The Path to Success: A Study of User Behaviour and Success Criteria in Online Communities

Erik Aumayr
Insight Centre for Data Analytics
National University of Ireland, Galway
Galway, Ireland
erik.aumayr@insight-centre.org

Conor Hayes
Insight Centre for Data Analytics
National University of Ireland, Galway
Galway, Ireland
conor.hayes@insight-centre.org

ABSTRACT

Maintaining online communities is vital in order to increase and retain their economic and social value. That is why community managers look to gauge the success of their communities by measuring a variety of user behaviour, such as member activity, turnover and interaction. However, such communities vary widely in their purpose, implementation and user demographics, and although many success indicators have been proposed in the literature, we will show that there is no one-fits-all approach to community success: Different success criteria depend on different user behaviour. To demonstrate this, we put together a set of user behaviour features, including many that have been used in the literature as indicators of success, and then we define and predict community success in three different types of online communities: Questions & Answers (Q&A), Healthcare and Emotional Support (Life & Health), and Encyclopaedic Knowledge Creation. The results show that it is feasible to relate community success to specific user behaviour with an accuracy of 0.67–0.93 F1 score and 0.77–1.0 AUC.

CCS CONCEPTS

• **Information systems** → **Social networks**; • **Computing methodologies** → *Machine learning*;

KEYWORDS

online communities, community success, user behaviour

ACM Reference format:
Erik Aumayr and Conor Hayes. 2017. The Path to Success: A Study of User Behaviour and Success Criteria in Online Communities. In *Proceedings of WI '17, Leipzig, Germany, August 23-26, 2017,* 9 pages.
DOI: 10.1145/3106426.3106469

1 INTRODUCTION

Businesses and non-commercial groups have long recognised the importance of online communities for knowledge sharing and support. The former discovered that they can increase revenue by maintaining an online presence that enables them to respond fast and on a global scale to their customers [2]. In the non-commercial case, people found that online communities provide help and support on any topic of personal interest, from gardening to healthcare (e.g. [52]). It is important for people and businesses that the communities are well maintained. It can be frustrating for their members if communities do not function well, e.g. by ignoring their requests, which in turn affects the communities' continuity and success.

As a result, community success has been discussed from different perspectives, e.g. people and organisations [13, 14], socialising and technology [35], and success requirements in different stages of the community life cycle [18]. But which success factors apply to which types of communities? Different communities strive for different goals and therefore must be measured by their fulfilment of these goals [27, 35, 50]. Moreover, a community's goals are often not directly measurable, and the relation between certain user behaviour and the achievement of these goals has not been proven by objective measures. This is crucial to understanding online community success, as community managers need to understand which user behaviour can be targeted in order to maintain or improve a community's facility to achieve its goals and serve its purpose.

This research presents an analysis of success in different types of online communities, each one with their own goals and hence with their own criteria for success. To analyse these success criteria, we first review existing literature on community success (Section 2), and then define success metrics for three different types of communities, namely Q&A communities (Section 3.1), Life & Health communities (Section 3.2), and Knowledge Creation communities (Section 3.3). Then, we collect user behaviour features that were proposed as success factors in the literature (Section 4.1), and extend this set with our own features (Section 4.2). Finally, we identify which user behaviour actually contributes to the success of the different community types by using the user behaviour features as predictors for the different success criteria (Section 5).

2 RELATED WORK

Endeavours to define and measure community success resulted in different perspectives and meanings of success. In the early 1990's, before the rise of Web 2.0 and social networks, DeLone and McLean devised a conceptual model to formalise aspects of information systems success, such as user satisfaction and impact on the individual and host organisation [13]. Subsequently, many articles investigated one or more of these components of success, including: **longevity**, e.g. survivability [37], member turnover [49], and member loyalty [20]; **commercial impact**, e.g. cost effectiveness [8], user acceptance [44], and commercial profit [20]; **user satisfaction**, e.g. member need satisfaction [42], sense of community

978-1-4503-4951-2/17/08. . . $15.00
DOI: 10.1145/3106426.3106469

[28], and aid for professional development [16]; **community performance towards a specific goal**, e.g. productivity [36], output quality [38], and impact on the scientific world [7]; **and other key performance indicators**, e.g. member participation and engagement [32] and responsiveness [48]. These different aspects show that the functionality of online communities is a multi-dimensional problem that has to be tackled from different perspectives. This is further underlined by the terminology that is used to describe the overall community functionality: Whereas early works used the term "success", referring to, among other things, acceptance by and satisfaction of the users [13, 25, 35], the term "health" has also become popular in the past decade. Community health is a term that has been used to describe the well-being and activity of a community in a more general sense, with the help of health indicators such as user participation and interactivity [30, 40, 48, 49].

In order to determine community success, one has to ask "Success for who?" [35], and identify the perspectives of the involved stakeholders, such as business managers (representing a company), community managers and community members [35]. Companies and organisations that set up an online community usually want to increase their revenue, e.g. by reducing monetary and time costs [8, 13, 14], improving product quality [13, 38], raising brand awareness and loyalty [20, 25], or achieving acceptance in the online community [44]. On the other hand, for communities that are owned by non-commercial entities, visibility and cost-effectiveness might not have a high priority, e.g. in the case of personal issues communities. The owner or manager of such a community may value user engagement and interactivity [22, 41, 52], or a community that is attractive to newcomers [9, 35] and lively and self-sustaining [37, 40, 55]. For moderators, it is important that users adhere to the social rules and policies [10, 43] and integrate newcomers well [18]. Finally, the community members themselves are the key group of stakeholders, as there would be no community without them. The satisfaction of their goals and wishes has high priority for retaining an active user base [28, 42, 54]. Users' expectations of the community include high quality and quantity of content [36, 38], professional development [16], and a maintenance of social connections through regular interactions with other users [32, 35, 48].

The specific purpose of a community further defines what criteria must be met in order to enable success. For example, an interest-based community must provide an environment where people can openly and transparently participate in discussions about the topic of interest, e.g. to achieve tangible impact in the field of interest [7]. In knowledge creation and other crowd-sourcing communities, the quality and quantity of user contributions are important [36, 53]. Software development communities aim for high quality as well, but also for functionality of the produced software [15], with a focus on timeliness of the results because of delivery deadlines [11]. Similarly, Questions & Answers communities have the primary purpose of providing information seekers with a platform to formulate their questions and receive timely and accurate solutions [17, 33].

The literature proposes many factors for community success, and while some articles make use of first-hand experiences by asking interviewees about their *personal perception* of community success [26, 42, 50], often these success factors are not evaluated against objectively measurable success variables. In this work, we aim to demonstrate that different success criteria are affected by different user behaviour, underlining the importance to find and evaluate the appropriate success factors for the specific kind of community success one is trying to assess. Therefore, we put together a collection of user behaviour features and evaluate them against several success metrics of three types of communities.

3 DATA AND DEFINITIONS OF SUCCESS

In this section, we introduce the three types of communities in our study, namely Questions & Answers (Q&A), Life & Health, and Knowledge Creation communities, and define their specific success criteria based on existing literature.

3.1 Q&A Communities

Online Q&A communities are not only a popular type of support community but also provide a tangible way to derive success metrics from their main goal: solving questions. We examine Stack Exchange (stackexchange.com) as a popular multi-purpose Q&A platform for hobbyists and professionals, where users discuss a wide variety of topics, from cooking and bicycle repair to software-related questions in the popular Stack Overflow site. From the Stack Exchange platform, we downloaded the publicly available data dump from June 2016, which contains 8 years of data: 152 forums that vary from a few hundred to almost three million users who produced up to 31 million posts in the most extreme case. Similar to existing literature, we refer to the forums in each data set as communities, each one defined by a specific topic [1, 17, 40, 41, 46, 48, 49, 54].

User satisfaction has been recognised as an important element to the success of online communities [13, 14, 28, 29, 42, 54]. While Q&A communities – like any other kind of online communities – are also subject to non-functional aspects, such as member integration, trust and interface usability [18], their main goal and their performance towards that goal can be measured as a function of successfully solved questions. Liu et al. [31] and Hiscock et al. [17] defined that information seekers are satisfied if their questions are sufficiently solved, which they indicate by explicitly marking the best answer. Hiscock et al. further added a time constraint to ensure that solutions arrive in a time that is relevant to the question asker [17]. We use the notion of asker satisfaction based on timely solved questions, and translate it to the level of the whole community by formalising *solved* and t_{solved} as follows:

solved: Measures the global proportion of solved questions for a given community (Equation 2). In our data, the range of solved questions lies between 17% and 78%. With Q as the set of questions:

$$S = \{q : q \in Q, q \text{ has an accepted answer}\} \tag{1}$$

$$solved = \frac{|S|}{|Q|} \tag{2}$$

$\boldsymbol{t_{solved}}$: Response time has been recognised as a relevant aspect of Q&A community success [17]. We measure the average time between a question and its accepted best answer (in hours), as per Equation 3. In our data, the average solving delay ranges from 15 hours to 45 days. Unresolved questions do not contribute to t_{solved}.

$$t_{solved} = \frac{\sum \{h : \text{time between } s \in S \text{ and its accepted answer}\}}{|S|} \tag{3}$$

3.2 Life & Health Communities

Healthcare and emotional support communities are similar to Q&A communities in the sense that the participants seek and provide help and support. They are also similar in that they both rely on a discussion-forum like structure, which allows us to regard sub-forums as communities, as discussed in Section 3.1. However, the nature of the problems, the help that is sought, and the way people interact are very different. Instead of a quick reply with factual and accurate information, people in Life & Health communities look for empathy and positive emotions, and to build emotional connections to other participants, who themselves often are or have been in a similar situation in life. From the popular Irish discussion forum Boards.ie, we collected a number of communities where people talk about their personal issues, such as disability, sexuality, health, parenting, and marriage. Because most communities of this kind are not accessible to the public, we rely on a relatively small number of 16 Life & Health communities for our analysis. They vary in size and activity from 71 users who created 251 posts in 2.5 months to over 9,000 users who created over 200,000 posts in six years.

The purpose of Life & Health communities is to provide users with a place to meet others and form personal relationships (and an overall sense of community) based on mutual experiences of personal issues [52]. As it is not feasible to directly measure the sense of community, we use a number of proxy metrics to capture whether the user interactions lead to a positive atmosphere and a well-connected community:

repliedTo: Inspired by Rowe and Alani, who used the relation between threads that received replies and threads that did not in their analysis of community health [40], we measure the proportion of threads in which posters received responses. In a successful Life & Health community, users are not being left alone but always receive replies to their enquiries for support and empathy. With T as the set of discussion threads:

$$repliedTo = \frac{|\{t : t \in T, t \text{ received responses from other users}\}|}{|T|} \tag{4}$$

connectedness: Another way to look at the interactive involvement of the users is to measure the network density of the user interaction graph [36], i.e. how well are the users connected and support each other. The network density is the number of actual connections divided by the number of possible connections in the user graph $G = V \times E$, where V represents the users and E the reply edges between the users:

$$connectedness = \frac{|E|}{\frac{|V|(|V|-1)}{2}} \tag{5}$$

replySentiment: The act of interaction between users might not be enough to gauge empathy and support in a community. Some community managers use sentiment analysis to assess the success of their communities [12]. We measure the amplitude of positive sentiment in replies as an indicator for a good atmosphere in the community, using the SentiStrength tool [45]. With the number of replies as R:

$$replySentiment = \frac{\sum \{\text{positive sentiment score of } r : r \in R\}}{|R|} \tag{6}$$

3.3 Knowledge Creation Communities

Knowledge creation platforms such as the Wikipedia online encyclopaedia (wikipedia.org) are concerned with collecting and curating knowledge, and hence aim for high quality [38] and a long-lasting value and impact [7] of the created content. Our data contains all article revisions of the Simple English Wikipedia (simple.wikipedia.org) up to December 2007. The Simple English Wikipedia is a subset of Wikipedia, where users work on describing topics in simple words that are aimed at non-native speakers, young people, and people with limited reading and comprehension abilities. The data has more than 480,000 revisions of over 32,000 articles.

As opposed to the other two community platforms, Wikipedia does not have a clearly defined discussion forum structure. Users collaborate on articles, and several articles may be related because they exhibit some similarity, e.g. based on topic and involved individuals. Instead, we utilise the community detection algorithm OSLOM [24] to cluster articles and the users who contributed to them. OSLOM is a graph-based algorithm, and we construct the graph by creating links between the articles, weighted by the number of collaborators each article pair has in common. With the aim to reduce noise and complexity of the article graph, we only consider content collaborations on main articles (Wikipedia namespace 0), and we exclude user profiles that are obviously automated bots, such as usernames that end in "-script". The smallest of the resulting 326 communities have only a handful of users, whereas the biggest one consists of 20,939 users who wrote 4,809 articles with 282,676 edits.

While it is difficult to measure the quality and value of created content in general, the purpose of the Simple English Wikipedia (being easy to read and understand) allows us to define success based on the information content and readability of articles [39, 47].

informativity: The information content of a text can be expressed as the number of content words in relation to function (i.e. non-content) words [39, 47]. From every posted text, we extract the number of content words (nouns, verbs, adjectives, adverbs, numbers and foreign language words) with the help of a Part-of-Speech tagger from the Python NLTK package [4]. With $p \in P$ representing all community posts, i.e. original articles and subsequent edits, we compute a normalised informativity score as follows:

$$informativity = \frac{\sum \left\{\frac{numContentWords}{numWords} : p \in P\right\}}{|P|} \tag{7}$$

complexity: The complexity of a text describes how difficult it is to read and understand, and there are several formulae to compute it based on sentence length and the number of complex words [21]. We use the original formula LIX (Swedish for "Läsbarhetsindex", readability index) that was introduced by pedagogics researcher Carl-Hugo Björnsson in 1968 [5]. As extremely long sentences can skew the distribution of LIX, we use a logarithmic transformation (logLIX, Equation 8) to reduce the impact of drastic outliers:

$$logLIX = log\left(\frac{numWords}{numSentences} + 100 * \frac{numLongWords}{numWords}\right), \tag{8}$$

where long words are defined as words with more than 6 letters.

$$complexity = \frac{\sum \{logLIX(p) : p \in P\}}{|P|} \tag{9}$$

4 USER BEHAVIOUR FEATURES

Community success is a multi-dimensional problem [13] and difficult to measure for many types of online communities. In a prominent article, Preece used the term "success determinants" to describe certain user behaviour that may indicate whether a community is successful [35]. In this section, we first collect user behaviour features that were proposed in the literature and then discuss our own additions. Some of the suggested user behaviour that could indicate success is not straightforward to measure, for example, the frequency of visits by unregistered users (i.e. lurkers) [35], trust among the members [26], and integration of new members [18]. We limit our experiments to user behaviour that is based on information available in our data, namely 1) user attraction and retention, 2) user activity, 3) user interaction, and 4) content creation.

Note that we adopt the definitions of success factors as they were given in the respective articles. Where concrete definitions are lacking, we adhere to the proposed concepts as closely as possible according to the available information in our data. Although the concept of discussion threads does not directly apply to online encyclopaedias, we will use the following terminology throughout the remainder of this paper: A thread is comprised of an original post, e.g. a question or the first version of a Wikipedia article, and zero or more subsequent responses to the original post. A response can be an answer/reply in discussion-forum type communities or a revision of the original article in Wikipedia.

4.1 From the Literature

4.1.1 User attraction and retention. The continuing growth of a community after its initiation has been recognised as an important factor for a successful community [18, 30]. Although mature communities grow slower than newly formed ones, a community is unlikely to remain functioning when the influx of new members in a given period of time is too low [9, 19, 48, 52], while at the same time users are leaving the community [9, 34, 38, 40, 52]. The churn of highly active participants (e.g. users with a high betweenness centrality) is especially dramatic and harmful to the community [36]. The resulting size of the community was often suggested to be an important factor as well [30, 35, 40, 46, 48, 51, 52], where a reasonably big community ensures that the critical mass of participants is achieved [37]. And finally, the age of a community could indicate success [51], as unsuccessful communities are expected to diminish and die sooner than later.

- Community age: The time between the first and last recorded community post in the resolution of seconds
- Community size: Total number of users
- Community growth: Newly joined users per day
- User churn: The average monthly ratio between users who have posted for the last time in the data and users who will continue posting in later months
- VIP churn: Proportion of top-10% contributors leaving the community, averaged over all months

4.1.2 User activity. A high user engagement indicates a successful community, which can be measured in the total number of posts [1, 30, 35, 41, 48, 49, 51] and posts per day (or other time units) [19, 35, 52], as well as the number of posts per user [35, 48, 52]. Others recognised that a good amount of new threads (or questions in the Q&A context, new articles in the Knowledge Creation context) is also required for a successful community, as they enable user engagement in the first place [46, 48], as well as threads or questions per day [19]. Then, of course, the number of contributors is important [46], measured as the number of people who respond with a reply, answer or content revision.

- Number of posts: Original posts and responses in total
- Number of original posts: Number of questions, threads or original articles, depending on the data set
- Posts per day: Number of total posts divided by the number of days the community has been active
- Original posts per day: Number of original posts divided by the community age in days
- Posts per user: Number of participation activity per user
- Number of responders: Authors of answers, replies, edits

4.1.3 User interaction. The number of posts per thread [9, 30, 35, 48, 49, 52], or thread length, and the number of unique users per thread [30, 48, 49] are measures for interactivity between users, where a high interaction between participants is considered beneficial for the community. In that respect, reciprocity is the relation between giving to and taking from the community, and can be measured as the number of seed posts (original posts that received at least one response from a different person) [49], non-seed posts (ignored original posts) [48], as well as the ratio between the two [40], and the users' response time [30, 49, 52] and reply effort (proportion of responses to original posts) [35]. Also, the connectedness between community members, e.g. the network density [36] or the clustering coefficient of the user graph [40], can indicate a well-functioning community based on the assumption that members benefit from a high information flow between them.

- Number of seed posts: Number of original posts that received at least one response from another person
- Number of non-seed posts: Ignored original posts
- Seed/non-seed ratio: Original posts with at least one response divided by the number of ignored original posts
- Thread length: Average number of posts in a thread
- Unique users per thread: Average number of distinct users that participate in a thread
- Reply effort: Per-user average of contributed responses divided by all their posts
- Response time: Average time between the original posts and the arrival of responses in hours
- Clustering coefficient: Average local clustering coefficient; affected by users responding to each other

4.1.4 Content creation. The quality of the submitted content shows how much effort the users put in participating, indicating that motivated users are a sign of a successful community. Preece considered the message length a measure of content quality [35]. Also, references to internal and external sources can indicate how much effort users are putting into providing additional information to their posts, such as references to other/earlier internal posts [48].

- Content length: Average number of words per post
- URLs in posts: Average number of references to internal or external sources per post

4.2 Additional Candidates

We extend the list of proposed success indicators to cover more aspects of user behaviour that might be relevant to community success. Most notably, we include the responses per user, the information spread in the community, as well as the standard deviation and normalisation of some user behaviour, which are to date not considered in the literature. The standard deviation of user behaviour can indicate particular risks in community functionality. For example, a high standard deviation of posts per user means that some few users carry most of the community activity, which is arguably less sustainable for a community than if the load is more equally distributed among all users. In the former case, a drop-out of a highly active participant can be devastating for a community. In general, a high standard deviation is not desirable as it indicates an unbalanced distribution of workload. For brevity, we omit the description of each individual standard-deviation based feature, as they all represent the same addition to the mean-based variables.

The normalisation of some factors allows us to look at user behaviour in a relative way, e.g. the proportion of original posts, rather than their absolute numbers that are depending on the size of the community and other factors. Further, we split content related factors into title, original post and response to investigate the effect of content features more fine-grained.

4.2.1 *User activity.*

- Original post proportion: Number of original posts in relation to all posts
- Responder proportion: Number of users that post responses in relation to all users
- Responses per user: The average number of responses per user could indicate sufficient user engagement

4.2.2 *User interaction.*

- Seed post proportion: Proportion of threads that received responses from others. This is analogue to the success criterion *repliedTo* and hence not used for its prediction.
- Information spread: Measures the average degree in the user graph, which is built by creating unweighted links between users who interact. A high average degree indicates that many users are involved in knowledge sharing.

4.2.3 *Content creation.*

- Original post length: The content length in Section 4.1 does not account for differences between original posts and responses. For example, shorter questions with few words in their description could indicate that they are easy to solve, which could motivate more users to participate
- Original post length ratio: Original post length in relation to the community's overall average content length
- Response length: Short responses, on the other hand, might not contain enough information and therefore decrease the community performance
- Response length ratio: Response length in relation to the community's overall average content length
- Title length: A clear and precise title that includes the necessary key terms to grasp the context of the thread could get other users' interest and foster interaction
- URLs in original posts: Similar to Section 4.1, original posts that contain links to internal or external sources show that the poster researched the issue, which could potentially increase the chance of receiving good responses
- Ratio of URLs in original posts: References in original posts in relation to the average of all posts
- URLs in responses: The average number of references per response may indicate additionally provided information
- Ratio of URLs in responses: References in responses in relation to the average of all posts
- Content length change (only Wikipedia): Difference in content length between revisions and the original article. Articles that receive additional content may show that the community is actively participating in content creation
- Edit distance (only Wikipedia): The Levenshtein edit distance between revisions and the original article indicates that contributors put effort into improving existing articles

5 PREDICTING COMMUNITY SUCCESS

In previous work, we investigated how well individual success indicators from the literature are related to Q&A community success, and we found that only a few have a strong correlation with the goals of Q&A communities [3]. Here, we extend this finding by studying how well the success criteria of different kinds of communities (introduced in Section 3) can be predicted, and which types and combinations of user behaviour are most important for the individual definitions of success. We employ regression, as well as binary classification, to predict community success (relative to the median for each success criterion in the case of classification). For this approach, we use the random forest predictor [6], a flexible prediction algorithm that is not restricted to linear relationships, and can perform both regression and classification. In our experiments, it outperformed other prediction algorithms such as support vector machines and linear and logistic regression. To assess the results of the regression, we report the root mean square error (RMSE), the normalised RMSE (divided by the mean), and the explained variance (R^2). For the binary classification, we report the F1 score – i.e. the harmonic mean between precision and recall, macro-averaged over positive and negative classes – as well as the area under the ROC curve (AUC); see Table 1.

Before we train the predictor, we pre-filter the 53 user behaviour features (for brevity we did not list the standard deviation features in Section 4) in order to decrease the search space for the predictor, which can greatly improve accuracy and reduce run time. We use the Boruta algorithm [23] for the feature selection. It selects and ranks features by comparing them to randomised versions of themselves, without over-pruning the feature space. The ranking and the retaining of minimal relevant features helps to understand their importance. In our experiments, Boruta led to consistently better prediction results compared to other methods, such as recursive feature elimination and correlation-based feature selection. We run the feature selection step for all data sets, all success criteria, and for each regression and classification. On average, the Boruta algorithm reduced the number of relevant features down to one-third of the available pool of features. In the extreme cases, the algorithm discarded a maximum of 47 (*replySentiment*) and a minimum of

	Q&A		Life & Health			Knowledge Creation	
Data	*solved*	t_{solved}	*repliedTo*	*connectedness*	*replySentiment*	*informativity*	*complexity*
Observed min	0.1747	15.2	0.7333	0.0013	1.8445	0.4660	0.5943
Observed mean	0.4387	227.5	0.8703	0.0112	2.1039	0.7036	1.6020
Observed max	0.7755	1072.8	0.9760	0.0447	2.4906	1.0	2.1560
Observed stdev	0.1061	166.6	0.0722	0.0124	0.1757	0.0645	0.1553
Regression							
RMSE (random baseline)	0.1058	166.1	0.0699	0.0120	0.1701	0.0644	0.1551
RMSE (prediction)	0.0698	109.7	0.0486	0.0067	0.1584	0.0511	0.1339
Normalised RMSE (pred.)	0.1590	0.4822	0.0559	0.5998	0.0753	0.0727	0.0836
R^2	0.5816	0.6065	0.7238	0.7996	0.5233	0.3628	0.2631
Classification							
F1	0.7553	0.8148	0.7385	0.9334	0.6730	0.6834	0.6894
AUC	0.8559	0.9092	0.9375	1.0	0.7917	0.7734	0.8020

Table 1: Generally, the prediction of the different success criteria results in low normalised error values of 0.06–0.16, except for t_{solved} (0.48) and *connectedness* (0.6). The good prediction accuracy is also reflected by regression results that are 7–44% above the random baseline, and by the good F1 (0.67–0.93) and AUC (0.77–1.0) values for the constructed binary classification.

12 (*informativity*) of the features. Some of the features that were rejected by the algorithm exhibit a correlation of more than 0.99, e.g. Number of posts and Number of users are too similar to Posts per day and Users per day, respectively.

Then, we construct a baseline by "predicting" the mean value for each success criterion (given in Table 1). This baseline gives us an indication of the worst case, in which the features would have no predictive power whatsoever. For the regression task, the random baseline "predicts" values with an error close to the standard deviation, whereas for the binary classification the baseline results in 0.5 for the AUC and F1 scores when averaged over both classes (not shown in Table 1). We then train the random forest predictor on the Boruta-filtered features to predict the seven individual success criteria from *solved* to *complexity* as listed in Table 1. In order to avoid over-fitting, we perform three repetitions of a 10-fold cross-validation (4 folds for the Health & Life communities because of their small number), where each random forest contains 100 trees.

The prediction models produce small normalised errors between 0.06 and 0.16 in most cases, with the exception of t_{solved} (0.48) and *connectedness* (0.6). Although the normalised RMSE is rather high in these two cases, their F1 and AUC scores look very promising. In summary, the relatively low errors and acceptable R^2 values for most predicted variables, as well as the improvement of 7–44% over the random baseline, show that it is feasible to explain different aspects of community success from observed user behaviour. The binary classification results show especially good accuracies, further supporting our findings. The high AUC values of up to 1.0 for the Life & Health communities are a good indicator for the feasibility of the prediction approach. However, our sample size of this community type is small, and we do not expect these extreme values for larger collections of communities.

5.1 Impactful User Behaviour

Not all the user behaviour features we described in Section 4 (including their standard deviations) are important for each individual success criterion. In the following, we discuss the features that the prediction algorithm deemed important for each success criterion.

5.1.1 Q&A Communities. For the question-solving performance (*solved*), the most impactful user behaviour is that which results in questions receiving many answers, in particular creating seed posts (Seed/non-seed ratio and Seed post proportion), and having many responses per user and users per thread. But also content features show some effect, such the length of questions and titles and the presence of URLs in replies, or rather their absence, as URLs in replies have a detrimental effect on the question-solving performance. A possible explanation for that could be that users prefer short and concise replies, without the need to read through additional resources. Furthermore, the Information spread, as well as the churn of users and VIP users, show an impact on the success of a Q&A community. The benefit of a high information spread is apparent, but our experiments also confirm that churn is a negative factor for Q&A performance, as churn leads to a decrease of knowledgeable experts who can solve questions.

The question solving time t_{solved} is similarly affected by user interaction features, such as Seed post proportion, Answers per user, as well as Reply effort and its standard deviation. In addition, a high standard deviation of VIP churn correlates to a short question-solving time. This is counter-intuitive, as it also implies an elevated level of churn, which has a negative impact on solving questions. Content features, such as Response length ratio and Content length, have a slightly negative impact on t_{solved}, as longer answers take longer to write. A strong negative feature is Community age, as old communities can receive answers to open questions much later because of their age. Note that we excluded Response time as a feature for t_{solved} for the obvious strong correlation.

5.1.2 Life & Health Communities. The frequency with which original posts are *repliedTo* in Life & Health communities is mainly predicted by thread-centred features, such as Unique users per thread, Thread length, Information spread and their standard deviations, as well as ORIGINAL POST PROPORTION. These features are apparent strong predictors, as *repliedTo* threads naturally have replies, which affects the interaction features. But also content features play a role, particularly RESPONSE LENGTH, CONTENT LENGTH and URLS IN POSTS. Surprisingly, they correlate to a decrease of *repliedTo* threads. We assume that a higher number of ignored posts implies that more attention is given to the posts that do receive replies. The *connectedness* between users is positively affected by NUMBER OF RESPONDERS and negatively by VIP CHURN. The NUMBER OF ORIGINAL POSTS, NUMBER OF SEED POSTS and ORIGINAL POSTS PER DAY correlate negatively to *connectedness*, showing that users are not well connected in communities that lack replies.

The third success criterion for Life & Health communities is *replySentiment*. It is best predicted by the length of submitted posts (in particular responses), which has a positive effect on the reply sentiment. However, while longer posts increase the probability for words with positive sentiment, they also tend to increase the probability for negative words. In this work, we focus on positive words to capture a good community atmosphere. We plan to better distinguish between positive and negative sentiment in the future.

5.1.3 Knowledge Creation Communities. *informativity* and *complexity* are metrics about content quality, hence it is not surprising that they largely dependent on content features, such as the length of original articles, edits and titles, as well as the number of URLs in original posts and responses. Interestingly, the length of original articles and edits has a detrimental effect on *informativity*, whereas the presence of URLs has a positive effect. We assume that URLs contain more content words than function words. *Complexity*, on the other hand, is equally affected by the content length and URLs, i.e. the longer the revisions, the more complex they are. EDIT DISTANCE and CONTENT LENGTH CHANGE contribute to the prediction results, but are not among the most impactful features. Beyond the content features, the prediction algorithm has found user activity and interaction features to be useful predictors of the success of Knowledge Creation communities. In particular, the number of unique contributors per article, the connectedness between contributors, the standard deviation of INFORMATION SPREAD, and the number of articles+edits per day are important for both success criteria *informativity* and *complexity*, as well-connected and actively contributing users improve the content quality.

For the same reason, the number of edits per user is an impactful positive predictor for *informativity*, whereas COMMUNITY AGE and RESPONSE TIME are strong negative factors. Older communities tend to have longer content, which is bad for *informativity*, as we already established. For *complexity*, impactful user interactivity features are INFORMATION SPREAD and its standard deviation, as well as the number of seed and non-seed posts. Further, COMMUNITY SIZE and the standard deviation of VIP CHURN are slightly correlated to *complexity*, which could indicate that communities that became too big and had VIP users leaving in irregular intervals created convoluted and overly complex content.

5.1.4 Summary. There are some obvious traits of user behaviour that have an impact on the individual success criteria, such as the effect of the answer rate on Q&A communities, content features on Knowledge Creation communities, and user interaction features on Life & Health communities. However, there is also less obvious user behaviour that a community manager should not disregard. In the case of Q&A communities, for example, well-written questions and question titles influence the question-solving performance, and too many references in answers should be avoided. Similarly, too many URLs in replies are bad for Life & Health communities, and long responses have a negative effect on the prediction of *repliedTo* threads. On the other hand, *replySentiment* benefits from longer responses. For Knowledge Creation communities, we discussed that short and revised articles are beneficial for the content quality. In addition, we found that user action and interaction features, such as unique contributors per article and well-connected contributors, improve the prediction of *informativity* and *complexity* of the created content over pure content features.

With regards to the investigated user behaviour features as predictors for community success, we can report that, although only a few of them were hugely impactful (in accordance with our findings in [3]), all features had at least some value for predicting one or more of the success criteria. That means that our additions to existing features from the literature proved helpful. Furthermore, some of our normalised features and standard deviations were rated as highly important by the prediction algorithm.

6 CONCLUSIONS

In this work, we investigate the notion of community success on a variety of different types of online communities and their particular goals and purposes, and hence with their own criteria for success. Based on existing literature, we define what success means for Q&A, Life & Health, and Knowledge Creation communities. Then, we collect and extend a range of user behaviour features that the literature suggests to be related to success, in order to predict the different success criteria for each type of community, and therefore identify which user behaviour is affecting the different notions of success and should be fostered by community managers. Prediction results show good accuracy (e.g. 0.67–0.93 F1 score and 0.77–1.0 AUC), which validates our findings of the relation between user behaviour and community success.

We find that, while some user behaviour is expectedly and apparently related to specific success criteria – such as user interaction features for predicting solved questions and content features for predicting content quality – there are also less obvious predictors. For example, URLs in answers have a negative effect on question-solving performance in Q&A communities, and long replies are related to more ignored user posts in Life & Health communities. In conclusion, there is no user behaviour or community characteristic that is a guaranteed indicator of success for all types of communities. For example, the often proposed community size [35, 40, 46, 48, 51] has little or no impact on the success criteria we investigated. It is therefore important to identify the concrete user behaviour that leads to success, given the community-specific goals and purposes.

7 LIMITATIONS AND FUTURE WORK

As we discuss in Section 2, there are many different types of communities, of which we only cover three types. In the future, a wider collection of community types with more data samples for each type will benefit our analysis on the impact of user behaviour on community success. That includes an analysis of the generalisability between success criteria in community platforms of the same type. Furthermore, the success criteria we use in this work must be understood as proxies, or assumptions, of what a successful community of each type needs, for the purpose of investigating the differences in user behaviour for different outcomes in community success. That is particularly true for Life & Health and Knowledge Creation communities because success cannot always be as clearly defined as the proportion of solved questions in Q&A communities. That means that their success criteria, such as $repliedTo$, $connectedness$ and $informativity$, need to be further studied and evaluated.

8 ACKNOWLEDGEMENTS

This research has been conducted with the financial support of Science Foundation Ireland (Grant Number SFI/12/RC/2289) and with data provided by Stack Exchange, Boards.ie and Wikipedia.

REFERENCES

[1] Sofia Angeletou, Matthew Rowe, and Harith Alani. 2011. Modelling and analysis of user behaviour in online communities. In *The Semantic Web-ISWC*. Springer.
[2] Arthur Armstrong and John Hagel III. 1996. The real value of online communities. *Harvard Business Review* 74 (1996).
[3] Erik Aumayr and Conor Hayes. 2017. Separating the Wheat from the Chaff: Evaluating Success Determinants for Online Q&A Communities. In *ICWSM*.
[4] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media.
[5] Carl Hugo Björnsson. 1968. *Läsbarhet*. Liber.
[6] Leo Breiman. 2001. Random forests. *Machine learning* (2001).
[7] Aidan Budd, Manuel Corpas, and others. 2015. A Quick Guide for Building a Successful Bioinformatics Community. *PLOS Computational Biology* (2015).
[8] Jacques Bughin and John Hagel III. 2000. The Operational Performance of Virtual Communities - Towards a Successful Business Model? *Electronic Markets* (2000).
[9] Brian S. Butler and Xiaoqing Wang. 2012. The Cross-Purposes of Cross-Posting: Boundary Reshaping Behavior in Online Discussion Communities. *Information Systems Research* (2012).
[10] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial Behavior in Online Discussion Communities. http://arxiv.org/abs/1504.00680
[11] Morakot Choetkiertikul, Hoa Khanh Dam, Truyen Tran, and Aditya Ghose. 2015. Characterization and Prediction of Issue-Related Risks in Software Projects. In *Working Conference on Mining Software Repositories*. IEEE Press.
[12] David Wagner and others. 2015. *Zum Status von Social-Media- und Community-Management in D-A-CH*. Technical Report. Bundesverband Community Management e.V. für digitale Kommunikation und Social-Media. https://www.bvcm.org/wp-content/uploads/2015/10/151026-BVCM-Studie-Report.pdf
[13] William H. DeLone and Ephraim R. McLean. 1992. Information systems success: The quest for the dependent variable. *IS research* (1992).
[14] William H. DeLone and Ephraim R. McLean. 2003. The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems* (2003).
[15] Vedat Diker. 2004. A Dynamic Feedback Framework for Studying Growth Policies in Open Online Collaboration Communities. *AMCIS Proceedings* (2004).
[16] Khe Foon Hew. 2009. Determinants of success for online communities: an analysis of three communities in terms of members' perceived professional development. *Behaviour and Information Technology* (2009).
[17] Philippa A. Hiscock, Athanassios N. Avramidis, and Jörg Fliege. 2015. Predicting Micro-Level Behavior in Online Communities for Risk Management. In *Data Science, Learning by Latent Structures, and Knowledge Discovery*. Springer.
[18] Alicia Iriberri and Gondy Leroy. 2009. A Life-cycle Perspective on Online Community Success. *ACM Computing Surveys (CSUR)* (2009).
[19] Aditya Johri, Oded Nov, and Raktim Mitra. 2011. Environmental jolts: Impact of exogenous factors on online community participation. In *ACM CSCW*.
[20] J. Koh and Y.-G Kim. 2004. Knowledge sharing in virtual communities: an e-business perspective. *Expert Systems with Applications* (2004).
[21] N. Korfiatis, E. García-Bariocanal, and S. Sánchez-Alonso. 2012. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *E-Commerce Research and Appl.* (2012).
[22] Robert E. Kraut, Paul Resnick, and others. 2012. *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press.
[23] Miron B. Kursa and Witold R. Rudnicki. 2010. Feature Selection with the Boruta Package. *Journal of Statistical Software* (2010).
[24] Andrea Lancichinetti, Filippo Radicchi, José J. Ramasco, and Santo Fortunato. 2011. Finding statistically significant communities in networks. *PlOS ONE* (2011).
[25] J. Lazar and J. Preece. 2002. Social considerations in online communities: Usability, sociability, and success factors. In *Cognition in A Digital World*.
[26] Jan Marco Leimeister, Pascal Sidiras, and Helmut Krcmar. 2004. Success factors of virtual communities from the perspective of members and operators: An empirical study. In *HICSS*. IEEE.
[27] Eldon Y. Li. 1997. Perceived importance of information system success factors: A meta analysis of group differences. *Information & Management* (1997).
[28] Hui Lin, Weiguo Fan, Linda Wallace, and Zhongju Zhang. 2007. An empirical study of web-based knowledge community success. In *HICCS*.
[29] Hsiu-Fen Lin. 2008. Determinants of successful virtual communities: Contributions from system characteristics and social factors. *Inform. & Mgmt* (2008).
[30] Lithium Technologies Inc. 2011. *Community Health Index for Online Communities*. Technical Report. http://www.lithium.com/pdfs/whitepapers/Lithium-Community-Health-Index_v1AY2ULb.pdf
[31] Y. Liu, J. Bian, and E. Agichtein. 2008. Predicting information seeker satisfaction in community question answering. In *ACM SIGIR conference on R&D in IR*.
[32] Xianghua Lu, Chee Wei Phang, and Jie Yu. 2011. Encouraging participation in virtual communities through usability and sociability development: An empirical investigation. *ACM SIGMIS Database* (2011).
[33] Kevin Kyung Nam, Mark S. Ackerman, and Lada A. Adamic. 2009. Questions in, knowledge in?: A study of naver's question answering community. In *SIGCHI Human Factors in Computing Systems*. ACM.
[34] Akshay Patil, Juan Liu, and Jie Gao. 2013. Predicting group stability in online social networks. In *World Wide Web*.
[35] Jenny Preece. 2001. Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology* (2001).
[36] Xiangju Qin, Michael Salter-Townshend, and Pádraig Cunningham. 2014. Exploring the relationship between membership turnover and productivity in online communities. In *ICWSM*.
[37] Daphne R. Raban, Mihai Moldovan, and Quentin Jones. 2010. An empirical study of critical mass and online community survival. In *Computer Supported Cooperative Work*. ACM.
[38] Sam Ransbotham and Gerald C. Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly* (2011).
[39] Inbal Ronen and others. 2011. *Metrics and Requirements Update for Employee Use Case*. ROBUST EU Project Deliverable. http://www.robust-project.eu/results/metrics-and-requirements-update-for-employee-use-case/at_download/file
[40] Matthew Rowe and Harith Alani. 2012. What makes communities tick? Community health analysis using role compositions. In *International Conference on Social Computing*. IEEE.
[41] Matthew Rowe, Miriam Fernandez, Sofia Angeletou, and Harith Alani. 2013. Community analysis through semantic rules and role composition derivation. *Journal of Web Semantics* (2013).
[42] Sunanda Sangwan. 2005. Virtual community success: A uses and gratifications perspective. In *Hawaii International Conference on System Sciences*. IEEE.
[43] Felix Schwagereit, Sergej Sizov, and Steffen Staab. Finding optimal policies for online communities with cosimo. In *Web Science Conference*.
[44] Trent J. Spaulding. 2010. How can virtual communities create value for business? *E-Commerce Research and Applications* (2010).
[45] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* (2010).
[46] S. L. Toral, M. Rocío Martínez-Torres, F. Barrero, and F. Cortés. 2009. An empirical study of the driving forces behind online communities. *Internet Research* (2009).
[47] Saskia Vola. 2010. *Simple Metrics for Text Mining*. Technical Report. Temis Deutschland GmbH. http://saskia-vola.com/simple-metrics-for-textmining/
[48] David Wagner, Alexander Richter, Matthias Trier, and Heinz-Theo Wagner. 2014. Towards a Conceptualization of Online Community Health. In *Inform. Systems*.
[49] Xiaoqing Wang and Shannon Lantzy. 2011. A Systematic Examination of Member Turnover and Online Community Health. In *ICIS*.
[50] G. Whyte, A. Bytheway, and C. Edwards. 1997. Understanding user perceptions of information systems success. *Journal of Strategic Information Systems* (1997).
[51] Anbang Xu, Jilin Chen, Tara Matthews, and others. 2013. CommunityCompare: visually comparing communities for online community leaders in the enterprise. In *SIGCHI Conference on Human Factors in Computing Systems*. ACM.
[52] Colleen Young. 2013. Community Management That Works: How to Build and Sustain a Thriving Online Health Community. *Medical Internet Research* (2013).
[53] Lixiu Yu, Paul André, Aniket Kittur, and Robert Kraut. 2014. A comparison of social, learning, and financial strategies on crowd engagement and output quality. In *Computer supported cooperative work & social computing*. ACM.
[54] Y. Zheng, K. Zhao, and A. Stylianou. 2009. Information quality and system quality in online communities: an empirical investigation. *SIGHCI* (2009).
[55] Haiyi Zhu, Robert E. Kraut, and Aniket Kittur. 2014. The impact of membership overlap on the survival of online communities. In *SIGCHI Conference on Human Factors in Computing Systems*. ACM.