Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM 2021)

Expertise and Dynamics within Crowdsourced Musical Knowledge Curation: A Case Study of the Genius Platform

Derek Lim, Austin R. Benson
Cornell University
dl772@cornell.edu, arb@cs.cornell.edu

Abstract

Many platforms collect crowdsourced information primarily from volunteers. As this type of knowledge curation has become widespread, contribution formats vary substantially and are driven by diverse processes across differing platforms. Thus, models for one platform are not necessarily applicable to others. Here, we study the temporal dynamics of Genius, a platform mainly designed for user-contributed annotations of song lyrics. A unique aspect of Genius is that the annotations are extremely local — an annotated lyric may just be a few lines of a song — but also highly related, e.g., by song, album, artist, or genre.

We analyze several dynamical processes associated with lyric annotations and their edits, which differ substantially from models for other platforms. For example, expertise on song annotations follows a "U shape" where experts are both early and late contributors with non-experts contributing intermediately; we develop a user utility model that offers one possible explanation for such behavior. We also find several traits appearing early in a user's lifespan of contributions that distinguish (eventual) experts from non-experts. Combining our findings, we develop a model for early prediction of user expertise.

1 Crowdsourced Lyric Annotation

"Lookin' around and all I see
Is a big crowd that's product of me"

— Kendrick Lamar, *A.D.H.D*

Online platforms for crowdsourced information such as Wikipedia and Stack Exchange provide massive amounts of diverse information to people across the world. While different platforms have varying information structure and goals, they share a fundamental similarity: the source of the content is a community of users who contribute their time and expertise to curate knowledge. Understanding the users that enable the success of these platforms is vital to ensure the continual expansion of their utility to the general public, but doing so requires special consideration of the particular structure and form of information that users contribute.

The activity and expertise distribution amongst users on these crowdsourced information platforms is often heavy-tailed, and there is substantial effort in understanding the sets of users that make meaningful and/or voluminous contributions (Pal et al. 2011; Movshovitz-Attias et al. 2013), and how contributions change over time (Anderson et al. 2012). One example problem is expertise detection (Zhang, Ackerman, and Adamic 2007). On platforms such as Quora and the TurboTax live community, experts are not clearly defined but longitudinal data can help identify experts (Pal et al. 2011; Patil and Lee 2015). In contrast, Stack Exchange users accumulate explicit reputation scores over time, and there is a focus on early determination of user expertise, where ex ante predictions leverage user behavior and their temporal dynamics (van Dijk, Tsagkias, and de Rijke 2015; Pal, Harper, and Konstan 2012).

Here, we take expertise detection and characterization as test problems to study the temporal dynamics of user contributions on *Genius* (`genius.com`), a platform primarily for crowdsourced transcription and annotation of song lyrics. Genius was launched as *Rap Genius* (after a brief stint as *Rap Exegesis*) in 2009, focusing on the lyrical interpretation of English rap songs.[1] Since then, Genius has grown to encompass lyrics in many genres and languages, and `alexa.com` ranks the web site 474th in terms of global engagement as of April 2021.[2]

On Genius, lyrics are originally sourced from users who work together to transcribe songs. After, users annotate the lyrics for interpretation, and the annotations are edited over time (Fig. 1). The structure of annotations on Genius differs substantially from other large, popular crowdsourced platforms such as Stack Exchange, Quora, Yelp, and Wikipedia. For one, Genius has no explicit question-and-answer format. Instead, the transcription of newly released songs offers an implicit question generation mechanism, namely, what is the meaning of these lyrics? Similar to Wikipedia, annotations are continually edited to provide a singular authoritative narrative. However, annotations are extremely localized in the song — an annotated lyric may be just a couple of bars in a rap song or the chorus of a folk song. Still, annotations are related within lyrics of the same song, album, artist, or genre. Finally, the content is much less formal than Wikipedia, as annotations often contain slang, jokes, and profanity.

The unique structure of Genius leads to user dynamics that

[1] https://genius.com/Genius-about-genius-annotated
[2] https://www.alexa.com/siteinfo/genius.com

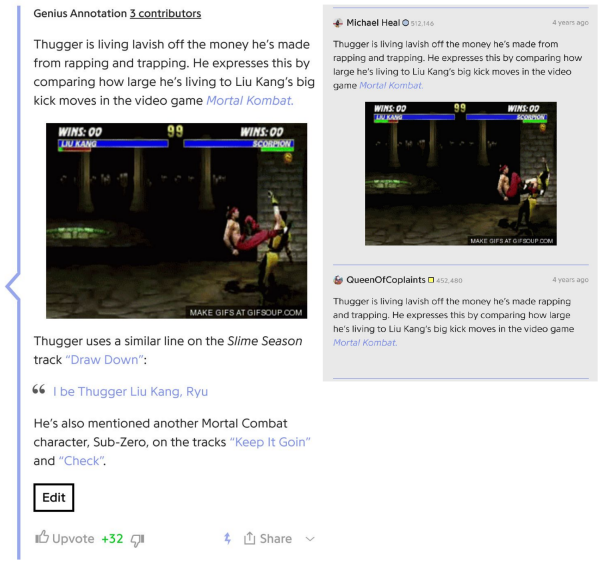

Figure 1: Screenshot of an annotation of the lyrics: "I'm livin' big, I swear to God I'm Liu Kang kickin'" in Young Thug's song *Digits* (https://genius.com/8869644). (Left) The annotation as of May 2020. (Right) Two edits of the annotation. The bottom is the second edit and the top is the third. This annotation has contributions from high-IQ users and contains rich structure such as hyperlinks, a quote, and an image — all examples of what we call *quality tags* (Section 3.1).

do not align with existing models for other crowdsourced information sites. In particular, the partitioning of a song's lyrics into smaller segments where users choose to annotate or edit content is unique. While two answers on the same Stack Overflow question may cover different aspects of the same question, they still result from the same prompt (the question). Likewise, two Amazon reviews on different properties of the same product originate from the same prompt (the product). However, two annotations on the same song on Genius have different prompts — namely, the two different lyric segments with which each annotation is associated. There is also some notion of competition and thus congestion in the choice of lyric segments to annotate, since there may only be one annotation per lyric segment, which contrasts with the unbounded number of possible answers on Stack Overflow or reviews on Amazon products. Moreover, while edits on the same annotation come from the same prompt (same lyrics), the edits are related by the past state of the given annotation.

One part of our analysis is on user *IQ* — Genius's aggregate measure of activity and experience of users that is analogous to reputation on Stack Exchange. We measure IQ at roughly one point in time and use it as a proxy for (eventual) expertise or experience when analyzing user behavior retrospectively. While studies of other crowdsourced information sites apply to user behavior and IQ on Genius, the structure of annotations also necessitates specialized metrics. Thus, we define metrics of annotation quality, coverage of lyrics by annotations, and lyric originality, which help uncover key traits of user contribution behavior that distinguish experts.

We find several patterns in the temporal dynamics of annotations and subsequent edits on Genius that have not been observed in prior studies of other crowdsourced information platforms. One distinguishing set of characteristics are "U-shaped" relationships in annotations made on a song over time. For example, early annotations on a song are made by high-IQ experts (or eventually high-IQ users), intermediate annotations are made by low-IQ or inexperienced users, and the most recent annotations are again made by eventual high-IQ users. We conceptualize this through an *IQ diamond* ($\langle\rangle$), in which a song lyric travels from top to bottom, being considered by users that are experts or eventual experts in the narrow top and bottom, and by the bulk of lower-expertise users in the wider middle. The quality of annotations and originality of annotated lyrics also follow similar patterns: the earliest and latest annotations are on average of higher quality and on more original lyrics.

Our IQ diamond model for annotation dynamics contrasts sharply with answer arrival dynamics on Stack Overflow. For example, Anderson et al. (2012) described a "reputation pyramid" model of user organization in which questions are first answered by higher-reputation users and then by lower-reputation users. This Stack Overflow model does not work as a model for Genius annotations, as it does not explain the increasing IQ in the later song annotations. Furthermore, the model does not agree with the editing dynamics; later edits tend to be made by more experienced users and increase the quality of the annotation. Similarly, review quality on Amazon and Yelp has a negative relationship with time (Bai et al. 2018), which disagrees with the "U-shaped" relationship of Genius annotation quality with time and the positive relationship of Genius edit quality with time.

To explain the materialization of the IQ diamond and to further understand latent factors inducing user annotation patterns, we develop a model of user utility based on network effects and congestion. In this model, users of different experience levels gain different utility from creating new annotations on a song, given the fraction of lyrics that are already annotated. Fitting simple parametric functions for network effects and congestion matches empirical differences in user behavior between low-IQ and (eventual) high-IQ users. This model offers one explanation for behavior, and we discuss alternative ideas such as loyalty, which is a driver of behavior on Reddit (Hamilton et al. 2017).

Similar to studies on Stack Overflow (Movshovitz-Attias et al. 2013) and RateBeer (Danescu-Niculescu-Mizil et al. 2013; McAuley and Leskovec 2013), we also analyze user evolution stratified by eventual IQ levels. We find inherent traits of eventual experts visible in their early contributions: even in their first annotations on the site, eventual experts create higher quality annotations, are more often an early annotator on a song, are more often an early editor of an annotation, and annotate lyrics that are more original. We use these features to design a simple discriminative model that successfully makes early predictions of eventual super experts — users with very high IQ on the site.

1.1 Additional Related Work

Basic statistics of Genius have been reported with models for the trustworthiness of annotations (Al Qundus 2018; Al Qundus and Paschke 2018). Additionally, the platform is used for African American literature courses (Rambsy 2018) and for understanding music history (Dawson 2018). Other music databases such as CDDB/Gracenote, MusicBrainz, and Last.fm have been analyzed in the context of recommender systems and information retrieval (Swartz 2002; Celma 2010; Koenigstein, Dror, and Koren 2011; Schedl, Gómez Gutiérrez, and Urbano 2014). More broadly, numerous types of annotation systems have been studied (Kalboussi et al. 2015).

In this paper, we also analyze Genius in the context of other well-studied crowdsourced information sites, such as Stack Overflow (Ravi et al. 2014; Posnett et al. 2012; Tian, Zhang, and Li 2013), Quora (Wang et al. 2013; Maity, Sahni, and Mukherjee 2015), Yahoo! Answers (Adamic et al. 2008), Amazon (Bai et al. 2018), and Wikipedia (Beschastnikh, Kriplean, and McDonald 2008; Mesgari et al. 2015). The temporal dynamics of user activity on such sites have been studied in several contexts (Anderson et al. 2012; Jurgens and Lu 2012; Paranjape, Benson, and Leskovec 2017; Patil and Lee 2015; Almeida, Mozafari, and Cho 2007).

Expertise plays a central role in the study of crowdsourced information. For example, employee-labeled "superusers" have been studied in the TurboTax live community (Pal et al. 2011), and Quora's manually identified "Top Writers" can be identified by temporal behavior (Patil and Lee 2015). On Stack Overflow, early user activity and textual features have been used to predict eventual experts (Movshovitz-Attias et al. 2013; van Dijk, Tsagkias, and de Rijke 2015; Pal, Harper, and Konstan 2012).

2 Data Description

We crawled `genius.com` from September 2019 to January 2020. The data spans over 10 years of user activity, with the earliest activity in October 2009. We refer to any page that can be annotated as a *song*, even though a small fraction of these pages correspond to other content such as transcriptions of podcasts or parts of a movie. Users create *annotations* on specific lyric segments within a song. In this work we use "lyrics" to refer to general words in a song, and "lyric segment" to refer to the sequence of words that belong to a single annotation. A lyric segment may consist of multiple lines of lyrics, or it may be as small as just a part of a line in a song. When there are unannotated lyrics on a song page, a Genius user can choose a subset of unannotated lyrics to annotate (thus creating a lyric segment). After a user makes an initial *annotation*, users can create an *edit* for an annotation, and the most recent version of the annotation (with edits) is displayed on the web page. We typically use "edits" to refer to the changes in annotations and reserve the word "annotation" for the initial annotation.

We collected lyrics, annotations, and annotation edit history from 223,257 songs. Of these, 33,543 songs have at least one annotation. In total, we collected 322,613 annotations and 869,763 edits made by 65,378 users. For each annotated lyric segment, we have the complete timestamped history of edits and content. The annotations and edits consist of text and HTML. We find that the distribution of the number of annotations made by a user and the number of annotations on a song are heavy tailed (Fig. 2).

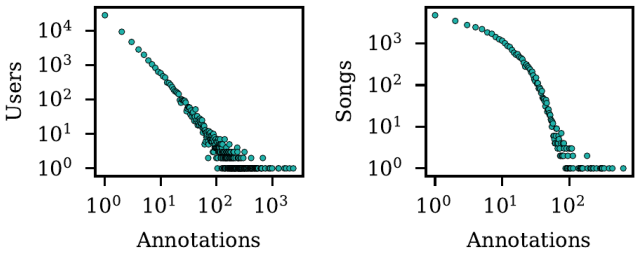

Figure 2: (Left) Distribution of annotation counts of users. (Right) Distribution of annotation counts on songs.

For each user, we have every annotation and edit that they have made on the collected songs. We also recorded the *IQ* of each user *at the time of our crawl*. IQ is an aggregate measure that accounts for various contributions on Genius, such as writing annotations and transcribing songs. As we only have snapshots of IQ and not the evolution of IQ over time, our analysis is based on the IQ that users had accumulated at the time of data collection, and we use this to study user behavior retroactively. Thus, when we refer to a high-IQ user in later sections, we mean that this user *eventually* accrues a large IQ score by the time that we collected the data. In other words, when analyzing user behavior retroactively, such users could have had high IQ at the time of an action, or they eventually earned high IQ by the time we collected the data. Users with low IQ have always had low IQ. When analyzing annotation dynamics and expertise in later sections, we often have to restrict to users with sufficient data (e.g., users with sufficiently many annotations) and these users are often old enough on the site so that their IQ has stabilized and their experience can be approximately measured by IQ (see user age statistics in Section 5). However, this type of analysis limits some of our user behavior findings, as we do *not* have IQ measurements at the time of the user activities, and this should be kept in mind when interpreting the results.

One way users accumulate IQ is when their annotation gets upvotes from other users.[3] Moreover, users can be given certain roles in the Genius community, such as Editor, Moderator, and Mediator; these roles come with certain permissions and signal certain accomplishments of the user.[4] The amount of IQ earned for an upvote is higher if that upvote comes from a user with one of these roles. An annotation made by a user can result in a net loss of IQ if it is sufficiently downvoted, but we do not know how much net IQ is gained per annotation or the users that vote on an annotation. We do not separate users by role, and we do not account for pressure against annotation due to possible loss of IQ.

In addition to annotations and edits, we also collected other Genius user actions such as suggestions, questions,

[3] https://genius.com/8839950

[4] https://genius.com/Genius-what-is-a-moderator-annotated

answers, comments, and transcriptions. We say that a user has *contributed* to a song if they have recorded some interaction with the song. However, our analysis focuses on annotations and edits. Although we do not have the exact fraction of user IQ that is obtained from annotations or edits, we believe that it accounts for a large portion of IQ for many users. User IQ and annotation counts are highly correlated ($r = 0.633$), and user IQ follows a similar distribution to that of user annotation counts.

All of the data, as well as code for reproducing the analyses in this paper, are available at:

https://github.com/cptq/genius-expertise.

3 Metrics for Annotations

The annotations that users create and edit on Genius are a unique type of crowdsourced contribution. In order to study user behavior and dynamics, we first define metrics for annotations related to quality, coverage, and originality.

3.1 Annotation Quality

To better understand the generation of content on Genius, we would like to quantify annotation quality. We simply take the number of HTML tags that indicate rich content creation as one proxy for quality. Specifically, we consider the following *quality tags*: `<a>`, `<img>`, `<iframe>`, `<blockquote>`, `<twitter-widget>`, `<ul>`, `<ol>`, and `<embedly-embed>`. The annotation in Fig. 1 has three unique quality tags: `<blockquote>`, `<img>`, and `<a>`. Through manual inspection, we found that the presence of these tags tends to indicate helpful or interesting information in annotations, making the number of such tags a useful and simple proxy for annotation quality on Genius. In Section 5, we show that higher quality tag counts in annotations distinguishes high-IQ users, even on their earliest annotations on the site. Moreover, in Section 6, quality tag counts are shown to be an informative feature for classifying super experts.

Many quality tags are associated with quality of user-generated content in other sites. For instance, featured articles on Wikipedia have more links and images than other articles (Stvilia et al. 2005), the probability of answer acceptance on Stack Overflow positively correlates with the number of URLs in the answer (Calefato et al. 2015), and the probability of retweeting on Twitter positively correlates with the presence of URLs in the tweet (Suh et al. 2010).

We also measure annotation quality by *length*, i.e., the number of characters in the content. This has been an effective metric for content quality on Wikipedia, Yahoo! Answers, and Stack Exchange (Blumenstock 2008; Stvilia et al. 2005; Adamic et al. 2008; Gkotsis et al. 2014). Later, we detail nuances in using annotation length as a proxy for annotation quality, as we find pressure to annotate early may cause the first annotations on a song to be shorter.

3.2 Annotation Coverage

The extent to which crowdsourced contributions satisfy the needs of information seekers is vital to a platform's success. To this end, we consider *coverage* — the amount of the information sought by visitors that is actually present on the site. Various notions of coverage have been used to analyze Stack Overflow discussions (Parnin et al. 2012) and the accuracy of Wikipedia content (Mesgari et al. 2015; Giles 2005; Samoilenko and Yasseri 2014; Brown 2011).

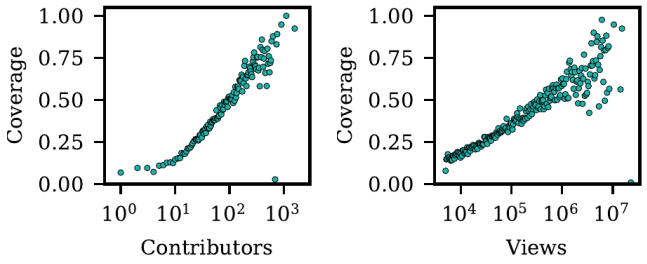

Figure 3: Mean annotation coverage as a function of number of contributing users on a song (left) and number of page views on a song (right), on songs with at least one annotation. Increased popularity with site contributors and visitors correlate with higher coverage. (Note: Genius only lists view counts for songs with at least 5,000 views).

As the fundamental function of Genius is to provide annotations on documents, we study coverage of lyrics by annotations. While some lyrics may be fillers or seemingly lack meaning, there is often still potential for interesting annotations. For example, annotations on such lyrics may provide references to other similar lyrics, references to related external social media content, or historical context.

Starting from all of the lyrics of a song, we compute its *annotation coverage* as follows. First, we remove lyrical headers (e.g., "[Verse 1: Kanye West]" or "[Refrain]"). This leaves L total text characters available for coverage. Next, let A be the total number of text characters in all lyric segments that have been annotated. Then the *annotation coverage* for the song is A/L. We note that this accounts for repeated sections of the song, since if for instance there are 3 repetitions of a chorus, then any annotation on a lyric segment in the chorus will usually also be considered an annotation on the 2 repetitions of this lyric segment in the other parts of the song. We find a positive relationship between annotation coverage and both the number of users contributing to the song as well as the number of views of a song (Fig. 3).

3.3 Lyric Originality

We find that annotation coverage also depends on the originality of the lyrics of a song. To measure originality, we first compute inverse document frequencies (idfs), where documents are songs, i.e., $\text{idf}(w) = \log(N/\text{df}(w))$, where N is the total number of songs and $\text{df}(w)$ is the number of songs containing word w. These words w are pre-processed by the same procedure as done for annotation coverage. Following ideas from Ellis et al. (2015), we define the *originality* of lyrics ℓ appearing in a song as the L-estimator

$$\text{originality}(\ell) = (p_{60} + p_{75} + p_{90})/3, \qquad (1)$$

where p_x denotes the xth percentile of the idf values of unique words in ℓ (with linear interpolation between percentiles if needed). We use large percentiles as such words are of interest to site visitors and annotators, and many words in song

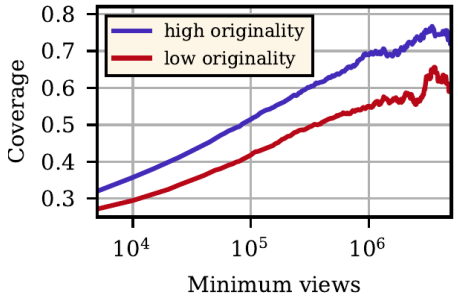

Figure 4: Mean annotation coverage over songs as a function of number of page views, stratified by songs in the upper third of originality (blue), and songs in the lower third of originality (red). Songs with more original lyrics tend to have higher annotation coverage.

lyrics may be fillers for aesthetic reasons (e.g., leading to a rhyme). Only computing percentiles over unique words prevents long, repetitive songs from achieving high originality scores. More-original songs tend to have higher annotation coverage (Fig. 4).

To illustrate, one lyric segment with low originality score is "I can never make him love me / Never make him love me" (Frank Ocean, *Bad Religion*), which has a score of 0.927. Note that each word of the second line is contained in the first line, so only the words of the first line are used to compute originality. Each word appears in many other songs, so the lyric has low originality. An example of a lyric segment with high originality is "a solo is now a poet / Hypnosis overdose on potions, adjustin' to the motions / And gettin' out all my emotions" (A$AP Rocky, *Everyday*), which has a score of 5.477. The words of highest idf are 'adjustin' at 8.06, 'hypnosis' at 7.91, and 'potions' at 7.50. The 60th, 75th, and 90th percentiles of idf are 3.47, 5.38, and 7.58.

4 Dynamics of Annotations and Edits

We now investigate relationships involving the temporal order of annotations and edits. We define the *time rank* R of an annotation on a song as the numerical position in which it was created; for example, the third earliest annotation on a song has a time rank of 3. Similarly, we define the *proportional time rank* q of an annotation for a song with at least two annotations by $q = (R-1)/(n-1)$, where n is the number of annotations on the song. This measurement allows for comparison of annotations with similar relative positions in a song's lifespan across songs with different numbers of annotations. The number of annotations on fully annotated songs (where every part of the lyrics is annotated) varies substantially, and some songs are not fully annotated yet have reached a type of equilibrium state of annotation coverage in which all meaningful lyric segments have been annotated. One of these two cases seems to hold for many songs in our dataset, although we cannot be sure if a song that is not fully annotated is truly "in equilibrium", as annotations may arrive after the point in time when we collected the data. Comparing annotations by proportional time rank is generally robust to all of this variation in the data. We use analogous definitions for edits on an annotation, and we define the edit at time rank 0 to be the (initial) annotation itself. Edits have similar variation, as there is no way to tell whether an annotation has reached its final edited state.

Below, we analyze how temporal orders relate to both contributing users and to the content itself. We find that annotator experience and measures of annotation quality exhibit a "U-shaped" pattern with respect to the proportional time rank of *annotations*. In other words, more users that already have or will eventually have high experience (i.e., have high IQ at the time we collected the data) and higher-quality content appear both early and late in time, with less experienced users (i.e., those with low IQ) and lower quality content in the intermediary. This is different from the negative relationship over time of user reputation for Stack Overflow answerers (Anderson et al. 2012), and the negative relationship over time of review quality on Amazon and Yelp (Bai et al. 2018). We develop an "IQ diamond" model of user behavior based on ideas of economic utility that offers one explanation for this behavior, and we also discuss alternative explanations. In contrast, for *edits*, eventual experience and quality grow over time.

4.1 Dynamics of Annotations

First, we analyze temporal dynamics of annotation creation (we do not consider edits until Section 4.3). Measuring the (eventual) IQ of a user making an annotation as a function of the proportional time rank (Fig. 5, top left), we find the aforementioned "U shape." Early annotations on a song are made by high-IQ users. After, the mean IQ decreases monotonically with proportional time rank until around half the total annotations have been made. After these middle annotations, the mean IQ increases monotonically. (Note that this is IQ at the time of data collection and not necessarily at the time of edits; high-IQ users either had high IQ at the time or eventually accrued high IQ, whereas low-IQ users always had low IQ.) This differs from other platforms such as Stack Overflow, where there is a monotonic decrease in user reputation over time (Anderson et al. 2012). The number of total annotations made by a user follows the same trend (Fig. 5, top right). Thus, the U shape is present if we measure the experience of users by an aggregate measure such as IQ or simply the total number of annotations made by a user.

We also consider annotation quality as a function of proportional time rank (Fig. 5, bottom row). Here, the number of quality tags follows the same U shape. Thus, not only are the earlier and later annotations on songs made by more able users, they are also of higher quality under this metric. However, our other quality metric — annotation length — largely increases with proportional time rank. A possible contributing factor is that early annotators may feel time pressure to annotate lyrics before others, incentivizing shorter annotations that are faster to create. This trend indicates that annotation length measures other factors beyond annotation quality. Such time pressure may also explain the somewhat lower number of quality tags for the earliest annotations compared to the latest annotations. We note that the earliest reviews on Amazon and Yelp are often more helpful and longer, with later reviews being of lower helpfulness and length (Bai et al. 2018). The positive trend of annotation length over time rank

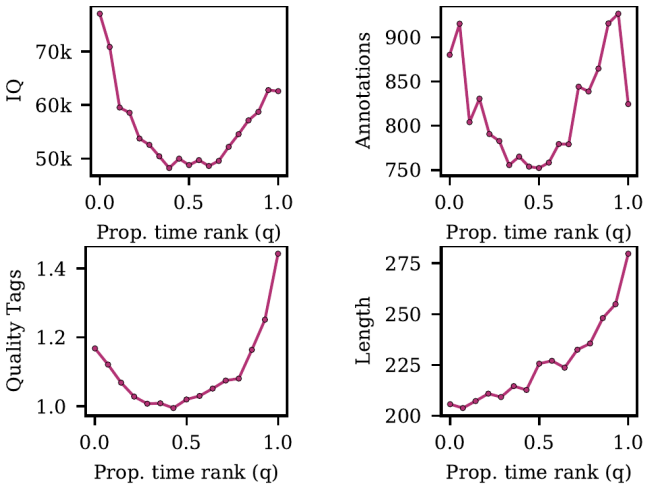

Figure 5: Various user and content statistics as a function of the proportional time rank of an annotation. Mean IQ (top left), total number of annotations made by the annotating user (top right), and number of quality tags (bottom left) follow a "U shape" with respect to proportional time rank.

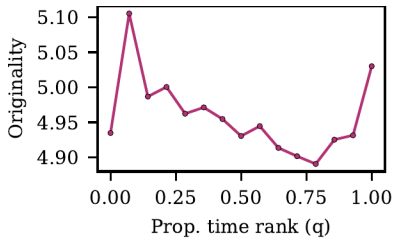

Figure 6: Lyric segment originality as a function of annotation proportional time rank. Apart from the first annotation, originality follows a "U shape."

on Genius has the opposite trend, and the "U-shaped" trend for quality tags is of a different nature.

Finally, we consider how annotated lyric segment originality relates to temporal annotation ordering. Again, for the most part, we see the familiar U shape (Fig. 6), indicating that the complex lyrics are annotated both at the beginning and end. However, the first annotations tend to be on lyrics with lower originality. This may be due to users' reluctance to annotate before there are any other annotations, leading them to annotate simpler lyrics to get the ball rolling.

4.2 The IQ Diamond (♢)

At a high level, annotation dynamics on Genius appear to be markedly different from dynamics on Stack Overflow, Amazon, and Yelp. Anderson et al. (2012) observed that user reputation decreases with later answers to a given question on Stack Overflow. From this, they developed a "reputation pyramid" model of user behavior, where new questions are first considered by high-reputation or experienced users before being considered by users with less experience. On Genius, we see the same initial descent but then a curious ascent producing the U shape. On average, early annotators are experienced users (or users that will eventually become experienced, i.e., high IQ by the time we collected the data), who quickly make annotations of high quality and on more novel lyrics. After, users with low experience levels make annotations of typically lower quality on lyrics that are less novel. Finally, the late annotations are again made by experienced users on the remaining lyrics, which tend to be more original. This behavior suggest an *IQ diamond* (♢) model for Genius, in which song lyrics are first processed by high-IQ users that form a narrow point of the diamond, then the song opens to a broader set of users to form the wide middle of the diamond, and finally narrows to the high-IQ users again.

To provide one explanation of the IQ diamond pattern, we develop a model of utility for user annotation. The utility of annotating a given song at any time depends on the proportion of the song that is already annotated at that time. More annotation coverage impacts utility both positively and negatively; one can gain more IQ by annotating songs with more activity, but higher coverage limits the choice of lyrics that a user may annotate. Thus, the users' utility functions may be modeled similarly to the utility of services with both (positive) network effects and (negative) congestion effects (Johari and Kumar 2009). Related models have been employed for users on crowdsourcing systems (Chen et al. 2019).

A Genius-specific feature of our model is that it considers user contribution to a set of prompts (lyrics) that are all related as lyrics of the same song. As mentioned in the introduction, this structure distinguishes Genius from other sites with different contribution dynamics. Any two separate annotations on the same song on Genius originate from different prompts (the two separate lyric segments) and can cause congestion (since others cannot annotate these lyric segments), whereas two reviews on an Amazon product answer the same prompt and do not prohibit others from adding related reviews (Gilbert and Karahalios 2010). Our model provides one possible and interesting explanation of the unique user behavior on Genius. We discuss possible alternative mechanisms and limitations of our model at the end of this section.

More formally, we first fix a song, and consider a population of N users labeled 1 to N. Let $\rho \in \mathbb{R}^N$ be the vector for which the kth entry ρ_k is the annotation coverage of the song by user k's annotations, so $\rho_k \geq 0$ and $\sum_{j=1}^{N} \rho_j \in [0, 1]$. Here we will refer to an annotation as an infinitesimal increase in coverage. We model the expected utility that user k would derive from adding an annotation by

$$u_k(\rho) = b_k + f_k\left(\sum_{j \neq k} \rho_j\right) - g_k\left(\sum_{j=1}^{N} \rho_j\right), \qquad (2)$$

where b_k, f_k, and g_k are characterized as follows.

- $b_k \geq 0$ is the expected a priori personal utility that user k derives from annotating a random lyric segment (assuming that users never have negative utility from annotating).
- $f_k(x) \geq 0$ is a nondecreasing function measuring the expected positive network effect when x proportion of the lyrics are covered by other users. The positive network effect arises because users tend to gain more IQ and have their contribution viewed by more people on songs that are more popular. Empirically, coverage is positively correlated with the number of song page views (Fig. 3).

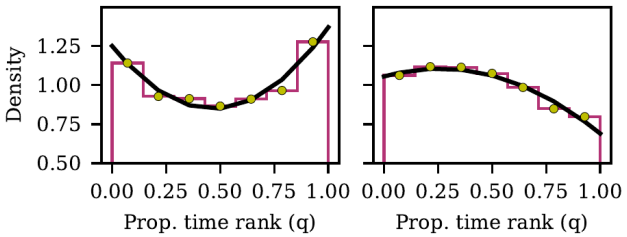

Figure 7: Distribution of proportional annotation time ranks for high-IQ (left) and low-IQ (right) users. The red bars are histogram bins where the height is the probability density. The black curves are our utility models that are fitted to the histogram bin midpoints (shown as yellow points).

- $g_k(x) \geq 0$ is a nondecreasing function that measures the expected congestion effects when x proportion of the lyrics are covered; lyrics that are already annotated cannot be annotated by user k.

For simplicity, we only consider two types of users: high-IQ (h) and low-IQ (l). For subsequent measurements, we consider high- and low-IQ users as those in the top third or bottom third in IQ (as measured at the time data was collected) over all users with at least 10 annotations.

Suppose that some user k has not yet annotated a song. Then $\sum_{j \neq k} \rho_j = \sum_j \rho_j$ is the annotation coverage and is just a proxy for the proportional time rank in our infinitesimal setting. Assuming the likelihood that user k makes an annotation is proportional to their utility $u_k(\rho)$, we would see that a user would make annotations at proportional time ranks corresponding to points at which their utility is high.

We measure this empirically by simply considering the distribution of proportional time ranks of annotations made by a high- or low-IQ user (Fig. 7). Next, we fit the model in Eq. 2 to these approximate utility curves. To this end, we make some assumptions on f_k and g_k. When there are no annotations, there are no network or congestion effects, so $f_k(0) = g_k(0) = 0$. Both of these functions are nonnegative and nondecreasing. We will assume that f_k and g_k are both concave, which could be due to diminishing returns (Chen et al. 2019). One extremely simple class of concave functions are quadratics, so we fit f_k and g_k as quadratic functions. Under our assumptions, the model parameters satisfy

$$f_k(x) = -a_1^{(k)} x^2 + a_2^{(k)} x, \quad a_1^{(k)} \geq 0, a_2^{(k)} \geq 2a_1^{(k)} \quad (3)$$

$$g_k(x) = -c_1^{(k)} x^2 + c_2^{(k)} x, \quad c_1^{(k)} \geq 0, c_2^{(k)} \geq 2c_1^{(k)}. \quad (4)$$

To determine these coefficients, we take the histogram midpoints $x_1, \ldots, x_t$ and histogram heights $y_1, \ldots, y_t$ from the proportional time rank distribution (Fig. 7) and solve the linear least squares problem

$$\min_{b_k, a_j^{(k)}, c_j^{(k)}} \sum_{j=1}^{t} (b_k + f_k(x_j) - g_k(x_j) - y_j)^2, \quad (5)$$

subject to the constraints (3), (4), and $b_k \geq 0$. We solve this problem for $k \in \{h, l\}$ for the two sets of users. Table 1 shows the fitted coefficients and Fig. 7 shows the resulting curves, which match the empirical distribution.

	b	a_1	a_2	c_1	c_2
high-IQ (h)	1.25	0.003	2.02	1.84	3.74
low-IQ (l)	1.06	0.79	1.83	0.04	1.44

Table 1: Fitted parameters of utility functions.

The coefficients in Table 1 match the IQ diamond model and give evidence for some more specific properties of user contribution behavior. First, $b_h > b_l$, which is sensible as the a priori utility that high-IQ users receive for annotating should be higher. Indeed, high-IQ users may derive extra benefits due to their status in Genius's social network or increased attention on other social media accounts linked to their Genius profile. Also, $a_2^{(h)} < c_2^{(h)}$ while $a_2^{(l)} > c_2^{(l)}$. Since $f_k'(0) = a_2^{(k)}$ and $g_k'(0) = c_2^{(k)}$, these inequalities imply that when a song only has a few annotations, high-IQ users are more influenced by the loss in utility from the few already existing annotations (congestion) while low-IQ users are more positively influenced by the presence of those few annotations (network effects).

Since $0 \approx a_1^{(h)} \ll a_1^{(l)}$, network effects in the fitted model approximately scale linearly for high-IQ users while they are most significant when there are few annotations for low-IQ users. This could arise if network effects for high-IQ users are due mostly to IQ gains from additional views and activity on a song. Recall that users earn IQ per upvote on their annotation, so the linearity of network effects may be due to the approximately linear positive relationship between song views and upvotes ($r = 0.515$). On the other hand, network effects for low-IQ users might come from social factors. The larger marginal network effects felt for early annotations may be due to desire of low-IQ users to achieve a baseline social validation that the song is worth annotating.

For congestion effects, $0 \approx c_1^{(l)} \ll c_1^{(h)}$. Thus, in our fitted model, congestion approximately scales linearly for low-IQ users while it is most significant when there are few annotations for high-IQ users. This could be caused by high-IQ users having more selectivity about the lyrics that they choose to annotate. This would agree with the behavior of experienced reviewers on Amazon products that are opinionated in their reviews (Gilbert and Karahalios 2010).

In our model, one could decompose the congestion effect by $g_k(x) = g_{k,s}(x) + g_{k,g}(x)$, where $g_{k,s}(x)$ is the expected utility lost due to lyrics that user k is qualified to annotate or is especially interested in annotating having already been annotated, and $g_{k,g}(x)$ is the utility lost due to other lyrics already being annotated. If $g_{k,g}(x)$ is proportional to the amount of general knowledge that user k has to annotate lyrics, we can assume it is linear. Then by concavity of $g_k(x)$, the function $g_{k,s}(x)$ is also concave. With this decomposition, a sufficient condition for the observed inequalities $c_1^{(l)} < c_1^{(h)}$ and $c_2^{(l)} < c_2^{(h)}$ is that high-IQ users have both more specific knowledge and general knowledge about lyrics than low-IQ users. The low value of $c_2^{(l)}$ would then suggest that low-IQ users have little specific knowledge and are not particularly

selective in which lyrics to annotate.

Additional considerations and alternative explanations. There are several additional considerations that may impact temporal dynamics for which our IQ diamond model does not take into account. One is a form of loyalty in which users interact with songs of their favorite artists, as loyalty is known to be a driver of user activity on other platforms (Hamilton et al. 2017). Artist pages on Genius have leaderboards that list users who have earned much IQ from annotating the lyrics of a particular artist. Users also have an artist leaderboard on their personal page, where they can display their positions on the artist leaderboards. Thus, users may be incentivized to focus their annotations on specific artists to acquire these accolades. Reddit users that are loyal to a particular subreddit contribute to many less-popular posts (Hamilton et al. 2017), so the analogy for Genius might be that users loyal to an artist may annotate many less-popular lyrics. Also, experienced reviewers on Amazon care strongly about their own brand or identity as a reviewer (Gilbert and Karahalios 2010), so high-IQ users on Genius may be more loyal when they have positions on leaderboards. The effect of loyalty on annotation dynamics may in fact result in differences between high-IQ and low-IQ users that are not accounted for in our model.

Also, our IQ diamond model only distinguishes users based on IQ and models each annotation as having the same impact on network effects and congestion, regardless of the user that made the annotation. Power dynamics between users may impact temporal dynamics of annotations and edits, as suggested by the positive relationships between edit time rank and user IQ (Fig. 8, top left). Moreover, some users have various "roles" in the Genius community (see Section 2), which may add to imbalanced power dynamics.

The IQ diamond model is also in some ways limited by the data that we have. Since the model fits the network effects function to data by approximating the coverage as $\sum_{j \neq k} \rho_j = \sum_j \rho_j$, we do not consider the behavior of a single user returning to edit an annotation or adding more annotations to a song. Also, we do not have the exact state of a user at the time of making each annotation, nor do we have sufficient data to determine the exact amount of IQ gained per annotation. As a user can lose IQ on a sufficiently downvoted annotation, potential IQ loss likely influences users in ways that are not explicitly captured by our model. More specifically, the assumption that the parameter b_k is always nonnegative might be unrealistic.

4.3 Dynamics of Edits on an Annotation

The temporal dynamics of edits are quite different from that of annotations. Around 95% of annotations have at most nine edits, so we directly study time ranks instead of proportional time ranks. We find that users often edit their own annotations, and users often make several consecutive edits in a row. Removing these edits gives qualitatively similar results, so we present results over all edits. Recall that in this section we consider the action at time rank 0 to be the initial annotation (the activity studied in the previous sections), and the actions at time ranks strictly greater than 0 are edits.

Figure 8 shows various properties of edits and the users making the edits as a function of time rank. Clear relationships emerge when considering the arrival of edits sequentially over the life of an annotation. Thus, we separate each plot into 10 strata, each one containing annotations that have k total edits, where $0 \leq k \leq 9$.

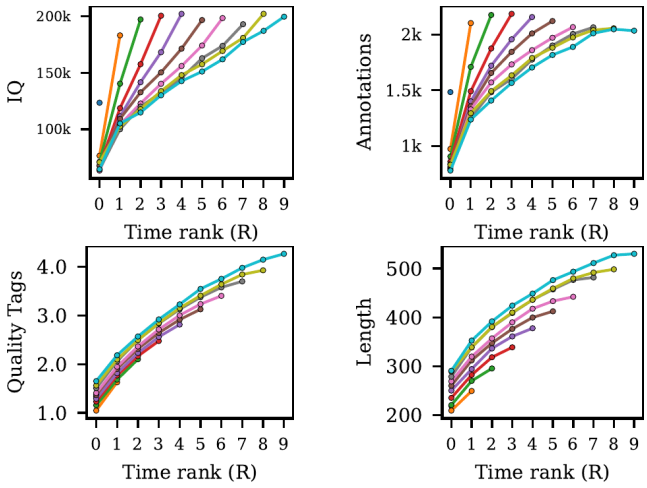

Figure 8: User and content statistics as a function of edit time rank. Each line corresponds to a different total number of edits for an annotation. For instance, the light blue line ends at time rank 9, so it is the strata with all annotations that have 9 total edits. The third dot on the light blue line in the top left figure displays the mean user IQ of users making the second edit on annotations with 9 total edits. For annotator experience, mean user IQ (top left) and number of annotations made by a user (top right) both increase with time rank. For quality, the mean number of quality tags (bottom left) and mean length (bottom right) increase with time rank.

We find that the (eventual) experience of a user — as measured by both mean IQ and mean total number of annotations at the time of data collection — increases as edit time rank increases, regardless of the number of edits on an annotation (Fig. 8, top row). This positive correlation could be explained by hesitancy of users with less experience to make edits on the content of a user with more experience. The plots in the top row of Fig. 8 have the opposite trend of similar plots for user reputation on Stack Overflow (Anderson et al. 2012). Annotations also have higher quality with more edits, as measured by both mean number of quality tags and length (Fig. 8, bottom row). Such a relationship is reasonable as edits are meant to augment content. The plots in the bottom row of Fig. 8 show the opposite trend of similar plots of review length and helpfulness on Amazon and Yelp (Bai et al. 2018).

There is nuance when comparing annotations with different numbers of total edits at a fixed time rank (points at a fixed x-axis value in Fig. 8). For a fixed edit time rank, the user making the edit tends to have higher IQ if there are fewer total edits of the annotation. However, the edits on annotations with fewer total edits have fewer quality tags and are shorter. We would expect this behavior in cases when the initial annotation has more quality tags and longer length, and the complexity may require more edits for the annotation to reach a final state.

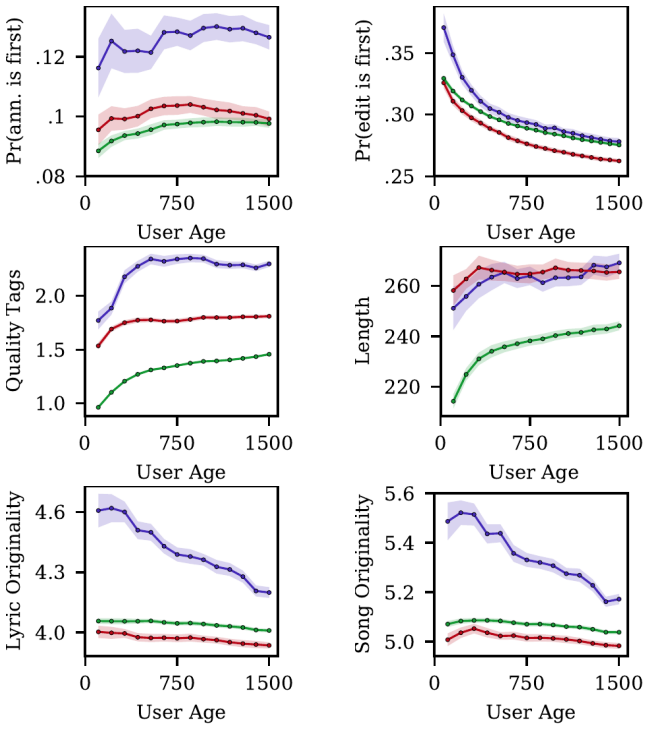

Figure 9: User behavior over days since first annotation or edit, stratified by high-IQ users ($\geq$ 100,000 IQ, blue), mid-IQ users (10,000–50,000 IQ, red), and all users (green). We show cumulative averages estimated by 100 bootstrapped samples at each point, with $\pm$ two standard deviations shaded in. For instance, the point at (750, 0.128) on the blue line in the top left figure indicates that over all annotations made by high-IQ users within 750 days of their first annotation, 12.8% are the first annotation on a song. (Top row) For high-IQ users, many annotations are the first one for a song and many edits are the first on an annotation. (Middle row) High-IQ users make annotations with higher quality tag count, and mid- or high-IQ users make longer annotations. (Bottom row) High-IQ users annotate more original lyric segments and songs.

5 Evolution of User Behavior

Having considered temporal dynamics of annotations and edits with respect to arrivals on a single song or annotation, we now analyze how user behavior changes over a user's lifespan on Genius. By studying behavior for users of differing eventual IQ levels, we observe how early and current behavior of an expert can be distinguished from that of other users. We use these ideas in Section 6 for early prediction of super experts. Similar analysis of user behavior evolution on Stack Overflow has proven useful for identifying experts (Movshovitz-Attias et al. 2013).

To analyze how user behavior changes over time, we measure cumulative averages of properties of the annotations and edits that users make over their lifespans. We also stratify users into three levels based on their IQ at the time of our data collection: users with at least 100,000 IQ (high-IQ users), users with 10,000–50,000 IQ (mid-IQ), and a group with all users. In the case of annotations, we include users with at least 10 annotations and for each user consider activity over the first 1,500 days after their first annotation. For edits, we consider users with at least 10 edits and study the 1,500 days after their first edit. There are 5,162 total users with at least 10 annotations, of which 176 are in the high-IQ strata and 904 are in the mid-IQ strata. There are 10,332 users with at least 10 edits, of which 176 are high-IQ and 1,116 are mid-IQ. The median user with at least 10 annotations made their earliest annotation in July 2013, so these accounts are sufficiently old to study behavior over time in relation to (eventual) IQ.

In agreement with Fig. 5 and the IQ diamond model in general, users with the highest IQ tend to make relatively more first annotations on songs (Fig. 9, top left). High-IQ users also make relatively more first edits on annotations early in their lifespan, although there is little difference with the general user population later in the lifespan (Fig. 9, top right). Thus, the positive relationship between IQ and edit time rank (Fig. 8, top left) is not likely caused by high-IQ users making relatively fewer early edits, but by later edits more likely coming from high-IQ users.

Quality of annotations in terms of quality tags and annotation length generally increases over a user's lifespan (Fig. 9, middle row). Users with high-IQ use relatively more quality tags, especially early on; however, annotation length from high-IQ users is comparable to those of mid-IQ users. Again, annotation length is nuanced. Annotation length has a mostly positive relationship with proportional time rank (Fig. 5, bottom right), and the discrepancy may come from high-IQ users that make more first-annotations on songs (Fig. 9, top left). As hypothesized above, there may be time pressure to annotate faster, driving down annotation length. Still, mid- and high-IQ users create substantially longer annotations than the rest of the users on the site, so there is some notion of quality or expertise that is marked by annotation length. Finally, high-IQ users have a tendency to annotate relatively more original lyric segments and songs (Fig. 9, bottom row), which may speak to their higher knowledge about lyrics as suggested by the IQ diamond model.

The above information is informative for early detection of expertise, so we reiterate the findings that certain properties of the early contributions of (eventual) high-IQ users on the site distinguish them from other users. From the start, high-IQ users make a higher proportion of first annotations and edits, make higher quality annotations as measured by number of quality tags, and annotate more original lyric segments and songs. These traits are likely beneficial to the Genius platform, and they appear to be inherent traits of high-IQ users upon entry to the site.

6 Early Prediction of Super Experts

Given the user behaviors discussed in earlier sections, we now turn to early user expertise prediction. To do this, we continue to use IQ as a proxy for expertise, as it is one simple metric that measures contributions of all types on the site, and we set up a classification problem that uses features derived from the first few annotations and edits of users.

First, we collected all users with at least 30 annotations and at least 30 edits. Of these users, we label the users in the

α_1	mean number of quality tags in first 15 annotations
α_2	mean time between first 15 annotations
α_3	number of first 15 annotations that are a song's first
α_4	mean originality on songs for first 15 annotations
e_1	mean time between first 15 edits
e_2	number of first 15 edits that are an annotation's first

Table 2: Predictors for classifying super experts and normal experts, derived from the analysis in Sections 4 and 5.

Predictor	mean regression coeff.	95% CI
α_1	0.8177	(0.590, 1.072)
α_2	0.7942	(0.548, 1.074)
α_3	0.5166	(0.349, 0.700)
α_4	0.2623	(0.098, 0.435)
e_1	-0.3832	(-0.586, -0.193)
e_2	0.2281	(0.054, 0.409)
intercept	0.1127	(-0.053, 0.283)

Table 3: Bootstrapped mean coefficients and confidence intervals of a logit model using the predictors in Table 2 for the outcome variable of super-expert.

highest third of IQ as "super experts," as these users are in the top 0.2% of IQ over all users at the time of data collection. We label the lowest third of IQ as "normal experts," as these users are still in the top 6.3% of IQ (having at least 30 annotations and edits leads to some accumulation of IQ). In total, we have 784 labeled users, whose mean number of annotations is 109 and mean number of edits is 537.

We use this coarse prediction framework for a couple of reasons. First, IQ is only a rough measurement of expertise. Second, our IQ data was scraped at different times, and we do not have data on the evolution of IQ over time for users. Nonetheless, since the distribution of contributions to the site is heavy-tailed (Fig. 2), the most active users contribute a large amount of content to the site, and we expect that our IQ splits are still meaningful; super experts have over three times higher mean number of annotations made than normal experts. Early predictions of super experts can be beneficial to a site operator. For instance, one can encourage these users with high potential to continue to contribute to the site, or to contribute content that requires expertise, such as annotations on lyrics of high originality (Anderson et al. 2013; Zhang, Ackerman, and Adamic 2007).

Next, we analyze a logit model for predicting super vs. normal experts from several content-based and edit-based predictors derived from our earlier observations of user behavior, such as the fact that experts use more quality tags, work on more original songs, and often make first edits (Table 2). These predictors were computed over the first 15 annotations and first 15 edits of the labeled users, and normalized to have zero mean and unit variance. We use a bootstrapped logistic model with 10,000 samples to estimate the mean and 95% confidence intervals of the regression coefficients (Table 3, with super-expert as the positive response variable).

Predictors	Accuracy	AUC
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, e_1, e_2$	$.673 \pm .029$	$.748 \pm .030$
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, e_1$	$.671 \pm .029$	$.744 \pm .030$
$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	$.659 \pm .030$	$.733 \pm .031$
$\alpha_1, \alpha_2, \alpha_3$	$.659 \pm .028$	$.727 \pm .031$
α_1, α_2	$.652 \pm .030$	$.715 \pm .033$
α_1	$.616 \pm .032$	$.674 \pm .034$

Table 4: Classification results for various feature subsets. Listed results are the mean and standard deviation over 1,000 random 75%/25% training/test splits. Guessing the most common test label has mean accuracy of 0.522.

The positive coefficients on α_1, α_3, α_4, and e_2 agree with our findings in Section 5 that the number of quality tags, proportion of first annotations, originality of annotated songs, and proportion of first edits, respectively, are higher early in the lifespan of high-IQ users. These results also substantiate our IQ diamond model, which agrees with the importance of first annotations in distinguishing expertise.

The fact that α_2 (time between annotations) has a positive coefficient while e_1 (time between edits) has a negative coefficient is surprising. Users with more time between their early annotations may face less time-pressure and may make higher quality annotations. On the other hand, users with less time between their edits may simply make more edits. If a user is making edits early in their lifespan, then they may have an eye for good contributions, which might provide a strong signal of expertise.

Finally, we evaluate these predictors in terms of out-of-sample predictions. We randomly split the data into 75% for training and 25% for test and averaged the test accuracy and AUC over 1,000 splits (Table 4). Over the splits, this classifier attains a mean accuracy of 0.673 and mean AUC of 0.748, whereas a majority-class baseline guess yields mean accuracy of 0.522. This substantial performance gain is remarkable given that we only use limited information from the first 15 annotations and edits.

7 Discussion

Crowdsourced information platforms provide a variety of information for the world. Here, we have analyzed various aspects of the temporal dynamics of content and users on the Genius platform that collects and manages crowdsourced musical knowledge. Genius has a substantially different knowledge curation process compared to Question-and-Answer platforms such as Stack Overflow or formal authoritative sources like Wikipedia. In turn, we found that the content and user dynamics have markedly different behavior from these other well-studied platforms.

We modeled one new type of dynamics with an IQ diamond ($\Diamond$) model, which captures the fact that eventual high-IQ users tend to annotate songs earlier and later on, with low-IQ users annotating in between. Even though the IQ diamond model is just one possible explanation for this user

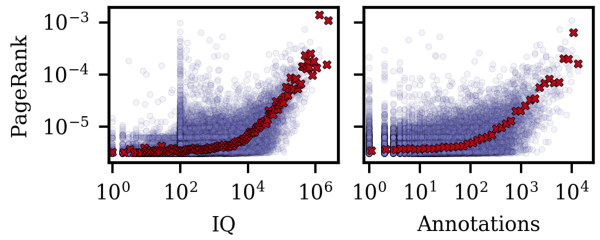

Figure 10: There is a positive relationship between user PageRank scores in the Genius social network and the IQ of the user (left) or number of annotations made by the user (right). Each blue point represents a user, and red X's are binned means. Many users have exactly 100 IQ, which is the amount awarded for adding a profile picture.

behavior, the ideas could potentially be used for mechanism design on platforms governed by dynamics similar to Genius. For example, to encourage annotation coverage of a song, one could incentivize (eventual) experts to make just a few annotations by, e.g., adjusting IQ incentives. Under the IQ diamond model, this will open the song to the bulk of users to create more annotations. Also, if higher quality annotations are desired, one may incentivize experts to edit intermediate annotations appearing at the bottom of the "U-shaped" curve (middle of the diamond). Furthermore, some users may desire to annotate a song but do not have the skills themselves. In this case, if they annotate some less original or "easy" lyrics, the induced network effects may incentivize experienced users to create high quality annotations.

Studying user behavior over time stratified by eventual expertise revealed several traits of eventual experts that can be identified by early behavior. We used these findings to develop simple and useful predictors of eventual experts. The set of predictors could be augmented with social features, as Genius manages a social network for its users. We collected the social network data of 782,432 users and 1,777,351 directed "following" relationships among them, although we do not have temporal information on the social network that would be useful for early prediction. Still, we found that certain social features are strongly correlated with user expertise. For instance, Fig. 10 shows that users with large PageRank scores have higher IQ and have made more annotations. Just using PageRank as a predictor for our experiments in Section 6 achieves a mean AUC of 0.972, and even just using the in-degree achieves a mean AUC of 0.977. While determining current expertise is often a useful task for crowdsourced-information sites, it is not a difficult task when we define expertise solely based on a public metric such as IQ.

Future directions. Although contribution dynamics on Genius are different compared to platforms such as Stack Overflow, Amazon, and Yelp, examining other types of sites with similar structure is a promising direction for future research. In particular, we hypothesize that platforms that have pages with separate but related open-ended contribution prompts may have similar contribution dynamics to that of Genius. One example is the Online Encyclopedia of Integer Sequences (`oeis.org`). The site has entries for integer sequences appearing in mathematics, and there are several open-ended ways to contribute to various structured "fields" of each sequence page (e.g., formulas, codes, or references to papers with the sequence). Such fields are similar to lyric segments on Genius, and contributions on a sequence might have similar dynamics to those of annotations on a song. Also, there may be related user contribution on sites that have been well-studied. For instance, on Stack Overflow, if there are multiple questions asked in quick succession about a specific topic, then the arrival of accepted answers to these questions may follow similar dynamics to the arrival of annotations to different lyric segments on a song on Genius.

There are also many avenues for future work based on the Genius data that we collected and the models that we developed. For one, we could design experiments based on the mechanism designs described earlier. Also, one could use the Genius data to augment other music datasets (Brost, Mehrotra, and Jehan 2019; Bertin-Mahieux et al. 2011). Conversely, other music data may improve the analysis of the content on Genius; Genius may be somewhat limited by its focus on lyrics since the musical context of the lyrics is important for both the analysis and experience of songs (Kehrer 2016). As described in Section 4.2, there are several possible extensions or improvements that one could make to the IQ diamond model that could be applicable outside of Genius.

Finally, there is plenty of information on Genius that we did not collect or analyze. In particular, "verified" annotations (annotations written by artists) and forms of user contribution besides annotations and edits are available. The rich linguistic information in the lyrics and annotations can also be analyzed in more depth. For instance, we did not consider the sequential and hierarchical structures in lyrics that have been used in music information retrieval (Tsaptsinos 2017). Such structure adds yet another layer of depth to the organization of contributions and content on Genius that distinguishes it from other crowdsourced information sites.

Acknowledgements

This research was supported in part by NSF Award DMS-1830274, ARO Award W911NF19-1-0057, ARO MURI, and JPMorgan Chase & Co.

References

Adamic, L. A.; Zhang, J.; Bakshy, E.; and Ackerman, M. S. 2008. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *WWW*, 665–674.

Al Qundus, J. 2018. Technical analysis of the social media platform genius. Technical report, Freie Universität Berlin.

Al Qundus, J.; and Paschke, A. 2018. Investigating the Effect of Attributes on User Trust in Social Media. In *Database and Expert Systems Applications*, 278–288. Springer International Publishing.

Almeida, R. B.; Mozafari, B.; and Cho, J. 2007. On the Evolution of Wikipedia. In *ICWSM*.

Anderson, A.; Huttenlocher, D.; Kleinberg, J.; and Leskovec, J. 2012. Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. In *KDD*.

Anderson, A.; Huttenlocher, D.; Kleinberg, J.; and Leskovec, J. 2013. Steering user behavior with badges. In *WWW*, 95–106.

Bai, T.; Zhao, W. X.; He, Y.; Nie, J.-Y.; and Wen, J.-R. 2018. Characterizing and predicting early reviewers for effective product marketing on e-commerce websites. *TKDE* 30(12): 2271–2284.

Bertin-Mahieux, T.; Ellis, D. P.; Whitman, B.; and Lamere, P. 2011. The Million Song Dataset. In *ISMIR*.

Beschastnikh, I.; Kriplean, T.; and McDonald, D. W. 2008. Wikipedian Self-Governance in Action: Motivating the Policy Lens. In *ICWSM*.

Blumenstock, J. E. 2008. Size Matters: Word Count as a Measure of Quality on Wikipedia. In *WWW*, 1095–1096. ACM.

Brost, B.; Mehrotra, R.; and Jehan, T. 2019. The Music Streaming Sessions Dataset. In *WebConf*, 2594–2600. ACM.

Brown, A. R. 2011. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science & Politics* 44(2): 339–343.

Calefato, F.; Lanubile, F.; Marasciulo, M. C.; and Novielli, N. 2015. Mining Successful Answers in Stack Overflow. In *MSR*.

Celma, O. 2010. Music recommendation. In *Music recommendation and discovery*, 43–85. Springer.

Chen, Y.; Wang, X.; Li, B.; and Zhang, Q. 2019. An Incentive Mechanism for Crowdsourcing Systems with Network Effects. *ACM Trans. Internet Technol.* 19(4). ISSN 1533-5399.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. In *WWW*, 307–318.

Dawson, T. M. 2018. Reinventing Genius in the .com Age: Austrian Rap Music and a New Way of Knowing. *Journal of Austrian Studies* .

Ellis, R. J.; Xing, Z.; Fang, J.; and Wang, Y. 2015. Quantifying Lexical Novelty in Song Lyrics. In *ISMIR*.

Gilbert, E.; and Karahalios, K. 2010. Understanding deja reviewers. In *CSCW*, 225–228.

Giles, J. 2005. Internet encyclopaedias go head to head. *Nature* .

Gkotsis, G.; Stepanyan, K.; Pedrinaci, C.; Domingue, J.; and Liakata, M. 2014. It's All in the Content: State of the Art Best Answer Prediction Based on Discretisation of Shallow Linguistic Features. In *WebSci*, 202–210.

Hamilton, W. L.; Zhang, J.; Danescu-Niculescu-Mizil, C.; Jurafsky, D.; and Leskovec, J. 2017. Loyalty in online communities. In *ICWSM*.

Johari, R.; and Kumar, S. 2009. Congestible Services and Network Effects.

Jurgens, D.; and Lu, T.-C. 2012. Temporal Motifs Reveal the Dynamics of Editor Interactions in Wikipedia. In *ICWSM*.

Kalboussi, A.; Omheni, N.; Mazhoud, O.; and Kacem, A. H. 2015. How to Organize the Annotation Systems in Human-Computer Environment: Study, Classification and Observations. In *INTERACT*.

Kehrer, L. 2016. Genius (formerly Rap Genius). Genius Media Group, Inc. genius.com. *JSAM* 10(4): 518–520.

Koenigstein, N.; Dror, G.; and Koren, Y. 2011. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In *RecSys*.

Maity, S. K.; Sahni, J. S. S.; and Mukherjee, A. 2015. Analysis and prediction of question topic popularity in community Q&A sites: a case study of Quora. In *ICWSM*.

McAuley, J. J.; and Leskovec, J. 2013. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *WWW*, 897–908.

Mesgari, M.; Okoli, C.; Mehdi, M.; Nielsen, F. Å.; and Lanamäki, A. 2015. "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *JASIST* .

Movshovitz-Attias, D.; Movshovitz-Attias, Y.; Steenkiste, P.; and Faloutsos, C. 2013. Analysis of the Reputation System and User Contributions on a Question Answering Website: StackOverflow. In *ASONAM*, 886–893. New York, NY, USA: ACM.

Pal, A.; Farzan, R.; Konstan, J. A.; and Kraut, R. E. 2011. Early Detection of Potential Experts in Question Answering Communities. In *UMAP*, 231–242. Berlin, Heidelberg: Springer.

Pal, A.; Harper, F. M.; and Konstan, J. A. 2012. Exploring Question Selection Bias to Identify Experts and Potential Experts in Community Question Answering. *ACM Trans. Inf. Syst.* 30(2).

Paranjape, A.; Benson, A. R.; and Leskovec, J. 2017. Motifs in Temporal Networks. In *WSDM*, 601–610. ACM.

Parnin, C.; Treude, C.; Grammel, L.; and Storey, M.-A. 2012. Crowd documentation: Exploring the coverage and the dynamics of api discussions on stack overflow. Technical report, Georgia Institute of Technology.

Patil, S.; and Lee, K. 2015. Detecting experts on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors. *Social Network Analysis and Mining* 6(1).

Posnett, D.; Warburg, E.; Devanbu, P.; and Filkov, V. 2012. Mining Stack Exchange: Expertise Is Evident from Initial Contributions. In *SocInfo*, 199–204. IEEE.

Rambsy, H. 2018. Becoming A Rap Genius. In *The Routledge Companion to Media Studies and Digital Humanities*. Routledge.

Ravi, S.; Pang, B.; Rastogi, V.; and Kumar, R. 2014. Great question! Question quality in community Q&A. In *ICWSM*.

Samoilenko, A.; and Yasseri, T. 2014. The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. *EPJ Data Science* 3(1).

Schedl, M.; Gómez Gutiérrez, E.; and Urbano, J. 2014. Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval. 8 (2-3): 127-261.* .

Stvilia, B.; Twidale, M. B.; Smith, L. C.; and Gasser, L. 2005. Assessing information quality of a community-based encyclopedia. In *ICIQ*, 442–454.

Suh, B.; Hong, L.; Pirolli, P.; and Chi, E. H. 2010. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In *SocialCom*, 177–184.

Swartz, A. 2002. Musicbrainz: A semantic web service. *IEEE Intelligent Systems* 17(1): 76–77.

Tian, Q.; Zhang, P.; and Li, B. 2013. Towards predicting the best answers in community-based question-answering services. In *ICWSM*.

Tsaptsinos, A. 2017. Lyrics-based music genre classification using a hierarchical attention network. In *ISMIR*.

van Dijk, D.; Tsagkias, M.; and de Rijke, M. 2015. Early detection of topical expertise in community question answering. In *SIGIR*.

Wang, G.; Gill, K.; Mohanlal, M.; Zheng, H.; and Zhao, B. Y. 2013. Wisdom in the Social Crowd: An Analysis of Quora. In *WWW*.

Zhang, J.; Ackerman, M. S.; and Adamic, L. 2007. Expertise Networks in Online Communities: Structure and Algorithms. In *WWW*.