Reputation Through Observation: Active Lurkers in an Online Community

Clemens Niemeyer and Mirco Schoenfeld

Abstract Lurkers are the invisibile majority in a typical online community: Users that silently observe, consume, and become accustomed to a community without interacting actively. At some point in time, a small fraction of lurkers decides to start taking part in a community in some way. In this paper, we investigate the implications of lurking for the interactions of such newly-active users or *active lurkers*. In our analysis, we focus on a sub-community of the well-known Online Social Network (OSN) Reddit and track linguistic development of users' comments as well as the development of user's reputation. We analyze and compare the complete lifecycles of two types of users – active lurkers and non-lurkers. Our work gives new insights into the effects of lurking with respect to linguistic adaption of community habits and to reputation active lurkers are able to gain. In general, most influential and innovative contributions were submitted by former lurkers.

Clemens Niemeyer
Bavarian School of Public Policy at Technical University of Munich, 80333 Munich, Germany
✉ clemens.niemeyer@gmail.com

Mirco Schoenfeld
University of Bayreuth, 95448 Bayreuth, Germany
✉ mirco.schoenfeld@uni-bayreuth.de

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 5, No. 1, 2018

DOI: 10.5445/KSP/1000087327/32
ISSN 2363-9881
(cc) BY-SA

1 Introduction

Lurkers are the invisible majority in a typical online community. In fact, only 10 % of all registered users actively participate leaving the vast majority of around 90 % as lurkers (Preece et al, 2004).

What lurkers actually do while lurking is still an open question. Some see lurkers as silent participants benefiting selfishly. Ohers consider lurking to be an equally worthy form of participation that is justified by personality traits of the people behind user profiles.

We argue that lurkers learn while lurking: For members of a community it is common to adapt to a characteristic language or *slang*. Speaking this slang then identifies the speaker as a member and distinguishes a community from others in a hurtless yet unmistakable way. Hence, learning a slang is crucial to becoming a part of a community. While some learn while speaking, lurkers learn through observation.

In this work, we will focus on lurkers who decide to actively participate at some point in time – the *active lurkers* what we will call them. We investigate if these active lurkers are initially better adapted to the slang of a community. We will further analyze what they contribute to a community, if they help innovating a community, and if they receive more attention than others.

Therefore, we employ the concept of user lifecycles (Wang and Yu, 2012) and divide the number of users' contributions over time into buckets of consistent levels of activity. This allows us to follow users over their whole active life, and, more importantly, compare the effects of lurking on all stages of users' active lifes.

We find that the knowledge gained while lurking distinguishes active lurkers substantially from users that immediately start taking an active part in a community. Also, active lurkers impel linguistic innovation, create influential content, and gain more reputation throughout their active lifes than others.

The remainder of this work is organized as follows. The following Section 2 illustrates our concept at a glance. Section 3 specifies our approach to identifying active lurkers and a community's characteristic use of language. Section 3 starts by describing our dataset compiled from the Online Social Network Reddit. Based on specific characteristics of that social network, we give an appropriate definition of lurkers and active lurkers. In Section 4, we describe our findings.

Before Section 6 concludes the paper, Section 5 mentions important related work in the field.

2 Concept

Our aim is to investigate the effect of lurking on users of Social Media throughout their lifecycle. There is evidence from the literature that lurking is a form of learning quietly (Dokhanchi et al, 2018; Panciera et al, 2010; Wang and Yu, 2012). We argue that lurkers on Social Media learn and adapt the characteristic language of a community – its *slang*. We further claim that it is this phase of learning that allows former lurkers to become accepted and reputable members of their community. Apart from Social Media, socialization through use of a specific language is a well-researched field (Goodwin and Kyratzis (2011); Leona (1978)).

This requires us to analyze two aspects of users' contributions: First, we have to evaluate contributions in terms of expressed language to see whether users have adapted the language of their community at all and to what extent, and, second, we have to evaluate the acceptance or reputation of these contributions to see which of them cause more reactions.

The aspect of adapting a certain language will be evaluated using perplexity as a distance measure between language models. Therefore, we sample a snapshot of the language expressed at a certain point in time and compare it to the language expressed in users' contributions. By obtaining perplexity between these two language models we can draw conclusions about the level of adaptation of a certain language. By measuring perplexity between such two language models of different points in time we can further conclude about the level of innovativeness or conservativeness of a user's language, i.e., individual language models being closer to the communitiy's language models of the future indicates users being more innovative because their language seems to be adapted by the rest of the community. The precise specification of our approach is given in Section 3.4.

The aspect of gaining reputation in a community will be evaluated using techniques of Social Network Analysis, namely the PageRank algorithm. The structure of discussions on Social Media often allows for obtaining network representations of interactions. From such interaction graphs, we are able to derive the reputation certain contributions are able to gain as PageRank centrality values. Aggregating these centrality values for users then yields a certain reputation a single user is able to gain. The specification of our approach is given in Section 3.5.

In order to focus on the effects of lurking, we focus on two types of users: The *active lurkers* and the non-lurkers. We consider non-lurkers to be users that start contributing immediately after registering with a user account. The definition of active lurkers is given in Section 3.2.

As stated above, our aim is to investigate the effect of lurking on the entire lifecycles of users. That means to focus on users who left the community at some point in time. By doing so, we are able to investigate the differences in the development of the users' language and reputation over time.

The next Section elaborates on the details of our approach starting with a description of our dataset and a precise definition of active lurkers.

3 Active Lurkers on Reddit

In this section, we give a precise definition of identifying active lurkers and measuring their involvement in a community in terms of language and reputation. We begin by introducing our dataset, subsequently describe how we define and identify active lurkers, and, finally, how we analyze their language.

3.1 Data Source: Reddit

In this work, we will consider user-generated texts taken from the Online Social Network (OSN) Reddit[1]. On Reddit, user-generated content is publicly available and usually consists of text which is organized into so-called subreddits. In a subreddit, users rendezvous with others of similar interest. There, they create posts, comment on other posts, rate, and organize content. Since we focus on adapting a certain community-specific language, we only consider posting and commenting as relevant form of interaction.
For the purpose of evaluation, we have chosen a subreddit targeting a community with a special interest in photography, called */r/photography*. We consider this a

[1] http://www.reddit.com/

community with a specialized vocabulary since discussions are mainly about technical details of equipment as well as about technical details of photographies.

From this subreddit, we crawled all user-generated posts and comments since foundation of the community in January, 2011 up to December, 2016. During that time we identified 18009 unique user accounts actively posting and commenting. Since during the first half of 2011 the community was considered too small (less than 500 different accounts were interacting) the first month of interest is July, 2011. Hence, our analysis is based on 5.5 years of user-generated content from */r/photography*.

During that time, we observed a total number of 1965664 posts and comments. Starting in July, 2011, we count a total number of 102865 posts and comments for the second half of the year. This number steadily increased over time to reach a total number of 460886 posts and comments in 2016.

3.2 The Active Lurker

As there is no single definition of lurking in the literature, we define a lurker as a registered user account who simply is not interacting in the community. An *active lurker* is a lurker who waited at least 60 days after registration before posting or commenting for the first time.

On Reddit, users do not have to register to certain subreddits but they register to the website in general. This means, that a user who we identify as an active lurker of */r/photography* may have been active in other subreddits in less than 60 days after registration. In the context of Reddit, we do not see this as contradicting our definition of lurkers or active lurkers: We consider subreddits as distinct communities spawning characteristic use of language (Tran and Ostendorf, 2016; Zhang et al, 2017). As such, turning towards a new subreddit means approaching a new community.

At this point, it is important to mention a limitation of our computational approach: We can not measure the extent to which an active lurker actually observed the new community prior to initial interaction. This would require at least a survey among the identified user accounts.
Figure 1 depicts distribution of days passed after registration until users posted or commented first in *r/photography*. This distribution does not cover all 18009

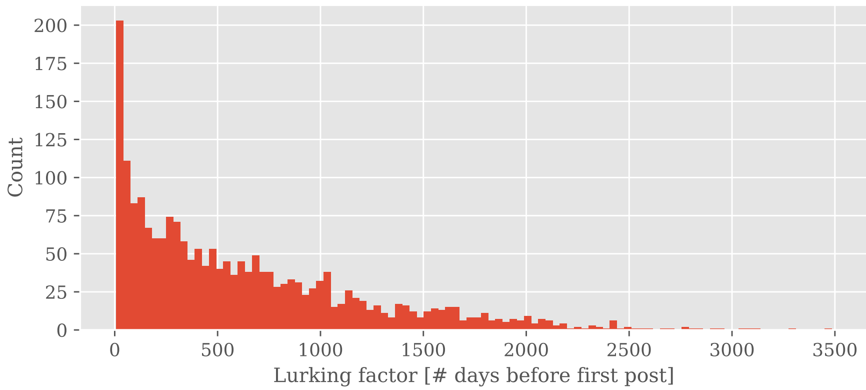

Figure 1: Distribution of observed lurked days. According to the median, 461.0 days passed before users posted or commented first in *r/photography* (mean: 620.7 days). More than one half of all user accounts of interest wait more than a year before posting.

identified user accounts: The following section will elaborate on how we selected users of interest.

3.3 User Lifecycles And Users of Interest

Our work aims at analyzing the complete active life of users to identify effects of lurking on all stages of a user's life. Hence, we want to concentrate on users who both began and stopped participating as active members in the above-mentioned community between July, 2011 and December, 2016.

The first interaction in a community is clearly identifiable, especially since our analysis of */r/photography* starts at its very beginning. What is considered to be the last interaction is often defined by researchers, though. Usually, a certain time has to pass after the last observed interaction to assume a user has abandoned a community. Here, we follow this procedure and define this time span to be 6 months. The timespan of activity is defined to be the time between a user's first and last interaction as long as a user has been posting or commenting in at least 6 different months. These restrictions ensure that we focus on users who have been (mentally) concerned with a community for a considerable amount of time.

Out of the above-mentioned 18009 user accounts 1624 fulfil our restrictions and were deemed relevant for our analysis: These 1624 accounts were created and abandoned during the observed time meaning they were created after July, 2011 and stopped interacting before end of June, 2016. In total, they provided 531450 posts during the period that our analysis is based on.

To account for different levels of activity of users we utilize the concept of lifecycles by averaging the number of interactions of a user over the number of months that the user was active. This yields a user lifecycle with consistent levels of activity. Most importantly, this allows for the comparison of users with different total life times.

3.4 Linguistic Innovativeness

A well-known indicator for a persons' integration into a community is language. Section 5.2 will elaborate this further.

In order to measure change in linguistic habits over time, we take regular *snapshots* of expressed language by randomly choosing 500 users each month. This sample is drawn from the entire set of users who posted in a particular month. Among these users, we randomly selected two of their posts as long as they had at least 4 visible interactions in that particular month and the chosen posts consisted of at least 30 words and at least two sentences. This yields a monthly set of 1000 comments with sufficient amount of text which is a procedure that has been proven to be appropriate before (Danescu-Niculescu-Mizil et al, 2013). Deleting URLs from these comments and replacing Reddit-specific formatting commands were both part of pre-processing that was applied both to comments building the monthly snapshot as well as comments used at other points in the process.

From these monthly snapshots, however, we build trigram-based language models with Kneser-Ney-Smoothing without pruning (Kneser and Ney, 1995). While several other procedures exist in the literature, Kneser-Ney-Smoothing is regarded as a de facto standard since, for one thing, it performs quite efficiently (Heafield et al, 2013) and, for another thing, this combination has been shown to yield very good results in general as well as for the amount of data that is expected here (Chen and Goodman (1996); Goodman (2001)).

From any such snapshot of linguistic state of the community we quantify the difference to any other post in the dataset using perplexity (Jelinek et al, 1977). The result is an information theoretic abstraction of how well a snapshot is able to predict a given post. In the context of this work, it can be interpreted as a quantification of how well a user has incorporated the language of the community. In general, the perplexity of a model q is given by

$$\mathrm{PP}(x_i) = b^{-\frac{1}{N}\sum_{i=1}^{N}\log_b q(x_i)}, \tag{1}$$

where customarily $b = 2$. In this general form, x_i represents a test sample drawn from a distribution and $PP(x_i)$ yields how well a model q is able to predict x_i. In Natural Language Processing, it is usually applied as perplexity per word $\mathrm{PP}_W(s)$ for a given sentence s

$$\mathrm{PP}_W(s) = \frac{\mathrm{PP}(s)}{|s|}, \tag{2}$$

with s consisting of words $\{w_0, \cdots, w_i\}$ and $|s|$ being the length of s. In this work, we compute a perplexity per post $\mathrm{PP}_P(p)$ as a perplexity per word $\mathrm{PP}_W(s)$ for every sentence s_i in post p that is then normalized with the number of sentences in p:

$$\mathrm{PP}_P(p) = \frac{\sum_i \mathrm{PP}_W(s_i)}{|p|}, \tag{3}$$

with p consisting of several sentences $\{s_0, \cdots, s_i\}$. For this work, we will utilize perplexity per post PP_P referring to it as perplexity if not stated otherwise.

By selecting models and posts from different points in time, perplexity can be interpreted as progressivity of language and, thereby, providing a measurement for innovativeness of posts. If, for example, perplexity between a post submitted at time t and a language model from a future snapshot $t+1$ is lower than perplexity between that post and the language model of the corresponding snapshot t, one could see that post as promoting linguistic innovation for that community. We derive linguistic innovativeness $PG(p_u)$ of a post p_u by considering 6 past and 6 future snapshots such that

$$PG(p_u) = \sum_{i=m(p_u)-6}^{m(p_u)-1} \begin{cases} -1 & \text{if}(PPL_i(p_u) < PPL_{m(p_u)}(p_u) \\ 0 & \text{otherwise} \end{cases} + \sum_{i=m(p_u)+1}^{m(p_u)+6} \begin{cases} +1 & \text{if}(PPL_i(p_u) < PPL_{m(p_u)}(p_u) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $m(p_u)$ is the month in which user u submitted post p_u, $PPL_{m(p_u)}(p_u)$ is the perplexity of post p_u compared to its original language model, and $PPL_i(p_u)$ is the perplexity of the same post compared to language model i. This yields a value of progressivity for a post ranging between $[-6, +6]$ where -6 would be interpreted as a very conservative post since it shows a lower perplexity to all language models from the previous 6 months. A value of +6 on the other hand would correspond to a very progressive post since it is closer to all 6 future language models in terms of perplexity.

3.5 A User's Reputation

To get a sense of a user's status in a community, it is convenient to analyze the structure of the community. As it will be pointed out in section 5.2, graph-based centrality measures can be applied to translate the complex structure of a community into a ranking of members. There exists a vast amount of different centrality measures all of which allow for different interpretation of the resulting rankings depending on how the underlying structure is calculated, respectively. For our work, we carefully selected the well-known PageRank algorithm (Page et al, 1999) because it fits best to our dataset and to ourto our overall aim.

The graph structure of a subreddit can be modelled from posts and comments resulting in an interaction graph. That is a graph $G(V, E)$ consisting of a set of vertices V and a set of edges E in which edges represent actual interactions (other than, for example, self-reported friendship ties as in social graphs) and vertices represent users responsible for the respective interaction. An edge $e(u, v)$ between two users u and v is only established if user u commented on a post (or a comment) of user v implicitly modeling the direction of interaction. Consequently, this directed graph supports a natural interpretation in terms of user-status: The more reactions a user receives and the more comment chains originate in a user's post, the higher that user's reputation. This is exactly what the PageRank algorithm expresses.

To account for the temporal dynamics of the community in question, we build separate interaction graphs for posts and comments from every month in the dataset and derive PageRanks of users from every resulting snapshot separately. This allows us to track the development of reputation users receive throughout their lifecycles.

4 Evaluation

In this section, we present our findings on how active lurkers drive linguistic innovation in a community and how lurking has an effect on the reputation active lurkers are able to gain throughout their user life.

4.1 Reputation Over Time

As covered in section 3.5, we model interaction graphs from monthly snapshots of the community in order to track a users' status throughout her lifecycle.

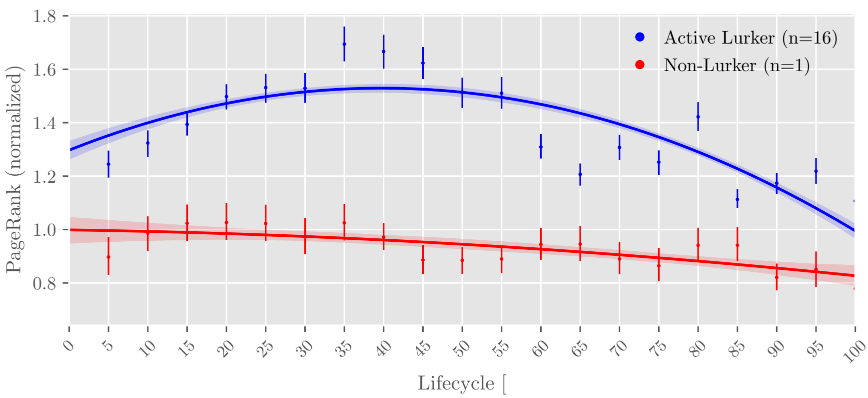

Figure 2: Development of normalized PageRanks for active lurkers and non-lurkers.

The generated graphs contain 196.24(±55.94) nodes on average as well as 1460.67(±471.76) edges, around 47 % of all nodes are connected in cohesive triangles, the strongest connected component contains 98.1 % of all nodes, and 71.6 % of all connections are symmetrical – all of which are typical characteristics of social network graphs (e.g. Granovetter (1973)). It can be concluded that the generated graphs are typical social networks that allow further analysis.

Figure 2 depicts the development of normalized PageRanks for active lurkers and non-lurkers separately. The Figure depicts deviations of the mean: A normalized PageRank of 2 corresponds to an actual PageRank twice as high as the average PageRank observed in that month. The dots and bars represent distributions of values describing their mean and standard deviation. To support the interpretation of the Figure, we added a smoothing curve. The smoothing is based on the assumption of unimodality among the development of reputation. This is indeed a very basic assumption of the development of reputation over time. However, even if there are several points in time at which users might gain a notable reputation, we consider it sufficient to describe the lifecycle as a unimodal course since this allows us to estimate a general trend. The blue curve in Figure 2 representing the active lurkers implies several peaks of reputation, namely at 35%, 80%, and 95% of the users' lifecycle. However, taking into account the smoothed curve a general trend of a period of increasing reputation and a period of decreasing reputation separated by a single peak appears as a valid assumption.

Further, former lurkers receive significantly higher reputation than non-lurkers which peaks at a factor of around 1.6 during the zenith of their life. Also, for non-lurkers, there is no visible phase of gaining reputation at all. Instead, non-lurkers are generally not successful in gaining reputation throughout their active life as the descending trend shows. These differences between active lurkers and non-lurkers are significant in terms of a t-test taking into account different sample sizes ($t = -51.79, p < 0.001$).

It is worth noting that the this general development can not be caused by different levels of activity such as a decreasing number of posts during later stages since the absolute number of posts is normalized throughout the active time of each users' life.

4.2 Linguistic Innovation

Innovativeness of language is expressed in terms of perplexity between a post and language models from that post's past and future. The closer a post is to language models from its past the more conservative, and the closer a post is to language models of its immediate future the more progressive it is.

Calculating progressivity for all posts of a user following formula (4), normalizing progressivities over the snapshots that user was active, and repeating both steps for all users in the dataset results in Figure 3. Again, users are classified into active lurkers and non-lurkers.

It can be seen, how, in their early stages of life, active lurkers indeed use a more progressive language than non-lurkers. But, beginning by around 15 % of their life, this progressiveness levels out. From that point on, progressivity of both active lurkers and non-lurkers oscillates rather similar yielding a negative trend.

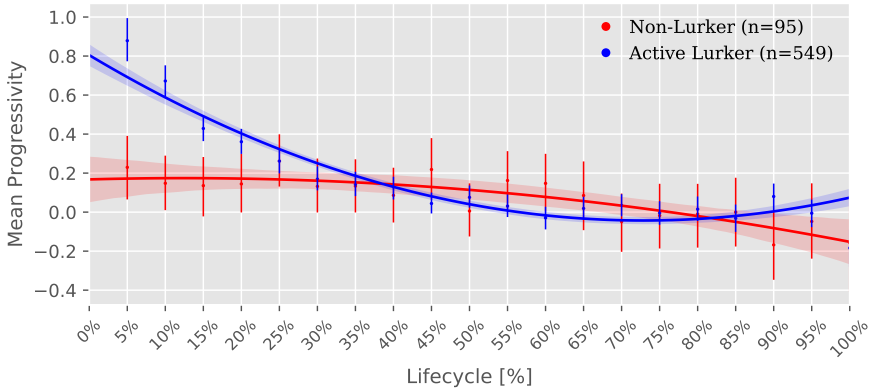

Figure 3: Progressivity of language of active lurkers and non-lurkers.

This observation slightly contradicts the expectation that active lurkers initially are more familiar with linguistic habits of communities than non-lurkers (Honeychurch et al, 2017). For that perspective, a progressivity of ≈ 0 would have been supportive. This means that active lurkers are mastering characteristic language right from the beginning. But, in fact, active lurkers show a mean progressivity of ≈ 1 meaning they rather speak a progressive language and have

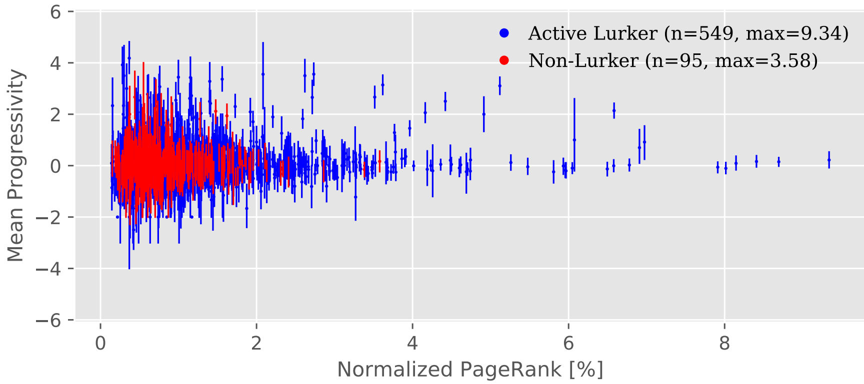

Figure 4: Progressivity and involvement of active lurkers in most influential posts.

a noticeable influence on the language of the community. This implies that long-time observation allows former lurkers to impel linguistic innovation in a community at least in the beginning of their active life.

For another view on linguistic innovation, Figure 4 depicts the share of active lurkers and non-lurkers in the most influential posts found in the dataset. It can be seen how posts with highest reported normalized PageRanks are rather progressive in terms of perplexity. Most importantly, all most influential and most progressive posts were contributed by active lurkers. The differences in progressivity between active lurkers and non-lurkers are significant in terms of a t-test taking into account different sample sizes ($t = 3.37$, $p < 0.001$).

5 Related Work

This section provides a brief overview over existing work regarding lurkers in online communities as well as Natural Language Processing (NLP) for online communities.

5.1 Lurking

So far, a lot of research has been devoted to comprehend a lurkers' motivation to passively participating in a community as well as to comprehend the effects of lurking (Sun et al (2014); Preece et al (2004); Lampe et al (2010)). It has been shown that lurkers actually benefit from observing a community in terms of information processing, learning, and even influencing their offline peers (van Uden-Kraan et al (2008); Honeychurch et al (2017); Takahashi et al (2003); Dokhanchi et al (2018); Wang and Yu (2012)).

Based on findings of importance of lurkers, first efforts can be seen to computationally encourage lurkers to turn into active participants by recommending most relevant content (Interdonato et al, 2015).

For this work, however, it is important to note that lurkers learn and benefit from observing, and extravert their positive community-experience to influence others.

5.2 Language And Community

In the context of community analysis, language has been shown to guide inclusion, characterize communities, and even form and solidify roles in communities. In fact, there is a distinct research area that focuses on interrelations between language and community called socio-linguistics.

According to influential works from that field, language reflects structural change in networks (Milroy and Margrain, 1980) and strong communities as stagnating networks lead to a conservative language (Lippi-Green, 1989). Such interrelations even hold in online communities (Paolillo (1999); Tagliamonte and Denis (2008)). Expressed language is even an indicator to identify roles of users in online communities (Buntain and Golbeck, 2014).

Another wide-spread approach in mining communities and relations between users is graph-based social network analysis (Wasserman and Faust, 1994). A vast amount of research utilizes generalized graph methods to identify important nodes, subgroups in communities, roles of users, and other measurements to comprehend embeddedness of users. For the context of this work, we will make use of the popular PageRank algorithm which can be used to translate a directed graph structure into a ranking of the reputation of nodes (Page et al, 1999).

6 Conclusion

An active lurker is a user who waited a considerable amount of time before starting to interact publicly in terms of posting or commenting. We found that lifecycles of active lurkers differ significantly from lifecycles of users who start interacting immediately after joining a community. First, throughout their active life, they manage to gain much more reputation than non-lurkers with a clear peak in zenith of their active lifes. Second, and this is an aspect worth investigating further, in the early phase of their active lifes, active lurkers drive linguistic innovation in a community. In general, most influential and innovative contributions were submitted by former lurkers. Hence, motivating lurkers to taking part actively could be a way to maintain vivacity and vividness of a community.

References

Buntain C, Golbeck J (2014) Identifying social roles in Reddit using network structure. In: Proceedings of the 23rd International Conference on World Wide Web, Broder A, Shim K, Suel T (eds), ACM, New York, NY, USA, WWW '14 Companion, pp. 615–620. DOI: 10.1145/2567948.2579231.

Chen SF, Goodman J (1996) An Empirical Study of Smoothing Techniques for Language Modeling. In: Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, Joshi A, Palmer M (eds), Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '96, pp. 310–318. DOI: 10.3115/981863.981904.

Danescu-Niculescu-Mizil C, West R, Jurafsky D, Leskovec J, Potts C (2013) No country for old members: User lifecycle and linguistic change in online communities. In: Proceedings of the 22nd international conference on World Wide Web, Moon S, Baeza-Yates R (eds), ACM, pp. 307–318. DOI: 10.1145/2488388.2488416.

Dokhanchi M, Kavanagh L, Reidsema C, et al (2018) Factors that influence peer learning in social media enhanced engineering courses. In: 29th Australasian Association for Engineering Education Conference 2018 (AAEE 2018), Engineers Australia, p. 183. ISBN: 978-1-925627-36-7.

Goodman JT (2001) A bit of progress in language modeling. Computer Speech & Language 15(4):403–434. DOI: 10.1006/csla.2001.0174.

Goodwin MH, Kyratzis A (2011) Peer language socialization. In: The handbook of language socialization, chap. 16, pp. 365–390. Wiley Online Library, Duranti A, Ochs E, Schieffelin BB (eds). DOI: 10.1002/9781444342901.ch16.

Granovetter MS (1973) The strength of weak ties. American journal of sociology 78(6):1360–1380. DOI: 10.2307/2776392.
Heafield K, Pouzyrevsky I, Clark JH, Koehn P (2013) Scalable Modified Kneser-Ney Language Model Estimation. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Schuetze H, Fung P, Poesio M (eds), Sofia, Bulgaria, pp. 690–696. URL: `https://www.aclweb.org/anthology/P13-2121`.
Honeychurch S, Bozkurt A, Singh L, Koutrapoulos A (2017) Learners on the periphery: Lurkers as invisible learners. European Journal of Open, Distance and E-Learning 20(1):191–211. DOI: 10.1515/eurodl-2017-0012.
Interdonato R, Pulice C, Tagarelli A (2015) "Got to Have Faith!": The DEvOTION Algorithm for Delurking in Social Networks. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Pei J, Silvestri F, Tang J (eds), ACM, New York, NY, USA, ASONAM '15, pp. 314–319. DOI: 10.1145/2808797.2809394.
Jelinek F, Mercer RL, Bahl LR, Baker JK (1977) Perplexity – a measure of the difficulty of speech recognition tasks. The Journal of the Acoustical Society of America 62(S1):S63–S63. DOI: doi.org/10.1121/1.2016299.
Kneser R, Ney H (1995) Improved backing-off for M-gram language modeling. In: 1995 International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 181–184vol.1. DOI: 10.1109/ICASSP.1995.479394.
Lampe C, Wash R, Velasquez A, Ozkaya E (2010) Motivations to Participate in Online Communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Fitzpatrick G, Hudson S, Edwards K, Rodden T (eds), ACM, New York, NY, USA, CHI '10, pp. 1927–1936. DOI: 10.1145/1753326.1753616.
Leona MH (1978) An examination of adolescent clique language in a suburban secondary school. Adolescence 13(51):495, Libra Publishers. ISSN: 0001-8449.
Lippi-Green RL (1989) Social network integration and language change in progress in a rural alpine village. Language in Society 18(2):213–234. DOI: 10.1017/S0047404500013476.
Milroy L, Margrain S (1980) Vernacular language loyalty and social network. Language in Society 9(1):43–70. DOI: 10.1017/S0047404500007788.
Page L, Brin S, Motwani R, Winograd T (1999) The pageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. URL: `https://www.semanticscholar.org/paper/The-PageRank-Citation-Ranking-to-Page-Brin/eb82d3035849cd23578096462ba419b53198a556`.
Panciera K, Priedhorsky R, Erickson T, Terveen L (2010) Lurking? Cyclopaths? A Quantitative Lifecycle Analysis of User Behavior in a Geowiki. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Fitzpatrick G, Hudson S, Edwards K, Tom R (eds), Association for Computing Machinery, New York, NY, USA, CHI '10, pp. 1917–1926. DOI: 10.1145/1753326.1753615.

Paolillo J (1999) The virtual speech community: Social network and language variation on IRC. Journal of Computer-Mediated Communication 4(4):0–0, Blackwell Publishing Ltd. ISSN: 1083-6101, DOI: 10.1111/j.1083-6101.1999.tb00109.x

Preece J, Nonnecke B, Andrews D (2004) The top five reasons for lurking: improving community experiences for everyone. Computers in Human Behavior 20(2):201–223. DOI: 10.1016/j.chb.2003.10.015.

Sun N, Rau PPL, Ma L (2014) Understanding lurkers in online communities: A literature review. Computers in Human Behavior 38:110–117, Amsterdam. DOI: 10.1016/j.chb.2014.05.022.

Tagliamonte SA, Denis D (2008) Linguistic ruin? Lol! Instant messaging and teen language. American Speech 83(1):3–34. DOI: 10.1215/00031283-2008-001.

Takahashi M, Fujimoto M, Yamasaki N (2003) The Active Lurker: Influence of an In-house Online Community on Its Outside Environment. In: Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work, Tremaine M, Simone C (eds), ACM, New York, NY, USA, GROUP ’03, pp. 1–10. DOI: 10.1145/958160.958162.

Tran T, Ostendorf M (2016) Characterizing the Language of Online Communities and its Relation to Community Reception. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Su J, Duh K, Carreras X (eds), pp. 1030–1035. DOI: 10.18653/v1/D16-1108.

van Uden-Kraan C, Drossaert C, Taal E, Seydel E, van de Laar M (2008) Self-reported differences in empowerment between lurkers and posters in online patient support groups. Journal of medical Internet research 10(2):e18. DOI: 10.2196/jmir.992.

Wang X, Yu Y (2012) Classify participants in online communities. International Journal of Managing Information Technology 4(1):1, Academy & Industry Research Collaboration Center (AIRCC). DOI: 10.5121/ijmit.2012.4101.

Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge. ISBN: 978-0-521387-07-1.

Zhang J, Hamilton WL, Danescu-Niculescu-Mizil C, Jurafsky D, Leskovec J (2017) Community Identity and User Engagement in a Multi-Community Landscape. In: Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM), NIH Public Access, Vol. 2017, p. 377.