Analysing the Usage of Wikipedia on Twitter: Understanding Inter-Language Links

Eva Zangerle
University of Innsbruck
eva.zangerle@uibk.ac.at

Georg Schmidhammer
University of Innsbruck
georg.schmidhammer@uibk.ac.at

Günther Specht
University of Innsbruck
guenther.specht@uibk.ac.at

Abstract

Wikipedia is a central source of information as 450 million people consult the online encyclopaedia every month to satisfy their information needs. Some of these users also refer to Wikipedia within their tweets. In this paper, we analyse links within tweets referring to a Wikipedia of a language different from the tweet's language. Therefore, we investigate causes for the usage of such inter-language links by comparing the tweeted article and its counterpart in the tweet's language (if there is any) in terms of article quality. We find that the main cause for inter-language links is the non-existence of the article in the tweet's language. Furthermore, we observe that the quality of the tweeted articles is constantly higher in comparison to their counterparts, suggesting that users choose the article of higher quality even when tweeting in another language. Moreover, we find that English is the most dominant target for inter-language links.

1. Introduction

The Wikipedia online encyclopaedia has shaped the way users seek to answer their information needs over the last decade and has become a central source of information for millions of users. Currently, Wikipedia is visited by 450 million users every month [39], making it the 6th most visited page on the web [2]. The Wikipedia project has shown that the collaborative efforts of a committed community can lead to a central platform, which allows users stemming from various backgrounds to actively contribute, communicate and engage [14].

The scientific community has intensively studied Wikipedia, its community, content and interactions between its members over the last decade. These analyses have been carried out from an internal point of view, concentrating on intrinsic factors influencing e.g., the quality of articles or the extent of Wikipedia and its 277 actively maintained languages [4]. Research on such intrinsic factors focuses mainly on the committed community as the driving force behind Wikipedia. The structure of the Wikipedia community and its effect on the information provided within articles is studied in e.g., [17], [18]. Furthermore, the quality and maturity of Wikipedia articles has also been studied and evaluated in e.g., [44], [16], [43]. An active community and high-quality articles are both crucial intrinsic factors for the popularity of Wikipedia. In contrast, extrinsic factors have hardly been studied yet. One of these extrinsic factors is how Wikipedia is referenced within other social media platforms. In previous work, we performed a first study on the usage of Wikipedia URLs on the microblogging platform Twitter [45]. In particular, we looked into inter-language links, i.e., links embedded in a tweet composed in language x and referring to a Wikipedia of different language y. This analyses showed that tweets refer to a Wikipedia other than the user's language with a probability of 20%—except for English and Japanese users. Following up on this research, we aim to deepen the understanding for inter-language links and investigate the underlying Wikipedia articles and tweets in detail. In particular, we are interested in the reasons why such inter-language links are facilitated by users. Therefore, we gather a dataset comprising more than 6 million tweets and perform a quantitative analysis of inter-language links regarding quality characteristics of the Wikipedia articles. We apply a set of 12 quality measures to the articles to be able to compare the quality of the article that was actually tweeted (e.g., Hawaiian_Islands on the English Wikipedia) and its counterpart in the tweet's original language (e.g., Archipiélago_de_Hawái on the Spanish Wikipedia).

In this paper, we aim to answer the following research questions (hereafter referred to as RQ):

- **RQ1:** How are inter-language links distributed among the different Wikipedias?
- **RQ2:** What are the causes for users to link to a Wikipedia other than the one of their language?

The main contribution of this paper is that it provides a detailed study on the usage and causes of inter-language links on Twitter. We find that English is the most dominant target for inter-language links, reaching a share of 62.68%. Our analyses show that for 84.92% of the articles referred to in tweets containing inter-language links, there is no counterpart in the tweet's original language. For the remaining 15.08% we find that the quality of the tweeted article is constantly higher than the quality of its counterparts, suggesting that users choose articles of higher quality even when tweeting in another language.

The remainder of this paper is structured as follows. Section 2 describes the background of our work and related work. Section 3 presents the dataset underlying the analyses. Section 4 describes the methods used for analysing the data. Section 5 presents our findings in detail and Section 6 discusses these findings in the light of the posed research questions. Section 7 concludes the paper and presents our future research plans.

2. Background and Related Work

Twitter and Wikipedia have been popular topics among researchers within the last decade. Aspects studied regarding the Twitter platform include the diffusion of information on Twitter [31], the credibility of information on Twitter [6] or the social structures behind the user network [3]. Furthermore, data retrieved from Twitter is facilitated for a multitude of applications. These range from real-time detection of events such as earthquakes [32] to the recommendation of news [1]. Similarly, Wikipedia has been subject to substantial investigations over the course of the last decade [12], [25]. The social interaction of users and editors of Wikipedia has also been examined [7], [22]. Previous research has also investigated quality measures for Wikipedia articles, which is of particular interest for our analyses. Warncke-Wang et al. classify previous research on this matter into editor-based assessment and article-based assessment approaches [38]. As for the approaches focussing on the features of articles, Dalip et al. present a study on a substantial set of textual, structural and editorial features of articles (including length, structure, style and readability, revision history and the social network of editors) [8]. The authors conclude that textual features lead to the best results in terms of their predictive power regarding article quality. Similarly, Blumenstock as well as Wu et al. have shown that the number of words used within an article serves as a good indicator for quality [4, 44]. As for the editor-based assessment of the quality of Wikipedia articles, [16] has shown that effective coordination between editors leads to higher-quality articles on Wikipedia. Similarly, Liu and Ram have shown that coordination between editors and in particular, the specific roles of users ("who does what?") in the process of editing an article influence the quality of articles [22]. Wilkinson and Huberman found that the number of distinct editors highly correlates with the quality of articles [43].

When it comes to research that includes both Wikipedia and Twitter, this research is mostly focused on how to exploit data gathered from Wikipedia to perform and improve research on Twitter. These approaches range from the classification of data to modelling user preferences. As for the classification of data, Parker et al. facilitate Twitter and Wikipedia to detect medical trends and issues. In particular, the authors map tweets to Wikipedia articles covering the same topics and subsequently match the introduction of the respective Wikipedia articles to medical dictionaries aiming to find tweets about medical issues [28]. Similarly, Genc et al. map tweets to a Wikipedia article covering the same topics to classify tweets. The authors compute the distance between these articles as the distance of the categories of the respective article and use this information as a similarity measure for tweets [11]. The Wikipedia category system is also exploited by Lim et al., who present an approach for Twitter user classification based on the celebrities that are followed by users. The authors make use of Wikipedia by extracting information about a celebrity's occupation and use this information to infer user interests [21]. Similarly, user interest classification based on the Wikipedia category system (and its hierarchy) has also been investigated in [13, 23]. Michelson and Macskassy also perform user interest classification by firstly detecting the topic of a tweet by applying Named Entity Recognition. In a second step, the authors match these entities against Wikipedia's category system to infer user interests, which are subsequently used for user classification.

For Named Entity Recognition (NER) approaches, Wikipedia data is frequently exploited. In particular, Wikipedia's disambiguation pages are a valuable source of information for resolving entities. This data is e.g., also used to perform NER on tweets. In particular, combining disambiguation pages with page titles and redirects extracted from Wikipedia allows for detecting named entities within tweets [20]. This data can also be used to compute semantic relatedness of topics and entities [10]. Page views for articles on Wikipedia are also exploited as e.g., Osborne et al. make use of the number of page views of articles and use these as an indicator for events. Generally, the authors aim to perform first story detection for tweets and use Wikipedia data to verify potential events [26].

We were the first to provide an analysis on the usage of Wikipedia links on Twitter and hence, the first study on extrinsic factors on Wikipedia [45]. We provide a quantitative study on three aspects of the usage of Wikipedia links within tweets and firstly find that besides English and Japanese, more than 20% of all links within a tweet referring to Wikipedia lead to a Wikipedia of another language. Secondly, we find that there are no particular categories or topical features of articles that are significantly more popular on Twitter than others. Thirdly, we provide an analysis on the correlation of the number of tweets about articles and a recent update/edit of this article. We conclude that there is no correlation between these factors. This study follows along these lines, however, we focus on the causes for inter-language links, which have not been studied previously. Furthermore, the dataset does not include bots that add a certain bias to the analyses.

3. Data

In the following, we present the crawling methods used for gathering the data underlying our study. Subsequently, we describe the steps taken to clean the dataset and finally, we provide the most important facts regarding the dataset.

3.1. Crawling Method

The basis for the analyses conducted in this paper is a dataset containing tweets about Wikipedia and more importantly, containing a link to a particular Wikipedia article. In the following section, we present the crawling methods facilitated for gathering such a dataset. Generally, we rely on the public Twitter Streaming API, which allows for retrieving tweets containing given filter keywords and associated metadata as JSON-objects [36]. As for the keywords used, we filter for tweets containing the term "wikipedia". In total, we were able to gather 6,415,762 tweets fulfilling the filter criterion between 2014/10/20 and 2015/04/28. Twitter restrains the amount of data that can be collected freely over its API. The number of delivered tweets matching the given keywords is capped by a rate limiting equal to the rate limiting of the public Streaming API (approximately 1% of all tweets) [37]. However, this rate limiting does not affect our crawling process as the number of tweets matching our query constantly is well below this limit (maximum number of tweets crawled per day: 63,795). Hence, we were able to crawl all tweets matching the given filter keywords during the given time period.

3.2. Dataset Cleaning

Performing the previously described crawling method naturally leads to the inclusion of false positives, i.e., tweets that contain the term "wikipedia" but are not particularly concerned with Wikipedia and hence, should be excluded from the dataset underlying our analysis. This cleaning operation is performed based on the JSON-representation of each tweet (as delivered by the Twitter API), which also provides information about the URLs contained in the tweets. In particular, it provides the fully expanded URL of possibly shortened URLs. Based on this information, we filter for all tweets containing a link to a Wikipedia by applying regular expressions.

A manual exploration of the tweets of the most active users in the dataset revealed that the distribution of the total number of tweets on a per-user-basis is dominated by bots, which are applications aiming to mimic human users to distribute spam on Twitter, spread information or to influence opinions of targets. A similar behaviour in regards to bots has also been detected in [45]. As the goal of this study is to analyse inter-language links of human users, we exclude bots from our dataset.

To detect and exclude bots from the dataset, we apply the following cleaning method to the dataset. Firstly, we compute the total number of tweets for each user. This distribution is heavy-tailed (mean number of tweet per user = 3.09, median = 1, 50^{th} percentile = 1). Previous research on bots on Twitter found that bots tweet and retweet more than human users [7, 9]. Therefore, we consider all outliers in the distribution of the number of tweets per user as bot candidates as these users tweet substantially more than other users. However, due to the heavy-tailed nature of the underlying distribution, we cannot detect outliers using traditional methods as e.g., SD around the mean, which is not robust for non-normal distributions [19]. In accordance with previous research dealing with outliers in heavy-tailed distributions [30], we compute all users within the 99^{th} percentile (i.e., all users with more than 130 tweets). This provides us with a set of 1,083 potential bot candidates. Subsequently, we make use of the BotOrNot bot detection service, which provides a web platform for bot detection [5]. The BotOrNot platform implements the bot detection mechanisms proposed by Ferrara et al. and has shown detection accuracy of 95% [9]. Using this platform allows for collecting a bot likelihood score for each given user. We gather this score for all bot candidates. Based on this score we consider all users with a likelihood score higher than 50% as a bot. I.e., all accounts for which the probability of being a bot is higher than the probability of being a human user are regarded as bots. We consider a likelihood threshold score of 50% as a conservative approach for the detection of bots within the dataset, which could potentially also lead to false positives (i.e., normal users that exhibit features that lead to a wrong classification result in regards to its bot likelihood). However, we argue that adopting a reasonably conservative and yet robust approach for outlier detection enables us to rule out bot-based bias within our data. As the BotOrNot service gathers all information relevant for the detection directly from the candidate's twitter account, the detection fails if the account does not exist anymore or the amount of data gathered is not sufficient. We are not able to detect the likelihood score for 274 accounts. For those users, who we could not directly identify as bots or humans, we again facilitate a conservative approach and therefore, exclude these users and their tweets to make sure not to include potential bots.

In a final cleaning step we exclude all accounts considered as bots and their tweets from our dataset. In total, we exclude a total of 404 bots and 264,897 tweets sent by these accounts to compute our final dataset, which is presented in the next section.

3.3. Dataset

We now give a brief description of the dataset underlying our analyses. Table 1 features a first summary of the dataset. The "Raw" column refers to the raw and uncleaned data gathered by applying the data collection method as described in Section 3.1, whereas the "Cleaned" column refers to the dataset after having performed the cleaning operations and bot removal as described in Section 3.2.

Table 1: Dataset overview

Feature	Raw	Cleaned
Tweets	6,415,762	2,844,399
Retweets	2,040,816	855,959
Distinct Users	2,287,430	1,092,732
Mentions	4,673,284	2,437,092
Distinct Hashtags	213,574	127,958
Hashtag Usages	2,283,535	788,210
Distinct URLs	1,976,479	1,179,288
URL Usages	4,825,230	3,130,420

4. Methods

In this section, we describe the methods facilitated to tackle the research questions described in Section 1. Generally, we firstly extract all inter-language links from the presented dataset. Based on this data, we perform the data analysis regarding inter-language links as presented in the following section. Figure 1 depicts an overview of the steps taken in the analysis process.

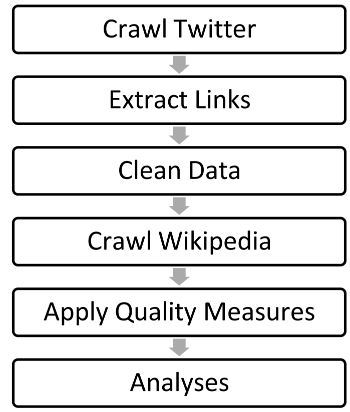

Figure 1: Analysis workflow

4.1. Extraction of Inter-Language Links

For a first analysis, we aim to extract the languages used by Twitter users and the Wikipedias that are referenced within the tweets contained in our dataset. Therefore, we firstly obtain the language of the tweet and secondly, extract the Wikipedia each tweet links to. For the analysis of the languages facilitated within the tweets, we rely on each tweet's metadata provided by the Twitter API. Based on a tweet's JSON representation gathered over the API, we extract the language of the tweet itself, which is computed by Twitter's language detection algorithm and subsequently provided within the tweet's JSON representation. We refer to this language as the "tweet"-language. This JSON-object provides the language of the tweet in the form of BCP 47 codes [29]. This information about the tweet language allows us to find so-called inter-language links, i.e., tweets for which the Twitter-language differs from the language used within the Wikipedia mentioned in the tweet. Therefore, we also obtain the language of this Wikipedia. The extraction of the Wikipedia edition based on a given URL extracted from a tweet is done by applying regular expressions. We refer to inter-language links as x/y where x is the language of the tweet and y is the language of the Wikipedia edition linked to. I.e., en/de refers to an English tweet referring to the German Wikipedia. To compute the distribution among these inter-language links, we group tweets according to their language and the Wikipedia they are linking to and add up the number of tweets of language x linking to Wikipedia y.

4.2. Quantitative Analysis of Wikipedia Inter-Language Links

In a second step, we aim to get a closer look at the causes of inter-language links. Our expectation is that people link to the Wikipedia of another language if the target article (or topic) they want to post on Twitter is either not available at all in their own language or is of lower quality. To investigate this matter, we firstly look into whether tweeted articles have a counterpart in the tweet's language (e.g., the Wikipedia article "Archipiélago_de_Hawái" is the Spanish counterpart to the English Wikipedia article "Hawaiian_Islands"). Secondly, we focus on those articles which do have a counterpart in the tweet's language and aim to find differences between the two Wikipedia articles in regards to their quality. Therefore, we look into previous research on the quality of Wikipedia articles to extract methods for detecting and assessing the quality of Wikipedia articles, which we later adapt to compare the quality of articles.

In principle, the Wikipedia community employs quality assessment for articles. This assessment is done manually by the community, i.e., the Wikipedia editors who categorize articles into seven quality assessment classes [40]. On the English Wikipedia, articles are assigned to classes based on a set of predefined criteria (cf. [41]). While this community-based classification of an article's quality serves a good indicator for the quality of articles of a single Wikipedia, Stvilia et al. found that the notion of quality of articles varies significantly among cross-contextual communities (in terms of cultural, social and economic aspects) as formed by the different language editions on Wikipedia [34].

To be able to compare the quality of articles stemming from different Wikipedias, we require a set of robust measures that are comparable across various Wikipedias. Therefore, we do not rely on Wikipedia's quality assessment classes as an indicator for the quality of articles among the different Wikipedia editions. Instead, we rely on the well-established measures developed by Stvilia [34] that have been further extended by Warncke-Wang et al. [38]. Warncke-Wang et al. looked into finding so-called actionable measures, i.e., measures that are able to directly indicate certain flaws within articles to be able to correct these and improve the overall quality of the article. The authors list the different features ordered by their predictive power in terms of classifying articles correctly into correct quality classes. We rely on this list of features for assessing the quality of articles. However, we constrain the set of measures to those directly related to the article (i.e., we exclude the measures Tenure, Completeness, Consistency, Authority/Reputation and Volatility). This is due to the fact that some of these measures require crawling the edit history and all metadata of all editors having contributed to any of the articles in our study (the study by Warncke-Wang featured a set of 4,454 articles while our study features a set of 199,552 articles). As the study by Dalip et al. suggests, article-based features serve as a good indicator for article quality [8]. Therefore, we restrain the evaluation of editor-based features to using the number of distinct editors (referred to as "Diversity" by Warncke-Wang et al.) having contributed to an article as proposed by Wilkinson and Huberman [43]. Hence, we propose to employ the following twelve measures for the quality of articles that we subsequently describe shortly (for a more detailed description please refer to [35, 38]):

1. Number of references/article length
2. Number of references
3. Article length
4. Diversity
5. Number of headings
6. Number of headings/article length
7. Informativeness
8. Number of images/article length
9. Number of wikilinks/article length
10. Currency
11. HasInfoBox
12. Complexity

Four of the introduced measures are computed relative to the length of the article to reflect that e.g., shorter articles naturally have less headings, references, etc. Generally, the length of articles is characterized as the number of words contained in the article (not including Wiki markup, etc.). The number of references is a measure for the number of citations, i.e., sources verifying the content of the article. The diversity of an article is the number of distinct editors of an article. The number of headings is considered as a notion of how structured the article is. The Informativeness measure describes how informative the content of an article is. It can be computed as Informativeness = 0*6 InfoNoise + 0.3 NumImages, where InfoNoise is the "proportion of text content remaining after removing MediaWiki code and stop words and stemming all words" [38] and NumImages is the number of images displayed on the article page. We adapt this notion of InfoNoise for our purpose to removing only MediaWiki code as the inter-language dataset features 41 different languages (62 languages when considering the whole dataset) and stemming and stop word removal is not feasible (for many languages, traditional stemming algorithms are not available [33]). We also incorporate the number of wikilinks, the number of links to other Wikipedia articles as an indicator for the profoundness of the article. From a temporal perspective, the currency of an article represents the up-to-dateness and is measured as the number of days passed since the last edit of the article. The "HasInfoBox" property determines whether a given article provides an infobox, a tabular summary of the most important facts given within an article. The complexity measure aims at detecting how difficult a given article is to understand and is computed by the Flesch-Kincaid Readability score [15].

For the extraction of the information listed above, we facilitate the MediaWiki-API [24]. To be able to analyse the state of the Wikipedias at the time a user actually sent a tweet, it is important to gather the required information for the date and time the tweet was actually sent. The MediaWiki API allows for specifying a so-called "starting timestamp" (rvstart), for which all edits of a given Wikipedia article are returned that occurred prior to the specified timestamp. For each article mentioned within our dataset, we retrieve all metadata (including links to all other language versions of the article to check their existence), the content of the article at that time and the last 500 edits of this particular article prior to the date and time the tweet was sent. Based on this data, we extract the measures listed above.

5. Results

In the following, we firstly present the findings of the inter-language link analysis to answer RQ1. Secondly, we look into these inter-language links and analyse the quality of the tweeted articles and their counterparts to answer RQ2.

5.1. Inter-Language Links

In the following section, we present our findings of the inter-language analysis. In a first step, we aim to get an impression on the distribution of languages used within tweets and the Wikipedia editions these tweets link to. Therefore, we apply the analysis method for inter-language links as described in Section 4.1 to the cleaned dataset. The distribution of links from a tweet's language x to a Wikipedia language y is depicted in a chord diagram in Figure 2[1]. This diagram displays the distribution of links from tweet languages to Wikipedia languages. I.e., it displays the fraction of links leading to a Wikipedia of the same language and the fraction of links referring to a Wikipedia of another language (expressed by an arc between these two languages, whereby the diameter of the arc represents the number of links between these languages). The diagram shows that for English and Japanese tweets, the majority links to English and Japanese.

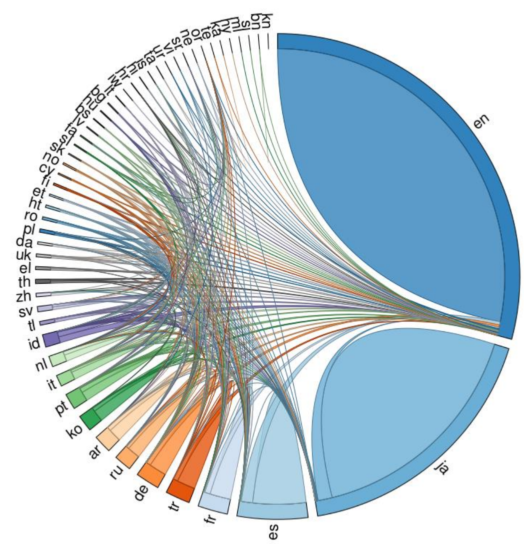

Figure 2: Chord diagram of inter-language links for tweet-languages with > 20 outgoing links

As for the distribution of inter-language links among the individual Wikipedias, we find that English is the predominant target of such links. In total, 691,424 tweets refer to a Wikipedia of a language different from the tweet's language. These results are in line with our previous findings [45]. In total, 62.68% of all links lead to an English Wikipedia page, followed by the Japanese Wikipedia (6.26%) and the Spanish Wikipedia (5.76%). Regarding RQ1, we therefore observe that English Wikipedia is the main target of inter-language links, whereas the share of all other Wikipedias remains substantially lower, which is also backed by the fact that the English Wikipedia is more comprehensive and maintained than other Wikipedias [42].

5.2. Quantitative Analysis of Causes

Based on the findings of the inter-language link analysis, we now aim to deepen our understanding for the causes of linking to the Wikipedia of another

[1] An animated version of this chord chart and a tabular presentation of the underlying data can be found online at http://ow.ly/RAoDS.

language. Therefore, we extract all inter-language links from the cleaned dataset. This leads to a set of 691,424 articles, which amounts to 23.66% of all articles within the dataset. As already carved out in the methods section, we firstly look into how many of the linked Wikipedia articles do actually have a counterpart in the Wikipedia of the tweet's language. We find that 84.92% of these articles do not have a counterpart in the tweet's original language at all. The remaining 99,776 articles (15.04%) do feature a counterpart in the tweet's language.

Based on these 99,776 articles featuring a counterpart in the tweet's language, we firstly evaluate the proposed quality measures and group the inter-language links by tweet language and Wikipedia language. Generally, the English, Spanish and German Wikipedias are the Wikipedias attracting most inter-language links in terms of links leading to this Wikipedia from a tweet of another language. The English Wikipedia's share of inter-language links (with an existent counterpart) is 84.04%, followed by the Spanish (6.60%) and German Wikipedias (3.38%).

When grouping the data by tweet language and Wikipedia language, we find that the highest number of inter-language links is facilitated for Spanish/English links (a link to the English Wikipedia embedded in a Spanish tweet), accounting for a total of 13.58% of all inter-language links. 12.27% of all inter-language links are performed for Japanese/English. The following inter-language link groups reach a share of more than 5% of all inter-language links: French/English (6.4%), Korean/English (6.19%), Italian/English (5.45%), English/Spanish (5.37%) and German/English (5.03%).

In a further analysis, we are interested in how the individual languages perform in terms of article quality in comparison to its counterparts. Therefore, we apply the following method for representing the performance of a given (directed) language combination (tweeted article language/counterpart language):

- We apply each quality measure to each article and its counterpart.
- For each article, we count for how many measures the article performs better than its counterpart. Similarly, we also compute this count for the counterpart article.
- Based on these aggregated values, we build two vectors; one for all aggregated values of the tweeted article and another vector for its counterpart, where each element in the vector represents the performance of a single article in the respective language.

This way, we obtain two vectors summarizing the results of all quality measures applied to all articles for a given language combination. Assume the following simplified example: we aim to analyse the behaviour of languages x and y. Our simplified dataset features three tweets of language x leading to an article of the Wikipedia of language y. Further assume we were to evaluate four different quality measures and x would perform better for all four of the measures for article 1, perform better in one of the measures for article 2 and in three measures for article 3. Hence, the resulting vector representing the aggregated performance of language x for the three given tweets would be $\vec{v}_t = \langle 4, 1, 3 \rangle$. Accordingly, the vector resulting for the counterpart articles in language y would be $\vec{v}_c = \langle 0, 3, 1 \rangle$. These vectors serve as our input for the analysis. Subsequently, we apply a Wilcoxon signed rank test to determine the statistical significance of the difference in the quality measures for all articles of each language combination. We restrain the set of language combinations to be evaluated to those featuring more than 30 inter-language links, resulting in 68 language combinations.

Table 2: Dominating languages (difference significant at p<0.05)**

Target	Better than**	Count
English	Spanish, Japanese, French, Korean, Italian, German, Arabic, Indonesian, Portuguese, Dutch, Turkish, Swedish, Thai, Polish, Romanian, Finnish, Danish, Norwegian, Farsi, Welsh, Hindi, Bulgarian, Latvian, Bosnian, Slovakish, Hungarian, Slovenian, Lithuanian, Bosnian	28
French	English, Japanese, Spanish	3
Spanish	English, Italian	2
Catalan	English, Portuguese	2
German	English	1
Japanese	German	1
Portuguese	Spanish	1
Turkish	English	1

We find that for 39 language combinations, the differences are statistically significant at the 0.05 level, i.e., languages for which the tweeted language performs significantly better than its counterpart language. Looking into the data, we observe that e.g., the language combinations English/Spanish as well as Spanish/English feature quality measures significantly different than the values of its counterparts (at the .05

level). Similarly, also English/German, German/English, French/English, English/French, Turkish/English and English/Turkish feature quality measures significantly higher than its counterparts. Table 2 presents a summary of these findings, displaying the languages users actually decided to tweet (target) and those counterpart languages these target languages perform significantly better (at the .05 level). We see that English performs significantly better for articles within the dataset in 28 languages. French performs better than English, Japanese and Spanish. Spanish and Catalan perform better than two other languages, notably English is among these.

Following up on these findings, we look into different languages and Wikipedias aiming to get a deeper understanding on how different languages perform for individual measures. Thus, we perform a Wilcoxon signed rank test to find whether there are statistically significant differences between the individual quality measures for the tweeted articles and their counterparts. Therefore, we again encode the results of each measure for each language combination into two vectors. These Boolean vectors contain a 1 entry if the respective article performed better than the article of the other language for the given quality measure, else the entry is 0. These vectors serve as input for the Wilcoxon test, which we perform for all language combinations with more than 30 inter-language links. The results show that for 65.79% (75 of 114) of all language combinations (tweeted language/Wikipedia language), all twelve quality measures facilitated the tweeted articles reached significantly higher values at the .001 level. For 96.49% of all language combinations, more than 50% of the quality measures reached significantly higher results for the tweeted article (at the .001 level). Looking into the language combinations featuring lower significance, we find that these are links between languages which both feature a comprehensive Wikipedia.

Comparing the individual quality measures, we find that for the HasInfobox measure and number of images relative to the length of the articles, 97.37% of all language combinations feature significantly higher quality values for the tweeted article (at the .001 level; 111 out of 114 language pairs show significant differences). The Currency measure reaches significantly higher values in 110 of 144 cases. For the number of sections, references and wikilinks (all relative to the article length) and the number of references, we also obtain highly significant differences in more than 92.98% of all cases. The length and informativeness measures reach the lowest value in terms of number of significantly different articles grouped by languages (77.19% of all language combinations).

To answer RQ2, we find that for 84.92% of all Wikipedia articles posted in another language, there is no counterpart in the tweet's language. When looking at the remaining 15.08%, we find that the differences between the articles posted and their counterparts in regards to quality are substantial and that tweeted articles feature higher quality. This fact suggests that users tweet the article of higher quality, even if it is in a language different from the tweet's language. We also observe that the dominating language in terms of article quality is English.

6. Discussion

In this section, we further discuss the findings presented in the previous section in the light of the research questions posed in the introductory section.

Our analysis of the Wikipedia articles referenced by inter-language links showed that 84.92% of all articles of a given language do not have a counterpart in the tweet's original language. Naturally, the non-existence of a Wikipedia article in a user's language is the primary cause for using the Wikipedia article of another language for this very issue.

The English, Japanese and Spanish Wikipedias are the top targets for inter-language links. These are also among the most popular and most comprehensive Wikipedias featuring a highly active community, which is a driving force for maintaining high quality across articles. In fact, these three Wikipedias are within the top 4 Wikipedias in terms of page views per month [27]. The English, Spanish and Japanese Wikipedias are within the top 13 Wikipedias in terms of number of articles and among the top 7 Wikipedias in terms of users and total number of edits. Such a highly active community is also reflected in the results of the Currency measure—a measure for the age the article, which is significantly higher for the articles posted (at the .001 level). These facts reflect the varieties of the different Wikipedias regarding the number of articles, edits, images and the size of the underlying community [42] as more comprehensive and well-maintained Wikipedias (i) feature more articles and (ii) feature articles of higher quality. We suspect that—besides the sheer existence of an article in a user's language—the quality of articles plays an important role for users in the process of posting Wikipedia links on Twitter. This can also be seen from the fact that for 71% of all languages, more than half of the measures are significantly higher for the tweeted article, implying that the quality of the tweeted article is substantially higher. Especially considering tweet languages for which the corresponding Wikipedia is relatively "small" (in terms of the number of articles and edits), the differences between the article and its counterpart are highly significant and users choose to tweet the article of higher quality. This is also backed by the finding that for English and the set of German, French and Turkish, the differences in the quality of articles are significant in both directions. This suggests that while both Wikipedias for these inter-language links generally feature a high quality, users still aim to tweet the article of higher quality in the individual cases. However, showing that users intentionally choose the article of higher quality is subject to further (qualitative) investigations.

We also observe that the proposed quality measures are subject to both cross-contextual issues and differences in the nature of languages. These factors limit the predictability of measures to a certain extent. Warncke-Wang et al. experienced similar problems. They found that some Wikipedia communities make use of references for verifying facts within the text while other communities tend to use a bibliography section and no inline citations. Such cross-contextual differences in terms of how communities structure their articles and how quality is assured and defined by these communities, cannot be reflected by a single measure. However, we argue that the mix of different measures as employed in this study provides a solid basis for our analyses as we incorporate a variety of measures for structure, length, edits, currency and the complexity of an article. Also, using measures computing structural and content-based factors relative to the length should reflect on the differences between articles as these allow shorter, but well-structured articles to be considered as of good quality as well. Also, the experiments showed that e.g., the diversity measure does not show results as significant as other measures do. It is subject to further investigation which quality measures may have a direct influence on the choice of articles.

7. Conclusion

In this paper, we study the interplay between Twitter and Wikipedia, i.e., how Wikipedia articles are referenced within tweets. In particular, we analyse the usage of inter-language links and look into possible causes for the usage of such links. We find that 84.92% of articles tweeted do not have a counterpart in the original tweet's language. As for the remaining 15.08%, we study the differences between the posted article and its counterpart in terms of article quality. We assessed the quality of both the posted articles and its counterpart articles using twelve quality measures covering textual, editorial and temporal aspects. This analyses show that the quality of the articles tweeted is constantly higher than the quality of their counterpart articles. Our findings suggest that users choose the article of higher quality even if it is in a language different from the tweet's language and the article would exist for the tweet's language.

Future work includes a qualitative study on the usage of inter-language links to investigate whether the observed user behaviour is intentional. Furthermore, we aim at extending the set of measures for assessing the quality of articles and looking into whether inter-language links may be used as an indicator for article quality.

10. References

[1] Abel, F., Gao, Q., Houben, G.J. and Tao, K. 2011. Analyzing User Modeling on Twitter for Personalized News Recommendations. *Lecture Notes in Computer Science* (2011), 1–12.

[2] Alexa: The top 500 sites on the web: *http://www.alexa.com/topsites/global*. Accessed: 2015-05-30.

[3] Bakshy, E., Hofman, J., Mason, W. and Watts, D. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. *Proc. of the 4th ACM Intl. Conf. on Web Search and Data Mining*. (2011), 65–74.

[4] Blumenstock, J.E. 2008. Size Matters: Word Count as a Measure of Quality on Wikipedia. *Proc. of the 17th Intl. Conf. on World Wide Web* (2008), 1095–1096.

[5] Bot or Not? *http://truthy.indiana.edu/botornot/*. Accessed: 2015-06-05.

[6] Castillo, C., Mendoza, M. and Poblete, B. 2011. Information Credibility on Twitter. *Proc. of the 20th Intl. Conf. on World Wide Web*. (2011), 675.

[7] Chu, Z., Gianvecchio, S., Wang, H. and Jajodia, S. 2010. Who is Tweeting on Twitter: Human, Bot, or Cyborg? *Proc. of the 26th Annual Conf. on Computer Security Applications* (2010), 21.

[8] Dalip, D.H., Gonçalves, M.A., Cristo, M. and Calado, P. 2009. Automatic Quality Assessment of Content Created Collaboratively by Web Communities: a Case Study of Wikipedia. *Proc. of the 9th ACM/IEEE-CS Joint Conf. on Digital Libraries* (2009), 295–304.

[9] Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. 2014. The Rise of Social Bots. *arXiv preprint arXiv:1407.5225*. (2014).

[10] Gabrilovich, E. and Markovitch, S. 2007. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. *IJCAI* (2007), 1606–1611.

[11] Genc, Y., Sakamoto, Y. and Nickerson, J. V 2011. Discovering Context: Classifying Tweets through a Semantic Transform based on Wikipedia. *Foundations of Augmented Cognition*. Springer. 484–492.

[12] Jullien, N. 2012. What we know about Wikipedia: A Review of the Literature analyzing the Project. *Available at SSRN*. 2053597, (2012).

[13] Kapanipathi, P., Jain, P., Venkataramani, C. and Sheth, A. 2014. User Interests Identification on Twitter Using a

Hierarchical Knowledge Base. *The Semantic Web: Trends and Challenges*. Springer. 99–113.

[14] Kaplan, A.M. and Haenlein, M. 2010. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*. 53, 1 (2010), 59–68.

[15] Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L. and Chissom, B.S. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*.

[16] Kittur, A. and Kraut, R.E. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. *Proc. of the Conf. on Computer Supported Cooperative Work* (New York, NY, USA, 2008), 37–46.

[17] Kittur, A., Suh, B., Pendleton, B.A. and Chi, E.H. 2007. He Says, She Says: Conflict and Coordination in Wikipedia. *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems* (New York, NY, USA, 2007), 453–462.

[18] Laniado, D., Tasso, R., Volkovich, Y. and Kaltenbrunner, A. 2011. When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages. *Intl. AAAI Conf. on Weblogs and Social Media* (2011).

[19] Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*. 49, 4 (Jul. 2013), 764–766.

[20] Li, C., Weng, J., He, Q., Yao, Y., Datta, A., Sun, A. and Lee, B.-S. 2012. TwiNER: Named Entity Recognition in Targeted Twitter Stream. *Proc. of the 35th Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval* (New York, NY, USA, 2012), 721–730.

[21] Lim, K.H. and Datta, A. 2013. Interest Classification of Twitter Users using Wikipedia. *Proc. of the 9th Intl. Symp. on Open Collaboration* (2013), 22:1–22:2.

[22] Liu, J. and Ram, S. 2011. Who Does What: Collaboration Patterns in the Wikipedia and Their Impact on Article Quality. *ACM Trans. Manage. Inf. Syst.* 2, 2 (2011), 11:1–11:23.

[23] Lu, C., Lam, W. and Zhang, Y. 2012. Twitter User Modeling and Tweets Recommendation based on Wikipedia Concept Graph. *Workshops at the Twenty-Sixth AAAI Conf. on Artificial Intelligence* (2012).

[24] MediaWiki API: *http://www.mediawiki.org/wiki-/API:Main_page*. Accessed: 2015-05-30.

[25] Nielsen, F.Å. 2012. Wikipedia Research and Tools: Review and Comments. *Available at SSRN 2129874*. (2012).

[26] Osborne, M., Petrovic, S., McCreadie, R., Macdonald, C. and Ounis, I. 2012. Bieber no more: First Story Detection using Twitter and Wikipedia. *Proc. of the Workshop on Time-aware Information Access. TAIA* (2012).

[27] Page Views for Wikipedia, Non-mobile, Normalized: *https://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm*. Accessed: 2015-06-14.

[28] Parker, J., Wei, Y. and Yates, A. 2013. A Framework for Detecting Public Health Trends with Twitter. *Proc. of the 2013 Intl. Conf. on Advances in Social Networks Analysis and Mining*. (2013), 556–563.

[29] Phillips, A. and Davis, M. 2006. BCP 47 -- Tags for Identifying Languages. IETF.

[30] Romero, D.M., Galuba, W., Asur, S. and Huberman, B.A. 2011. Influence and Passivity in Social Media. *Lecture Notes in Computer Science* (2011), 18–33.

[31] Romero, D.M., Meeder, B. and Kleinberg, J. 2011. Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political hashtags, and Complex Contagion on Twitter. *Proc. of the 20th Intl. Conf. on World Wide Web* (2011), 695–704.

[32] Sakaki, T., Okazaki, M. and Matsuo, Y. 2010. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. *Proc. of the 19th Intl. Conf. on World Wide Web* (2010), 851.

[33] Stemming algorithms for various European languages: *http://snowball.tartarus.org/texts/stemmersoverview.html*. Accessed: 2015-06-14.

[34] Stvilia, B., Al-Faraj, A. and Yi, Y.J. 2009. Issues of Cross-Contextual Information Quality Evaluation-The Case of Arabic, English, and Korean Wikipedias. *Library and Information Science Research*. 31, 4 (2009), 232–239.

[35] Stvilia, B., Twidale, M.B., Smith, L.C. and Gasser, L. 2005. Assessing Information Quality of a Community-based Encyclopedia. *Proc. of the Intl. Conf. on Information Quality* (2005), 442–454.

[36] Twitter API: POST statuses/filter: *https://dev.twitter.-com/streaming/reference/post/statuses/filter*. Accessed: 2015-06-06.

[37] Twitter: Monitoring and Analyzing Tweets: *https://blog.twitter.com/2014/connecting-to-the-pulse-of-the-planet-with-twitter-apis*. Accessed: 2015-05-30.

[38] Warncke-Wang, M., Cosley, D. and Riedl, J. 2013. Tell Me More: An Actionable Quality Model for Wikipedia. *Proc. of the 9th Intl. Symp. on Open Collaboration* (2013), 8.

[39] Wikimedia Foundation: Keep Wikipedia Free: *http://wikimediafoundation.org/wiki/Keep_Wikipedia_Free*. Accessed: 2015-06-09.

[40] Wikipedia: Article Grading Scheme: 2015. *http://en.wikipedia.org/wiki/Template:Grading_scheme*. Accessed: 2015-06-10.

[41] Wikipedia: Featured Article Criteria: 2015. *http://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria*. Accessed: 2015-05-30.

[42] Wikipedia: List of Wikipedias: *http://en.wikipedia.org-/wiki/List_of_Wikipedias*. Accessed: 2015-06-05.

[43] Wilkinson, D.M. and Huberman, B.A. 2007. Cooperation and quality in wikipedia. *Proc. of the 2007 Intl. Symposium on Wikis*. (2007), 157–164.

[44] Wu, K., Zhu, Q., Zhao, Y. and Zheng, H. 2010. Mining the Factors Affecting the Quality of Wikipedia Articles. *2010 Intl. Conf on Information Science and Management Engineering* (Aug. 2010), 343–346.

[45] Zangerle, E., Schmidhammer, G. and Specht, G. 2015. Wikipedia on Twitter: Analyzing Tweets about Wikipedia. *Proc. of the 11th Intl. Symp. on Open Collaboration* (2015).