EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

CrossMark

On the predictability of the popularity of online recipes

Christoph Trattner[1]* , Dominik Moesslang[2] and David Elsweiler[3]

*Correspondence:
christoph.trattner@uib.no
[1]Department of Information
Science and Media Studies,
University of Bergen, Bergen,
Norway
Full list of author information is
available at the end of the article

Abstract

Popularity prediction has been studied in diverse online contexts with demonstrable
practical, sociological and technical benefit. Here, we add to the popularity prediction
literature by studying the popularity of recipes on two large and well visited online
recipe portals (Allrecipes.com, USA and Kochbar.de, Germany). Our analyses show
differences between the platforms in terms of how the recipes are interacted with
and categorized, as well as in the content of the food and its nutritional properties.
For both datasets, we were able to show correlations between recipe features and
proxies for popularity, which allow popularity of dishes to be predicted with some
accuracy. The trends were more prominent in the Kochbar.de dataset, which was
mirrored in the results of the prediction task experiments.

Keywords: Online recipes; Food; Popularity

1 Introduction

The traces users leave behind when interacting with items online, combined with properties of the items themselves can be used to predict how popular individual items will become with users of a service. This concept—known as *popularity prediction*—has been studied in diverse contexts including with social media content [1], online news articles [2], and posted videos [3, 4]. Successfully predicting which items will gain popularity is useful because placing popular content on entry pages can drive users to the site and maintain engagement [5], can influence the content systems recommend to individual users [6] and allow interests and cultural trends to be monitored over time [7]. If the popularity prediction task is formulated such that future popularity is estimated using only data available at upload time, there can be different advantages. For example, we are investigating one means of addressing the cold-start recommendation problem [8]. Moreover, based on the results, users can be advised to modify their content or its description in some way to improve its reception and visibility in the community. Successfully predicting popularity can offer additional, technical benefits, such as improved efficiency in how caching resources are allocated [9, 10].

In this work we use data acquired from two different cooking platforms associated with different geographical regions to study the popularity of online recipes—an item yet to receive detailed attention in the popularity prediction literature. A further benefit of studying popularity in this context is that because there is a strong relationship between the online recipes people view and bookmark and the food actually consumed [11–13], achieving

an understanding of and the ability to predict which recipes become popular can offer a lens through which eating habits can be studied at a societal level.

Here we examine and compare two popular food portals from divergent food cultures. The first, Allrecipes.com, is primarily used by people living in the United States and, at the time of writing, claims to be the world's largest food-focused social network [14]. The second, Kochbar.de, is a popular service in Germany offering similar functionality to the American site. It has been reported that Kochbar.de is the second largest food community platform in Europe.[a] By studying popularity in the context of two platforms with separate communities, we are able to understand how robust the predictive features are for popularity prediction in diverse geographical and socioeconomic contexts.

Our research is driven by following research questions:

- *RQ1.* To what extent can popularity patterns be identified in the online food communities Allrecipes.com and Kochbar.de?
- *RQ2.* To what extent do the two online communities (Allrecipes.com and Kochbar.de) differ from or are similar to each other, with respect to features derived from the food and nutritional psychology literature?
- *RQ3.* To what extent can potential correlations be found that might be useful in a prediction task?
- *RQ4.* To what extent is the popularity of online recipes predictable and which are the most useful predictive features for this prediction task?

In the following sections we review appropriate background literature, which serves to motivate the investigation of the above research questions. We continue to introduce the datasets used and methodology chosen to achieve our aims, as well as present and discuss our findings. In the final sections, we discuss what our results mean when set against the context of the limitations of our study and finally, we propose future research directions.

2 Background

In this section we briefly review three bodies of related work. First, we summarize the popularity prediction problem, showing how it has been tackled and in which contexts. We then turn our focus to the unit of study in this work by summarizing research from the fields of information retrieval and recommender systems relating to online recipes. Finally, we review research from psychology and nutritional anthropology, which provides insight into the factors influencing human food choice. This work informs the feature engineering decisions when deriving predictive models.

2.1 Research on popularity prediction on the web

Popularity prediction as a scientific problem is, to some extent, a response to the age-old problem of information saturation [15]. Since its development, the World Wide Web has only served to intensify Herbert Simon's dictum that an overload of information leads to a poverty of attention [16]. Deciding which sources to attend to is a daily struggle for users [17, 18] and is open to numerous biases including subconscious human, as well as algorithmic biases [19].

Patterns in web page accesses are well known to be highly skewed and Zipfian distributed [20, 21]. Online videos, for example, which make up a significant proportion of Internet traffic, have been studied extensively [22–24], revealing that the interest generated by items shared on the web is ephemeral and complex, and is influenced by numerous, changing factors, all of which combine to make prediction difficult [22, 25].

Summarizing the literature reveals that typically moderate correlation is found between popularity metrics derived from various user interaction traces, such as likes, views, shares etc. [7, 26, 27]. This is most likely explained by the fact that metrics relate to different aspects, ranging from attractiveness to utility to quality, and are often associated with different user aims or actions (e.g., consumption, sharing, the provision of feedback, indirect communication etc. [28]).

Although popular items tend to be highly accepted as user recommendations, e.g., [29], it should be clear that popularity does not equate to quality. Past research has demonstrated that popularity is a complex, multifaceted concept, which is well known to be difficult to predict [30]. Nevertheless, by studying how diverse metrics relate to each other, e.g. [31, 32], researchers can achieve a broader and better understanding of what the popularity content actually means, as well as what can be achieved with accurate predictive models for popularity.

Typical foci of investigation for researchers studying popularity prediction are diverse information items shared on the web. These include movies (Simonoff and Sparrow 2000) or songs (Pachet and Sony 2012), online news articles [2], as well as various social network content (Facebook: [33], Twitter [34–36], Weibo [37] and YouTube [3, 4]).

The problem of predicting popularity has been formulated in various ways. Shulman [30] distinguishes between efforts to predict apriori based on item content and meta-data and modified versions of the problem where researchers are allowed to peek into early adoption activity for an item to inform the prediction. Peeking formulations have been proven to lead to predictions with greater accuracy, e.g., [33–35]. A second dimension by which formulations of the problem can be distinguished relates to the aspect being predicted. In regression formulations the exact popularity of an item measured on some scale is predicted, whereas in classification formulations popularity is discretized into distinct classes, again based on some specified criteria [38]. The literature suggests that the latter makes the problem more tractable [30].

The literature summarized thus far emphasizes that popularity prediction has been studied in various contexts, using divergent methods, with heterogeneous motivations. One context, however, which has yet to be studied in detail are popularity metrics relating to online recipes. We believe this would offer utility for a number of reasons relating to understanding food culture, eating habits and links to epidemiology research, as well as the design of technology to assist people nourish themselves more healthily. We explain these points in more detail in the following section by summarizing relevant related work.

2.2 Research on online recipes

The way people interact with recipes online can provide clues regarding user food preferences and eating habits. Kusmierczyk et al. and Trattner et al. analyzed data from the German community platform Kochbar.de and found clear seasonal and weekly trends in online food recipe production, both in terms of nutritional value (fat, proteins, carbohydrates, and calories) [39, 40] and in terms of ingredient combinations and experimentation [41]. Similar patterns were observed by Wagner et al. [42] and West et al. [43]. West and colleagues also found correlations between recipes accessed via search engines and incidence of diet-related illness, which resemble findings reported recently by Said & Bellogin [44], De Coudhury et al. [45] and Abbar et al. [12] in the context of Allrecipes.com, Instagram and Twitter respectively. Rokicki et al. [46] investigated differences in nutritional

values between user recipes created by different user groups, finding, for example, that recipes from females are, on average, richer in carbohydrates. The carbohydrate content of recipes seems to decrease with increasing user age, which mirrors the advice given by most nutrition advice centers. Thus, the literature provides strong evidence for the utility of using interactions with online food items to understand eating habits and for epidemiological purposes.

These published analyses are all very related to work in the recommender systems literature, which attempts to model user food preferences to present individual users with meal recommendations that they would like to eat [47, 48] or to assist users make healthier food choices [49]. A recent review of research in this area can be found in [50]. The literature emphasizes that food choice is complex and context sensitive, with user, situational, temporal, and social factors playing a role [51].

The research reviewed in this section is relatively new and data-driven. The research has, moreover, been largely performed without being related to an extensive body of work from psychology, dating from the 1950s, which has investigated the factors influencing the food choices people make. In the following section we provide a brief review of important work in this domain, which later informs the feature engineering procedures in our work.

2.3 The psychology of food choices

People typically make around 200 food choices every day [52]. Choosing which food to eat is a complex process influenced by a number of context factors at biological, personal, situational, social and socio-economic levels [53]. Among the most commonly investigated factors in the food literature are taste or sensory appeal, health-related issues, ethical concerns, convenience, price, and weight control considerations [54]. Food choices reflect mood, with people having been shown to eat to receive emotional comfort or improved mood, evoke past experiences or experience something new when choosing food [54–56].

Individual differences can be found with respect to the importance placed on these diverse attributes and this can depend on other factors including age, gender, race, lifestyle, socioeconomic status, cultural background, and education [57, 58]. However, the evidence suggests that for most people, the driving factors seem to be the taste of food and how it appeals to one's other senses, followed by concerns about health, weight control, nutritional value, and cost [59, 60] and the recently published studies summarized above with online recipes only seem to confirm this [48, 49]. That being said, the food choices people make can be biased in countless ways. For example, people make poor decisions when stimulated (e.g., when hungry and surrounded by the sights and smells of calorie rich food) [52] or when emotional [61] or stressed [62]. People, moreover, adapt their behaviour to the social context with obese individuals more likely to be friends with other obese individuals [63] and people consume more when they eat in groups, rather than alone [52].

Thus, popularity prediction in the domain of online recipe portals is related to food recommender systems, which is, in turn, related to studies in psychology to understand the processes of food choice. All three bodies of research point to the prediction of the popularity of recipes being a challenging task. That being said, by explaining some of the factors influencing food choices, the literature also offers hints as to features, which may be helpful for predicting popularity. We use these hints to drive our analyses, as we experiment with popularity prediction in the context of online recipes.

2.4 Building on prior research

The research presented in this article extends previous work in at least three distinct ways. Firstly, we are the first, to our knowledge, to examine the popularity prediction problem in the context of online recipes. We argue that there is a number of potential benefits to this, including the possibility of improved recommendations, user advice on how to improve their content, and an instrument for learning about societal eating habits. Secondly, although both datasets presented have been studied to some extent in the past, neither have been studied with respect to popularity prediction, nor does any publication exist, which compares the two datasets directly. This is advantageous because it offers insight into how two divergent food cultures compare and contrast. Thirdly, we collate and evaluate an extensive set of features from diverse related work including from papers in the popularity prediction literature and from papers in the online recipe literature. We also derive new features from the food psychology literature, which we believed may offer utility for the task at hand. Although many of the features have been investigated to some extent in the past, this article presents an evaluation of a comprehensive feature set, which goes beyond any previous investigation.

3 Materials

In the following two sub-sections we describe how the data we use to address our research aims were obtained and present some descriptive statistics about the data acquired and analyzed. In addition to this we provide insight into how the two platforms function and supplement our descriptions with screenshots taken of recipe presentation and how users are able to upload recipes in both platforms.

3.1 Allrecipes.com

The first dataset used and that shall be described is Allrecipes.com. According to Ebizmba.com[b] it is the most popular online food platform on the Web and has global website rank[c] of 885 and 246 inside the USA (July 2017). The dataset was obtained between 20th and 24th of July 2015 by implementing a standard Web crawler [64]. It contains 60,983 recipes published between the years 2000 and 2015 on the Allrecipes.com website. Of these recipes, 58,263 contain nutrition information. The recipes were published by 25,037 users during this time period. The recipes and associated user profiles were obtained via the sitemap available in the robots.txt file.

In the following paragraphs, we describe the 'recipe profile view' and the 'upload view' which are available to the user when clicking on a recipe or when uploading a new recipe to Allrecipes.com. In addition, we provide information about the 'user profile view' as we also extract features from this view based on the available entities (meta-data), see Table 1.

Recipe Profile View. As presented in Fig. 1 and sub-figures (a)–(c), the recipe profile view presents the user with all relevant features about the recipe and allows other users to rate and comment on the recipe. Besides typical entities such as average rating, number of ratings, image, recipe ingredients and cooking directions, this view also features the nutritional information of the meal. This information is only available for public and reviewed recipes and not for personal recipes uploaded to the user's private space. To estimate the nutritional information of the recipes uploaded by the community, Allrecipes.com relies on the ESHA Research's nutrient database,[d] which allows Allrecipes.com to provide information for 18 different nutrients including energy (kCal), protein (g), carbohydrate (g),

Table 1 Entities (meta-data) available in Allrecipes.com and Kochbar.de recipe and user profile views as employed for the study. As shown not all types of entities are available for both sites as indicated with a × symbol

Entity	Allrecipes.com	Kochbar.de
Recipe Profile View		
Recipe Name (ID)	×	×
Cook Name (ID)	×	×
Recipe Title	×	×
Recipe Image	×	×
Average Recipe Rating	×	×
Ingredient List	×	×
Preparation Time	×	×
Instruction Text	×	×
Nutrition Table	×	×
Recipe Rating List	×	×
Recipe Comment List	×	×
Recipe Categories[†]	×	×
Users Profile View		
User Name (ID)	×	×
First/Last Name		×
Uploaded Recipes	×	×
Cooking Interests	×	
Biography		×
Gender		×
Date of Birth		×
Location of Residence (City, Country)	×	×
Bookmarked Recipes	×	
Friends (User Names)		×
Followers (User Names)	×	
Member Since (Date)	×	

Note: [†] In Allrecipes.com and Kochbar.de the recipe view only provides a category bread crumb. To obtain all possible category labels for a recipe in Allrecipes.com and Kochbar.de we needed to crawl the 'Category view', which provided a navigable category tree and links to recipes.

sugar (g), sodium (g), fat (g) saturated fat (g), and fiber (g). User comments and ratings are displayed at the bottom of the view. Additionally, there is the "rate and review" button, when pressed it shows a window with the rating feature (based on 5 stars) and a text area for the comment. Important to note is that these screenshots were taken in August of 2017. However, the dataset was crawled in 2015 when providing a rating was mandatory when commenting on a recipe. This is no longer the case.

User Profile View. We do not provide screenshots for this view as although publicly available, these contain user personal data. The user view in Allrecipes.com comprises information about a user's uploaded, favoured recipes, comments and rated recipes. Until 2015 this view also contained information about where the cook is currently living, member since information, cooking interests, hobbies and some additional text describing the user. We made use of the users' "current home location" information, which is provided on city and county level, in our analyses.

Recipe Upload View. Sub-plot (d) in Fig. 1 presents the upload view. In order to upload a recipe to Allrecipes.com, one has to provide multiple recipe parameters. These include the recipe title, a description, the preparation steps, a list of ingredients, preparation and cooking time and the number of servings. Also, the privacy level must be defined. When submitted as a public recipe and "Kitchen Approved", the editorial staff will review the recipe to ensure quality and calculate the nutritional information based on the ingredients. A picture of the recipe is optional. Furthermore, it is important to note that the ingredients
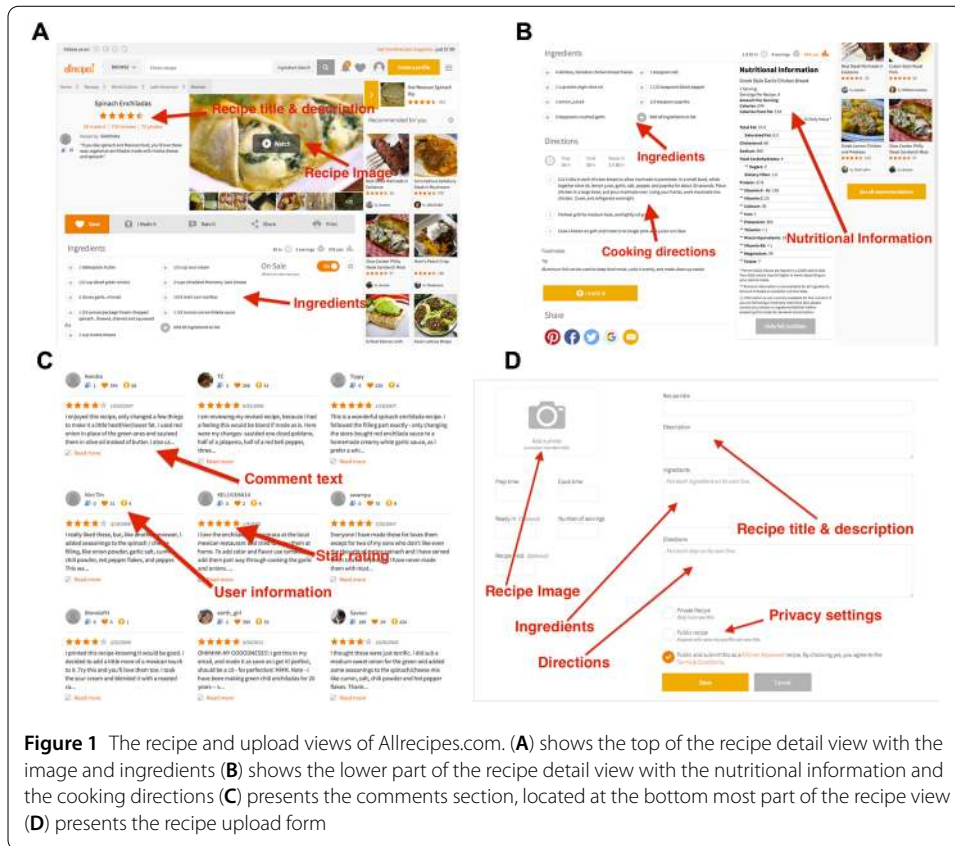
Figure 1 The recipe and upload views of Allrecipes.com. (**A**) shows the top of the recipe detail view with the image and ingredients (**B**) shows the lower part of the recipe detail view with the nutritional information and the cooking directions (**C**) presents the comments section, located at the bottom most part of the recipe view (**D**) presents the recipe upload form

Table 2 Basic statistics of the Allrecipes.com and Kochbar.de dataset

Feature	Allrecipes.com	Kochbar.de
Number of published recipes	60,983	405,864
Recipes containing nutrition information	58,263	309,360
Users with published recipes	25,037	18,212
Recipes rated/commented	46,713[†]	400,155/360,668
Ratings/comments provided to recipes	1,032,226[†]	7,796,004/2,751,820
Users who provided ratings/comments	125,762[†]	19,444/21,951
Distinct Ingredients	3842	2028
Distinct recipe categories	939	246

Note: [†] Number of ratings and comments are the same per recipe in Allrecipes.com as one can only rate and comment at the same time.

are captured with free form text with no spellcheck or guiding feature. This is the reason for the many misspellings and word variants found in the data set (see Sect. 4.1).

A very important feature of Allrecipes.com is that the platform also provides categorical information which was extracted from the platform's category tree which is as available through the platform's 'category view'.[e]

Table 2 gives an overview of the basic statistics of the Allrecipes.com dataset and Table 1 provides an overview of the available entities used later for feature extraction.

3.2 Kochbar.de

The second dataset we employ in this work is from the online food website Kochbar.de. It was crawled by [41] in 2014. According to [65] Kochbar.de attracts more than 6.6 million

Figure 2 The recipe view of Kochbar.de. Plot (**A**) highlights the top part of the view, showing the main parameters of the recipe and the rating feature (**B**) presents the bottom part of the recipe detail view with the nutritional information on the lower left side (**C**) shows the comments section located at the bottom part of the detail view

unique users every month, is ranked among the top-50 most popular websites in Germany and is the second most popular food website in Europe.

Kochbar.de hosts about half a million recipes by users from all over the world. In 2014 Kusmierczyk et al. [41] crawled over 400,000 recipes uploaded between the years 2008 and 2014. These recipes were uploaded by nearly 200,000 unique users. Kochbar.de is very similar to Allrecipes.com in terms of how the platform works as well as the functionalities provided to users.

As in Allrecipes.com, Kochbar.de offers two main views to users when inspecting a recipe or when uploading a recipe to the platform. These are described below along with the user profile view.

Recipe Profile View. As presented in Fig. 2 and sub-figures (A)–(C), the recipe view shows almost the same information as the corresponding view in Allrecipes.com. It displays the main recipe information such as title, description, preparation steps/time, and ingredients.

Only 4 nutrients are presented to the user, which are estimated using the German Nutrient Data Base.[f] In particular, calories, protein, carbohydrates and fat content measured in 100 g per meal are presented. The rating and commenting functionalities are slightly different to those offered in Allrecipes.com. As mentioned above, in Allrecipes.com, users

could only rate and comment on recipes together at least until 2015. In Kochbar.de these two features are separated.

User Profile View. Again we do not provide screenshots for this view as it contains personal data. As with Allrecipes.com, the user view in Kochbar.de information about the user's uploaded recipes and gives the user the opportunity to provide information about the town/country the user is currently living in, and some specific text fields related to cooking interests. In the same way as with the Allrecipes.com data, we made use of the users' "current home location" information in our analyses.

Recipe Upload View. Fig. 2 and sub-plots (A)–(B) presents the upload view, which is more complex than the one provided by Allrecipes.com. Besides the standard recipe parameters such as title, description and preparation time, difficulty level or price level, it also features an elaborate ingredients definition widget. The ingredients widget supports the user with ingredient suggestions based on the input provided. Along with the ingredient name, it has two separate input fields for the amount and the unit. Presumably, this improves the automatic ingredient parsing, on which the nutrient information is based on. Nevertheless, there are still many misspellings and word variants in this data set (see Sect. 4.1). This could either be explained with incorrect usage of the ingredient widget, misspelled ingredients that are already in the Kochbar.de database, or that the widget has not been always been part of the upload form. The cooking instructions are also captured in a more structured manner. The form provides separate text fields for every step. Photographs of the meal can also be uploaded. In contrast to Allrecipes.com, the Kochbar.de upload form allows the user to manually overwrite the automatically calculated nutritional information. It is also possible to assign descriptive tags and categories to the recipe, which is not possible in Allrecipes.com.

Similar to Allrecipes.com, Kochbar.de contains a dedicated sub-page with a category tree[g] relating recipes to categories from which we obtained the category labels of a recipe.

Table 2 gives an overview of the basic statistics of the Kochbar.de dataset and Table 1 provides and overview of the available entities used later for feature extraction.

4 Methodology

In this section we describe the methodology applied in 4 distinct sub-sections: Sect. 4.1 explains how we isolate and pre-process the recipes from the collection; Sect. 4.2 explains how we measure popularity; Sect. 4.3 outlines the features investigated and, finally; Sects. 4.4 and 4.5 present the statistical approaches used to compare and model predictions, respectively.

4.1 Data selection and pre-processing

For the purpose of our analyses we made use of recipes that were (1) main meals, (2) provided nutrition information and (3) had at least one image of the prepared meal available. In Allrecipes.com 11,194 recipes fulfill these criteria while for Kochbar.de the number was 81,232. We chose to focus mostly on main dishes because, as shown in Fig. 4, the frequency for which recipes in different categories are uploaded varies strongly across the platforms. Main dishes is a popular category common to both platforms, and thus provides a fair basis from which to compare the platforms. We do provide additional analyses using all of the recipes in the datasets, including the prediction experiments and we emphasize this to the reader as appropriate.
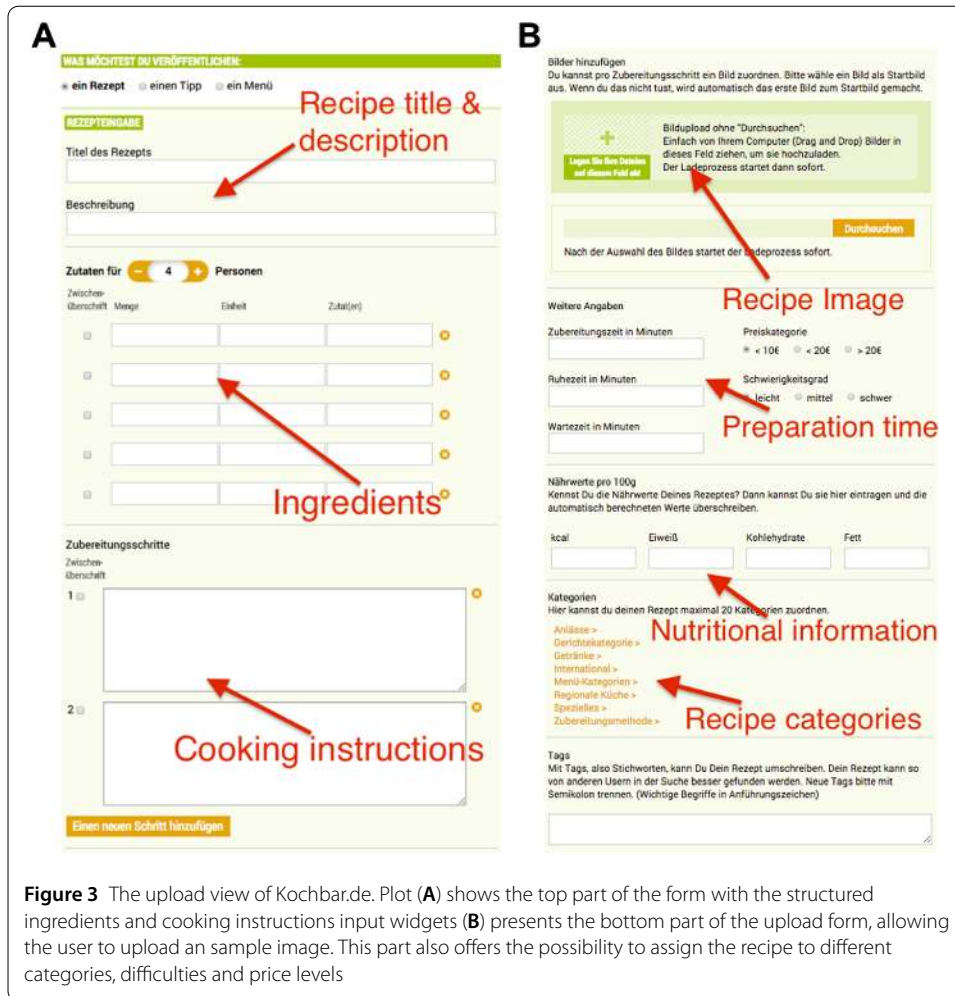
Figure 3 The upload view of Kochbar.de. Plot (**A**) shows the top part of the form with the structured ingredients and cooking instructions input widgets (**B**) presents the bottom part of the upload form, allowing the user to upload an sample image. This part also offers the possibility to assign the recipe to different categories, difficulties and price levels
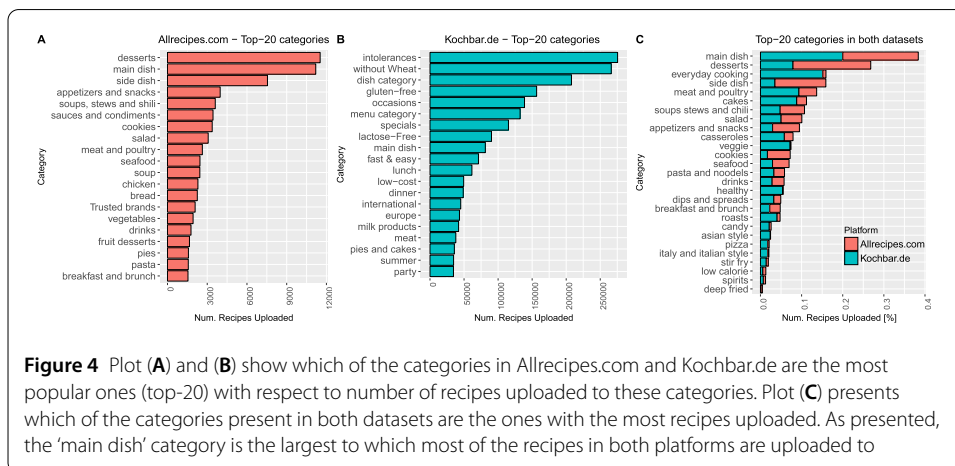
Figure 4 Plot (**A**) and (**B**) show which of the categories in Allrecipes.com and Kochbar.de are the most popular ones (top-20) with respect to number of recipes uploaded to these categories. Plot (**C**) presents which of the categories present in both datasets are the ones with the most recipes uploaded. As presented, the 'main dish' category is the largest to which most of the recipes in both platforms are uploaded to

Since both datasets had been used for other purposes in the context of previous studies [41, 64], the data had already been structured and partially cleaned. That being said, the ingredient lists for both datasets were noisy due to misspellings and natural vocabulary variation that occur when recipes are uploaded by users. To tackle these issues, we

made use of the Spoonacular.com Web API,[h] which offers services for analyzing food and recipes.[i] The API is free to use for academic purposes and allows the extraction of normalized ingredients from free-form text. As Spoonacular.com only supports English text, the Google translate API was employed to translate German to English text. Text that could not be recognized by the Spoonacular API (such as misspellings) was converted manually.

Although this pre-processing pipeline is not perfect, manual inspection of the results demonstrated to us that in the vast majority of cases, the output produced was accurate. Taking a sample of 1000 matched ingredients revealed only five results we considered inappropriate, e.g., "tomatoes dried in oil" was matched as "dried tomatoes" and "flour to roll out" was matched to "roll". The other two mismatches would make very little nutritional difference, e.g., "China spice" was matched as "BBQ Spice". Many of the results were impressive. For example, even though the German phrase for free-range eggs "Freilandeier" was translated as "eggs Freiland", this still resulted in a usable match of "eggs".

By taking this approach we were able to reduce the total number of unique ingredients from 723,911 to 3842 for Allrecipes.com and from 302,126 to 2028 for Kochbar.de.

4.2 Analyzing popularity

As indicators for popularity we chose the *number of comments* and the *number of ratings* for each recipe as these were available for both datasets. It made little sense to analyze ratings applied or the sentiment of comments on recipes as neither varies to a great extent across recipes. Over 99% of the recipes in the Kochbar.de dataset were given a 5-star rating and while in Allrecipes.com the trend is not so pronounced, most of the recipes with ratings were rated as 4 or 5.

We also plot average number of comments/ratings over time, both cumulatively and non-cumulatively. Further insight is provided by calculating mean and median popularity for each time slot, that is for 1 day, 7 days (one week), 30 days (one month) and 365 days (one year).

4.3 Feature engineering for recipe popularity prediction

To investigate which factors help explain popularity, we derived a set of features relating to aspects highlighted in the literature summarized in Sects. 2.2 and 2.3.

The features relate to a recipe's content, presentation, nutrition and healthiness, complexity, seasonality and innovation (7 feature sets in total). We also derive features capturing aspects relating to the authoring user and the context surrounding the interaction. Although many of these features have been used in previous investigations in various diverse contexts, it is the first time they have been examined together, in exhaustive detail. It is also the first time that these features have been tested in the context of predicting popularity.

We briefly describe the main groupings below and explain how these were calculated. The numbers in the brackets denote the number of features in each feature set.

Recipe Nutrition (4). We derived 4 features representing the nutritional properties of a recipe, which past work has been shown to influence how people perceive online recipes [51]. The features are:

- *Kcal (per 100 g)*, measuring the amount of energy in Kcal per 100 g in a recipe.
- *Protein (per 100 g)*, capturing the amount of protein measured in grams per 100 g in a recipe.

- *Carbohydrates (per 100 g)*, capturing the amount of carbohydrates measured in grams per 100 g in a recipe.
- *Fat (per 100 g)*, capturing the amount of fats measured in grams per 100 g in a recipe.

Recipe Healthiness (1). To establish the healthiness of a recipe—often cited as a motivator for food choice [59, 60]—we derive the WHO health score, reported in [64, 66], which defines healthy ranges for the daily intake of macro- and micro-nutritional components as recommended by the World Health Organization (WHO). The nutrients used to derive the score are proteins, fat and carbohydrates. As in previous studies using the metric, the score ranges from 0 if none of the ranges were met, to 3 if all were met. The feature is labeled as *WHO health score* throughout this paper.

Recipe Complexity (5). Complexity are concepts associated with food preferences in the food recommender systems literature [51].

In particular, we derived 5 features describing various aspects relating to recipe complexity:

- *Preparation Time (Min.)* captures the amount of time needed in minutes to prepare a meal based on a given recipe.
- *Num. Preparation Steps* captures the number of steps to prepare a meal from a recipe.
- *Num. Servings* captures the number of meals servings a recipe provides.
- *Num. Ingredients* relates to the number of ingredients mentioned in a recipe to prepare a meal.
- *Num. Categories* details the number of category labels, which have been applied to a recipe.

Recipe Presentation (21). 21 features describe diverse aspects of a recipe's presentation. We have further grouped these into sub-sets capturing visual and textual aspects. Visual features include the 10 features proposed by San Pedro and Siersdorfer [67] to capture the attractiveness of an image. These were originally shown to work well in the context of photographs on the platform Flickr, but our recent work demonstrates that a sub-set of these features also work well in gauging the attractiveness of photographs associated with online recipes [68]. Concretely, the features derived include: sharpness, contrast, saturation, colorfulness, entropy and naturalness, all of which are defined formally below. All low level image features are measured with the freely available OpenIMAJ Java Framework[j] in version 1.3.5. OpenIMAJ is a collection of tools for analysis multimedia content such as images or video and was developed by the University of Southampton.[k]

- *Image: Sharpness.* This image metric measures the clarity and level of detail of an image. It is related to the brightness contrast of edges in an image. The algorithm utilizes the images Laplacian, divided by the locale average luminance (μ_{xy}) around pixel (x,y):

$$image\_sharpness = \sum_{x,y} \frac{L(x,y)}{\mu_{xy}}, \quad \text{with } L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{1}$$

- *Image: Sharpness variation.* Similar to the saturation variation, sharpness variation is calculated via the standard deviation of all pixel sharpness values.
- *Image: Contrast.* Contrast is the relative difference in brightness or color of local features in an image. In [69] contrast is defined as the "assessment of the difference in appearance of 2 or more parts of a field seen simultaneously or successively". There

are many metrics for contrast, but the root mean square contrast (RMS-contrast) is often used to compare images [67]. We calculate RMS-contrast as follows:

$$image\_contrast = \frac{1}{N} \sum_{x,y} (I_{xy} - \bar{I}), \tag{2}$$

where I_{xy} is the intensity of a pixel, $\bar{I}$ represents the arithmetic mean of the pixel intensity and N is the number of pixels in the image. OpenIMAJ offers the RMSContrast class for this measurement.

- *Image: RGB Contrast.* The RGB contrast is almost identical to the basic contrast calculation, explained above. However, it is extended to the three-dimensional RGB color space.

- *Image: Saturation.* According to the International Commission on Illumination [70] the image saturation is defined as the "colourfulness of an area judged in proportion to its brightness". It describes the quality of the color effect or vividness. In the HSV color space the saturation estimation can be calculated via the RGB approximation of

$$image\_saturation = \frac{1}{N} \sum_{x,y} S_{xy}, \quad \text{with}$$

$$S_{xy} = \max(R_{xy}, G_{xy}, B_{xy}) - \min(R_{xy}, G_{xy}, B_{xy}), \tag{3}$$

where N is the amount of pixels in an image and R_{xy}, G_{xy} and B_{xy} are the coordinates of the color of the pixel in sRGB space.

- *Image: Saturation variation.* This method estimates the variation in saturation via the sample standard deviation[l] of all pixel saturations of the image

$$image\_saturation\_variation = \sqrt{\frac{\sum_{x,y}(S_{xy} - \bar{S})^2}{N-1}}, \tag{4}$$

where N is the number of pixels, S_{xy} is the list of pixel saturations and $\bar{S}$ represents the arithmetic mean of the pixel saturations.

- *Image: Brightness.* The average brightness of an image attempts to measure the subjective visual perception of the energy output of a light source. The brightness of the recipe images was extracted with the AvgBrightness class[m] with the default NTSC weighting scheme and no mask. It uses a standard luminance algorithm

$$image\_brightness = \frac{1}{N} \sum_{x,y} Y_{xy}, \quad \text{with}$$

$$Y_{xy} = (0.299 * R_{xy} + 0.587 * G_{xy} + 0.114 * B_{xy}), \tag{5}$$

where Y_{xy} denotes the luminance value and N is the amount of pixels in an image. R_{xy}, G_{xy} and B_{xy} are the three RGB color space channels of pixel(x,y).

- *Image: Colorfulness.* The International Commission on Illumination [71] has defined colorfulness as an "attribute of a visual perception according to which the perceived color of an area appears to be more or less chromatic". Colorfulness can be calculated via the individual color distance of the pixels. Therefore, the image needs to be

transferred in to sRGB color space using $rg_{xy} = R_{xy} - G_{xy}$ and $yb_{xy} = 1/2(R_{xy} + G_{xy}) - B_{xy}$ and subsequently, colorfulness can be measured as

$$image\_colorfulness = \sigma_{rgyb} + 0.3 \cdot \mu_{rgyb}, \quad \text{with}$$

$$\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}, \qquad \mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2}, \tag{6}$$

where R_{xy}, G_{xy} and B_{xy} are the color channels of the pixels and σ is the standard deviation, respectively μ the arithmetic mean. The colorfulness of the recipe images was measured with corresponding class of OpenImaj.

- *Image: Entropy.* In information theory, entropy is known as a measure of randomness or the amount of information content provided by a source. The entropy of an image is often used to determine how much information needs to be encoded by a compression algorithm. As an example, an image of moon craters will have a very high edge contrast, which leads to a high entropy, meaning it cannot be compressed very well. This suggests that it can be used to measure an image's texture [72]. We used Shannon entropy as follows: First we converted the image to grey scale, where each pixel has only a intensity value. Secondly, we count the occurrences of each distinct value. Then, we apply the following formula:

$$image\_entropy = - \sum_{x \in [0..255]} p_x \cdot \log_2(p_x), \tag{7}$$

where p_x is the probability of finding the gray-scale value x among all the pixels in the image.

- *Image: Naturalness.* The concept of naturalness describes the difference (or similarity) between an image and the human visual perception of the real world, with respect to colorfulness and dynamic range. Although subjective, it is an important image quality metric when it comes to color image design [73] and according to San Pedro and Siersdorfer [67] it can be measured as follows: First transfer the image color space, if not already, to HSL. Then use only pixels within the thresholds $20 \leq L \leq 80$ and $S \geq 0.1$. In the next step, pixels are grouped in to one of the three sets 'Skin', 'Grass' or 'Sky', based on their H coordinate (hue). In order to calculate the naturalness of each set, the average saturation value of the group (μ_S) is used:

$$N_{\text{Skin}} = e^{-0.5\left(\frac{\mu_S^{\text{Skin}} - 0.76}{0.52}\right)^2}, \quad \text{if } 25 \leq hue \leq 70$$

$$N_{\text{Grass}} = e^{-0.5\left(\frac{\mu_S^{\text{Grass}} - 0.81}{0.53}\right)^2}, \quad \text{if } 95 \leq hue \leq 135 \tag{8}$$

$$N_{\text{Sky}} = e^{-0.5\left(\frac{\mu_S^{\text{Sky}} - 0.43}{0.22}\right)^2}, \quad \text{if } 185 \leq hue \leq 260$$

In the final step, the naturalness index can be calculated using

$$image\_naturalness = \sum_i \omega_i N_i, \quad i \in \{\text{'Skin', 'Grass', 'Sky'}\}, \tag{9}$$

where ω_i represents the fraction of pixels of the specific group in the whole image. Naturalness ranges from 0 (a unnuatural image) to 1 (a natural image).

Textual presentation relates to 5 features, which capture aspects of the recipe title and 6 features capturing the presentation of instruction text [74].

For the instruction text we obtained the following features:

- *Instruction: Num. Chars* captures the number of characters in the instruction text block of a recipes.
- *Instruction: Num. Words* captures the number of words in the instruction text.
- *Instruction: Num. Sentences* captures the number of sentences in the instruction text.
- *Instruction: Readability Score* measures, on a scale from 0 to 100, the extent to which the instruction text is readable. We employed LIX as proposed by [75] who showed that LIX works well across different languages including English, French and German. In addition to the advantage LIX can be used for both Allrecipes.com and Kochbar.de, it is simple to compute as it bypasses difficulties of other readability algorithms, which need to calculate the number of syllables in advance. It is mainly based on simple text measures:

$$LIX = 100 \times RLW + ASL, \quad \text{with } RLW = \frac{n_{lw}}{n_w}, ASL = \frac{n_w}{n_s}, \tag{10}$$

  where n_{lw} is the number of long words (words > 6 characters), n_w is the word count and n_s is the sentences count. Hence, the higher the value the more difficult the text is to read. Typically values of around 20 mean easy to read and greater than 60 hard to read [75].
- *Instruction: Entropy* captures the amount information contained in the text measured as $H = -\sum_{i=0}^{N} p_i \cdot \log_2(p_i)$, where p_i captures the probability of a certain character available in the whole instruction text string. N stands for the number of unique character available in the text.
- *Instruction: Sentiment* captures the amount of sentiment used to in the instruction text. Sentiment for the title as well as for the description text was attained with SentiStrength,[n] which provides two measures to analyze text in terms of the sentiment expressed by the user: positivity $\phi^+(t)$ (between +1 and +4, +4 being more positive) and negativity $\phi^-(t)$ (between −1 and −4, −4 being more negative). Following the approach of Kucuktunc et al. [76], we derived one single metric based on the values of positivity and negativity provided by Sentistrength, namely 'attitude', which provides the predominant sentiment of a text [77]. It is calculated as $\phi(t) = \phi^+(t) + \phi^-(t)$.

The same metrics were then also applied to the recipe titles, the exception being that we did not measure the number of sentences. Thus, the title-based presentation features (5 in total) can be summarized as follows:

- *Title: Num. Chars* captures the number of characters in the title text.
- *Title: Num. Words* captures the number of words in the title text.
- *Title: Readability Scores* captures to what extent the title text is readable measured via LIX (see instructions).
- *Title: Entropy* captures the amount information contained in the text measured via entropy (see instructions).
- *Title: Sentiment* captures the amount of sentiment available used to in the title text measured via SentiStrength (see instructions).

Recipe Seasonality (4). Previous work, such as [39], has shown that strong seasonal trends can be observed in online recipe preferences and upload behavior [78]. To capture these effects we derived in total 4 seasonal features which we computed as follows:

- *Upload Month.* Captures the month in the year when the recipe was uploaded.
- *Day of Month.* Captures the day in the month in which the recipe was uploaded.
- *Day of Week.* Captures the day the week in which the recipe was uploaded.
- *Within Season.* Captures the extent to which the recipe was in season at time of upload. We capture this through the recipes ingredients, i.e. a recipe should be in season when its ingredients are in season. Therefore distribution probabilities for each distinct ingredient in the entire ingredient database were estimated for each month. Finally, we calculated the mean of the ingredient probabilities for all ingredients of a specific recipe, using its month of upload. More specifically, the process was as follows: For each ingredient we (1) count the occurrences of the ingredient for each month over all recipes; (2) use the occurrences list to calculate the density function with an univariate kernel density estimator, such as the KDEUnivariate[o] method of StatsModels, which sets the optimal bandwidth automatically; (3) evaluate the point density[p] of the estimated density function for all twelve months. We then calculate for each recipe:

$$within\_season = \frac{1}{N} \sum_{i=0}^{N} \rho_i^m,$$ (11)

where N is the amount of ingredients of the recipe and ρ_i^m is the seasonality probability of ingredient i at the recipe upload month m. The seasonality value ranges from 0 (none of the ingredients were in season at upload date) to 1 (all ingredients were in season).

Recipe Innovation (6). The innovation of recipes is captured by 6 features metrics which are calculated as follows:

- *Ingredients rank.* Captures the average popularity rank of the ingredients in a recipe. It is calculated as follows:

$$ingredients\_rank = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{rank}(I_i)}{N_I},$$ (12)

where N is the total number of ingredients used in the recipe, N_I the number of ingredients in the whole corpus (until recipe upload date) and the function rank() returns the frequency rank of the ingredient I_i in the whole recipe corpus (until recipe upload date).

- *Categories rank.* The feature "categories rank" is calculated in the same way. However, it are based on the ranks of the categories with respect to the complete recipe corpus (until recipe upload date) and is calculated as follows:

$$categories\_rank = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{rank}(C_i)}{N_C},$$ (13)

where N is the total number of categories used to label a recipe, N_C the number of categories in the whole corpus (until recipe upload date) and the function rank() returns the frequency rank of the categories C_i in the whole recipe corpus.

- *Title Words rank.* Similar to the above mentioned ingredient and category rank features, this feature captures popularity of title words and is calculated as follows:

$$title\_words\_rank = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{rank}(TW_i)}{N_T}, \tag{14}$$

where N is the total number of title words used in a recipe, N_T the number of title worlds in the whole corpus and the function $\text{rank}()$ returns the frequency rank of the title words C_i in the whole recipe corpus.

In addition to these simple ranking features we also calculated three features which capture the innovation of a recipe based on all previous uploads at an ingredient level.

- *Recipe Innovation Jaccard.* The first feature is called 'recipe innovation jaccard' and was introduced by Kusmierczyk et al. [41]. As indicated by the name, the metric employs Jaccard's index to calculate the similarity of two recipes r and r' as follows:

$$recipe\_innovation\_jaccard = 1 - \max_{r' \prec r} \frac{|\{i : i \in r \wedge i \in r'\}|}{|\{i : i \in r \vee i \in r'\}|}, \tag{15}$$

where the Operator $\prec$ shows the temporal precedence of the recipes (r') compared to upload date of recipe r and where the parameter i denotes the ingredients of the recipes.

- *Avg. Recipe Innovation Jaccard.* In addition to the above mentioned metric we calculate the average recipe innovation of a recipe. The previous metric finds only the most similar recipe while the metric describes the average similarity when the recipe r is compared to each recipe r' in the collection:

$$avg\_recipe\_innovation\_jaccard = 1 - \max_{r' \prec r} \frac{|\{i : i \in r \wedge i \in r'\}|}{|\{i : i \in r \vee i \in r'\}|}. \tag{16}$$

- *Recipe Innovation IDF.* A final innovation metric measures the innovation of a recipe based on the inverse ingredient frequency. Hence the rarer an ingredient in a recipe, i.e., the lower the counter, the higher the innovation. The metric is an adopted of Kerne et al.'s innovation metric that captures the novelty of an idea based on employing an inverted index showing which creators had the same answer [79]. To do so, first, a function that returns all ingredients of a specific recipe r needs to be defined: $I_r = \{i_1, \ldots, i_n\}$. Thereafter, we define a inverse function, which returns all recipes r^* a particular ingredient i is used in, as follows:

$$ing\_occurrences_i = \{r^* : r \in R \wedge i \in I_r\}, \tag{17}$$

where R is our case the set of all recipes previously uploaded (in time) by the users in the recipe collection. The rareness of an ingredient can then be computed as follows:

$$ing\_rareness_i = \frac{|ing\_occurrences_i|}{|R|}, \tag{18}$$

where the value ranges from 0 (the ingredient was used in none of the previously uploaded recipes) to 1 (the ingredient was used in all recipes previously uploaded).

Finally, the innovation IDF value for a recipe is calculated as the sum of all its ingredient rareness, normalized by the amount of all ingredients used in the recipe:

$$recipe\_innovation\_IDF = \frac{1}{|I_r|} \sum_{i \in I_r} ing\_rareness_i \qquad (19)$$

User Activity and Context (16). A further set of features describes aspects relating to activities performed by the recipe author and the context within which these take place. In particular the features (16 in total) are the following:

- *Num. Recipes uploaded until Upload.* The number of recipes a user had uploaded at the upload time of a given recipe.
- *Num. Comments written until Upload.* The number of comments a user had written for other recipes at the upload time of a specific recipe.
- *Num. Comments received until Upload.* The number of comments a user had received for her/his recipes at the upload time of a specific recipe.
- *Num. Recipes uploaded per day/week/month/year.* These are 4 features which capture the number of recipes uploaded on average by a particular user over the course of one day, one week, one month or one year at the upload time of a specific recipe.
- *Num. Ratings provided per day/week/month/year.* These are 4 features which capture the number of ratings provided on average by a user for other users' recipes over the course of one day, one week, one month or one year at the upload time of a specific recipe.
- *Num. Comments provided per day/week/month/year.* These are 4 features which capture the amount of comments provided by a user on average over the course of one day, one week, one month or one year to other people's recipes.
- *Cook Living in Germany/USA.* Finally, a binary feature captures whether a user is located in the origin country of the food websites. This means that if a Allrecipes.com user is living in the USA, the value will be 1 otherwise 0. The same approach is applied to Kochbar.de with the value being 1 if the user is located in Germany and 0 if not.

Recipe Popularity (8). To capture recipe popularity and appreciation we employ the number of ratings and comments provided by the users within a given time range after upload date. In total we derived 8 features which are as the following:

- *Num. Comments received within day/week/month/year.* These are 4 features which capture the amount of comments a recipe has received within a one day, one week, one month and one year period after the recipe has been published.
- *Num. Ratings received within day/week/month/year.* These are 4 features which capture the amount of ratings a recipe has received within a one day, one week, one month and one year period after the recipe has been published.

4.4 Comparative statistical analysis

To compare the two communities we computed standard descriptive statistics, such as mean, median, minimum and maximum, for all features on both datasets and used significance tests to establish differences. Due to the differences in the feature distributions it was necessary to employ various tests. A Brown–Forsythe test for finding statistically equal variance was utilized. The Brown–Forsythe test is used for group comparison based on median absolute deviations (MAD) and is more robust against outliers compared to the

Leven's test, which uses the mean. In the case of equal variance of the two feature distributions ($p < 0.05$), a Wilcoxon Rank sum test was performed. When the test rejected the equal variance hypothesis, a two-sample Kolmogorov–Smirnov (KS) test was used. We report p-values.

"While a p-value can inform whether an effect exists, the p-value will not reveal the size of the effect" [80]. As such, we also computed the effect size r for each statistical test [81]. The effect sizes were calculated as $\frac{Z}{\sqrt{n_x+n_y}}$, where Z is the Z statistic, and n_x the size of sample x and n_y the size of sample y.

We mark effect sizes $0.1 \leq r < 0.3$ with an ∘ symbol (small), effect sizes $0.3 \leq r < 0.5$ with an ∘∘ symbol (medium) and effect sizes great than $r \geq 0.5$ with an ∘∘∘ symbol (large) according to Cohen's criteria [82]. According to Cohen [82] effect sizes smaller than 0.1 are irrelevant and as such are not highlighted.

4.5 Predictive modelling

The literature review presented in Sect. 2 illustrates that the task of popularity prediction can be set up in diverse ways. In this work the prediction task is preceded by a correlation analysis, which informs on possible correlations between predictors (features) and outcome variables (popularity proxies). As a metric we chose Spearman's rank correlation coefficient, since the metric assesses both linear and non linear relationships between two variables and can cope with continuous and discrete variables simultaneously. Subsequently, we operationalize and evaluate a prediction task. In line with recent literature, we chose a classification setup to predict apriori whether a recipe will become popular in the future. We establish binary classification variables to determine whether a recipe belongs to the popular or non popular class and attempt to validate models with test data. Reflecting this we calculated medians of the popularity metrics, following the approach in [30]. Recipes below the median are considered as negative and above as positive examples. While it is true that a median split may result in many recipes being close to the boundary between the two classes, we chose to remain with the standardized approach based on the findings of Hofman and colleagues [83], who highlight the danger of non-standardization in prediction tasks, such as ours. They show empirically that individually defensible choices can arrive at qualitatively different answers to the same question.

The experiments were performed on complete and balanced datasets. To ensure completeness, potential missing data was imputed with the help of the R library Hmisc.[q] To ensure balance, the majority classes were undersampled randomly employing R's sample procedure. The classification experiment itself was conducted with 3 different classifiers using the R library Caret.[r] Besides a Random Forest (RF) classifier, Logistic Regression (LOG) and Naive Bayes (NB) were employed. RF and NB have been successfully applied in similar 'before publication' popularity classification studies [30]. As evaluation protocol five fold-cross validation was chosen and 'Accuracy' was used as the main performance metric. Furthermore, the variable importance for each feature set was reported employing Caret's 'VarImp' feature, to capture the importance of the variables after the models have been built, as well as FSelector's[s] InfoGain feature to find the Top-10 features before model construction.

5 Results

5.1 Popularity analysis (RQ1)

In a first step we investigate popularity across datasets, looking at how popularity of recipes differs over time and across the two platforms. In this step we look for patterns over the full datasets i.e., using recipes in all categories.

A first view on temporal patterns in popularity can be seen in Figs. 5 (A)–(I). Figures 5 (A), (D), (G) display the number of comments and ratings assigned applied non-cumulatively over time. Figures 5 (B), (E), (H) show the same information cumulatively. Finally, Figs. 5 (C), (F), (I) use violin plots to demonstrate how the same popularity aspects evolve over time for the two recipe platforms. Whereas in the case of Allrecipes.com a constant increase can be observed in the number of ratings and comments over time, a different picture is shown for Kochbar.de. Here we see that most of the comments and ratings are provided within the first 7 days after publishing and little development is seen thereafter. One can observe a noticeable decay pattern for Kochbar.de whereby after a period of 10 days after publication any popularity tends to have dissipated. A different pattern is observed for Allrecipes.com where popularity is, on the whole, a much more consistent phenomenon, in terms of how and when recipes are attended to.

Figure 5 Allrecipes.com and Kochbar.de ratings and comments over time. The first 3 plots ((**A**), (**D** and (**G**)) in the column to the left represent the mean number of ratings and comments applied to the recipes (all categories) over time from the time on it was uploaded. In both platforms most activity is recorded on the day the recipe was published (day 0). The three plots in the middle ((**B**), (**E**) and (**H**)) and to the right ((**C**), (**F**) and (**I**)) present similar statistics calculated cumulatively. While the number of ratings and comments in Kochbar.de saturate, interaction with recipes in terms of ratings and comments constantly increases in Allrecipes.com. The diamond symbols in the violin plots to the right denote median values and overlap means which are denoted with black dots (only partly visible)

Figure 6 Plots (**A**) and (**B**) present distributions of comment sentiment and ratings in both datasets for recipes (all categories). As presented, in both datasets the ratings applied to recipes are extremely positive ($M = 4.96$ (Allrecipes.com) vs $M = 4.45$ (Kochbar.de)). Similar trends can be observed for the comment sentiment which is also very on the positive (M=2.02 (Allrecipes.com) vs M=1.87 (Kochbar.de), Scale: -4 = very negative to $+4$ = very positive). Plots (**C**) and (**D**) present mean rating and comment sentiment of recipes over time. The lines represent the linear regression of the observations and the lighter colored hulls show the confidence interval of the regression. As presented on average the trends are rather constant and there are neither strong up or downwards trends
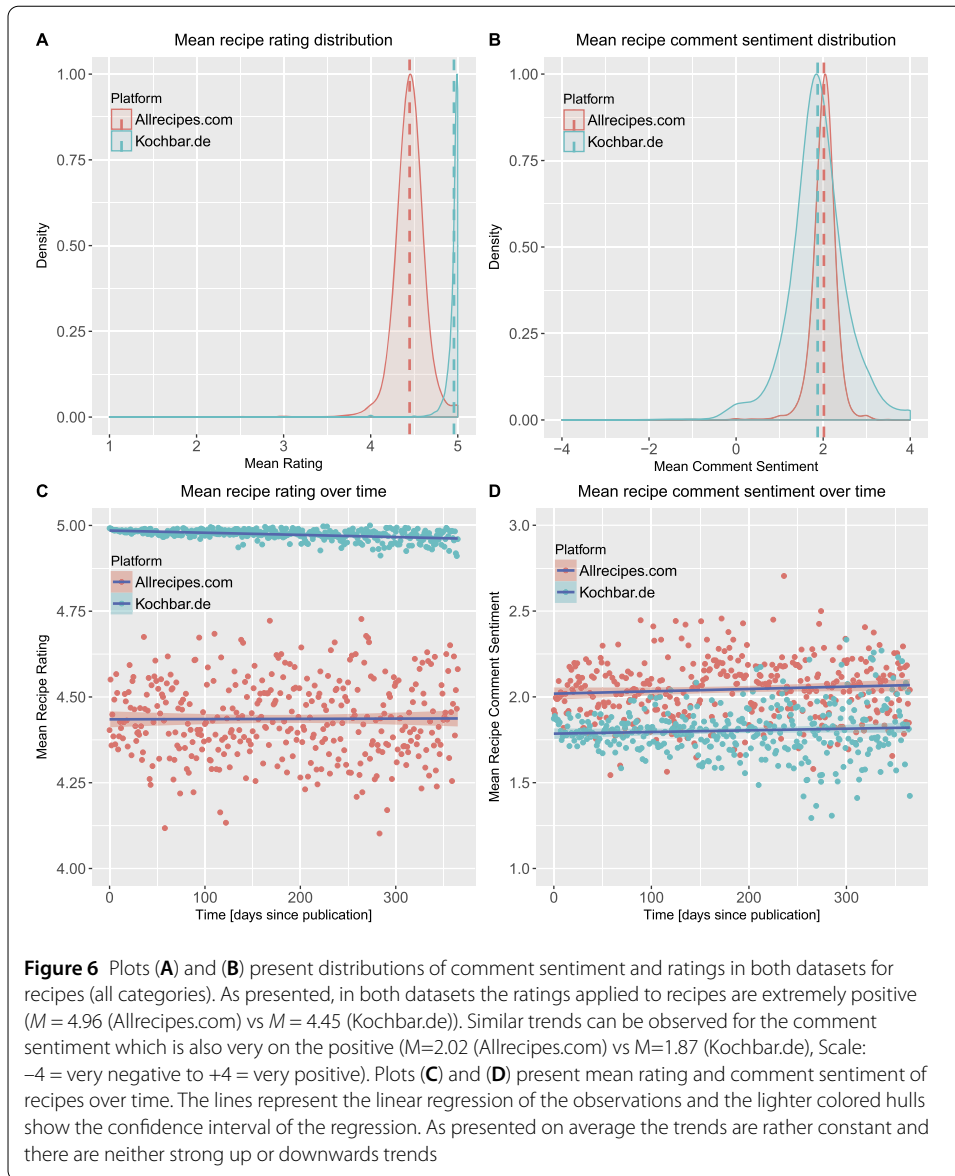
Figure 6 provides a different view of recipe appreciation by showing how recipes are commented on and rated over time after the date of publication. The recipes on both platforms receive overwhelmingly positive ratings with a low standard error for ratings being found in both datasets. While Kochbar.de recipes receive on average a rating of 5 with very little spread (nearly all rated recipes are given 5 stars), Allrecipes.com recipes obtain a slightly lower average rating of 4.41. On the other hand, the comment sentiment is higher for Allrecipes.com recipes when compared to the Kochbar.de recipes (2 vs 1.8). The analyses show limited variation in recipe appreciation, both generally across recipes and for the same recipe over time (see Fig. 5). This means that the ratings provided for a recipe are stable and do not seem to be too heavily influenced by fashions or trends. This is unsurprising for Kochbar.de, where the ratings tend to be applied in the short period after a recipe is published, but more so for the Allrecipes.com, where recipes are discovered and evaluated for much longer periods.

Figure 4 shows the distribution of the top-20 categories in Allrecipes.com and Kochbar.de (considered together and separately), revealing variance in the popularity of different recipe categories. The most popular category in terms of number of uploaded recipes (measured in percentages),[t] when considering both portals, is the 'main dish' category (18.35% vs 20.01% (Kochbar.de)), followed by 'desserts' (18.90% vs 7.90% (Kochbar.de)) and 'side dishes' (12.37% vs 3.51% (Kochbar.de)). The dessert category is the most popular in Allrecipes.com (N = 11,526), if no overlapping categories are considered, while the 'intolerance' (N = 11,526) is the most prominent one in Kochbar.de (N = 275,478), followed by 'without Wheat' and other health related categories. The popularity of recipes in these categories demonstrates dietary trends in Europe [84, 85].[u]

Thus, these initial analyses show that patterns in popularity do exist. It seems that recipe appreciation differs across platforms (recipes on Allrecipes.com continue to gain in popularity for longer than on Kochbar.de) and across categories within the two platforms (different categories are popular on the two sites with dietary trends featuring heavily on Kochbar.de, but not on Allrecipes.com). We build on these results by statistically comparing the two services with respect to the features outlined in Sect. 4.3. As inconsistent patterns were observed here across categories and because categories are disjoint across platforms, we restrict our comparison to main dishes, a popular category common to both Allrecipes.com and Kochbar.de.

5.2 Comparative statistical analysis (RQ2)

Table 3 provides an overview of the statistics (Mean, Median, Standard Deviation, Min. and Max.) values of the features grouped in 7 different feature sets as used to predict popularity in this work. The table serves two purposes. First it gives an overview of the statistical properties of the features and second, it allows the comparison of the two platforms based on these features. As such the last two columns in Table 3 report p-values and effect sizes (r). The table shows that the feature distributions vary significantly between the platforms, which is also reflected by the differences in means and medians and significant p-values. Observing low p-values is, however, not unsurprising given the large quantities of data available (N = 11,194 (Allrecipes.com) vs N = 81,232 (Kochbar.de)).

Notable exceptions are 'Image Sharpness Variation' (M = 0.29 vs 0.29 (Kochbar.de); $p <$ 0.1), 'Day of Month' when the recipe was uploaded (M = 15.71 vs 15.72 p = 0.541) and the feature 'Living in Germany/USA' (M = 0.88 vs 0.90 (Kochbar.de); $p <$ 0.1).

Due to the frequency with which significant results are found, we additionally report effect sizes. When examining the r values in Table 3 we find the features with the highest effect sizes for 'User Activity & Context' and 'Recipe Popularity & and Appreciation' sets. A good example is that 'Num. Ratings received within week' shows significantly lower values for Allrecipes.com than for Kochbar.de (M = 0.38 vs 2.39 (Kochbar.de); $p <$ 0.001; r = 0.28). In the User Activity & Context set the feature 'Num. Comments Received until Upload' showed the highest effect size (M = 73.36 vs 2472.59; $p <$ 0.001; r = 0.44).

In the other feature sets, such as 'Recipe Nutrition', Protein (per 100 g) showed the largest effect size and was significantly higher for Allrecipes.com recipes (M = 9.44 vs 6.49 (Kochbar.de); $p <$ 0.001; r = 0.23). In the 'Recipe Complexity' set the feature 'Num. Categories' showed the highest effect size with significantly more categories applied to recipes in Kochbar.de (M = 4.31 vs 12.61 (Kochbar.de); $p <$ 0.001; r = 0.50). In the 'Recipe Presentation' set we find the highest effect size for the feature 'Instruction: Readability

Table 3 Feature statistics in terms Means, Medians, Standard Deviations (Sd.Dev.), Minima (Min.) and Maxima (Max.) for the Allrecipes.com and Kochbar.de platforms (main dishes). The last two columns report statistics in terms of p-values and effect sizes (r) which aim at comparing the feature distributions between the platforms. r ranges from 0–1 and gives an impression about the magnitude of the observed differences (0 = no difference, 1 = very large differences). The first 7 feature sets capture recipe and user specific properties, while the last feature set captures recipe popularity, the outcome variable we aim to predict

Feature	Allrecipes.com (N = 11,194)					Kochbar.de (N = 81,232)					p-value	r
	Mean	Median	Std.Dev.	Min.	Max.	Mean	Median	Std.Dev.	Min.	Max.		
Recipe Nutrition (4)												
Kcal (per 100 g)	162.24	153.00	65.02	4.00	697.00	213.08	172.00	158.71	1.00	998.00	<0.001	0.06
Protein (per 100 g)	9.44	9.08	4.00	0.13	27.62	6.49	4.70	5.51	0.00	36.00	<0.001	0.23°
Carbohydrates (per 100 g)	11.54	10.23	8.44	0.00	79.74	11.88	5.70	15.26	0.00	100.00	<0.001	0.11°
Fat (per 100 g)	8.66	7.64	5.43	0.03	53.97	15.47	10.00	17.39	0.00	99.90	<0.001	0.10°
Recipe Healthiness (1)												
WHO Health Score	0.35	0.00	0.66	0.00	3.00	0.36	0.00	0.63	0.00	3.00	<0.001	0.01
Recipe Complexity (4)												
Preparation Time (Min.)	86.22	50.00	113.54	3.00	1740.00	41.43	30.00	32.70	1.00	1180.00	<0.001	0.24°
Num. Preparation Steps	3.75	4.00	1.63	1.00	19.00	4.15	4.00	2.89	1.00	25.00	<0.01	0.01
Num. Servings	6.31	6.00	3.95	1.00	72.00	3.74	4.00	1.73	1.00	70.00	<0.001	0.31°°
Num. Ingredients	10.00	9.00	4.00	1.00	35.00	11.41	11.00	4.02	1.00	72.00	<0.001	0.12
Num. Categories	4.31	4.00	0.88	3.00	6.00	12.61	12.00	5.75	1.00	36.00	<0.001	0.50°°°
Recipe Presentation (21)												
Image: Sharpness	0.18	0.17	0.09	0.02	0.81	0.18	0.15	0.10	0.01	1.21	<0.001	0.07
Image: Sharpness Variation	0.29	0.26	0.15	0.00	1.50	0.29	0.26	0.17	0.01	2.00	<0.1	0.01
Image: Contrast	0.05	0.05	0.02	0.00	0.16	0.05	0.04	0.02	0.00	0.19	<0.001	0.00
Image: Contrast RGB	0.17	0.16	0.07	0.00	0.49	0.16	0.15	0.07	0.01	0.57	<0.001	0.00
Image: Saturation	0.27	0.26	0.10	0.02	0.75	0.23	0.21	0.10	0.00	0.80	<0.001	0.18°
Image: Saturation Variation	0.15	0.15	0.05	0.00	0.41	0.15	0.15	0.05	0.00	0.41	<0.001	0.02
Image: Brightness	0.49	0.49	0.10	0.12	0.89	0.50	0.49	0.11	0.06	0.92	<0.001	0.03
Image: Colorfulness	0.24	0.24	0.07	0.03	0.54	0.23	0.22	0.08	0.00	0.68	<0.001	0.08
Image: Entropy	7.47	7.53	0.32	3.16	7.96	7.37	7.43	0.34	4.25	7.96	<0.001	0.13°
Image: Naturalness	0.83	0.85	0.12	0.00	1.00	0.79	0.80	0.13	0.00	1.00	<0.001	0.11°
Instruction: Num. Chars†	682.57	616.00	327.94	86.00	2866.00	775.61	690.00	417.15	10.00	5761.00	<0.001	0.08

Table 3 (*Continued*)

Feature	Allrecipes.com (N = 11,194)					Kochbar.de (N = 81,232)					p-value	r
	Mean	Median	Std.Dev.	Min.	Max.	Mean	Median	Std.Dev.	Min.	Max.		
Instruction: Num. Words†	115.99	104.00	56.07	14.00	498.00	110.20	97.00	60.89	1.00	889.00	<0.001	0.00
Instruction: Num. Sentences†	9.73	9.00	4.34	1.00	40.00	10.48	10.00	5.91	1.00	64.00	<0.001	0.03
Instruction: Readability Score†	30.42	30.45	5.50	8.00	54.49	48.84	45.92	16.62	1.00	379.91	<0.001	0.53°°°
Instruction: Entropy†	4.35	4.35	0.08	3.96	4.63	4.51	4.51	0.10	2.85	5.52	<0.001	0.46°°
Instruction: Sentiment†	0.15	0.00	0.99	−3.00	4.00	0.28	0.00	1.13	−4.00	4.00	<0.001	0.04
Title: Num. Characters†	25.43	24.00	9.93	3.00	110.00	50.94	44.00	31.56	3.00	400.00	<0.001	0.33°°
Title: Num. Words†	4.13	4.00	1.85	1.00	19.00	7.72	6.00	5.61	1.00	70.00	<0.001	0.24°
Title: Readability Score†	39.99	36.33	27.61	1.00	105.00	54.72	52.00	22.61	1.00	107.00	<0.001	0.19°
Title: Entropy†	3.72	3.76	0.33	1.50	4.54	4.04	4.10	0.37	1.50	5.18	<0.001	0.30°°
Title: Sentiment†	0.03	0.00	0.33	−3.00	4.00	0.39	0.00	0.99	−4.00	4.00	<0.001	0.09
Recipe Seasonality (4)												
Upload Month	5.96	6.00	3.28	1.00	12.00	6.12	6.00	3.45	1.00	12.00	<0.001	0.01
Day of Month	15.71	16.00	8.74	1.00	31.00	15.72	16.00	8.74	1.00	31.00	0.541	0.00
Day of Week	4.23	5.00	2.01	1.00	7.00	3.84	4.00	2.01	1.00	7.00	<0.001	0.00
Within Season	0.08	0.09	0.01	0.04	0.25	0.08	0.08	0.01	0.05	0.16	<0.001	0.06
Recipe Innovation (6)												
Ingredients rank	0.08	0.06	0.05	0.00	1.00	0.08	0.07	0.05	0.00	1.00	<0.001	0.01
Categories rank	0.05	0.02	0.06	0.00	1.00	0.07	0.06	0.03	0.00	1.00	<0.001	0.24°
Title Words rank	0.07	0.04	0.10	0.00	1.00	0.05	0.03	0.08	0.00	1.00	<0.001	0.11°
Recipe Innovation Jaccard	0.57	0.60	0.14	0.00	1.00	0.51	0.53	0.13	0.00	1.00	<0.001	0.00
Avg. Recipe Innovation Jaccard	0.96	0.97	0.02	0.00	1.00	0.94	0.94	0.02	0.00	1.00	<0.001	0.33°°
Recipe Innovation IDF	0.02	0.01	0.05	0.00	1.00	0.00	0.00	0.02	0.00	1.00	<0.001	0.37°°
User Activity & Context (16)												
Num. Recipes uploaded until Upload	16.96	0.00	78.71	0.00	1386.00	258.32	94.00	380.71	0.00	5795.00	<0.001	0.42°
Num. Comments written until Upload	9.24	0.00	55.26	0.00	2027.00	1582.92	103.00	4000.69	0.00	38,124.00	<0.001	0.39°°
Num. Comments received until Upload	73.36	0.00	514.37	0.00	6310.00	2472.59	317.00	5470.63	0.00	52,878.00	<0.001	0.44°°
Num. Recipes uploaded per day	0.05	0.00	0.16	0.00	1.00	0.34	0.16	0.58	0.00	11.25	<0.001	0.28°
Num. Recipes uploaded per week	0.38	0.02	1.11	0.00	7.01	2.39	1.14	3.94	0.00	63.00	<0.001	0.28°

Table 3 (*Continued*)

Feature	Allrecipes.com (N = 11,194)					Kochbar.de (N = 81,232)					p-value	r
	Mean	Median	Std.Dev.	Min.	Max.	Mean	Median	Std.Dev.	Min.	Max.		
Num. Recipes uploaded per month	1.64	0.10	4.83	0.01	30.48	10.38	4.96	17.15	0.01	270.00	<0.001	0.28°
Num. Recipes uploaded per year	19.68	1.16	57.98	0.07	365.90	124.55	59.54	205.54	0.16	3285.00	<0.001	0.28°
Num. Ratings provided per day	0.02	0.00	0.05	0.00	0.91	5.93	1.70	10.01	0.00	95.28	<0.001	0.28°
Num. Ratings provided per week	0.11	0.02	0.35	0.00	6.35	41.49	11.87	70.07	0.00	666.98	<0.001	0.28°
Num. Ratings provided per month	0.46	0.10	1.48	0.00	27.23	177.80	50.89	300.30	0.00	2858.49	<0.001	0.28°
Num. Ratings provided per year	5.49	1.23	17.55	0.00	323.18	2109.90	603.89	3563.61	0.00	33,920.80	<0.001	0.29°
Num. Comments provided per day	0.02	0.00	0.05	0.00	0.91	2.25	0.48	4.31	0.00	50.43	<0.001	0.25°
Num. Comments provided per week	0.11	0.02	0.35	0.00	6.35	15.76	3.37	30.15	0.00	350.88	<0.001	0.25°
Num. Comments provided per month	0.46	0.10	1.48	0.00	27.23	68.52	14.64	131.14	0.01	1535.51	<0.001	0.25°
Num. Comments provided per year	5.49	1.23	17.55	0.00	323.18	822.10	175.63	1573.07	0.16	18,295.90	<0.001	0.25°
Cook Living in Germany/USA	0.88	1.00	0.33	0.00	1.00	0.90	1.00	0.29	0.00	1.00	<0.1	0.01
Recipe Popularity (8)												
Num. Comments received within day	0.09	0.00	0.30	0.00	5.00	5.14	2.00	7.11	0.00	75.00	<0.001	0.42°°
Num. Comments received within week*	0.20	0.00	0.50	0.00	7.00	7.83	4.00	9.29	0.00	92.00	<0.001	0.49°°
Num. Comments received within month*	0.43	0.00	0.97	0.00	24.00	8.34	5.00	9.53	0.00	96.00	<0.001	0.48°°
Num. Comments received within year	2.30	1.00	5.43	0.00	212.00	9.19	6.00	9.86	0.00	123.00	<0.001	0.37°
Num. Ratings received within day	0.09	0.00	0.30	0.00	5.00	11.73	6.00	14.46	0.00	127.00	<0.001	0.47°
Num. Ratings received within week*	0.20	0.00	0.50	0.00	7.00	19.23	11.00	20.85	0.00	169.00	<0.001	0.50°°°
Num. Ratings received within month*	0.43	0.00	0.97	0.00	24.00	20.60	12.00	21.66	0.00	175.00	<0.001	0.50°°°
Num. Ratings received within year	2.30	1.00	5.43	0.00	212.00	23.13	15.00	22.37	0.00	206.00	<0.001	0.47°

Note: † Language dependent features. * Indicates the 4 dependent variables features which we predict in this paper. We do not predict the number of ratings/comments received within one day/year as they are very similar to the results obtained from the week and month prediction task. Higher mean values (M) in terms of compared features are underlined. ° indicates low effect size, °° medium effect size and °°° high effects. Effect sizes below r < 0.1 are not marked.

Score' with a sign higher score for Kochbar.de recipes (M = 30.42 vs 48.84 (Kochbar.de); $p < 0.001$; $r = 0.53$). Lastly, for the 'Recipe Innovation' set we find the highest effect size for the feature 'Recipe Innovation IDF' which is significantly larger for Allrecipes.com (M = 0.02 vs 0.00 (Kochbar.de); $p < 0.001$; $r = 0.37$). It is worth mentioning that for the 'Recipe Seasonality' and 'Recipe Healthiness' sets we did not find any effect sizes above $r < 0.1$.

In this section we have demonstrated several differences in the properties of the recipes sourced from two different platforms. The Allrecipe.com are more protein rich, whereas in terms of innovation, it seems Kochbar.de is more conservative. Ratings and categorization behavior seems to differ with recipes in Kochbar.de being assigned to more categories and receiving more comments; and drop-off rates being seen in ratings on this service. In the following section we examine these properties to establish a relationship between these properties and popularity. This forms a useful precursor to the modeling work in Sect. 5.4.

5.3 Correlation analysis (RQ3)

The results presented in Sect. 5.1 demonstrate the existence of patterns in popularity with respect to platform over time and across categories. Building on this, in Sect. 5.2 we demonstrated that the two datasets differ in terms of the developed features. The next step is to discover if popularity prediction can be achieved and if so, which features offer utility for this purpose. In a first step, we perform a correlation analysis with this aim in mind. Figures 7 and 8 show the outcome of the correlation analysis for Allrecipes.com and Kochbar.de. The two figures present a correlation matrix based on Spearman's rank correlation where non-significant features ($p < 0.05$) are marked with black dots and significant features in respect to popularity metrics are marked with a red asterisk.

When comparing both matrices, it is obvious that more and stronger correlations exist for Kochbar.de than Allrecipes.com.

For example, high positive and significant correlations (ρ up to 0.94) are found in Kochbar.de between the user *activity features* (number of recipes uploaded, comments written or ratings provided) and the outcome variables (number of comments/ratings received within time-span). The fact that these features show strong positive correlations supports the assumption that the previous user activity may relate to the future popularity of recipes these users upload.

Further observations of note relate to *recipe innovation*. In the Kochbar.de as well as in the Allrecipes.com dataset the outcome variables significantly negatively correlate with the innovation feature 'Recipe Innovation IDF' (ρ up to −0.35 for Allrecipes.com and −0.15 for Kochbar.de). Hence, innovative recipes on the both platforms tend to receive more ratings and comments.[v]

In terms of the *presentation of recipes*, the recipe instruction readability score feature (ρ up to −0.13 for Allrecipes.com and Kochbar.de) is negatively and significantly correlated with the outcome variables (lower numbers in readability means the less complex and the more readable the text). Other than that positive correlations with the outcome variables can be found in Kochbar.de also for the image feature 'Image:Brightness' (ρ up to 0.13) or the number of title words or characters (ρ up to 0.26).

The remaining correlations are less surprising. For example, as one would expect, the number of preparation steps is positively correlated with the size of the instruc-
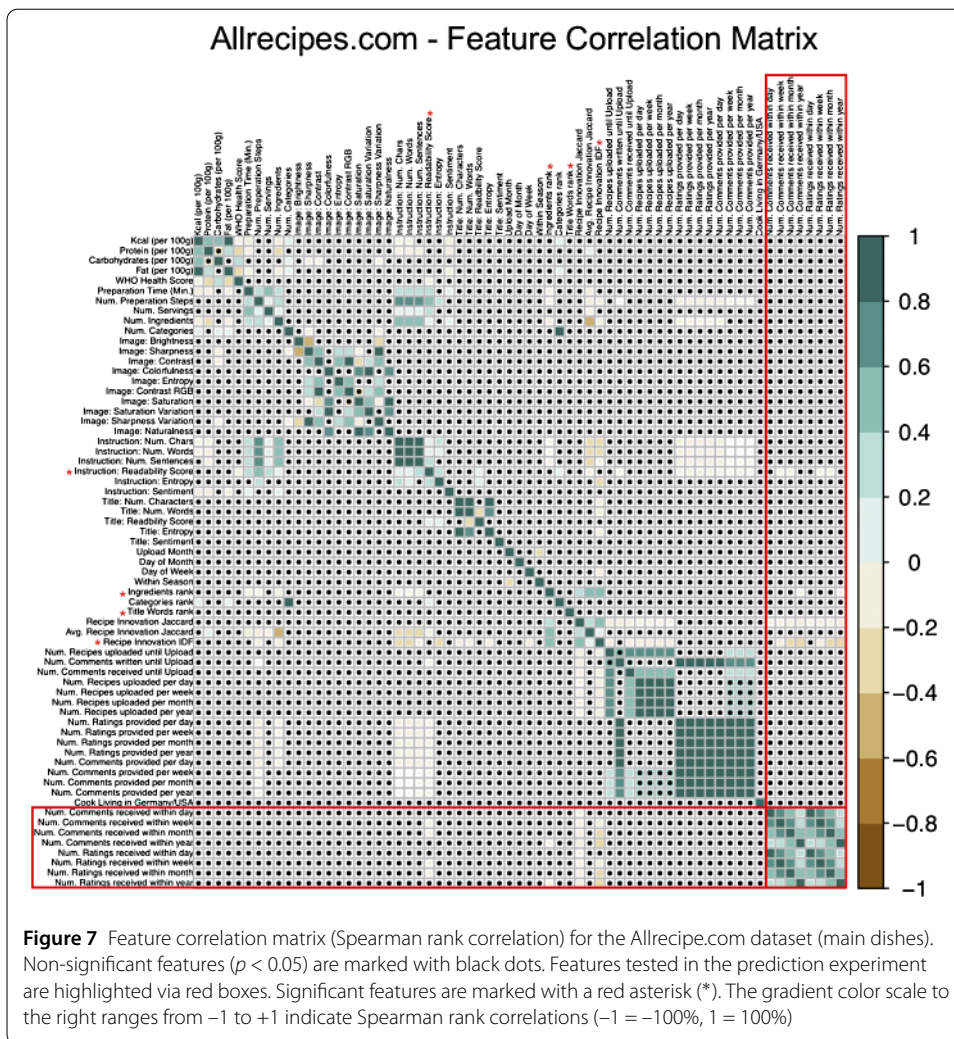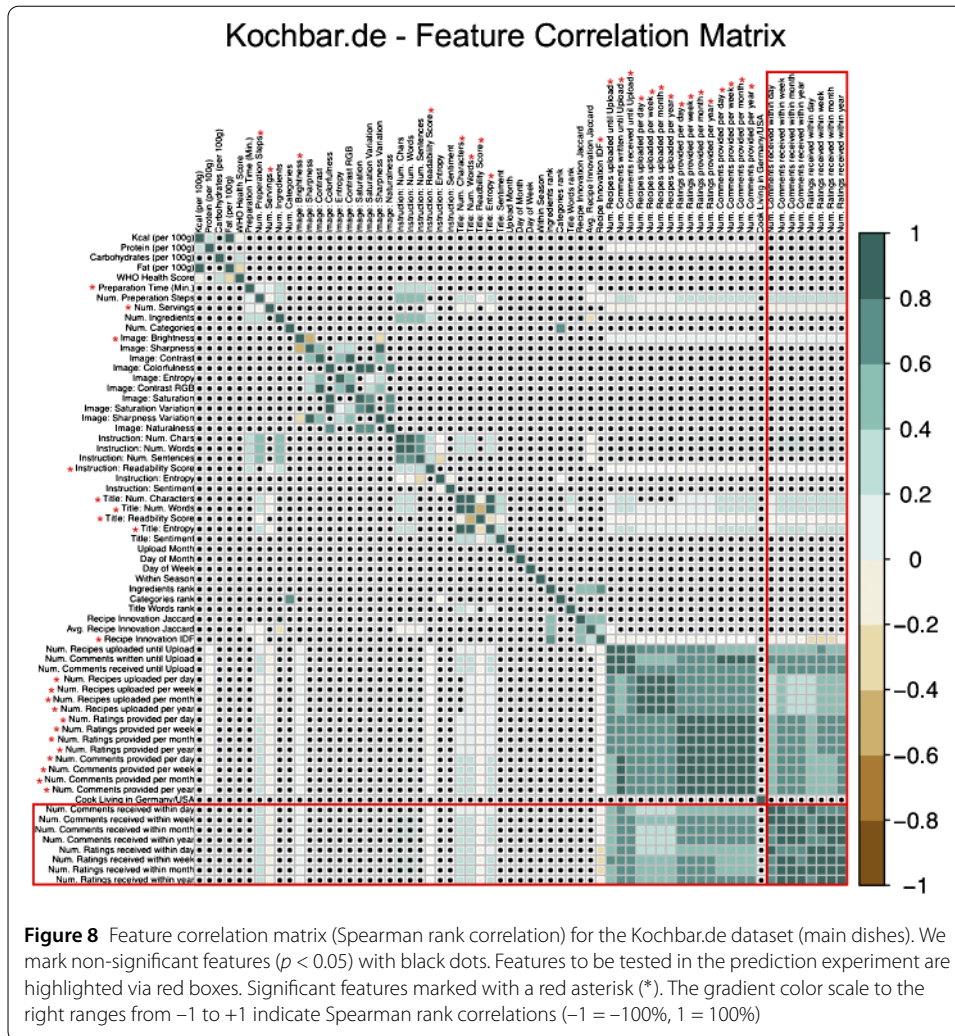
Figure 7 Feature correlation matrix (Spearman rank correlation) for the Allrecipe.com dataset (main dishes). Non-significant features ($p < 0.05$) are marked with black dots. Features tested in the prediction experiment are highlighted via red boxes. Significant features are marked with a red asterisk (*). The gradient color scale to the right ranges from −1 to +1 indicate Spearman rank correlations (−1 = −100%, 1 = 100%)

tion text and correlations can be found when comparing image features are with each other.

The correlation analyses reveal some hints that prediction may be possible with Kochbar.de, in particular correlations were revealed between features and outcome variables. The findings suggest that features measuring the user activity, recipe innovation and presentation would be descriminative when predicting popularity in Kochbar.de. The analyses suggest that prediction may be more difficult with the Allrecipes.com dataset as the patterns found were not so pronounced (less significant correlations). However, they are comparable to Kochbar.de and it seems that recipe innovation and recipe presentation (instruction readability) are the most important ones. Next, we run experiments to determine the potential for popularity prediction.

5.4 Predictive modeling (RQ4)

Table 4 presents the primary results for the prediction experiments whereby popularity after one week and after one month are estimated. These choices reflect the patterns discussed above, which show that several days are required for interaction patterns for recipes to stabilize. Moreover, as ratings behaviour thereafter is relatively stable over time on All-

Figure 8 Feature correlation matrix (Spearman rank correlation) for the Kochbar.de dataset (main dishes). We mark non-significant features ($p < 0.05$) with black dots. Features to be tested in the prediction experiment are highlighted via red boxes. Significant features marked with a red asterisk (*). The gradient color scale to the right ranges from −1 to +1 indicate Spearman rank correlations (−1 = −100%, 1 = 100%)

recipes.com and most of the ratings and comments for a recipe are collected within a month after publishing on Kochbar.de, very similar results are achieved when predicting popularity after one year. For economy of space and ease of communication, we omit our results for the popularity after a year experiments.

Our analyses thus far have focused on the main dishes category for the reasons outlined above. However, to examine the generalizability of the approach, we additionally report the results of experiments predicting the future popularity of all recipes in both collections. The top-half of the Table 4 shows the outcomes for the category 'main dishes' for both popularity proxies employed ('number of ratings' and 'number of comments'), while the bottom half shows the results for the same experiments over the whole dataset (i.e. over all categories). Table 4 reveals that regardless of the classifier employed (NB, LOG, and RF) and proxy studied (comments or ratings), the feature sets offering best performance are consistently 'Recipe innovation' features for Allrecipes.com and 'User Activity & Context' for Kochbar.de. In most cases, when predicting the future popularity of main meals, the best prediction results are achieved with Logistic Regression (LOG) for Allrecipes.com and with Random Forests (RF) for Kochbar.de.

Table 4 Prediction accuracy results for the number of ratings and number of comments based on user and recipe feature sets within one week and within one month after a recipe (Main dish or all categories of recipes) was published, for Allrecipes.com and Kochbar.de. The 'All' feature set refers to when all 7 feature sets (Recipe Nutrition, Healthiness, Complexity, Presentation, Seasonality, Innovation and User Activity & Context) are combined, 'Top-10' denotes the best 10 features selected via InfoGain and the 'Top-20' categories refer to the categories in the dataset to which the most recipes have been uploaded. Note that the number of ratings and comments are identical in Allrecipes.com and thus the results are not mirrored again for the number of comments

Feature Set	Accuracy												Num.Feat.
	Allrecipes.com						Kochbar.de						
	Within one Week			Within one Month			Within one Week			Within one Month			
	RF	LOG	NB	RF	LOG	NB	RF	LOG	NB	RF	LOG	NB	
Main dishes													
Num. Ratings	Class size: 1653			Class size: 2772			Class size: 38,573			Class size: 37,789			
Recipe Nutrition	51.90%	54.43%	52.90%	50.23%	52.04%	51.70%	51.34%	50.77%	51.02%	50.75%	50.57%	50.87%	4
Recipe Healthiness	48.28%	48.73%	47.74%	50.87%	50.83%	51.05%	49.91%	50.18%	49.90%	49.79%	49.68%	49.94%	1
Recipe Complexity	52.18%	53.70%	48.10%	51.82%	49.68%	53.01%	62.71%	62.89%	64.30%	62.76%	62.73%	64.44%	5
Recipe Presentation	55.70%	52.89%	53.71%	55.01%	56.78%	55.77%	71.10%	61.94%	71.92%	71.54%	61.94%	72.26%	21
Recipe Seasonality	53.44%	53.89%	50.90%	52.85%	53.72%	53.07%	62.68%	53.57%	54.44%	61.59%	53.36%	54.23%	4
Recipe Innovation	56.69%	61.39%	58.13%	57.47%	58.62%	58.66%	64.62%	64.92%	57.59%	63.45%	64.11%	57.08%	6
User Activity & Context	55.34%	56.24%	49.55%	56.82%	48.85%	49.80%	89.95%	82.56%	80.66%	89.44%	81.91%	79.79%	16 (12)
All	59.22%	58.14%	51.54%	60.23%	59.09%	54.36%	88.45%	84.51%	80.50%	88.09%	83.91%	80.00%	57 (53)
Top-10 Features	56.69%	59.49%	57.32%	59.90%	58.01%	52.22%	89.32%	82.60%	82.75%	88.88%	81.96%	82.12%	10
Num. Comments	Same as ratings						Class size: 35483			Class size: 37800			
Recipe Nutrition							51.15%	51.06%	51.27%	51.08%	51.19%	51.37%	4
Recipe Healthiness							50.31%	50.34%	50.21%	50.11%	50.22%	49.76%	1
Recipe Complexity							61.67%	62.32%	63.57%	61.33%	61.63%	62.56%	5
Recipe Presentation							69.43%	62.08%	70.16%	69.07%	61.58%	69.86%	21
Recipe Seasonality							55.26%	51.87%	52.66%	54.79%	51.47%	52.31%	4
Recipe Innovation							57.30%	59.09%	52.62%	56.09%	57.83%	55.23%	6
User Activity & Context							86.04%	79.67%	80.21%	85.04%	78.40%	78.74%	16 (12)
All							85.50%	80.60%	79.15%	84.23%	79.36%	80.70%	57 (53)
Top-10 Features							86.15%	79.63%	81.12%	84.89%	78.42%	79.72%	10

Table 4 (*Continued*)

Feature Set	Accuracy												Num.Feat.
	Allrecipes.com						Kochbar.de						
	Within one Week			Within one Month			Within one Week			Within one Month			
	RF	LOG	NB	RF	LOG	NB	RF	LOG	NB	RF	LOG	NB	
All kinds of dishes (all categories)													
Num. Ratings	Class size: 7948			Class size: 12583			Class size: 14428			Class size: 142,011			
Recipe Nutrition	50.58%	52.13%	51.50%	51.33%	52.91%	52.79%	51.83%	50.56%	51.74%	51.72%	50.75%	51.71%	4
Recipe Healthiness	49.84%	50.26%	49.75%	49.97%	50.27%	49.90%	49.99%	50.09%	50.04%	49.99%	50.13%	50.08%	1
Recipe Complexity	51.79%	52.85%	53.14%	53.00%	53.45%	51.22%	63.99%	62.56%	58.57%	63.49%	62.32%	59.91%	5
Recipe Presentation	53.49%	54.65%	55.16%	55.12%	56.49%	56.65%	69.94%	62.11%	70.75%	70.15%	61.93%	70.96%	21
Recipe Seasonality	52.66%	52.77%	53.03%	53.16%	54.16%	53.95%	67.61%	53.35%	53.81%	66.26%	52.76%	53.35%	4
Recipe Innovation	55.20%	**56.73%**	**56.50%**	56.47%	**57.96%**	**58.09%**	68.77%	68.01%	62.63%	67.05%	66.74%	62.23%	6
User Activity & Context	**55.20%**	53.46%	52.19%	**58.12%**	53.80%	54.33%	**89.24%**	**81.13%**	**78.59%**	**88.62%**	**80.51%**	**77.77%**	16 (12)
All	57.77%	58.61%	53.24%	60.78%	60.57%	62.29%	87.87%	83.57%	78.42%	87.17%	82.67%	78.75%	57 (53)
Top-10 Features	56.42%	54.76%	53.43%	59.37%	56.88%	59.90%	88.39%	81.17%	80.35%	87.76%	80.52%	79.87%	10
Top-20 Categories	52.94%	53.21%	51.84%	54.22%	54.20%	50.02%	58.77%	58.10%	50.03%	58.61%	57.89%	50.03%	20
Num. Comments	Same as ratings						Class size: 127005			Class size: 138,568			
Recipe Nutrition							51.40%	51.28%	52.07%	51.51%	51.21%	52.03%	4
Recipe Healthiness							50.00%	49.88%	50.05%	50.17%	49.90%	50.14%	1
Recipe Complexity							61.29%	60.84%	55.75%	60.44%	60.11%	58.54%	5
Recipe Presentation							67.68%	61.67%	68.51%	67.17%	60.99%	66.62%	21
Recipe Seasonality							57.07%	52.45%	52.54%	55.88%	52.47%	52.55%	4
Recipe Innovation							58.15%	59.51%	57.71%	56.42%	57.74%	56.21%	6
User Activity & Context							**83.89%**	**76.15%**	**78.14%**	**82.73%**	**74.77%**	**76.51%**	16 (12)
All							83.30%	77.61%	76.61%	81.87%	76.01%	79.21%	57 (53)
Top-10 Features							83.92%	76.21%	79.53%	82.46%	74.80%	77.77%	10
Top-20 Categories							57.26%	56.84%	50.23%	56.61%	56.28%	50.20%	20

Note: RF = Random Forrest, LOG = Logistic Regression, NB = Naive Bayes. The highest accuracy values in each set are underlined, whereas the second and third highest are bolded.

The highest performance is achieved using number of ratings as a proxy for popularity. When combining all features (All set) we are able to accurately predict popularity in up to 60.23% of cases for Allrecipes.com data and up to 88.45% of cases in Kochbar.de. When considering only the Top-10 features, as determined using Information Gain, slightly improved performance can be attained for the Kochbar.de dataset. Given the difficulty of problem being tackled, we feel the performance is remarkably high. Bearing the stronger and more plentiful correlations found in Sect. 5.3 in mind, the fact that better results were achieved with the Kochbar.de recipes is less surprising. Overall, we observe that performance when predicting popularity after 1 week and 1 month is comparable.

When number of comments is used as a proxy for popularity, the results demonstrate similar patterns for both platforms. Overall the task becomes slightly more difficult with the accuracy values of typically 3–4% less being achieved. Taking number of comments as the popularity measure, in Allrecipes.com we achieve maximum prediction accuracy of 61.39% when predicting for the one week (7 days) and 58.62% when predicting one month in advance (30 days). In Kochbar.de, again we achieve higher accuracy. Predicting popularity one week ahead on the German dataset, we achieve a maximum accuracy of 86.04% and for one month ahead the best result was 85.04%. The second and third best feature sets using comments for Allrecipes.com are 'User Activity & Context' and the 'Recipe Presentation', with an accuracy up to 56.82% being achieved. A similar trend is observed when investigating the second and third best feature sets in Kochbar.de. Here we also find that the 'Recipe Presentation' feature set works remarkably well, as well as the 'Recipe Complexity' set with Accuracy values up to 70.16% for comments.

When predicting popularity using the full dataset (bottom-half of Table 4), in general the same trends are observed, with 'Recipe Innovation' and 'User Activity & Context' being the most important feature sets, although, the overall predictions are typically 1–3% lower in terms of accuracy, which is to be expected given the inconsistent patterns observed across categories as shown in Sect. 5.1.

To study the effects of differences in popularity across categories in more detail, we trained models using only category information (Top-20 Categories). To do this we represented recipes by a binary vector of size 20, one element for each of the twenty most popular categories for that platform, where 1 denotes that the category had been applied to the recipe and 0 denotes this not to be the case. Surprisingly, given the differences observed across categories, none of these models could outperform any of the other full models (all), although in some of the cases, model trained on 'Recipe Nutrition' and 'Recipe Healthiness' sets were outperformed. The best performance attained by a "Top-20-Categories" model is achieved Kochbar.de dataset for the proxy 'number of ratings' with an accuracy value of 58.77%. In Allrecipes.com this model reaches no higher than 54.22%. This indicates that there is a signal in categories but weaker as the other signals derived directly form the recipes.
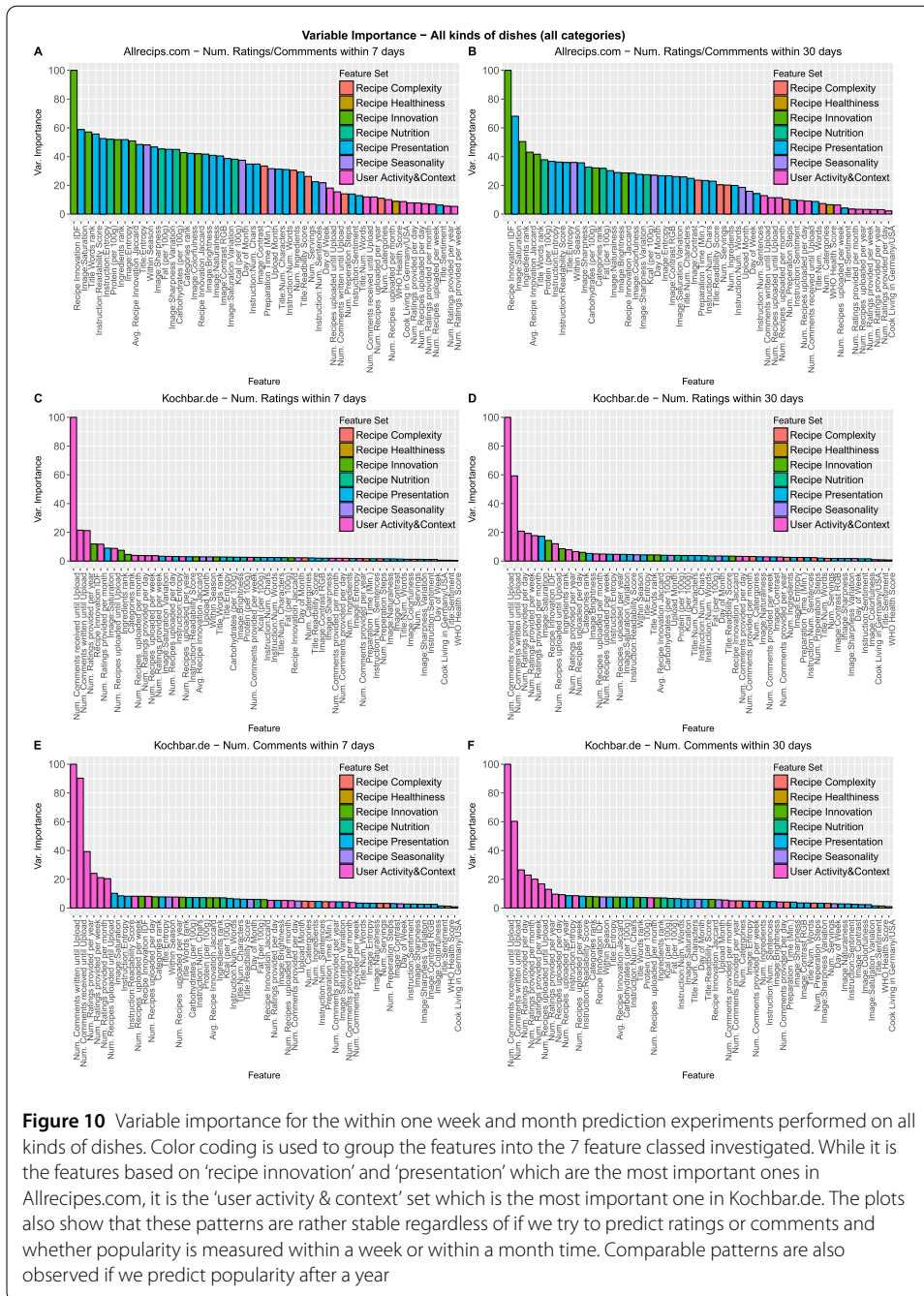
As a final step, we investigate the contribution of individual features to the predictive models. Figures 9 and 10 illustrates this graphically, showing the variable importance, measured by Breiman's method,[w] for the highest performing model for each dataset/prediction task combination, for main dishes and for all dishes, respectively.

It seems that while 'recipe innovation' and 'recipe presentation' were the most important predictive features in Allrecipes.com, for Kochbar.de, the most influential features were

Figure 9 Variable importance for the within one week and month prediction experiments performed on main dishes. Color coding is used to group the features into the 7 feature classes investigated. While it is the features based on 'recipe innovation' and 'presentation', which are the most important ones in Allrecipes.com, it is the 'user activity & context' set, which is the most important one in Kochbar.de. The plots also show that these patterns are rather stable regardless of if we try to predict ratings or comments and whether popularity is measured within a week or within a month time. Comparable patterns are also observed if we predict popularity after a year

related to user activity, that is the recipes posted by prominent users in the community typically go on to be the most popular. In line with the prediction results obtained before the best individual feature seems to be 'Recipe Innovation IDF' for Allrecipes.com while it is 'Num. Comments received until upload' and 'Num. Comments Written until Upload' for Kochbar.de. Comparing Figs. 9 and 10 suggests that the overall patterns regarding the influential features for each service is similar regardless of whether "main dishes" or "all categories" are being studied. One slight difference that can be observed is that innovation factors become more important when you predict popularity over all categories on Allrecipes.com.

Figure 10 Variable importance for the within one week and month prediction experiments performed on all kinds of dishes. Color coding is used to group the features into the 7 feature classed investigated. While it is the features based on 'recipe innovation' and 'presentation' which are the most important ones in Allrecipes.com, it is the 'user activity & context' set which is the most important one in Kochbar.de. The plots also show that these patterns are rather stable regardless of if we try to predict ratings or comments and whether popularity is measured within a week or within a month time. Comparable patterns are also observed if we predict popularity after a year

6 Summary & discussion

In this section we first summarize our findings with respect to our research questions before discussing the findings in the context of the literature and with respect to future work.

6.1 The primary findings

RQ1. While investigating popularity patterns we uncovered significant differences between the two platforms and across categories within platforms. Whereas for All-recipes.com the most popular categories reflected the traditional distinction between

mains, sides and desserts, on Kochbar.de popular recipes categories often relate to health and dietary preference, such as allergies and intolerance. We, moreover, observed differences in temporal patterns in popularity between the two sites with Allrecipes.com having stable patterns of appreciation whereas Kochbar.de had steep drop-off rates, most likely reflecting differences in how recipes are accessed and promoted on the two sites.

RQ2. We uncovered differences between the platforms when investigating the features outlined in Sect. 4.3. Recipes from Allrecipes.com were found to be higher in protein content than those on Kochbar.de, however, no effect sizes were observed for the WHO score, which is a measure of overall nutritional healthiness was close to zero. Thus, we cannot claim that the recipes of one service are healthier than the other. A further distinction between the recipes on the two services relates to recipe novelty. The recipes on Kochbar.de seem to be more homogeneous with the innovation scores being significantly higher for the recipes in the Allrecipes.com dataset.

RQ3. The correlation analysis presented in Sect. 5.3 revealed relationships among the predictive features and between the predictive features and the outcome variables (proxies for popularity). More and stronger correlations were found when analyzing the Kochbar.de dataset. The strongest and significant correlations included user activity features, which represent how active a user is within the community, such as the number of comments or ratings received within a given time span, and the outcome variables. Other strong correlations for both datasets were observed for recipe innovation and the complexity of a recipe.

RQ4. Mirroring the findings of the correlation analysis, the performance of the derived predictive models was consistently better for the Kochbar.de dataset. Typical accuracy achieved was $\sim 80\%$ for the Kochbar.de experiments whereas an accuracy of 56–60% was achieved in the Allrecipes.com experiments. Overall, given the difficulty of the task, we were pleasantly surprised by the predictive performance achieved across the experiments for both portals, with stable results being provided by the various feature sets studied. The stability in performance using different popularity proxies and time windows for predictions provides evidence supporting the robustness of the predictive features. 'Recipe innovation' and 'recipe presentation' were the most important classes of predictive features in Allrecipes.com but for Kochbar.de, the features with most predictive power were related to user activity. In other words, on Kochbar.de the recipes posted by prominent users in the community typically go on to be the most popular.

6.2 Discussion

The findings can be related to past and future work. Firstly, we have shown that there are signals in the recipe content and user interaction patterns with recipes, which allow popularity to be predicted to some extent and to a large extent on Kochbar.de. This is a contribution to the popularity prediction literature, which has shown prediction to be possible on many web-based items, but not online recipes. In doing so we open up the benefits that have been offered in other contexts, such as driving and maintaining user engagement [5], informing recommendations [6] and influencing caching strategies [9, 10] to online food portals. A large contribution of our work here is the feature engineering efforts. By collating and documenting features from literature from diverse fields, we provide a strong platform for researchers to continue our work and build on popularity prediction in online recipe settings, perhaps by using different modeling approaches or by studying other food portals.

In our work we studied two different platforms and by doing so we were able to identify similarities and differences across the communities. We wish to note that because of several, sometimes subtle differences in the platforms, their interfaces and contexts in which they are used, caution should be emphasized in interpreting the findings. We identified similarities, which include the healthiness of recipes, the number of preparation steps, image properties for uploaded food photos and the percentage of users of the service resident in the primary country (Germany or the United States). Differences identified include the longevity of popularity, with Kochbar.de recipes only having a short period of user interest as previously reported with social media posts [86] and Allrecipes.com recipes offering utility for longer periods, as can be observed with web pages [87]. Further differences between the services studied relate to the content of the recipes themselves with the Kochbar.de recipes being more homogeneous in terms of the ingredients used and the Allrecipes.com recipes being more innovative according to existing metrics [41]. The Allrecipes.com recipes contained more protein and because we know that posted and interacted with recipes relate to those recipes actually consumed [11–13], such comparisons are a useful complement to more traditional dietary comparisons, such as [88, 89].

In terms of predicting popularity, recipe innovation was the most important factor in Allrecipes.com with more innovative recipes tending to be more popular overall. Interestingly, opposite trend was true for Kochbar.de. The most important predictive features for Kochbar.de related to user behaviours. A strong signal common to both platforms, however, comes from attributes of the uploaded images of recipes. This aligns with the findings of past work from the food recommendation literature [68] and indeed the psychology literature, which also emphasizes the visual nature of food choice [90, 91].

Our findings suggest that further work should be done on the visual features of food images building on the efforts started by Yang and colleagues [92] and Salvador and colleagues [93]. Other future work would be to investigate the usage of features, which assist users improve their content and its presentation based on models, such as those which we have presented. Past work has shown how subtle changes in presentation can influence how online recipes are perceived [49]. It would be interesting to know how suggestions would be used and how this would influence interaction dynamics with online food portals generally.

7 Conclusions

In this work we have tackled the popularity prediction problem in the context of online recipes, investigating two large scale data collections from a service based in the United States (Allrecipes.com) and Germany (Kochbar.de). We observed differences between the platforms in terms of how the recipes are interacted with and categorized, as well as in the content of the food. However, for both datasets, we were able to show correlations between recipe features and proxies for popularity, which allowed popularity of dishes to be predicted to a surprisingly high degree. The trends were more prominent in the Kochbar.de dataset, which was mirrored in the results of the prediction task experiments.

List of Abbreviations
Not applicable.

Availability of data and materials
The aggregated datasets with calculated features and popularity proxies used in this work can be made available by request. The original data analyzed was obtained using the sitemap files available in the robots.txt file, in line with the terms and conditions of the services at the time of crawling. We cannot share the data in this form as it would mean breaking the terms of service.

Ethics approval and consent to participate
Does not apply as we rely on public available data.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Authors' contributions
Conceived and designed the experiments: CT. Performed the experiments: CT, DM. Acquired the data: CT, DM. Analyzed the data: CT, DM, DE. Wrote the paper: CT, DE, DM. All authors read and approved the final manuscript.

Author details
[1]Department of Information Science and Media Studies, University of Bergen, Bergen, Norway. [2]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria. [3]Chair of Information Science, University of Regensburg, Regensburg, Germany.

Endnotes
[a] http://meedia.de/2016/03/17/die-neuesten-agof-zahlen-rekorde-fuer-chefkoch-kochbar-media-markt-und-saturn/
[b] http://www.ebizmba.com/articles/recipe-websites
[c] https://www.alexa.com/siteinfo/allrecipes.com
[d] http://dish.allrecipes.com/customer-service/nutrition-information/
[e] Allrecipes.com: https://www.allrecipes.com/recipes/
[f] https://www.blsdb.de/
[g] https://www.kochbar.de/kochen/
[h] https://spoonacular.com/food-api
[i] We also tested other options such as the ingredient parser from the NYTimes https://github.com/NYTimes/ingredient-phrase-tagger which provided poorer results.
[j] http://www.openimaj.org
[k] http://www.southampton.ac.uk/
[l] https://people.richland.edu/james/lecture/m170/ch03-var.html
[m] http://openimaj.org/apidocs/org/openimaj/image/feature/global/AvgBrightness.html
[n] http://sentistrength.wlv.ac.uk/
[o] http://www.statsmodels.org/stable/generated/statsmodels.nonparametric.kde.KDEUnivariate.html
[p] http://www.statsmodels.org/stable/generated/statsmodels.nonparametric.kde.KDEUnivariate.evaluate.html#statsmodels.nonparametric.kde.KDEUnivariate.evaluate
[q] https://www.rdocumentation.org/packages/Hmisc/versions/4.0-3
[r] https://www.rdocumentation.org/packages/caret/
[s] https://cran.r-project.org/web/packages/FSelector/FSelector.pdf
[t] Percentages are computed by counting the number of recipes in a category divided by the number total recipes in the recipe platform.
[u] This is still true if the distribution of recipes uploaded to Allrecipes.com is examined over the same period, for which Kochbar.de recipes were available.
[v] Note: The lower the recipe IDF factor the higher the innovation.
[w] https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Bao P, Shen H-W, Huang J, Cheng X-Q (2013) Popularity prediction in microblogging network: a case study on sina weibo. In: Proceedings of the 22nd international conference on world wide web. ACM, New York, pp 177–178

2. Tatar A, Antoniadis P, De Amorim MD, Fdida S (2014) From popularity prediction to ranking online news. Soc Netw Anal Min 4(1):174
3. Pinto H, Almeida JM, Gonçalves MA (2013) Using early view patterns to predict the popularity of YouTube videos. In: Proceedings of the sixth ACM international conference on web search and data mining. ACM, New York, pp 365–374
4. Figueiredo F (2013) On the prediction of popularity of trends and hits for user generated videos. In: Proceedings of the sixth ACM international conference on web search and data mining. ACM, New York, pp 741–746
5. Arapakis I, Lalmas M, Cambazoglu BB, Marcos M-C, Jose JM (2014) User engagement in online news: under the scope of sentiment, interest, affect, and gaze. J Assoc Inf Sci Technol 65(10):1988–2005
6. Steck H (2011) Item popularity and recommendation accuracy. In: Proceedings of the fifth ACM conference on recommender systems. ACM, New York, pp 125–132
7. Figueiredo F, Benevenuto F, Almeida JM (2011) The tube over time: characterizing popularity growth of YouTube videos. In: Proceedings of the fourth ACM international conference on web search and data mining. ACM, New York, pp 745–754
8. Lika B, Kolomvatsos K, Hadjiefthymiades S (2014) Facing the cold start problem in recommender systems. Expert Syst Appl 41(4):2065–2073
9. Chen X, Zhang X (2003) A popularity-based prediction model for web prefetching. Computer 36(3):63–70
10. Famaey J, Iterbeke F, Wauters T, De Turck F (2013) Towards a predictive cache replacement strategy for multimedia content. J Netw Comput Appl 36(1):219–227
11. West R, White RW, Horvitz E From cookies to cooks: insights on dietary patterns via analysis of web usage logs. In: Proc. of WWW'13, pp 1399–1410. International World Wide Web Conferences Steering Committee
12. Abbar S, Mejova Y, Weber I You tweet what you eat: studying food consumption through Twitter. In: Proc. of CHI'15
13. Trattner C, Parra D, Elsweiler D (2017) Monitoring obesity prevalence in the United States through bookmarking activities in online food portals. PLoS ONE 12(6):e0179144
14. Allrecipe.com Press report (2016) Available at http://press.allrecipes.com/. Accessed 20 June 2016
15. Tatar A, de Amorim MD, Fdida S Antoniadis P (2014) A survey on predicting the popularity of web content. J Internet Serv Appl 5(1):8
16. Simon HA (1971) Designing organizations for an information-rich world
17. Bawden D, Robinson L (2009) The dark side of information: overload, anxiety and other paradoxes and pathologies. J Inf Sci 35(2):180–191
18. Schwarz J, Morris M (2011) Augmenting web pages and search results to support credibility assessment. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 1245–1254
19. White R (2013) Beliefs and biases in web search. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. ACM, New York, pp 3–12
20. Cunha CR, Bestavros A, Crovella ME (1995) Characteristics of WWW client-based traces. Technical report, Boston University Computer Science Department
21. Breslau L, Cao P, Fan L, Phillips G, Shenker S (1999) Web caching and Zipf-like distributions: evidence and implications. In: INFOCOM'99. Eighteenth annual joint conference of the IEEE computer and communications societies. Proceedings. IEEE, vol 1. IEEE, pp 126–134
22. Cherkasova L, Gupta M (2004) Analysis of enterprise media server workloads: access patterns, locality, content evolution, and rates of change. IEEE/ACM Trans Netw 12(5):781–794
23. Gill P, Arlitt M, Li Z, Mahanti A (2007) YouTube traffic characterization: a view from the edge. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, New York, pp 15–28
24. Chesire M, Wolman A, Voelker GM, Levy HM (2001) Measurement and analysis of a streaming media workload. In: USITS, vol 1, pp 1
25. Cha M, Kwak H, Rodriguez P, Ahn Y-Y, Moon S (2009) Analyzing the video popularity characteristics of large-scale user generated content systems. IEEE/ACM Trans Netw 17(5):1357–1370
26. Bernstein MS, Bakshy E, Burke M, Karrer B (2013) Quantifying the invisible audience in social networks. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 21–30
27. Agarwal D, Chen B-C, Wang X (2012) Multi-faceted ranking of news articles using post-read actions. In: Proceedings of the 21st ACM international conference on information and knowledge management. ACM, New York, pp 694–703
28. Meier F, Elsweiler D, Wilson ML (2014) More than liking and bookmarking? Towards understanding Twitter favouriting behaviour. In: ICWSM
29. Schaller R, Harvey M, Elsweiler D (2013) RecSys for distributed events: investigating the influence of recommendations on visitor plans. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. ACM, New York, pp 953–956
30. Shulman B, Sharma A, Cosley D (2016) Predictability of popularity: gaps between prediction and understanding. In: ICWSM, pp 348–357
31. Castillo C, El-Haddad M, Pfeffer J, Stempeck M (2014) Characterizing the life cycle of online news stories using social media reactions. In: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing. ACM, New York, pp 211–223
32. Chatzopoulou G, Sheng C, Faloutsos M (2010) A first step towards understanding popularity in YouTube. In: INFOCOM IEEE conference on computer communications workshops, 2010. IEEE, pp 1–6
33. Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J (2014) Can cascades be predicted? In: Proceedings of the 23rd international conference on world wide web. ACM, New York, pp 925–936
34. Lerman K, Hogg T (2010) Using a model of social dynamics to predict popularity of news. In: Proceedings of the 19th international conference on world wide web. ACM, New York, pp 621–630
35. Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J (2015) Seismic: a self-exciting point process model for predicting tweet popularity. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1513–1522
36. Tsur O, Rappoport A (2012) What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In: Proceedings of the fifth ACM international conference on web search and data mining. ACM, New York, pp 643–652

37. Yu L, Cui P, Wang F, Song C, Yang S (2015) From micro to macro: uncovering and predicting information cascading process with behavioral dynamics. In: Data mining (ICDM), 2015 IEEE international conference on. IEEE, pp 559–568

38. Romero DM, Tan C, Ugander J (2013) On the interplay between social and topical structure. In: ICWSM

39. Kusmierczyk T, Trattner C, Nørvåg K Temporality in online food recipe consumption and production. In: Proc. of WWW'15

40. Trattner C, Kjetil Nørvåg TK (2016) FOODWEB—studying food consumption and production patterns on the web. ERCIM News 104:50

41. Kusmierczyk T, Trattner C Nørvåg K Temporal patterns in online food innovation. In: Proc. of WWW'15 Companion

42. Wagner C, Singer P, Strohmaier M (2014) The nature and evolution of online food preferences. EPJ Data Sci 3(1):1

43. West R, White RW, From HE Cookies to cooks: insights on dietary patterns via analysis of web usage logs. In: Proc. of WWW'13, pp 1399–1410

44. Said A Bellogín A You are what you eat! Tracking health through recipe interactions. In: Proc. of RSWeb'14

45. De Choudhury M, Sharma SS Characterizing dietary choices, nutrition, and language in food deserts via social media. In: Proc. of CSCW '16

46. Rokicki M, Herder E, Demidova E What's on my plate: towards recommending recipe variations for diabetes patients. In: Proc of UMAP'15 LBRS

47. Freyne J, Berkovsky S (2010) Recommending food: reasoning on recipes and ingredients. In: User modeling, adaptation, and personalization, pp 381–386

48. Harvey M, Ludwig B, Elsweiler D (2013) You are what you eat: learning user tastes for rating prediction. In: International symposium on string processing and information retrieval. Springer, Berlin, pp 153–164

49. Elsweiler D, Trattner C, Harvey M (2017) Exploiting food choice biases for healthier recipe recommendation

50. Trattner C, Elsweiler D (2017) Food recommender systems: important contributions, challenges and future research directions. arXiv preprint. arXiv:1711.02760

51. Harvey M, Ludwig B, Elsweiler D Learning user tastes: a first step to generating healthy meal plans? In: Proc. of LIFESTYLE'12, p 18

52. Wansink B, Cashman M (2006). Mindless eating

53. Bellisle F (2005) The determinants of food choice. EUFIC Review 17:1–8

54. Steptoe A, Pollard TM, Wardle J (1995) Development of a measure of the motives underlying the selection of food: the food choice questionnaire. Appetite 25(3):267–284

55. Biloukha OO, Utermohlen V (2000) Correlates of food consumption and perceptions of foods in an educated urban population in Ukraine. Food Qual Prefer 11(6):475–485

56. Zandstra EH, De Graaf C, Van Staveren WA (2001) Influence of health and taste attitudes on consumption of low-and high-fat foods. Food Qual Prefer 12(1):75–82

57. Glanz K, Basil M, Maibach E, Goldberg J, Snyder DAN (1998) Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. J Am Diet Assoc 98(10):1118–1126

58. Prescott J, Young O, O'neill L, Yau NJN, Stevens R (2002) Motives for food choice: a comparison of consumers from Japan, Taiwan, Malaysia and New Zealand. Food Qual Prefer 13(7):489–495

59. Rozin P, Zellner D (1985) The role of Pavlovian conditioning in the acquisition of food likes and dislikes. Ann NY Acad Sci 443(1):189–202

60. Stafleu A, de Graaf C, van Staveren WA, Schroots JJ (1991) A review of selected studies assessing social-psychological determinants of fat and cholesterol intake. Food Qual Prefer 3(4):183–200

61. Macht M (2008) How emotions affect eating: a five-way model. Appetite 50(1):1–11

62. Oliver G, Wardle J, Gibson EL (2000) Stress and food choice: a laboratory study. Psychosom Med 62:853–865

63. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. N Engl J Med 357(4):370–379

64. Trattner C, Elsweiler D (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, pp 489–498. International World Wide Web Conferences Steering Committee

65. Schroeder J (2017) AGOF: Chefkoch erstmals mit mehr als 20 Mio. Nutzern. Rekorde auch für kochbar, essen und trinken und Lecker. http://meedia.de/2017/03/09/agof-chefkoch-erstmals-mit-mehr-als-20-mio-nutzern-rekorde-auch-fuer-kochbar-essen-trinken-und-lecker/

66. Howard S, Adams J, White M (2012) Nutritional content of supermarket ready meals and recipes by television chefs in the United Kingdom: cross sectional study. BMJ, Br Med J 345:e7607

67. San Pedro J, Siersdorfer S (2009) Ranking and classifying attractiveness of photos in folksonomies. In: Proceedings of the 18th international conference on world wide web. WWW '09. ACM, New York, pp 771–780

68. Elsweiler D, Trattner C, Harvey M (2017) Exploiting food choice biases for healthier recipe recommendation. In: Proceedings of ACM SIGIR conference, SIGIR'17, Tokyo, Japan, p 11

69. International Commission on Illumination: 17–251 contrast colourfulness. http://eilv.cie.co.at/term/251. Accessed 26 May 2017

70. International Commission on Illumination: 17–1136 saturation. http://eilv.cie.co.at/term/1136. Accessed 26 May 2017

71. International Commission on Illumination: 17–233 colourfulness. http://eilv.cie.co.at/term/233. Accessed 26 May 2017

72. Cornell University: image entropy. http://www.astro.cornell.edu/research/projects/compression/entropy.html

73. Huang K-Q, Wang Q, Wu Z-Y (2006) Natural color image enhancement and evaluation algorithm based on human visual system. Comput Vis Image Underst 103(1):52–63

74. Pitler E, Nenkova A (2008) Revisiting readability: a unified framework for predicting text quality. In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 186–195

75. Anderson J (1981) Analysing the readability of English and hon-English texts in the classroom with lix

76. Kucuktunc O, Cambazoglu BB, Weber I, Ferhatosmanoglu H (2012) A large-scale sentiment analysis for Yahoo! answers. In: Proceedings of the fifth ACM international conference on web search and data mining. ACM, New York, pp 633–642

77. Parra D, Trattner C, Gómez D, Hurtado M, Wen X, Lin Y-R (2016) Twitter in academic events: a study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences. Comput Commun 73:301–314

78. Kusmierczyk T, Trattner C, Nørvåg K (2016) Understanding and predicting online food recipe production patterns. In: Proceedings of the 27th ACM conference on hypertext and social media. ACM, New York, pp 243–248

79. Kerne A, Webb AM, Smith SM, Linder R, Lupfer N, Qu Y, Moeller J, Damaraju S (2014) Using metrics of curation to evaluate information-based ideation. ACM Trans Comput-Hum Interact 21(3):14

80. Sullivan GM, Feinn R (2012) Using effect size—or why the p value is not enough. J Grad Med Educ 4(3):279–282

81. Pallant J (2007) SPSS survival manual: a step-by-step guide to data analysis using SPSS version 15. McGraw-Hill, Nova Iorque

82. Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Earlbaum Associates, Hilsdale

83. Hofman JM, Sharma A, Watts DJ (2017) Prediction and explanation in social systems. Science 355(6324):486–488

84. Fashion for gluten- and dairy-free continues to grow. http://www.bbc.com/news/business-39488047. Accessed 21 December 2017

85. Grain drain: should everyone adopt a gluten-free diet? https://www.theguardian.com/lifeandstyle/2017/aug/24/grain-drain-should-everyone-adopt-a-gluten-free-diet. Accessed 21 December 2017

86. Asur S, Huberman BA, Szabo G, Wang C (2011) Trends in social media: persistence and decay. In: ICWSM

87. Teevan J, Alvarado C, Ackerman MS, Karger DR (2004) The perfect search engine is not enough: a study of orienteering behavior in directed search. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 415–422

88. Keys A, Mienotti A, Karvonen MJ, Aravanis C, Blackburn H, Buzina R, Djordjevic BS, Dontas AS, Fidanza F, Keys MH et al (1986) The diet and 15-year death rate in the seven countries study. Am J Epidemiol 124(6):903–915

89. Janssen I, Katzmarzyk PT, Boyce WF, Vereecken C, Mulvihill C, Roberts C, Currie C, Pickett W (2005) Comparison of overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns. Obes Rev 6(2):123–132

90. Krajbich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci 13(10):1292–1298

91. Schur EA, Kleinhans NM, Goldberg J, Buchwald D, Schwartz MW, Maravilla K (2009) Activation in brain energy regulation and reward centers by food cues varies with choice of visual stimulus. Int J Obes 33(6):653–661

92. Yang L, Hsieh C-K, Yang H, Pollak JP, Dell N, Belongie S, Cole C, Estrin D (2017) Yum-me: a personalized nutrient-based meal recommender system. ACM Trans Inf Syst 36(1):7

93. Salvador A, Hynes N, Aytar Y, Marin J, Ofli F, Weber I, Torralba A (2017) Learning cross-modal embeddings for cooking recipes and food images. In: Proc. of CVPR'17