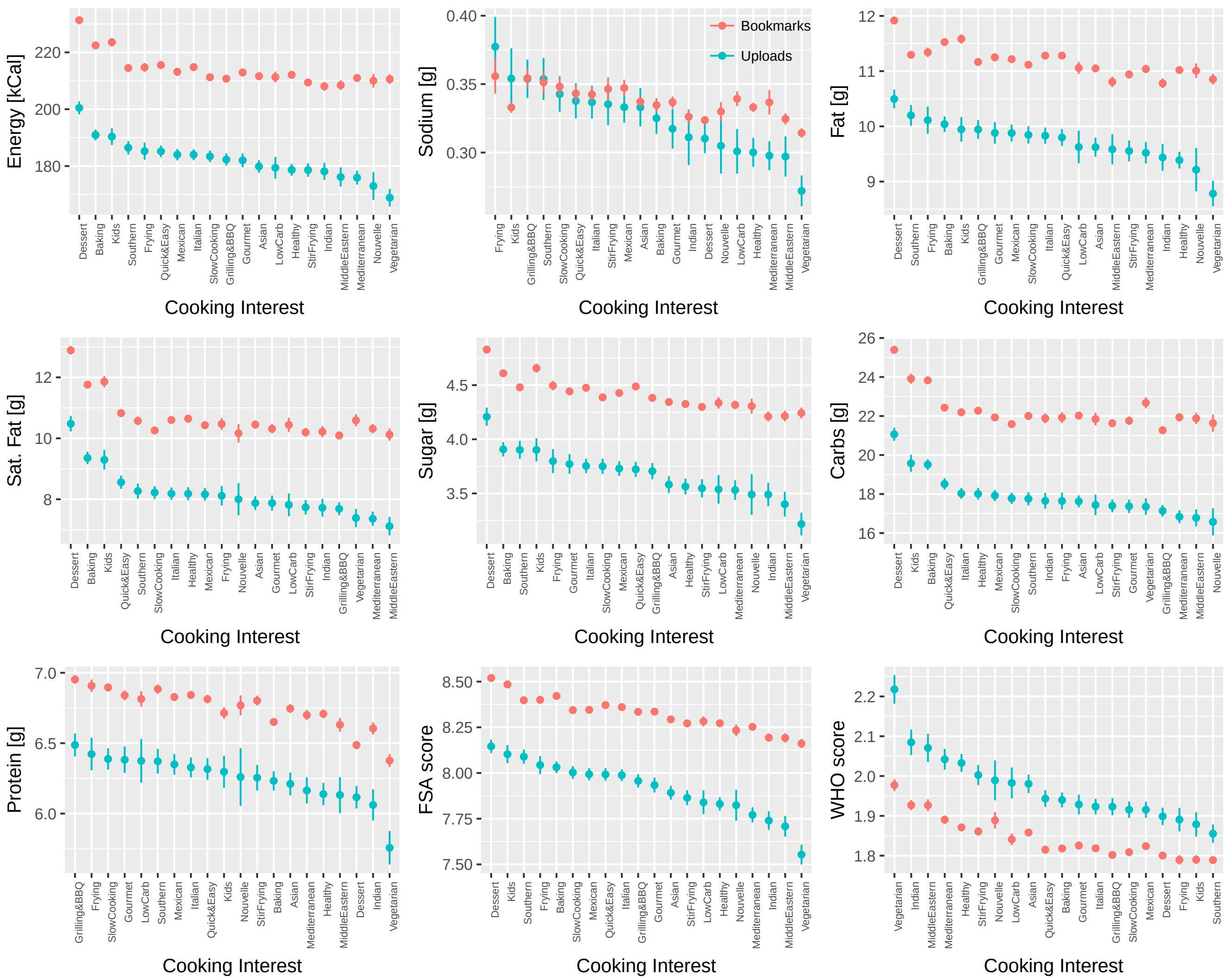

Figure 3: Means and std. errors in terms of macro nutrients (normalized to 100g per recipe) and health scores for recipes bookmarked and uploaded in the context of cooking interests denoted by the users in Allrecipes.com. Red dots highlight bookmarks and bluish green dots uploads. A Kruskal-Wallis test performed for each macro-nutrient and health score reveals there are sign. differences within the two groups - uploads and bookmarks ($p < .001$).

we separate users with respect to their *cooking interests* and their *hobbies*. Users could choose these elements from a list to be included in their user profile. The distribution of cooking interests and hobbies is displayed in Figure 1. As expected, general hobbies such as reading, music and walking are far more prevalent than more specialized hobbies such as quilting or hunting. Regarding cooking interests, 'Baking', 'Italian' and 'Mexican' food are mostly represented, which reflects a largely female and largely Western-American user population. 'Low Carb' and 'Nouvelle' (Cuisine) are at the bottom of the list.

Cooking Interests. Figure 3 shows the nutritional values for the uploaded (bluish green) and bookmarked (red) recipes. For each diagram, the cooking interests are sorted in descending order with respect to the values for uploaded recipes.

In almost all cases, the values for bookmarked recipes are higher than for uploaded recipes (remember that the scale is opposite for the WHO scores) and distances are about similar for cooking interests on each measure. A notable exception is sodium, with relatively lower values for bookmarked recipes in the categories 'Frying' and 'Kids' - but in absolute numbers these are the meals with highest sodium values.

A disturbing observation is that the cooking interest 'Kids' has high (unhealthy) scores on the FSA scale and low (unhealthy) scores on the WHO scale. The 'Kids' category scores consistently high on energy, sodium, fat, saturated fat, sugar and carbs. Only for protein, the Kids category has average values.

On the healthy side, 'Vegetarian', 'Middle-Eastern', 'Indian' and 'Mediterranean' are among the top, according to FSA and WHO,

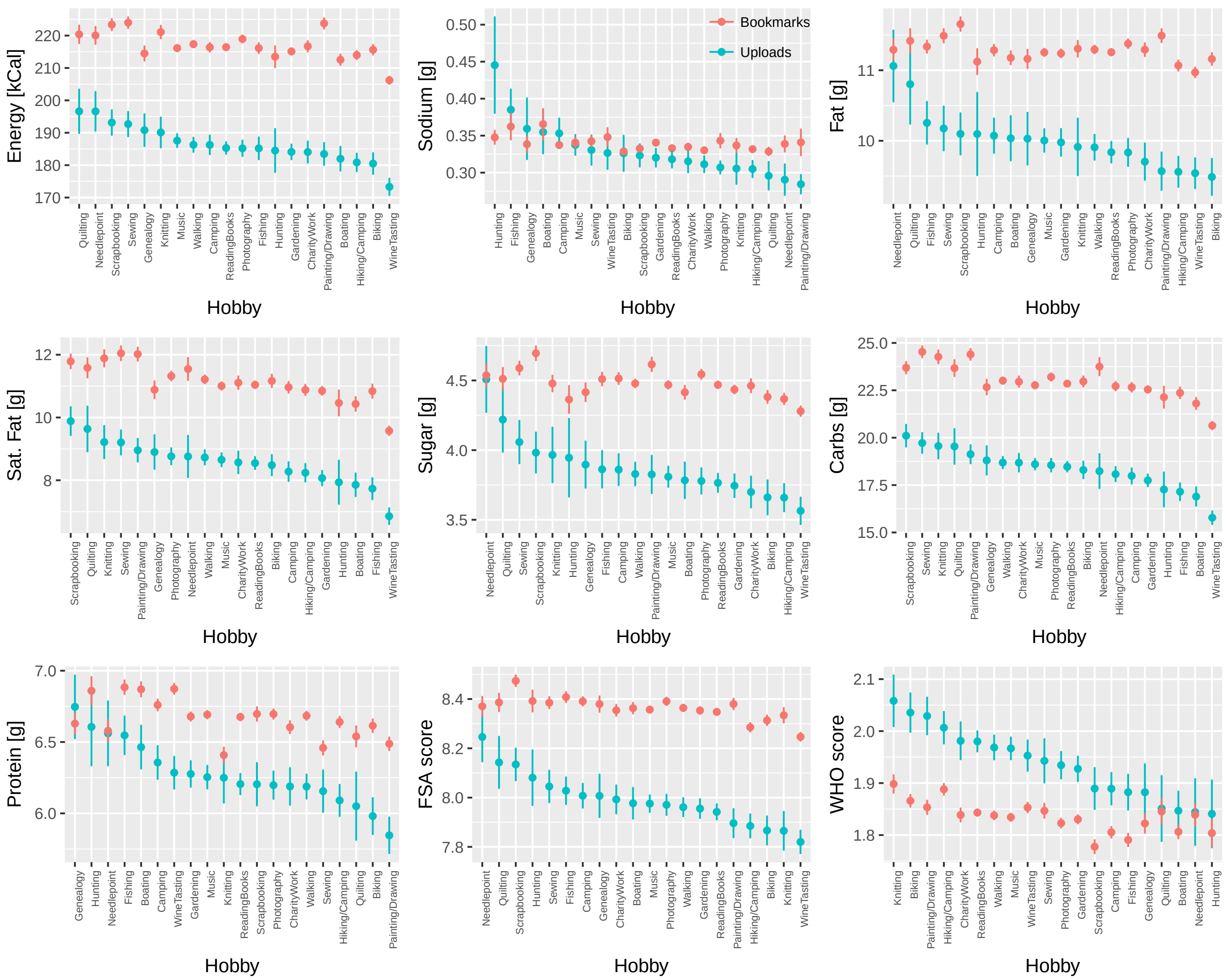

Figure 4: Means and std. errors in terms of macro nutrients (normalized to 100g per recipe) and health scores for recipes bookmarked (red dots) and uploaded (bluish green dots) in the context of hobbies denoted by the users in Allrecipes.com. A Kruskal-Wallis test performed for each macro-nutrient and health score reveals there are sign. differences within the two groups - uploads and bookmarks ($p < .001$).

only then followed by 'Healthy'. This confirms that these four cuisines are, on average, healthier than other cuisines. By contrast, the category 'Southern' (Southern United States)[3], which is known to be rich in gravy[4] consistently is on the unhealthy side of the spectrum.

Hobbies. The average nutrition values for uploaded and bookmarked recipes for each hobby category are depicted in Figure 4, similar as in the previous figure. From the less regular patterns for the bookmarks, it can be observed that the relation between hobbies and eating patterns is less straightforward than for cooking interests. Still, some trends can be observed.

Active hobbies, such as 'Biking', 'Hiking' and 'Boating' are associated with lower intake of energy, fat and carbs. People with an interest in 'Hunting' and 'Fishing' score high on protein and sodium, which was to be expected. Creative 'feminine' hobbies such as knitting and sewing are associated with high fat, sugar and carbs, which is associated with baking (as discussed in more detail in [17], baking is a female-dominated activity).

6 DISCUSSION AND CONCLUSIONS

In this paper, we showed several differences in users' food preferences. Explicitly indicated preferences, as expressed by the recipes they upload themselves, are generally healthier than implicitly indicated preferences, as expressed by the bookmarked recipes. Furthermore, we have shown that the average nutritional values - and

[3] https://en.wikipedia.org/wiki/Cuisine_of_the_Southern_United_States
[4] http://www.prevention.com/food/healthy-eating-tips/southern-diet-might-be-least-healthy-nationwide

the differences between uploaded and bookmarked recipes - vary quite a bit between cooking interests and corresponding cuisines. Some hobbies are related to specific eating patterns as well. We believe that these findings provide a basis for more targeted and persuasive recommendations.

As we previously argued in [17], food preferences are not only guided by taste, but also by other aspects, such as dietary needs, seasonality, availability of ingredients, and societal conventions and expectations. Relatively successful attempts have been made to implement health-aware food recommenders [7, 9, 11, 23], but mainly making use of simple content-based or collaborative methods. As stated in the conclusions of Freyne and Berkovsky [10], generating recipe recommendations needs a more complex approach and particularly diversity and user satisfaction need to be guarded.

Tailored health communication is an approach in which data about a specific user are used to determine the most appropriate strategy to meet that person's unique needs [15]. Persuasive techniques play an important role in the process. We believe that – particularly given the observed differences between user groups – taking the user's culinary interests and lifestyle (in terms of hobbies) into account would form a basis for improved explanations and persuasive strategies in future recipe recommender systems [21]. A 'Grilling and BBQ' enthusiast with hobbies such as 'Hunting' and 'Fishing' arguably needs different motivations than an 'Indian Food' lover who enjoys 'Knitting' and 'Painting'. Particular attention should be paid to user groups – and recipe categories – that consistently and disturbingly perform bad in terms of nutrition, such as 'Kids'.

Health-aware recipe recommendation usually makes use of the same (content-based or collaborative) strategies throughout the whole recipe collection [20]. Given the differences in terms of nutritional values, as well as in ingredients and preparation methods (such as the use of oil and fat, or salt) that we found in this paper, it would be very worthwhile to investigate cuisine-specific strategies – and explanations – for providing healthy recommendations that convince and satisfy the user.

Our analysis is based on a representative, large dataset drawn from Allrecipes.com, the world's largest food-focused social network, with users from a large number of countries. However, the number of cooking interests and hobbies that users could indicate in their profiles is fixed, and it seems that Allrecipes' selection of cooking interests and hobbies to be included did involve some subjective choices. For this reason, a deeper analysis of the impact of user lifestyle on dietary choices has not been possible using this dataset. We plan to continue this analysis with a different dataset of a recipe website that contains full-text user profiles.

REFERENCES

[1] 2014. FSA. Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets. available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300886/2902158_FoP_Nutrition_2014.pdf. Last accessed on 27.6.2016. (2014).

[2] 2016. Allrecipe.com Press report. available at http://press.allrecipes.com/. Last accessed on 20.6.2016. (2016).

[3] 2016. ESHA, Nutrition Labeling Software. available at http://www.esha.com/. Last accessed on 20.6.2016. (2016).

[4] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. You Tweet What You Eat: Studying Food Consumption Through Twitter. In *Proc. of CHI'15*.

[5] Yong-Yeol Ahn, Sebastian E Ahnert, James P Bagrow, and Albert-László Barabási. 2011. Flavor network and the principles of food pairing. *Scientific reports* 1 (2011).

[6] Munmun De Choudhury and Sanket S Sharma. Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media. In *Proc. of CSCW '16*.

[7] David Elsweiler and Morgan Harvey. Towards automatic meal plan recommendations for balanced nutrition. In *Proc. of RecSys'15*. ACM, 313–316.

[8] David Elsweiler, Morgan Harvey, Bernd Ludwig, and Alan Said. Bringing the "healthy" into Food Recommenders. In *Proc. of DRMS'15*. 33–36.

[9] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting Food Choice Biases for Healthier Recipe Recommendation. In *Proceedings of the ACM SIGIR Conference*.

[10] Jill Freyne and Shlomo Berkovsky. Recommending food: Reasoning on recipes and ingredients. In *Proc. of UMAP'10*. 381–386.

[11] Mouzhi Ge, Francesco Ricci, and David Massimo. Health-aware Food Recommender System. In *Proc. of RecSys '15*. 333–334.

[12] Simon Howard, Jean Adams, Martin White, and others. 2012. Nutritional content of supermarket ready meals and recipes by television chefs in the United Kingdom: cross sectional study. *BMJ* 345 (2012).

[13] Tomasz Kusmierczyk, Christoph Trattner, and Kjetil Nørvåg. Temporality in online food recipe consumption and production. In *Proc. of WWW'15*.

[14] Ferda Ofli, Yusuf Aytar, Ingmar Weber, Raggi al Hammouri, and Antonio Torralba. 2017. Is Saki #Delicious?: The Food Perception Gap on Instagram and Its Relation to Health. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 509–518. DOI: http://dx.doi.org/10.1145/3038912.3052663

[15] Barbara K Rimer and Matthew W Kreuter. 2006. Advancing tailored health communication: A persuasion and message effects perspective. *Journal of Communication* 56, s1 (2006).

[16] Markus Rokicki, Eelco Herder, and Elena Demidova. 2015. What's on my plate: Towards recommending recipe variations for diabetes patients. *Proc. of UMAP'15 LBRS* (2015).

[17] Markus Rokicki, Eelco Herder, Tomasz Kusmierczyk, and Christoph Trattner. Plate and Prejudice: Gender Differences in Online Cooking. In *Proc. of UMAP'16*. 207–215.

[18] Gary Sacks, Mike Rayner, and Boyd Swinburn. 2009. Impact of front-of-pack 'traffic-light'nutrition labelling on consumer food purchases in the UK. *Health promotion international* 24, 4 (2009), 344–352.

[19] Alan Said and Alejandro Bellogín. You are What You Eat! Tracking Health Through Recipe Interactions. In *Proc. of RSWeb'14*.

[20] H. Schaefer, S. Hors-Fraile, R. Karumur, A. Calero Valdez, H. Torkamaan, T. Ulmer, A. Said, and C. Trattner. Towards Health (Aware) Recommender Systems. In *Proc. of DH'17*.

[21] Nava Tintarev and Judith Masthoff. 2007. A survey of explanations in recommender systems. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*. IEEE, 801–810.

[22] Christoph Trattner and David Elsweiler. 2017. Estimating the Heathiness of Internet Recipes: A Cross sectional study. *Frontiers in Public Health* (2017).

[23] Christoph Trattner and David Elsweiler. 2017. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th International Conference on World Wide Web*. 489–498.

[24] Claudia Wagner and Luca Maria Aiello. Men eat on Mars, Women on Venus? An Empirical Study of Food-Images. In *Proc. of WebSci'15 Posters*.

[25] Claudia Wagner, Philipp Singer, and Markus Strohmaier. 2014. The nature and evolution of online food preferences. *EPJ Data Science* 3, 1 (2014), 1–22.

[26] Robert West, Ryen W. White, and Eric Horvitz. From Cookies to Cooks: Insights on Dietary Patterns via Analysis of Web Usage Logs. In *Proc. of WWW'13*. 1399–1410.

[27] Joint Who and FAO Expert Consultation. 2003. Diet, nutrition and the prevention of chronic diseases. *World Health Organ Tech Rep Ser* 916, i-viii (2003).