Understanding and Predicting Online Food Recipe Production Patterns

Tomasz Kusmierczyk
NTNU, Trondheim, Norway
tomaszku@idi.ntnu.no

Christoph Trattner
Know-Center, Graz, Austria
ctrattner@know-center.at

Kjetil Nørvåg
NTNU, Trondheim, Norway
noervaag@idi.ntnu.no

ABSTRACT

Studying online food patterns has recently become an active field of research. While there are a growing body of studies that investigate how online food in consumed, little effort has been devoted yet to understand how online food recipes are being created. To contribute to this lack of knowledge in the area, we present in this paper the results of a large-scale study that aims at understanding how historical, social and temporal factors impact on the online food creation process. Several experiments reveal the extent to which various factors are useful in predicting future recipe production.

Keywords

online food; creational patterns; recipe creation; ingredient usage; predictive modeling; food recommender systems

1. INTRODUCTION

Investigating online user patterns does not only help us in understanding and learning about what people want and need but also how to improve user experiences. In the context of food, and in particular nutrition research, a huge body of literature exists that tries to understand how we consume or produce food in our daily lives. Previous studies were typically performed offline in a survey-based format and were capturing only a small fraction of a population, failing to elicit data objectively. Recent innovations in research follow a more pragmatic way by mining patterns users leave behind in the World Wide Web. The main advantage of such a method is that behavior can be computed without the direct involvement of the user. As such, it allows to learn user behavior or the behavior of a whole population fast, objective and on a large scale.

Problem Statement. While current research in this area is mostly devoted to understand how people consume food online, i.e. how people search, view or rate recipes in online food community forums and how this, e.g., correlates to real-life health related issues, little attention has yet been devoted yet to understand online food production patterns. To the best of our knowledge, *no other work has yet been devoted to the problem of understanding and predicting online food production patterns such as type of recipe being created or ingredients used.*

HT '16, July 10-13, 2016, Halifax, NS, Canada
© 2016 ACM. ISBN 978-1-4503-4247-6/16/07...$15.00
DOI: http://dx.doi.org/10.1145/2914586.2914632

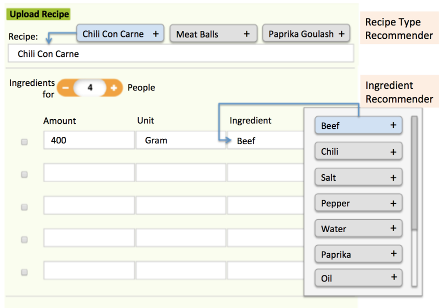


Figure 1: Example of a recommender system that tries to support the user in his food production process in the online food community *kochbar.de.*

The main goal of this research is to (i) *shed light on online food production patterns*, (ii) *the predictability of these patterns*, and in particular (iii) *the impact of information consumption as well as historical, social and temporal factors*. The two production patterns we are interested in investigating in this paper are (i) *type of recipe being created* by a user in an online food community system and (ii) the *types of ingredients being used*.

Application. The topic of this paper not only enables us to understand, how online food communities evolve and how this might impact our health, but could also help in the design of food recommender systems that could support people in the food production process. Currently, this process is a time-consuming task, and as shown in many other domains such a system would help the user not only to perform this kind of task more efficient but could also increase the user experience or increase the quality of the content being created. Figure 1 shows such a system, that aims at predicting and recommending type of recipe the user is likely going to create, as well as the ingredients used to create the recipe.

Research Questions. To drive our research, we have identified two research questions. The first one is on understanding the factors and their corresponding correlations with the online food creation process while the second one is on the usefulness of these correlations in a prediction task. In particular, the questions are as follows:

- **RQ1.** To what extent do historical, social and temporal factors have an impact on the online food production patterns such as type of recipe being created and ingredients used?

- **RQ2.** To what extent are these factors useful to predict online food recipe production?

Table 1: Basic statistics of the dataset.

#users	199k	#ingredients	1,483	#recipes	406k
#publishing	18k	#food types	2,523	#ratings	7,795k
#rating	19k	#categories	246		

2. RELATED WORK

Studying online food patterns is a relatively new strand of research and only a few related studies exist so far in this context.

From the consumption side, the most prominent work is the study of West et al. [18]. In their work, they analyzed seasonal trends and correlation between heart diseases and online food consumption using website log files from allrecipes.com.[1] Similar studies exploiting online recipe ratings were performed by Said and Bellogin [13], by Abbar et al. in the context of Twitter [1], and recently by De Choudhury and Sharman [5] in the context of Instagram. Another recent study in this context of consumption is the work of Wagner et al. [17], who investigated the dynamics of online food consumption in an European food community platform based on data from log files.

The most popular work in this strand of research in the area of recommender systems was done by Berkovsky and Freyne [3, 7], who were the first to study online food recipe consumption patterns and preferences, with the purpose of building systems for recommending recipes. Another relevant work is the experiments of Teng et al. [14] where they try to induce a number of features to train a statistical model that is able to recommend recipes to users. Recent studies also worth mention here are the works of Trevisiol et al. [16], Ge et al. [9], and Elsweiler & Harvey [6], who studied intelligent meal planning and health-aware food recommender systems. Finally, there is the study of Rokicki et al. [12] proposing an interesting approach to recommend healthy recipes to diabetes patients.

As mentioned in the introduction, from the producer side, very little research has been performed yet. Apart from our own preliminary research on temporality in online food recipe production [10, 11, 15], to the best of our knowledge, the only other related work is the one by Ahn et al. [2], who mined and analyzed three different online food community platforms, in order to unveil patterns in recipe creation across different food cultures.

3. DATASET

Our work relies on a dataset obtained from the German online food community website kochbar.de[2]. The basic statistics of the dataset are presented in the Table 1.

The dataset contains more than 400 thousand recipes from the years 2008-2014, and recipes are labeled with 246 categories such as 'Desserts' or 'Christmas'. Additionally, for each recipe, metadata covering information about ingredients, publication time and title is provided. In the initial dataset, ingredients were lists of arbitrary strings provided as free-form text by the users, and several standard pre-processing steps such as filtering, names translation and unification were performed in order to clean the data.

Furthermore, based on the recipe titles, we mined the types of recipes that we denote further in the paper as "food types". We identified those as titles appearing at least 5 times, i.e., if some title, for example 'apple pie', appears for 5 or more recipes we assume that it represents a common food type. Then, using substring inclusion, we matched all recipe titles to extracted types. For example, 'Magic *Apple Pie* by Mrs Schultz' was identified as a special variant of '*apple pie*'.

[1]http://allrecipes.com
[2]http://kochbar.de

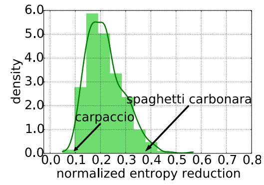


Figure 2: Distribution of reduction in ingredients entropy when food type is known.

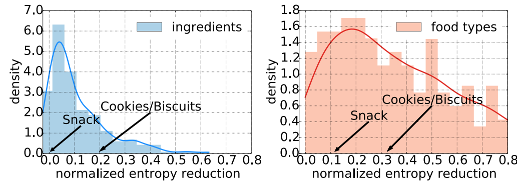


Figure 3: Distributions of reduction in entropies of ingredients and food types when category is known.

The second important entity in our dataset, next to a recipe, is the user. The almost 200k users established 195k friendship relations. 18 thousand users were active publishers that uploaded at least one recipe and around 19 thousand were actively rating recipes, providing in total 7 million ratings. Most of the ratings (99.1%) are 5-star ratings, so in our analysis we ignore the value and just consider the fact of the rating itself.

4. UNDERSTANDING ONLINE FOOD PRODUCTION PATTERNS

In this section, we study the extent to which particular factors have an impact on the online food production process (RQ1).

4.1 Food Types vs. Ingredients

In our study we are analyzing both food types and ingredients, and the initial expectation was that some ingredients should be more typical for particular food types than for the others. However, the quantitative evaluation shows this is not the case, i.e., in most cases we are not able to say to what extent there is a correlation between a particular pair of food type and ingredient.

Figure 2 provides a deeper insight into quantitative dependencies between food types and ingredients. As a measure for evaluating the discriminative power of food type we use the normalized entropy reduction (horizontal axis) $\frac{H(X)-H_{type}(X)}{H(X)}$, where $H(X)$ is the entropy of ingredients measured over all recipes and $H_{type}(X)$ is the entropy measured only over recipes of the particular type. Two illustrative examples are 'spaghetti carbonara' that determines well both set of ingredients and their frequencies distribution (entropy is reduced by almost 40%.), while on the other hand, 'carpaccio' is a very general type and not helpful for ingredients prediction (entropy is reduced only by 10%). The average reduction over all recipes is 22%.

4.2 Categories vs. Types and Ingredients

Users tend to have their own sets of favorite categories, and in order to validate the extent to which the preference towards some categories can be useful in a prediction task, we measured how well categories determine ingredients and food types. The result-

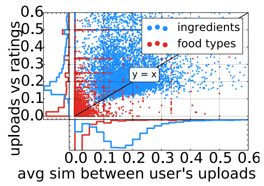


Figure 4: Similarities (measured over ingredients and food types) between recipes uploaded and rated by users.

ing plots are shown in Figure 3, where the horizontal axis represents normalized entropy reduction, in this case: $\frac{H(X)-H_{category}(X)}{H(X)}$, where $H(X)$ is the entropy measured over all recipes and $H_{category}(X)$ is the entropy measured only over recipes from the particular category. The value is highest for the most discriminative categories, indicating that they are also the best for predicting other items.

Average reduction over all recipes for ingredients entropy (first plot) is equal to 11%. For food types (second plot) entropy distribution is more flat and skewed towards high values, e.g., the mean is 40%. However some categories are more discriminative than the others, for example 'Cookies/Biscuits' vs. 'Snack'.

4.3 Production vs. Consumption

Before investigating online food production patterns in more detail, we wanted to understand better differences between online food consumption (recipe rating behavior) and production (recipe creation behavior).

Figure 4 provides insights into the relations between recipes consumed (rated) and produced (uploaded) by users. The horizontal axis measures the average cosine similarity between recipes published by the user u: $x(u) = avg_{r,r' \in U_u} sim(r,r')$. The vertical axis measures the average cosine similarity between recipes published and rated by the user u: $y(u) = avg_{r \in U_u, r' \in R_u}\ sim(r,r')$. U_u is the set of uploads by u and R_u the set of ratings by u.

We observe that users upload and rate in different ways, i.e., similarities along horizontal and vertical axes have different values and distributions. Users under the equal similarity line $y = x$ (with higher similarities within uploaded recipes than between uploaded and rated recipes) follow their own publishing style. Users above the line (with higher similarity between uploads and ratings) have publishing styles strongly influenced by information acquired from ratings.

When similarity is measured over common ingredients, for approximately 32% of the users their uploads are significantly less similar to other uploads than to ratings. For them ratings seem to be a better predictor than historical uploads. Only for 11% the opposite is true. For the rest 57%, the observed differences between similarities were not found statistically significant, i.e., t-test with $\alpha = 0.001$ was not able to reject the hypothesis that $x(u) = y(u)$.

For food types we find similar patterns as for ingredients. In the first group, where uploads to ratings similarities dominate, we observed 48% of the users. For 51% we did not observe statistically significant (according to t-test) differences. Only for less than 1% of the users similarities between uploads dominate over uploads to ratings similarities.

4.4 Historical Factors

Historical information is in many cases a very useful source of information for predicting the future. Henceforth, we were interested in investigating how useful user's historical uploads can be in future content prediction.

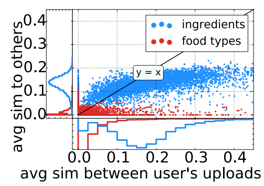


Figure 5: Similarities (measured over ingredients and food types) between user's own recipes and to other users' uploads.

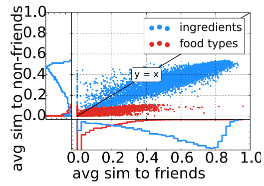


Figure 6: Similarities to user friends in comparison to non-friends.

Figure 5 presents a comparison of recipes from a single user to other users' recipes, and show mean cosine similarities measured over ingredients and food types. Each user is represented by single data point (one for ingredients and one for food types) and characterized by two values: mean similarity between recipes he uploaded (horizontal axis) and mean similarity to recipes uploaded by the other users (vertical axis).

The plots are biased towards the horizontal axis, i.e., towards high similarities between single user's recipes. However, our observations indicate that historical factors might be useful for the prediction of future user production only to some extent. For 41% of users (for ingredients) and 6% of users (for food types), the mean similarity between their uploads is higher than to uploads by others. For 42% (for ingredients) and 93% (for food types), we were not able to distinguish between single user's uploads and uploads by the other users (t-test was not able to reject means equality hypothesis). Finally, we found a group of users that behave in the opposite of expected way, i.e., they avoid repeating themselves and their uploads resemble more what others produce.

4.5 Social Factors

Social factors have been found to be a very useful source of information in many prediction tasks. In the context of online food communities, social connections are expressed by explicit friendship relations.

Figure 6 compares the mean cosine similarity of users to their friends (horizontal coordinate) and to other people that they are not connected to (vertical coordinate). The difference to previous plots where similarity is measured on recipe level should be noted. Here we focus on users' general preferences, i.e., all recipes from a user are merged together and similarities are averaged over users and not over recipes.

We also observe that from social connections such as friendship we can obtain useful information about user's preferences and biases. For example, approximately 95% of the users prefer the same items (items/food types) as their friends, while the opposite is true for only 5%.

4.6 Temporal Factors

People in their daily lives follow regular patterns that are related to time. For example, we behave differently on working days than during weekends. This periodic factors can also influence online

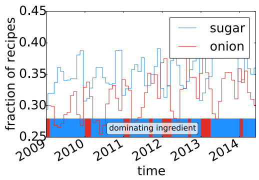


Figure 7: Uploading popularity plot of two sample ingredients. Typically, one of the items (blue one) is dominating, however, due to the shift in temporal patterns the other (red one) seasonally becomes more popular.

food production patterns.

In this study, we focus on the popularity of various ingredients and food types, and find that most of them follow seasonal and weekly trends. We noticed that some items are used in recipes more often in some periods than in the others. Our findings also confirm the hypothesis that these changes are significant enough to change relative popularities of different items. For example, Figure 7 presents a sample situation where seasonality determines ingredients usage (sugar vs. onion) and relative popularity.

5. PREDICTING ONLINE FOOD PRODUCTION PATTERNS

While the previous section focused on understanding the impact of various factors in the online food recipe creation process, the focus of this section is the predictability of this process (RQ2). The two problems we are trying to tackle now are the following: (1) *Given a target user u, what type of recipe r is he going to produce?* (2) *Given a target user u, which of the available ingredients i is he going to use?* The second problem we study in two settings: without recipe type given and in a context-aware setting when recipe type is already known (extracted from the recipe title entered by the user).

In addition to set of users $u \in \mathscr{U}$, the set of recipes $r \in \mathscr{R}$, the set of types of recipes $t \in \mathscr{T}$, ingredients $i \in \mathscr{I}$, and categories $c \in \mathscr{C}$ we consider influential factors, such as temporal context T (seasonal and weekly trends), friendship F. The task is then to propose a scoring function $S(u,e)$ (where $e \in \{t,i\}$) that assigns a preference score (predicts ranking) for candidate recipe type t or ingredient i for user u.

5.1 Evaluation Protocol

The evaluation protocol we follow in this paper is the one usually used in order to evaluate predictive models and recommender system offline in a time-based manner [4].

We split the dataset in training and test samples according to the time line, employing the leave-one-out strategy. Hence, the training set contains all the recipes published by user apart from the last published (this one is put into the test set). In our evaluation, we considered only users who have at least one recipe produced. Users who have uploaded exactly one recipe were considered as cold-start users (their only recipe was moved to the test set). In order to determine the quality of our predictors we used the *nDCG@k* measure ($k = 3$ for food types and $k = 10$ for ingredients used).

5.2 Predictors

Historical Predictors. The first proposed scoring function, based on findings from Section 4.4, depends on the popularity of the item e (either ingredient i or type t) in historically uploaded recipes and is defined as following:

$$\text{MPU}(u,e) = \sum_{r \in U_u} [e \in r]$$

where $[condition]$ takes 1 if *condition* is true and 0 otherwise and the expression $e \in r$ means that e (either ingredient, food type or category) is assigned to the recipe r. U_u is the set of recipes uploaded by the user u in the past.

Similarly, relying on findings from Section 4.3 where we showed that uploads often strongly correlate with ratings, we can define:

$$\text{MPU-R}(u,e) = \sum_{r \in R_u} [e \in r]$$

where R_u is the set of recipes rated by the user u in the past.

For some users the historical data may be very sparse. Hence, it might be better to smooth the scores by incorporating the information from recipes of other users that are somehow related, for example through common categories (Sections 4.2 and 4.4). The predictor that measures the popularity of the item e in categories used by the user u is defined as following:

$$\text{C}(u,e) = \sum_{c \in \mathscr{C}} w(u,c) \cdot \left(\sum_{r \in \mathscr{R}} [c \in r \wedge e \in r] \right)$$

where $w(u,c) = \sum_{r \in U_u} [c \in r]$ measures the popularity of category c in user u's recipes. The second part weights the popularity of the item e in recipes from the category.

Food types and ingredients are strongly correlated (Section 4.1). When the set $T \subset \mathscr{T}$ (typically of size of only one or two) of types assigned to the recipe is known (extracted from the recipe title already typed by the user) the above scoring function can be adjusted in the following way:

$$\text{C}[T](u,i) = \sum_{t \in T, c \in \mathscr{C}} w(t) w(u,c) \cdot \left(\sum_{r \in \mathscr{R}} [c \in r \wedge i \in r \wedge t \in r] \right)$$

where $w(t)$ if the relative weight of a type t. We define $w(t) = \frac{1}{\sum_{r \in \mathscr{R}} [t \in r]}$ to give a higher importance to less popular and more specific types.

Similar to the case of uploads, we define a set of predictors that reflect the popularity in recipes rated by the user u: respectively C-R and C-R$[T]$ where U_u is replaced with R_u in the formulas.

Social Predictors. In Section 4.5 we have shown that social factors have a strong impact on food production. Hence, we propose to exploit friendship relations in the following scoring function:

$$\text{F}(u,e) = \sum_{f \in F_u} \sum_{r \in U_f} [e \in r]$$

where F_u is the set of direct friends of the user u. Assuming context (set of types T) to be known above scoring function is adjusted in the following way:

$$\text{F}[T](u,i) = \sum_{t \in T} \sum_{f \in F_u} \sum_{r \in U_f} w(t) \cdot [i \in r \wedge t \in r]$$

By replacing U_u with R_u in the above formulas we get the scoring functions over ratings instead of uploads, respectively F-R and F-R$[T]$.

Temporal Predictors. Temporal impact on the online food production process was observed (Section 4.6) on both seasonal and weekly (to a lesser extent) level implying two variants of the time-dependent scoring function:

$$\text{T}[time](u,e) = \sum_{u' \in \mathscr{U}} \sum_{r' \in U_{u'}} [tm(r') = tm(u) \wedge e \in r']$$

where *time* can be either a month or a week day and $tm(.)$ is a function that returns respectively month or week day of the web site access or recipe upload.

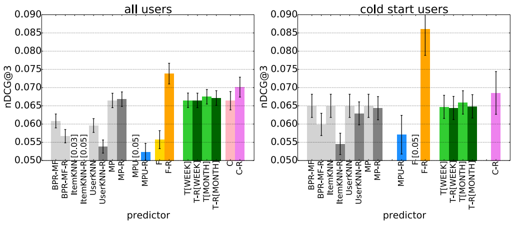


Figure 8: Food types prediction quality (means and standard errors) for cold start and all users. Colors are used to mark different groups of methods.

Similarly, we can also consider temporal factors for ratings T-R[*time*] and with the context known T[$time, T$](u,i).

5.3 Baseline Methods

For the first baseline we used a non-personalized scoring function, the so-called *most popular* approach:

$$\mathrm{MP}(u,e) = \sum_{r \in \mathscr{R}} [e \in r]$$

Additionally, we introduced *the most popular based on ratings* score:

$$\text{MP-R}(u,e) = \sum_{u \in \mathscr{U}} \sum_{r \in R_u} [e \in r]$$

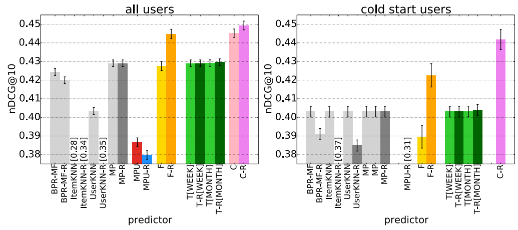
Similar to above, we also define context-aware baselines MP[T] and MP-R[T] which measure popularity when family type is known.

Apart from these naive methods, we also compare to state-of-the-art methods from the literature, namely BPR-MF, ItemKNN, and UserKNN (only applicable in context-blind case) in two variants. The first variant relies on uploads and the second (suffix R, e.g, BPR-MF-R) on ratings. We used the popular implementations from *MyMediaLite* library with the default settings [8].

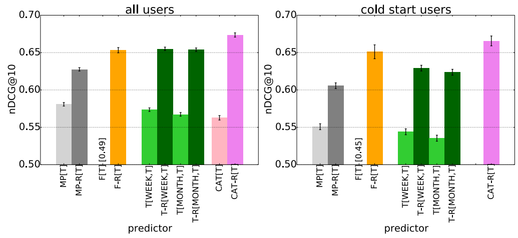
5.4 Results and Discussion

Food Type Prediction. Figure 8 summarizes the results of our food types prediction experiment, considering all users and only cold-start users. As shown, the prediction of food types is a hard task and henceforth the obtained values on both plots are low, and baselines achieve relatively high scores and few approaches overpower them. One predictor, namely F-R (popularity in ratings by user friends), performs significantly better than the others. The strength of the method is especially notable on the second plot (for cold start users). The second best results are obtained via C-R (popularity in often rated categories). Time-based functions perform no better than baselines. The worst among our approaches are MPU and MPU-R that rely entirely on user's previous uploads and ratings, suggesting that one of the key factors influencing results quality is the data sparsity, e.g., just by regularizing with categorical information we can improve from one of the worst approaches (MPU-R) to almost the best one (C-R).

Ingredient Prediction. Figure 9 summarizes the results of the ingredients prediction experiment in two cases: (a) recipe type is unknown and in (b) recipe type is given. What is notable is that when context (recipe type) is included, the prediction is much easier, resulting in higher values ($nDCG@10 \sim 0.7$). Furthermore, we observe that rating-based predictors perform remarkably better than those based on uploads. Similarly, as previously observed, we note that the best prediction quality is obtained for scoring functions based on social factors such as friendship (F-R and F-R[T])



a) Prediction of ingredients when food type is unknown



b) Prediction of ingredients when food type is known

Figure 9: Ingredients prediction quality (means and standard errors) for cold start and all users.

and for the methods that exploit categories to overcome data deficiency (C-R and C-R[T]). On the other hand, we hypothesize that due to the lack of sufficient evidence, recommenders such as MPU and MPU-R, as well as popular methods such as UserKNN and ItemKNN, perform poorly. Time-based methods perform better than these baselines but their prediction quality depends on food type information availability.

6. CONCLUSIONS

In this paper, we extended our understanding on the factors that have an impact on the users in their content creation process in online food communities. We approached this by conducting several analytical and predictive experiments on a large-scale dataset obtained from one of the largest online food community platforms available on the Web, namely *kochbar.de*. As our empirical analysis reveals, factors such as user history, social relations and temporality have indeed a significant impact on how recipes are created and why certain types of recipes and ingredients are being used by the user. In particular, our findings were as follows:

- Information consumption often correlates with future information production.
- Food types, ingredients and categories show strong correlations.
- Only some fraction of users follow their own styles and have strong preferences towards ingredients, food types and categories.
- There are strong correlations between social factors such as friendship and content being produced.
- Ingredients and food types popularities are season-dependent influencing their relative prevalence.
- It is easier to predict ingredients than exact food type.
- Approaches relying on social factors and applying ratings and categorical information are the most useful to predict future production.

7. REFERENCES

[1] S. Abbar, Y. Mejova, and I. Weber. You tweet what you eat: Studying food consumption through twitter. In *Proc. of CHI'15*, 2015.

[2] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási. Flavor network and the principles of food pairing. *Scientific reports*, 1, 2011.

[3] S. Berkovsky and J. Freyne. Group-based recipe recommendations: Analysis of data aggregation strategies. In *Proc. of RecSys '10*, pages 111–118, 2010.

[4] P. G. Campos, F. Díez, and I. Cantador. Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. *UMUAI*, 24(1-2):67–119, 2014.

[5] M. De Choudhury and S. S. Sharma. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proc. of CSCW '16*, 2016.

[6] D. Elsweiler and M. Harvey. Towards automatic meal plan recommendations for balanced nutrition. In *Proc. of RecSys '15*, pages 313–316, 2015.

[7] J. Freyne and S. Berkovsky. Intelligent food planning: Personalized recipe recommendation. In *Proc. of IUI '10*, pages 321–324, 2010.

[8] Z. Gantner, S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. MyMediaLite: a free recommender system library. In *Proc. of RecSys '11*, pages 305–308, 2011.

[9] M. Ge, F. Ricci, and D. Massimo. Health-aware food recommender system. In *Proc. of RecSys '15*, pages 333–334, 2015.

[10] T. Kusmierczyk, C. Trattner, and K. Nørvåg. Temporal patterns in online food innovation. In *Proc. of WWW '15 Companion*, pages 1345–1350, 2015.

[11] T. Kusmierczyk, C. Trattner, and K. Nørvåg. Temporality in online food recipe consumption and production. In *Proc. of WWW'15*, 2015.

[12] M. Rokicki, E. Herder, and E. Demidova. What's on my plate: Towards recommending recipe variations for diabetes patients. In *Posters, Demos, Late-breaking Results and Workshop Proc. of UMAP'15*, 2015.

[13] A. Said and A. Bellogín. You are what you eat! tracking health through recipe interactions. In *Proc. of RSWeb'14*, 2014.

[14] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic. Recipe recommendation using ingredient networks. In *Proc. of WebSci'12*, 2012.

[15] C. Trattner, T. Kusmierczyk, and K. Nørvåg. FOODWEB - studying food consumption and production patterns on the web. *ERCIM News*, 2016(104), 2016.

[16] M. Trevisiol, L. Chiarandini, and R. A. Baeza-Yates. Buon appetito: recommending personalized menus. In *Proc. of HT'14*, pages 327–329, 2014.

[17] C. Wagner, P. Singer, and M. Strohmaier. The nature and evolution of online food preferences. *EPJ Data Science*, 3(1):1–22, 2014.

[18] R. West, R. W. White, and E. Horvitz. From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In *Proc. of WWW'13*, 2013.