

1 Neural Age Screening on Question Answering Communities

2 Mohan Timilsina^{a,*}, Alejandro Figueroa^{b,2}

3 ^aData Science Institute, Insight Centre for Data Analytics, National University of Ireland Galway, Ireland

4 ^bDepartamento de Ciencias de la Ingeniería, Facultad de Ingeniería, Universidad Andres Bello, Antonio Varas 880, Santiago, Chile

5 ARTICLE INFO

6 ABSTRACT

7
8
9 *Keywords:*
10 graph
11 embedding
12 regression
13 prediction
14 activity
15 users

For online social networks, demographic analysis is absolutely essential for improving their services in many ways. It is instrumental in understanding their different audiences, members and competitors. As well as that, it is pivotal in designing effective personalization and contextualization strategies, especially for displaying and creating better content. There is, for this reason, a great bulk of research into how demographic variables are characterized and how they impact online platforms such as Facebook and Twitter. But surprisingly, only a handful of works delve into their characterization and effect on community Question-Answering (cQA) websites. In this particular context, the subject of age demographics remains largely unexplored.

This paper takes the lead on interpreting automatic age recognition on CQAs (a.k.a. age screening) as a regression task. To this effect, it compares state-of-the-art graph-based neural network regression and embedding models on a massive activity-graph encompassing ca. 16 and 837 million nodes (members) and edges, respectively. For this study, a large-scale subset of ca. 657,000 community fellows was automatically associated with their age via aligning their profile texts with a limited number of linguistic patterns.

In short, our results show that Node2vec significantly outperforms other embeddings regardless of the regression model used for casting predictions. When this embedding is combined with Artificial Neural Network Regressions, we obtained our best configuration scoring a Root Mean Square Error (RMSE) of 8.39. An interesting qualitative feature of this embedding space is that age-based centroid vectors tend to form a trail ordered by age. Lastly, our outcomes also signal that activity graph based models can rival its counterparts based on image and textual inputs, paving the way for constructing effective multi-modal approaches.

31 1. Introduction

32 Essentially, cQA services provide a versatile online framework whereby their members can satisfy their information
33 needs by publishing questions expecting to be answered by other community peers. This dynamic is a decisive
34 advantage over traditional web searches since every fellow can ask and retrieve responses that are not readily found
35 across web documents. In many cases, these websites also allow for conducting opinion polls and surveys, making the
36 interaction between humans of all ages and walks of life much easier. This versatility is a determining factor when
37 seeking and sharing information.

38 Undoubtedly, demographic analysis is essential for enhancing user experience and enforcing terms of service and
39 local laws. To exemplify, as per their terms, Facebook and Instagram require people to be at least thirteen years old to
40 register, and this minimum age is higher in some countries. Moreover, age verification is in the spotlight of several new
41 and proposed regulatory frameworks on data protection, online harm, and child safety. Nevertheless, age screening is
42 not as simple as it might sound since platform abusers, criminals, and young people get around it by misrepresenting
43 their age.

44 By and large, people can provide their demographic information when registering by filling out a form composed
45 of mandatory (e.g., email and alias) and optional fields (e.g., age, gender, and postal codes). But most of the time,
46 optional fields are left blank as people are increasingly jealous of their privacy and see it as another annoying form in
47 their lives. Nevertheless, their activity in the community can unwittingly yield clues about some of this missing data.

48 Unquestionably, almost all demographic analyses deal with age as one of their principal variables since it typically
49 determines behavioral patterns like buying habits and our responses to advertising. However, age differences can also
50 be perceived across specific online services, including recommender systems, where suggestions are more frequently
51 clicked by elderly users [1]. In addition, people at different stages of their lives have distinct interests and ways of

✉ mohan.timilsina@insight-centre.org (M. Timilsina); alejandro.figueroa@unab.cl (A. Figueroa)
ORCID(s): 0000-0002-6317-728X (A. Figueroa)

expressing themselves, and often they spend their time on separate platforms. In the case of cQAs, the absence of this demographic information makes it harder to enhance safety and user experience by, for instance, managing content effectively and discovering members that can readily respond to unresolved questions.

As far as we can tell, automatic age detection remains unexplored mainly in the context of cQA sites, where a handful of prior studies have addressed this as a multi-class classification problem on textual and/or visual inputs [2]. Hence, the novelty of this paper lies in exploring the performance of several frontier regression models working on activity graphs for age prediction [3, 4, 5, 6]. The underlying motivation behind using regression is two-fold: a) most of the time, decision boundaries corresponding to a particular age cohort are arbitrary; and b) for some tasks, such as age screening and fake profile detection, you need predictions much more fine-grained than a generational period. With this in mind, our contributions to this body of knowledge are summarized as follows:

- First and foremost, automatic age recognition is perceived as a regression instead of a classification task. Therefore, we provide the graph-based approach for the age prediction of the users by only analyzing their activities in Q&A communities.
- We direct our attention to infer informative patterns from user activity graphs rather than from their texts or avatars. More precisely, our object of study is a massive graph comprising the activity of 16 million community fellows, where 657,805 were automatically labeled with their age.
- As for graph modeling, we considered several continuous feature representations for community members (nodes), including state-of-the-art embedding algorithms such as Graph Convolutional Network and Node2vec. We then capitalized on these low-dimensional spaces for testing several regression strategies such as Linear Regression, Support Vector Regression, Random Forest Regression, k-Nearest Neighbors, and Artificial Neural Networks.
- Node2Vec with a smaller space dimension close to 80 contributes the best age recognition error.

In a statement, our empirical outcomes unveil that integrating Node2Vec with Artificial Neural Network regressions yields the best results. Furthermore, in this continuous space, we also discover that age-based centroid vectors form a path ordered by age. Lastly, our outcomes show the promise of activity graph-based models since they suggest that this strategy can compete with approaches built on top of texts and images. The remainder of this paper is organized as follows. For starters, section 2 outlines the related work. Next, sections 3 and 4 present our research questions and methods, respectively. Subsequently, sections 5 and 6 report and discuss our experiments and findings. Eventually, section 7 draws some conclusions and sketches some future work.

2. Related Work

As far as we know, our work pioneers the idea of profiting from regression models for predicting the age of cQA members. We start this effort by casting this problem as a regression task, and in so doing, we capitalize on state-of-the-art neural networks. Additionally, this paper introduces the use of their activity graphs to build these models.

In the broader sense, our study is at the crossroads of research areas, including cQA platforms, user analysis, age prediction, and neural network regression models on large-scale graphs. In particular, our focus is on the relationship between age demographics and question-answer activity patterns.

2.1. Community Question Answering

Recently, several challenging aspects of cQA websites have caught the attention of researchers concerning the application of word embeddings and neural networks to question analysis [7, 8, 9], expert finding [10, 11, 12, 13, 14], question routing [15, 16] and answer selection [8, 17, 18, 19, 20].

Word embeddings have been utilized for vectorizing questions, whereby semantically equivalent elements are fetched from the archives every time a new question is posted [7, 21]. Likewise, semantic embedding spaces have been devised as a foundation for inference strategies, especially for mapping factoid questions onto logical representations of correct answers guided by knowledge bases [22]. Another approach scores word co-occurrence likelihoods across question-answer pairs via a word embedding-based correlation strategy that capitalizes on translation models [23]. Also, semantic embeddings have been employed to find experts across cQA sites [11], especially coupled with Bayesian models in multi-modal multi-view frameworks [24].

99 One way of refining the understanding of questions is separating informational from conversational elements
 100 [25, 26]. It has been discovered that the performance substantially increases when exploiting semi-supervised learning,
 101 pre-trained Yahoo! Answers embeddings, and label propagation operating on Recurrent Neural Networks (RNNs)
 102 together with LSTM (Long Short Term Memory) [26]. From another viewpoint, neural network architectures have
 103 been designed for automatically distinguishing useful clarification questions across cQA threads [27]. Moreover, by
 104 jointly learning representations of questions, new elements can be recommended to potential answerers grounded on
 105 a heterogeneous information network embedding algorithm and LSTM models [15].

106 Among other tasks, neural networks have shown to be most fundamental when devising techniques for choosing the
 107 best answers [17]. More concretely, they were proven to be instrumental for constructing hybrid models that mix outputs
 108 of different components: a) built on content, and non-content attributes [28]; and b) that jointly capture the semantic
 109 relevance and member expertise [29]. Other hybrid approaches have proposed attention mechanisms to semantically
 110 model question-answer pairs individually and interactively [30], and asymmetric multi-faceted ranking networks to
 111 jointly exploit the deep semantic relevance between question-answer pairs and the authority of the answerers [31]. On
 112 the same subject, the work of [32] built a multi-task learning framework for re-ranking answer candidates, where the
 113 main task models question-answer pairs, whereas the auxiliary sub-task is the partial order between pairs of answer
 114 candidates.

115 Due to their versatility, neural networks are utilized for dealing with a wide diversity of cQA challenges. For
 116 instance, Convolutional Neural Networks (CNNs) and RNNs are combined for modeling the question-answer semantic
 117 matching and the semantic correlation in chronological sequences of answers [33]. Further, they have been used for
 118 learning novelty-related properties from unlabeled data as a way of emphasizing diversity in answer ranking [34] and
 119 for detecting promising answers [35]. Finally, in a different light, the strategy of [36] represented the cQA archive
 120 by constructing a knowledge map where questions and answers are its two main dimensions, and their intersection
 121 comprises a cluster of the related question-answer pairs.

122 2.2. User Analysis in cQA Platforms

123 Researchers have studied cQA members from several angles, like their evolution, motivations, and expertise. For
 124 instance, their evolution can be analyzed according to their reaction to changing norms at distinct stages of their
 125 community lives [37], or in conformity to their shifts between the asking and answering roles over time [38]. On
 126 the one hand, some factors make newcomers more excellent contributors [39], but also, factors that push askers to
 127 switch away from free-based to payment-based platforms [40].

128 The motivations behind the participation of cQA members have also grabbed some attention [41, 42, 43]. While it is
 129 true that many fellows post and obtain information periodically, it is also true that few peers take the time to contribute
 130 knowledge to the community consistently. Notably, it has been shown that psychological, social, and functional factors
 131 differ across lurkers and askers, answerers as well [42]. At significant, positive votes motivate knowledge sharing;
 132 and conversely, negative votes are significantly detrimental. Most importantly, comments moderate the relationships
 133 between voting and knowledge contribution [43].

134 Community peers, who play a pivotal role as producers of high-quality material, are known as experts or
 135 authoritative sources [44, 45], whereas reputation collectors are primary sources of low-quality answers as well as of
 136 answers to duplicate questions [46]. For detecting these individuals, general measures of expertise have been proposed,
 137 primarily based on the theory of belief functions that account for member's attributes and the time they spend on the
 138 website [47]. For the same purpose, the probability of becoming an expert can be computed on the grounds of four
 139 property groups, namely, user behavior, topic similarity, emerging topics, and topic transition [48]. Another aspect to
 140 consider when seeking suitable answerers is their intimacy over different topics [49]. Many studies have explored the
 141 inherent relationship between user criteria and answer quality [18, 50, 51].

142 Over the past few years, neural networks have been employed to automatically identify experts. Take, for instance,
 143 the work of [29], who encoded latent user expertise into answer representation by hierarchical attention mechanisms
 144 within LSTM networks, capturing more informative and implicit topic interests. Parallel with this, a cognitive-based
 145 lightweight method was devised for simultaneously assessing the expertise and reliability of community peers [52]. In
 146 the same light, by modeling users, one can bridge the semantic gap between distinct members, as similar peers may
 147 use similar wordings across their answers [53]. This way, user representations can be learned in an end-to-end neural
 148 network in a way that best explains user expertise on question topics.

149 Several additional active research topics concerning user modeling, such as [54], discovered a homophilous
 150 behavior across undetected abusive fellows by examining their flagging patterns. Further, the propensity to become

loyal has shown to be apparent from initial interactions in the community, and loyal members employ language that signals collective identity and engage with more esoteric, less popular content [55]. From what and how individuals post, one could predict with high accuracy if they are willing to answer an unknown peer when the question fits their expertise [56, 57]. Along the same lines, the work of [58] showed that identity-based trust, feedback from previous knowledge contributions, opportunities for social exposure, word-of-mouth marketing, and pressure from norms of reciprocity have a positive impact on the continued knowledge contribution behavior of the members.

2.3. Age Recognition on cQA Platforms

Few studies have dug deeper into demographic factors and their impact on cQA services [2, 59, 60]. Instead, most try to discover discriminative patterns by inspecting texts authored by community fellows [2, 61]. For instance, the research of [59] touched on age demographics when examining their attitude and sentimentality. Interestingly, they revealed that members are likely to respond to peers of the same age more positively; and that sentimentality decreases with age.

In Stack Overflow¹, programmer reputation scores were observed to increase relative to age well into the 50's, that programmers in their 30's tend to focus on fewer areas relative to those younger or older in age, and that there is not a strong correlation between age and scores in specific knowledge areas [61]. Since older people continue to trail Gen Z and Millennials in the adoption of technologies, there is an under-representation of these groups that has a harmful influence on demographic analysis. By casting age recognition as a classification task, the work of [2] delved into this impact on three distinct modalities (i.e., texts, images, and metadata), in particular, focusing on the effects of the gradual evolution from one cohort to the next; while [62] dealt only with texts.

2.4. Embedding Models for Large-scale Graphs

In general, graphs are categorized into four different groups: (a) homogeneous graphs, (b) heterogeneous graph, (c) graph with auxiliary information, and (iv) graph constructed from non-relational data [63]. Each of these groups aims at dealing with different types of challenges related to the construction of continuous graph feature spaces. These continuous feature spaces are vector representations of a graph's entities and, in some cases, relations. Thus, generating graph embeddings involves finding a function, f , which maps each node to the embedding space. Defined formally, $f : \mathcal{V} \rightarrow \mathbb{R}^{|\mathcal{V}| \times d}$ where d is the dimensionality of the embedding space, \mathcal{V} is the nodes such that $d \ll \mathcal{V}$ and \mathbb{R} denotes the set of all real numbers.

In the case of our work, the object of study is a large-scale homogeneous graph. In this kind of network, all nodes and edges belong to one type only. There are different embedding approaches proposed for different variants of this class of graph, i.e., unweighted-undirected [64], weighted [65], directed [66], and weighted-directed [67]. The fundamental role of all these strategies is capturing structural information and preserving the connectivity patterns observed in the input networks through the embedding [68]. These low-dimensional representations enable exploiting similarities based on neighboring associations to solve various graph mining tasks including node classification [69, 70], link prediction [71] and graph reconstruction [72].

A classic approach to infer embeddings in homogeneous networks is by conducting random walks over its structure. Some promising methods in this category include DeepWalk [64] and Node2vec [67]. Recently, some techniques, such as APP [73] and VERSE [74], have improved the quality of the inferences by refining the learning step by personalized page-rank-oriented random walks. However, the main caveat of this sort of strategy has to do with its computing costs since they will be proportional to the number of random walks, which can be prohibitive for large graphs [75]. In addition, non-random graph-based embedding methods have also been proposed. These approaches use deep neural networks instead of performing random walks, but their training also demands a lot of computational resources, especially for large graphs [74].

Random walk enables the extraction of node structures by learning their vector representations based on some notion of topological similarity (or proximity). The construction of an embedding starts by selecting a random-walk method such as rooted PageRank [76]. Then, a similarity function maps realizations of the random-walk process into real values that represent the notion of node proximity. Likewise, non-random graph-based embedding strategies employ deep learning techniques where there is a specific task to optimize. In the case of Structural Deep Network Embedding (SDNE), there is a first unsupervised component that has an encoder-decoder network architecture for reconstructing the input signal [65]. This reconstruction is believed to preserve the proximity of a network. The

¹stackoverflow.com

supervised component is the second element of the loss function and this will preserve the first-order proximity of the graph. The third and final component is a regularization term.

Incidentally, Graph Convolutional Networks (GCN) has proven to be very promising for learning node embeddings and for performing their classification in semi-supervised approaches [77]. Conversely, Graph Auto Encoders (GAE) can be integrated with GCN and auto-encoder models to infer these contiguous feature spaces [77, 78]. From another perspective, recent techniques, including GraphGAN [79], ANE [80] and DWNS [81], have adopted Generative Adversarial Networks (GAN) to model node connectivity probabilities accurately. Unfortunately, all these methods do not scale to large graphs [75].

Another novel approach to dealing with homogeneous graphs is factorizing their adjacency matrix. For instance, some studies directly compute the adjacency matrix in order to obtain node embeddings [82, 83, 84]. More recently, this matrix factorization strategy enhanced with spectral propagation (known as ProNE) has shown its efficacy in the context of node classification tasks [85], but on the flip side, its performance has been less competitive for link prediction and graph reconstruction [75]. It is worth noting here that this embedding has not been tested in node regression settings. Lastly, it should be highlighted that GRA (Global Resource Allocation), which iteratively tunes the proximity matrix, enables reaping effective results in multi-label node classification problems at the expense of high computational costs [86].

2.5. Graph-based Age Screening on Other Social Networks

There is a vast bulk of research into age demographics on Facebook [87] and Twitter [88, 89, 90, 91]. For instance, word embeddings have reaped poor performance on Twitter when training Wikipedia-based CBOWs mixed with SVMs (RBF kernels) [92].

As for age screening based on graph analysis, we find the work of [93], who used regression to predict six variables (i.e., gender, age, ethnicity, education, income, and child status) of ca. 46k Twitter users based solely on their network activity (i.e., whom they follow). Later, they extended this study to political preferences [94]. Concerning node embeddings, these have been recently utilized for discovering similar users [95] and experts [24, 96].

3. Research Questions

With prior works as a foundation, we pioneer the attempts at conducting an in-depth study into age predictors built on top of regression models and node contiguous feature spaces in the context of question-answers interchanges. To this end, we conducted a comprehensive analysis of state-of-the-start graph embeddings devised mainly for coping with massive activity graphs. Specifically, we aim primarily to answer the following three research questions:

RQ1: What are effective kinds of machine learning regression approaches for predicting age in a large-scale setting? In section 5, our empirical results indicate that the most effective configuration integrates Artificial Neural Network Regression with Node2vec embeddings.

RQ2: Is there any pattern that characterizes birth years in a low-dimensional space?

In section 6.1, we discovered that age-based centroid vectors tend to form a sequence ordered by age from youngest to oldest (i.e., forty years old).

RQ3: How does age prediction based on graph regressions perform in compared to prior works on other modalities such as texts and images?

In section 6.2, experimental outcomes suggest that activity graphs are a better option than profile images and are competitive to textual models.

4. Methodology

Figure 1 highlights the framework of our study. Each step in this pipeline is described in the following subsections.

4.1. CQA Graph Creation

In order to fetch question-answers pages (see Figure 2a) and user profiles (see Figure 2b) from Yahoo! Answers, we took advantage of the web scraper implemented in [97], which ran for around three years (September 2015–2018). According to community guidelines, askers must categorize their questions in conformity to a taxonomy of up to three levels (e.g., Home & Garden → Decorating & Remodeling in Figure 2a) whenever they publish a new question. Hence, top-level entries are broad and embody a constantly growing number of questions and answers, whereas most

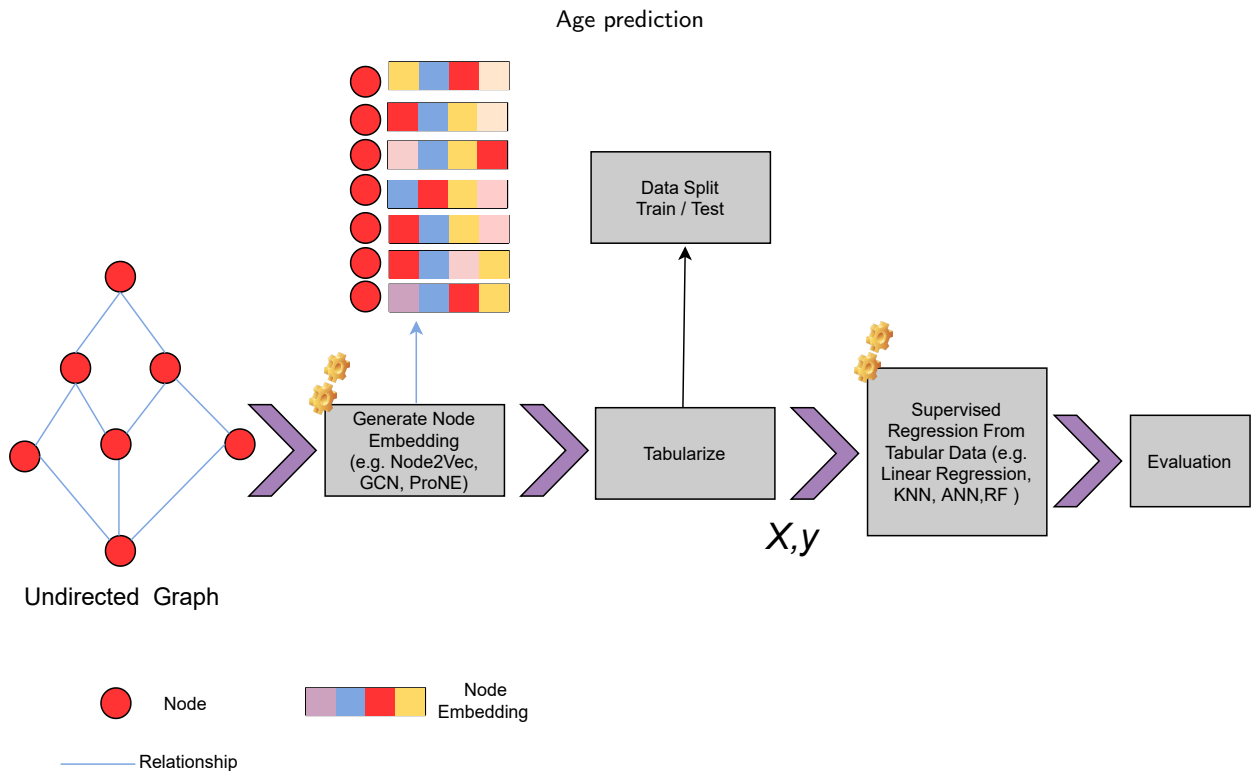


Figure 1: Machine learning pipeline to predict the user's age exploiting graph embeddings.

248 fine-grained categories at the bottom (third-level) are more specific, providing narrow coverage, and ergo most of them
 249 seldom get new questions.

250 Therefore, this crawler browses through questions contained only in the first two levels, retrieving the top ten
 251 questions displayed on each category page. Each of these category pages was visited several times during the three
 252 years. As a result, about 53 million question-answers pages were downloaded, where all question titles, bodies, and
 253 answers were extracted accordingly, and, like manner, ca. twelve million profile pages were fetched. From them, self-
 254 descriptions and questions were extracted, which were used later to get more questions and reach additional users using
 255 recursively navigating a network.

256 In order to determine the age, all content predominantly written in English was singled out via a text-based language
 257 detector², and subsequently, we looked for valid putative birth years (1910–2008) and ages (10–99) at the paragraph
 258 level. We split matched paragraphs into sentences via CoreNLP³, and accordingly, we kept every sentence that starts
 259 with any of following case-insensitive surface patterns: (a) [I am | I m | I'm | Im | I turned] [| an | now | only | age |
 260 turning] NUMBER; (b) I was born [in/on] [DATE | YEAR]; and (c) My age is NUMBER.

261 When aligning, we checked if we could find an occurrence of a unit such as kg and mg or a time such as weeks or
 262 days. This way making sure that these matched numbers did not correspond to commonly used metrics/units. Every
 263 time these alignments failed, a POS-based analysis was conducted by basically ensuring that: (a) there was only one
 264 pronoun and no additional verb before the first identified number/year; and (b) there was no negation before the first
 265 number/year.

266 From the overall 16,580,746 members involved in the corpus, we could automatically discover the birth year of
 267 657,805 (3.97%) community peers. Note here that the tag for an individual was preserved if and only if all references
 268 fell into a gap of two years, where the earliest year was kept as the final label. Then, we created an undirected graph by
 269 linking every asker with all his/her answerers, resulting in a total of 836,513,013 edges. In detail, each question-answer
 270 page was mapped into a small graph where its involved users are nodes, and every edge connects an answerer to the
 271 respective asker (see Figure 3). Eventually, all these small graphs were merged into one final activity graph. In so

²code.google.com/archive/p/language-detection/

³stanfordnlp.github.io/CoreNLP/

Dborg2 asked in Home & Garden > Decorating & Remodeling · 2 days ago

What's the best way to increase kitchen space?

20 Answers · Relevance ▾

Sarah · 22 hours ago

Favorite Answer ★★★★★

You can create the illusion of more space by painting your kitchen with white or light colors. If you're on a budget, this is a cheap technique you can try. Good luck!

Common Sense · 1 hour ago

Paint your walls and cabinets a light color to create a more open feeling of space.

Butto "increase kitchen space" is to have well organized Kitchen storage ...Show more

? · 3 hours ago

Keep the kitchen equipment neat and tidy

(a) Question-answers page discussing “What’s the best way to increase kitchen space?”

Lv7 381,598 points

Ashley M

Favorite Answers 39% · Answers 55,289

Block · Follow

I love Yahoo Answers. I try to be completely fair and honest even if I know that my answer may not be liked by everyone. I don't claim to be right 100% of the time, but I do like to think that my honesty and frankness does more good than bad. ...Less

Questions · Answers · Followed Questions · Following Users

How do they make eye glasses for small children?
Specifically, how do they determine what prescription the child needs?
Because older children an adults can do the "better, or worse" thing...but what about the infants you ...Show more
2 Answers · Optical · 5 years ago

I'd like to learn the basics of programming and coding, please help?
I'm aware that learning programming and coding is a lengthy process, and there's LOTS of different languages to learn. I know there are kits out there designed to teach kids the basics, and they wind up building an actual little program or something in the process. ...Show more
2 Answers · Programming & Design · 5 years ago

Tell me your experiences with septum piercings please?
I want to get my septum pierced. A friend of mine who is a college professor recently got hers done and has been able to flip it up into her nose without issue. That's what is making me want to do it more, because I like the idea of being able to hide it without taking it out. ...Show more
1 Answer · Tattoos · 5 years ago

(b) Profile page belonging to “Ashley M” Here, a self-description and a list of asked questions are illustrated.

Figure 2: Sample pages from Yahoo! Answers.

272 doing, we considered all 16 million members. For empirical purposes⁴, labelled samples were randomly divided into
273 394,745 training (60%), 131,519 testing (20%) and 131,541 evaluation (20%).

274 4.2. Generating Node Embeddings

275 To learn continuous feature representations of community members (Node2Vec [67] and PRONE [85]), we
276 explored the fast and scalable suite of graph embedding models implemented by Matt Ranger’s library⁵. By and large,
277 the algorithms provided by this toolkit are prevalent since they are proven to be helpful in coping with various kinds
278 of graph mining tasks like node classification, link prediction, sub-graph mining, and community analysis. For matrix
279 factorization-based embeddings, we used the scikit API⁶. Similarly, for Graph Convolutional Networks (GCNs), we
280 used the open-source library known as Stellargraph⁷. In general, graph embeddings are believed to yield highly accurate
281 representations of the underlying latent structures of a given graph.

282 From this suite, we capitalized on the following strategies to capture the graph, edge and node topology, and other
283 relevant information as well:

- 284 • **Graph Convolutional Network (GCN):** It consists of graph convolution layers, whereby embeddings are
285 constructed for each node via mixing its neighbors from the previous layer in the network [77].

$$286 H^{l+1} = A'X^{(l)}W^{(l)}, X^{(l+1)} = \sigma_{ReLU}(H^{(l+1)}), \quad (1)$$

287 In this equation, $X^{(l)} \in R^{N \times F_l}$ is the embedding at the l^{th} layer for all the N nodes and $X^{(0)} = X$; A' is the
288 normalized and regularized adjacency matrix and $W^{(l)} \in R^{F_l \times F_{l+1}}$ is the feature transformation matrix which

⁴The full activity graph and the labels used for this study will be publicly available online upon acceptance.

⁵github.com/VHRanger/nodevectors

⁶scikit-learn.org/stable/modules/classes.html

⁷github.com/stellargraph/stellargraph

Age prediction

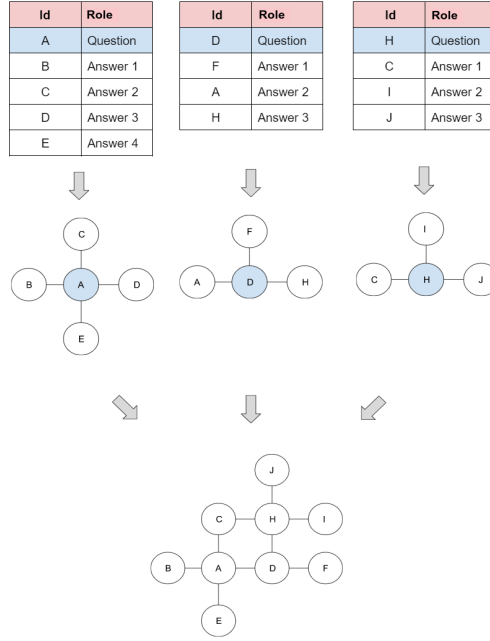


Figure 3: The activity graph creation process for three illustrative question-answer pages.

will be learnt for the downstream tasks. The activation function $\sigma_{ReLU}(\cdot)$ is usually set to be the element-wise Rectified Linear Unit (ReLU).

- Node2Vec:** This method integrates Deep Learning with random walks to automatically sample paths from the graph employing exploiting neighborhood structures [67]. To describe this random walk process, let c_i denote the i^{th} node in the walk, starting with $c_0 = u$ where u is the source node. Then, the probability corresponding to the node c_i is generated by the following distribution:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} = a_{i+1} & \text{if } (v, x) \in E \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where π_{vx} is the unnormalized transition probability between nodes v and x , and Z is the normalizing constant. It is worth noting here that the generation of these random walks is defined by two input parameters (cf. [67]).

- ProNE:** This approach learns embeddings via matrix factorization with the enhancement of spectral propagation [85]. ProNE leverages the simplest edge structure to represent a node-context pair. The edge set then forms a node context pair set $\mathcal{D} = E$. Here, E represents edge set and \mathcal{D} is the set of node pairs (v_i, v_j) . Formally, the probability of occurrence of a context v_j given a node v_i is defined as:

$$p^{i,j} = \sigma(r_i^T, c_j) \quad (3)$$

In this mathematical formula, $\sigma(\cdot)$ is the sigmoid function and $r_i, c_i \in \mathbf{R}^d$ represent the embedding and context vectors of node v_i respectively. Similarly, $p_{ij} = A_{ij}/D_{ii}$ indicates the weight of (v_i, v_j) in \mathcal{D} and A is the adjacency matrix of the graph.

- Non-Negative Matrix Factorization (NMF):** Since the adjacency matrix A is a non-negative, where each column represents a particular node, the NMF can be utilized for obtaining two non-negative matrices M and

G such that the dimensions of M and G are $q \times k$ and $k \times q$, respectively ($k < q$) [98]. In this decomposition, the product of these two new matrices (i.e., $M \times G$) is very close to the original A :

$$A = M_{q \times k} G_{k \times q} \quad (4)$$

Here, k is a parameter and denotes the dimension of the latent space. M is referred to as the base matrix, since it consists of the bases of the latent space. Conversely, G is called the coefficient matrix, since it represents the combination of coefficients of bases necessary for reconstructing A . Generally, this decomposition task is modelled as the following Frobenius norm optimization problem:

$$\min_{m,g} \| (A - MG) \|_F^2 \quad \text{such that} \quad M, G \geq 0 \quad (5)$$

- **Singular Vector Decomposition (SVD):** This factorization aims at computing a low (k) rank representation A_k of the adjacency matrix $A_{q \times b}$ via the following mathematical relation:

$$A = M \Sigma G^T \quad (6)$$

where M is an $q \times k$ unitary matrix, Σ is an $k \times k$ diagonal matrix and V is an $k \times b$ unitary matrix.

- **GGVec:** It is a new embedding algorithm for undirected graphs available in the Nodevector library⁸. The algorithm is based on sampling edgewise local methods which update the embedding weights on each step. The advantage of GGVec is that, it easily scales for large graphs.

4.3. Supervised Regression

In order to make a fair comparison of different graph embedding models, supervised regressions require data in tabularized form. In so doing, we first set the number of dimensions of the embedding space to 128 (default value). Then, after learning, each vector in this space is labeled with an outcome variable, in this case, the age of the user. Once the data was tabularized, it was divided into training, evaluation (parameter tuning) and testing in conformity to the splits described in section 4.1. As for learning regressions, we profited from the following array of approaches:

- **Linear Regression:** This is the most widely used analytic technique for predicting quantitative responses. Let Y denote the “dependent” variable (the one we are interested in forecasting), and let X_1, \dots, X_p represent the “independent” variables, wherefrom we want to build a prediction model for Y . Then the linear regression model takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_p + \epsilon \quad (7)$$

where β is the set of regression coefficients and ϵ is the error term in the regression model.

- **k-Nearest Neighbor (k -NN) Regression:** It is another classical method employed for determining quantitative outcomes. To assign a new vector with a label, this strategy finds the k nearest neighbors using a pre-defined distance metric (e.g., Euclidean or Mahalanobis). This distance is computed for all samples across the training material, and the final estimate is given by averaging the labels of the k closest training instances. More concretely, the k -NN regression function η for an unseen point \vec{v} is as follows:

$$\hat{\eta}_{k-NN}(\vec{v}) = \frac{1}{k} \sum_{i=1}^k Y(\vec{v}_i) \quad (8)$$

where Y returns the class label of a point \vec{v} within the seen samples (\vec{V}). In this approach, the number of nearest neighbors is a parameter, and $\hat{\eta}$ is the predicted label for the new data point \vec{v} .

⁸github.com/VHRanger/nodevectors

- 343 • **Support Vector Regression (SVR):** For a given training set consisting of n data points \vec{v}_i and their respective
344 labels y_i ($i = \{1, 2, 3, \dots, n\}$), the support vector regression model is expressed as:

$$345 \quad Y(\vec{v}_i) = \sum_i^n \beta_i \phi(\vec{v}_i, \vec{v}_i) + b \quad (9)$$

346 where β is defined as the weight vector which corresponds to a nonlinear mapping function (ϕ) and a constant
347 threshold b . The parameters β and b need to be estimated externally in the model.

- 348 • **Ridge Regression:** This technique is similar to the linear regression, but enriched with a regularization term. The
349 goal of this term is shrinking the coefficients of the model, which are produced in an effort to reduce the variance.
350 In this way, this strategy aims at achieving much lower prediction errors. In particular, the ridge regression
351 coefficients estimates $\hat{\beta}$ are the values that minimize:

$$352 \quad \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (10)$$

353 Here, $\lambda > 0$ is a tuning parameter that needs to be determined separately and the term $\lambda \sum_{j=1}^p \beta_j^2$ is called a
354 *shrinkage penalty*.

- 355 • **Lasso Regression:** It is conceptually similar to Ridge Regression, but this approach applies the penalty by using
356 an absolute value:

$$357 \quad \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (11)$$

358 Likewise, $\lambda > 0$ is a tuning parameter that needs to be set separately, and $\lambda \sum_{j=1}^p |\beta_j|$ denotes the absolute value
359 of the *shrinkage penalty*.

- 360 • **Random Forest Regression:** This is a non-parametric ensemble regression method, which comprises a set of
361 M trees: $\{T_1(X), T_2(X), \dots, T_M(X)\}$, where $X = \{x_1, x_2, \dots, x_p\}$, is a p -dimension input feature vector. The
362 ensemble produces M outputs for each tree $\hat{Y}_1 = T_1(X), \dots, \hat{Y}_M = T_M(X)$, where $\hat{Y}_m, m = 1, \dots, M$ is the m^{th}
363 tree output. The final regression output is computed as the average of all predictions given by the tree.
- 364 • **Artificial Neural Network Regression (ANNR):** It learns a function $f(\cdot) : R^n \rightarrow R^o$ by training on a dataset. In
365 this function, n denotes the number of dimensions for the input vectors; whereas o is the number of dimensions
366 for the output. For a set of feature vectors $X = \{x_1, x_2, x_3, \dots, x_p\}$ where p is total number of features and a
367 continuous variable y as a target, this function can learn a non-linear approximator that performs the regression.

368 Note that hyperparameters for every regression model, their respective search spaces, and their scoring functions
369 are shown in Table 1.

370 4.4. Evaluation Metric

371 The Root Mean Square Error, or RMSE for short, is typically used for evaluating supervised regression algorithms.
372 Two distinct advantages of RMSE over other metrics are: a) it penalizes large errors, and b) it avoids absolute values,
373 which are highly undesirable in many mathematical calculations [99, 100]. More precisely, RMSE is calculated as the
374 square root of the averaged squared differences between each prediction \hat{y}_i and its actual observation y_i :

$$375 \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

376 Note that in this formula, n is the total number of observations, and the “squared” error prevents the cancellation
377 of positive and negative values.

Model	Hyperparameters and Search Space	Scoring Function
Linear Regression	no parameter required	–
Ridge Regression	alpha: [1e-15,1e-10,1e-5,1e-2,1e-1,1,5,10]	Mean squared error
Lasso Regression	alpha: [1e-15,1e-10,1e-5,1e-2,1e-1,1,5,10]	Mean squared error
K-Nearest Neighbor Regression	neighbors: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	Mean squared error
Artificial Neural Network Regression (ANNR)	number of hidden layers: [2, 4, 6, 8, 10, 20, 30, 40] number of hidden units :[8, 16, 32, 64] dropout rate [0.2, 0.5] epochs [50, 100] batch size [32, 64]	Mean squared error
Support Vector Regression	C: [1e-3, 1e-2, 1e-1, 1, 10] gamma: [1e-5, 1e-3, 1,10]	Mean squared error
Random Forest Regression	Estimators: [100, 200] max Depth: [3, 4, 5, 6, 7, 8, 9]	Mean squared error

Table 1

Hyperparameters search space used for tuning the different regression models.

378 Similarly, we also used the Mean Absolute Percentage Error (MAPE) metric. It is scale independent and provides
379 meaningful error as a percentage. It can be used, therefore, to compare prediction performance over different datasets.
380 MAPE is computed as:

$$381 \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (13)$$

382 5. Results

383 Table 2 displays the results obtained by each combination of regression and embedding models. In light of these
384 figures, we draw the following conclusions:

- 385 1. Continuous feature spaces built by Node2Vec outperform all its counterparts regardless of the regression strategy
386 used for the predictions.
- 387 2. It is worth noting that GCN learns node embeddings by incorporating node features into the graph. There are no
388 external node features in this task, which might have been detrimental to its performance.
- 389 3. Most important, ANNR trained with Node2Vec and ProNE embeddings accomplished the best performance,
390 namely the two lowest RMSE scores. The difference between these two values is 4.7%. Conversely, when trained
391 with GCN embeddings, the RMSE decreased by 20.51% with respect to the best configuration.
- 392 4. As for Support Vector Regressions, they quickly converge but yield significant training errors. It might be due
393 to the fact that the training data is still sparse with respect to the labels, despite the high number of embeddings.

394 As for embeddings, Figure 4 confirms the superiority of Node2Vec as it also reaps the lowest MAPE value. In this
395 embedding, Support Vector Regression and K-Nearest Neighbour Regression finish with tie scores, and the gap with
396 ANNR is very narrow. To be more concrete, Support Vector Regression accomplishes marginally lower MAPE values
397 wrt. ANNR when operating on top of Node2Vec, GGVec, ProNE, NMF, and SVD. Regarding RMSE (see Table 2),
398 ANNR performs better under Node2Vec, GGVec, ProNE, GCN, and NMF. Since we prefer minimizing the error to
399 minimizing the relative error, using RMSE is a much more suitable option. It is worth recalling here that RMSE is
400 more useful for showing bigger deviations than MAPE, and it heavily penalizes outliers [101].

401 To check statistical significance, we applied the Wilcoxon signed rank test between the results produced by ANNR
402 and all other regression techniques (see Table 3). Overall, significant differences were found between predictions cast
403 by ANNR and most of the other approaches (p -values < 0.05).

404 Figure 5 highlights the RMSE score obtained by varying the dimensions for the Node2vec space from 10 to 128.
405 In short, a larger space is capable of storing more information. In Figure 5, we observe that when the embedding
406 dimension is greater than 80, the RMSE starts to decrease. Eventually, Node2Vec finished with its best performance
407 using its default embedding parameter (128). However, this improvement is marginal compared to lower amounts of
408 dimensions, namely 110 and 120. It also means that enlarging this space does not necessarily entail a significantly
409 better performance.

Age prediction

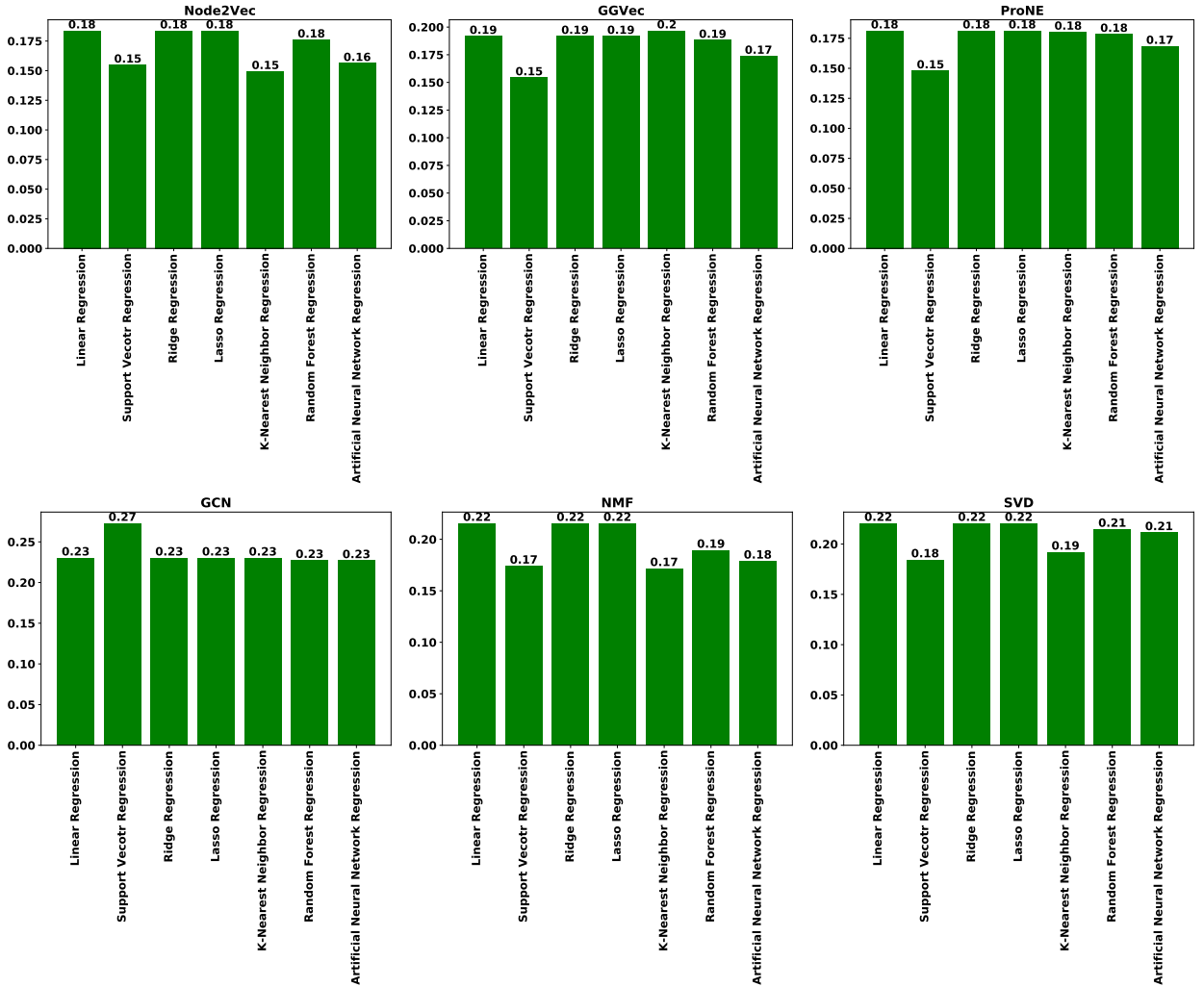


Figure 4: Performance of different regression algorithm in predicting age. X-axis is the regression algorithms and Y-axis is the MAPE score.

410 Node2Vec learns feature representations of nodes in the network. Generally speaking, as the number of target
 411 features increases, the node representation is likely to improve, leading to the construction of more effective regression
 412 models. In our case, from 10 to 80 dimensions, Node2Vec did not help to reduce the RMSE score significantly.
 413 However, from 80 on, it begins to have a greater impact on the performance of the regression models, namely
 414 notoriously decreasing the RMSE value. We also observe that the performance starts to asymptotically converge when
 415 the size of the feature vectors reaches a value of around 100. This result confirms the observations discussed in the
 416 original work of Node2Vec (cf. [67]). On the whole, we conjecture that this is also, in part, a result of the sensitivity of
 417 the various hyperparameter combinations found in ANNR. Thus, the optimal performance of the Node2vec+ANNR
 418 configuration depends on all its factors: the hyperparameters, the network, and the regression task.

419 6. Discussion

420 6.1. Node2Vec Embeddings

421 The best regression models were built on top of Node2Vec embeddings (see Table 2). On average, they score an
 422 RMSE of 8.84 ($\sigma = \pm 0.287$), and the best fit in this group finishes with a value of 8.39. Curiously enough, k -NN
 423 regression is the runner-up reaping an RMSE of 8.56. Given that, in this model, a predicted age is the average of its k

Regression Models	Embedding Models					
	Node2Vec	GGVec	ProNE	GCN	NMF	SVD
Linear Regression [102]	8.95	9.31	9.10	10.21	9.8	9.89
Support Vector Regression [103]	9.26	9.75	9.47	10.47	10.31	10.18
Ridge Regression [104]	8.95	9.31	9.10	10.21	9.8	9.89
Lasso Regression [105]	8.95	9.31	9.11	10.20	9.8	9.88
K-Nearest Neighbor Regression [106]	8.56	8.91	8.92	10.17	8.9	9.85
Random Forest Regression [107]	8.84	9.26	9.03	10.11	9.11	9.69
Artificial Neural Network Regression [108]	8.39	8.81	8.79	10.11	8.82	9.72

Table 2

Results achieved by each graph embedding algorithm coupled with different supervised machine learning regression models. Figures denote the respective RMSE scores.

	Linear Regression	Support Vector Regression	Ridge Regression	Lasso Regression	K-Nearest Neighbor Regression	Random Forest Regression
Node2Vec	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
GGVec	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
ProNE	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
GCN	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	0.00441
NMF	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
SVD	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16

Table 3

P -values corresponding to the Wilcoxon Signed Rank Test ($\alpha = 0.05$). Figures in bold indicate significant p -values.

424 nearest neighbors, we can conclude that there is a significant degree of homogeneity among fellows of the same age in
425 terms of their activity patterns in the community.

426 As a means of finding the optimal amount of neighbors, we carried out experiments considering the following
427 values for k : {2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60}. It is worth stressing that we used the Euclidean distance
428 since it is well-known to be the best method for continuous feature vectors [69, 109]. Note that Figure 7 shows that the
429 optimal value for k is 30. Until $k = 30$, the RMSE decreases when the number of neighbors increases, and after that,
430 the RMSE starts to increase slowly. It is because by adding more neighbors, the prediction starts to approach the mean
431 value of the user's age. Hence, the identification of the optimal quantity of neighbors is crucial for reliable predictions.
432 That is to say, each birth group has its own array of informative characteristics that are present across their members
433 to the extent that these traits define them. Consequently, the corresponding centroid vector can capture the pertinence
434 of each discriminative feature to a specific cohort.

435 Since these regression models were constructed in the 128-dimensional embedding space, we capitalized on PCA
436 (Principal Component Analysis) to infer the three main components and plot the respective centroids on a 3D-graph
437 afterward (see Figure 6a).

438 From eighteen to forty years old, these centroid vectors form a trail ordered by age, which gradually and
439 systematically moves from one age cluster to the other. To be more precise, the prior and succeeding age groups
440 are likely to be the closest neighbors of a particular cluster. This move changes its direction in conformity to the
441 distinct components that start to become, to a certain degree, less and more informative. Therefore, there is a gradual,
442 continuous change of these three latent attributes when transitioning from one age cohort to the other. We conjecture
443 that the minimum age requirement to enroll in the platform (i.e., twelve years old) is, in part, one of the main causes
444 for the noise observed before eighteen years old since younger people sometimes misrepresent their age as a way of
445 circumventing the policies of the platform.

446 In the case of members older than forty, we conjecture that there is less diversity in this segment due to the fact
447 that most of these people trail in the adoption of cQA technologies. For this reason, the data represents a segment that
448 belongs to a specific profile of people who is more likely to enroll in this sort of service. Another reason might be
449 that additional, but relatively insignificant variables are needed to draw a more apparent distinction. Similarly, we find

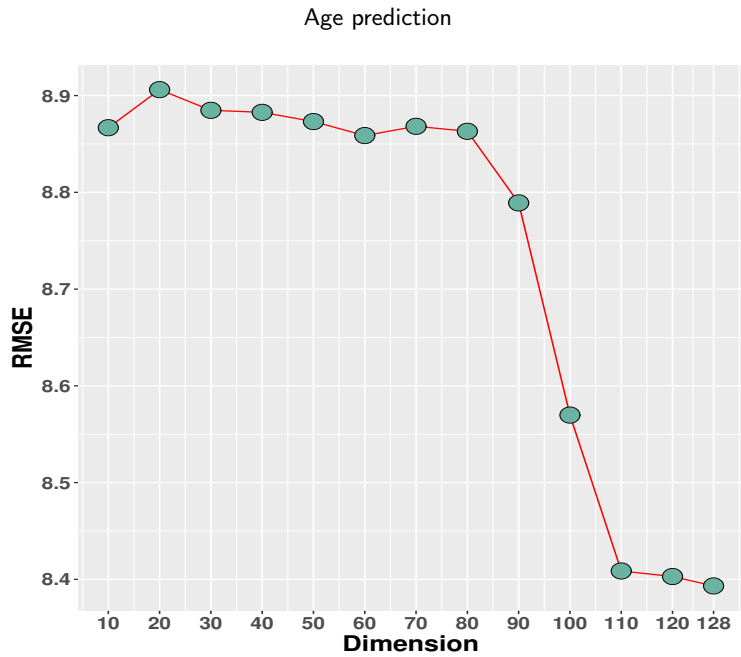
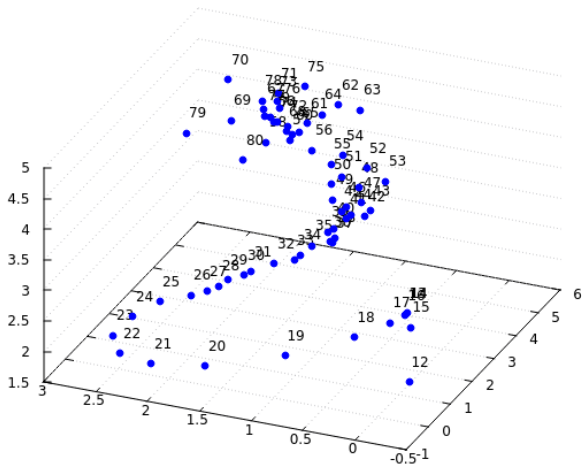
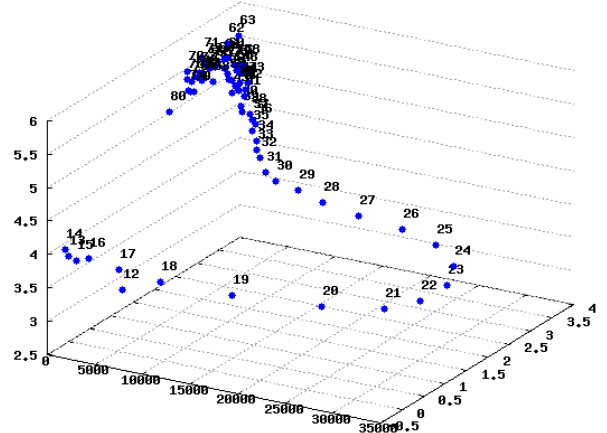


Figure 5: Impact of the Node2vec space dimensionality on the performance of ANNR.



(a) The three components were obtained via PCA on the 128-dimensional Node2Vec embedding used by the best models (cf. Table 2).



(b) The three components were obtained by directly training three-dimensional Node2Vec embeddings.

Figure 6: Three-dimensional centroid vectors. Each point represents the centroid of an age group from twelve to 80 years old.

450 similar qualitative results when training three-dimensional embeddings straightforwardly (see Figure 6b). Given these
 451 findings, we built an ANN regression model on top of this new embedding, and accordingly, the final RMSE value
 452 achieved by this configuration was 9.849. Thus, learning the remaining embedding dimensions is necessary since they
 453 bring about a tangible performance improvement.

454 Another attractive aspect of this space is that centroids are linearly separable to a great extent (i.e., older/younger
 455 than a particular age). Note that a centroid vector here is the geometric center of an age group; that is to say, it is the
 456 arithmetic mean position of all points (community members) within that particular group. It opens up the opportunity
 457 of exploiting well-studied linear classification models, which are also typically computationally efficient. Furthermore,
 458 in the event of casting this problem as a classification task, it would be interesting to investigate if more considerable

Age prediction

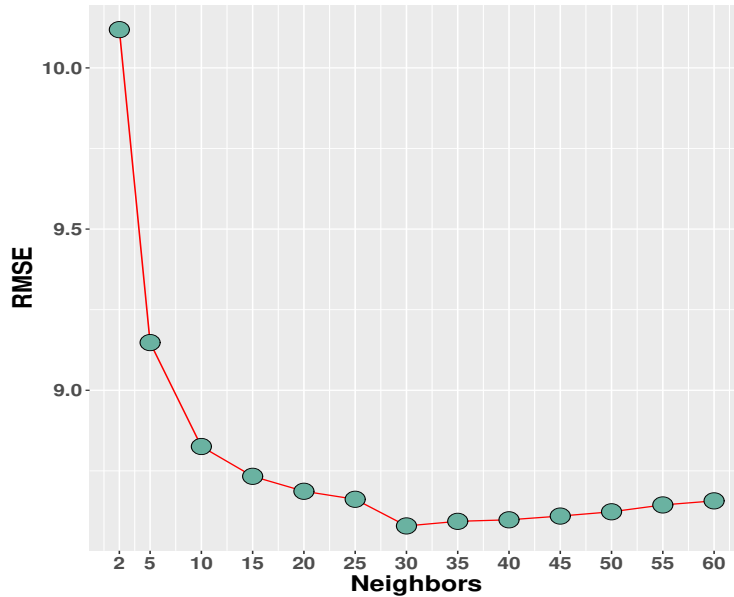


Figure 7: Impact of the number of neighbors on the Node2vec space on the performance of ANNR.

459 distances between consecutive centroids are better candidates for class (generational) boundaries, following a similar
460 principle to Support Vector Machines (SVMs) [110].

461 Lastly, Figure 6b suggests that these centroids might form some kind of loop or spiral, which might give support to
462 the recurring generation cycle (i.e., four archetypes of generations) theorized by William Strauss and Neil Howe [111].

463 6.2. Graph-based versus Text/Image-oriented Models

464 From another standpoint, the outcomes reaped by the best regression model (i.e., Node2Vec + ANN) can be
465 mapped into their respective age cohorts, this way we can compute its corresponding Accuracy and F1-Score, and
466 consequently, we can have a rough estimate of how this graph-based model compares to systems that have been modeled
467 as a classification task.

468 As a reference, we can use the five age groups proposed by Strauss and Howe for obtaining both classification
469 metrics [111]. Accordingly, our best graph regression finished with Accuracy and F1-Score of 62.27% and 0.3423,
470 respectively. If we consider the suggestion of [2] on reducing these five age cohorts to three, our best regression achieves
471 an Accuracy of 63.52% and an F1-Score of 0.5705.

472 Our study benefited from the same datasets as from [2], but in juxtaposition, this prior research focused on the text
473 view of this collection instead of the underlying graph. Interestingly enough, the results accomplished by exploiting
474 graph activity patterns rival their counterpart's neural networks for texts. More precisely, their best F1-Score is 0.3444
475 in the case of five cohorts, while in the event of three age groups, it varies from 0.4361 to 0.6350 (cf. [2]).

476 The work of [2] also tried models based on profile images, but in this case, their models were less effective. They
477 concluded that it is hard to find visual patterns related to age, even by an eyeball inspection. For this modality, their
478 best Accuracy was 60.6% for three classes. It is worth stressing here that this empirical setting was different from the
479 one used in our graph and their text models due to the fact that many community members do not have a profile image.
480 Therefore, this array of image samples belongs to a subset of the community fellows utilized for devising graph and
481 text-based models.

482 Overall, our results signal that graph-based neural network regression models hold promise, especially since they
483 seem to be a better option than profile images for this task. Moreover, they are competitive with textual models and
484 can be used for constructing classifiers that take advantage of the synergy between these two modalities, leveraging
485 age-related activity and language patterns. A good example is tackling data sparseness, which characterizes fellows
486 that produce few texts and/or exhibit low activity in the community.

487 7. Conclusion

488 In a statement, this paper dissects the application of assorted regression models and continuous feature spaces to the
 489 prediction of age across cQA service members. In a nutshell, Node2vec significantly outperforms other embeddings
 490 regardless of the regression strategy exploited for automatically assigning an age to a community fellow. Furthermore,
 491 it is worth highlighting that our best prediction rate (i.e., $RMSE = 8.39$) was reaped when Node2vec was amalgamated
 492 with Artificial Neural Network Regressions. It is also worth experimenting in the future about the synergy between
 493 deep-learning based regression model [112] with graph-embeddings.

494 In the Node2vec space, we also discovered that age-based centroid vectors tend to form a trail ordered by age. This
 495 finding brings up several research questions that are worth investigating. For instance, the magnitude of the distance
 496 between consecutive points (e.g., 17 and 18 years old) can signal good cohort borders. The answer to this question
 497 would significantly impact the design of cost-efficient linear classifiers. Another important discovery is that activity
 498 graph-based models can compete with strategies built on top of images and texts.

499 As for future research directions, we envisage that these activity patterns will make the design of practical multi-
 500 modal approaches and linear age predictors easier. We also envision that graph, image and text based models can help
 501 each other in multi-view, multi-task or transfer learning settings.

502 Acknowledgment

503 This work was partially supported by the project Fondecyt “Multimodal Demographics and Psychographics for
 504 Improving Engagement in Question Answering Communities” (1220367) funded by the Chilean Government. We
 505 would also like to acknowledge Science Foundation Ireland (SFI/12/RC/2289_P2) for supporting this research.

506 CRedit authorship contribution statement

507 **Mohan Timilsina:** M.T conducted experiments, analysed the results and revised the draft. **Alejandro Figueroa:**
 508 A.F provided conceptualization of this study, methodology and prepared the original draft.

509 Declaration of Competing Interest

510 The authors declare that they have no known competing financial interests or personal relationships that could have
 511 appeared to influence the work reported in this work.

512 Data Availability Statement

513 The full activity graph and the labels used for this study will be publicly available online upon acceptance.

514 References

- 515 [1] Joeran Beel, Stefan Langer, Andreas Nürnberger, and Marcel Genzmehr. The impact of demographics (age and gender) and other user-
 516 characteristics on evaluating recommender systems. In Trond Aalberg, Christos Papatheodorou, Milena Dobрева, Giannis Tsakonas, and
 517 Charles J. Farrugia, editors, *Research and Advanced Technology for Digital Libraries*, pages 396–400, Berlin, Heidelberg, 2013. Springer
 518 Berlin Heidelberg.
- 519 [2] Alejandro Figueroa, Billy Peralta, and Orietta Nicolis. Coming to grips with age prediction on imbalanced multimodal community question
 520 answering data. *Information*, 12(2), 2021.
- 521 [3] Abdelghani Bouziane, Djelloul Bouchiha, Nouredine Doumi, and Mimoun Malki. Question answering systems: Survey and trends. *Procedia*
 522 *Computer Science*, 73:366 – 375, 2015. International Conference on Advanced Wireless Information and Communication Technologies
 523 (AWICT 2015).
- 524 [4] Ivan Srba and Maria Bielikova. A comprehensive survey and classification of approaches for community question answering. *ACM Trans.*
 525 *Web*, 10(3), August 2016.
- 526 [5] Arshad Ahmad, Chong Feng, Shi Ge, and Abdallah Yousif. A survey on mining stack overflow: question and answering (q&a) community.
 527 *Data Technol. Appl.*, 52:190–247, 2018.
- 528 [6] Zhengfa Yang, Qian Liu, Baowen Sun, and Xin Zhao. Expert recommendation in community question answering: a review and future
 529 direction. *International Journal of Crowd Science*, 3:348–372, 09 2019.
- 530 [7] Nouha Othman, Rim Faiz, and Kamel Smaïli. Enhancing question retrieval in community question answering using word embeddings.
 531 *Procedia Computer Science*, 159:485–494, 2019. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the
 532 23rd International Conference KES2019.

- 533 [8] Barun Patra. A survey of Community Question Answering. *ArXiv e-prints*, May 2017.
- 534 [9] Mohammad Sadegh Zahedi, Maseud Rahgozar, and Reza Aghaeizadeh Zoroofi. Hca: Hierarchical compare aggregate model for question
535 retrieval in community question answering. *Information Processing & Management*, 57(6):102318, 2020.
- 536 [10] Harry Stokhof, Kalliopi Meli, and Konstantinos Lavidas. To answer, or not to answer, that is the question: Experts' contribution to question-
537 answering platforms. 11 2020.
- 538 [11] Negin Ghasemi, Ramin Fatourehchi, and Saeedeh Momtazi. User embedding for expert finding in community question answering. *ACM*
539 *Trans. Knowl. Discov. Data*, 15(4), March 2021.
- 540 [12] Dipankar Kundu, Rajat Kumar Pal, and Deba Prasad Mandal. Preference enhanced hybrid expertise retrieval system in community question
541 answering services. *Decision Support Systems*, 129:113164, 2020.
- 542 [13] Dipankar Kundu, Rajat Kumar Pal, and Deba Prasad Mandal. Topic sensitive hybrid expertise retrieval system in community question
543 answering services. *Knowledge-Based Systems*, 211:106535, 2021.
- 544 [14] Lingfeng Dong, Jie Zhang, Liqiang Huang, and Yuan Liu. Social influence on endorsement in social q&a community: Moderating effects of
545 temporal and spatial factors. *International Journal of Information Management*, 61:102396, 2021.
- 546 [15] Zeyu Li, Jyun-Yu Jiang, Yizhou Sun, and Wei Wang. Personalized question routing via heterogeneous network embedding. In *The Thirty-
547 Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference,
548 IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27
549 - February 1, 2019*, pages 192–199. AAAI Press, 2019.
- 550 [16] Mohan Timilsina, Brian Davis, Mike Taylor, and Conor Hayes. Predicting citations from mainstream news, weblogs and discussion forums.
551 In *Proceedings of the International Conference on Web Intelligence*, pages 237–244, 2017.
- 552 [17] Jiahui Wen, Hongkui Tu, Xiaohui Cheng, Renquan Xie, and Wei Yin. Joint modeling of users, questions and answers for answer selection
553 in cqa. *Expert Systems with Applications*, 118:563 – 572, 2019.
- 554 [18] Hengyi Fu and Sanghee Oh. Quality assessment of answers with user-identified criteria and data-driven features in social q&a. *Information
555 Processing & Management*, 56(1):14 – 28, 2019.
- 556 [19] Min Yang, Lei Chen, Ziyu Lyu, Junhao Liu, Ying Shen, and Qingyao Wu. Hierarchical fusion of common sense knowledge and classifier
557 decisions for answer selection in community question answering. *Neural Networks*, 132:53–65, 2020.
- 558 [20] Leandro Amancio, Carina F. Dorneles, and Daniel H. Dalip. Recency and quality-based ranking question in cqas: A stack overflow case
559 study. *Information Processing & Management*, 58(4):102552, 2021.
- 560 [21] Yongliang Wu, Shuliang Zhao, and Ruiqiang Guo. A novel community answer matching approach based on phrase fusion heterogeneous
561 information network. *Information Processing & Management*, 58(1):102408, 2021.
- 562 [22] Min-Chul Yang, Do-Gil Lee, So-Young Park, and Hae-Chang Rim. Knowledge-based question answering using the semantic embedding
563 space. *Expert Systems with Applications*, 42(23):9086–9104, 2015.
- 564 [23] Yikang Shen, Wenge Rong, Nan Jiang, Baolin Peng, Jie Tang, and Zhang Xiong. Word embedding based correlation model for
565 question/answer matching. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI '17*, page 3511–3517. AAAI
566 Press, 2017.
- 567 [24] Lei Sang, Min Xu, ShengSheng Qian, and Xindong Wu. Multi-modal multi-view bayesian semantic embedding for community question
568 answering. *Neurocomputing*, 334:44–58, 2019.
- 569 [25] Daniel Palomera and Alejandro Figueroa. Leveraging linguistic traits and semi-supervised learning to single out informational content across
570 how-to community question-answering archives. *Information Sciences*, 381:20 – 32, 2017.
- 571 [26] Ido Guy, Victor Makarencov, Niva Hazon, Lior Rokach, and Bracha Shapira. Identifying informational vs. conversational questions on
572 community question answering archives. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining,
573 WSDM '18*, pages 216–224, New York, NY, USA, 2018. ACM.
- 574 [27] Sudha Rao and Hal Daumé III. Learning to ask good questions: Ranking clarification questions using neural expected value of perfect
575 information. *CoRR*, abs/1805.04655, 2018.
- 576 [28] Dalia Elalfy, Walaa Gad, and Rasha Ismail. A hybrid model to predict best answers in question answering communities. *Egyptian Informatics
577 Journal*, 19(1):21 – 31, 2018.
- 578 [29] Shanshan Lyu, Wentao Ouyang, Yongqing Wang, Huawei Shen, and Xueqi Cheng. What we vote for? answer selection from user expertise
579 view in community question answering. In *The World Wide Web Conference, WWW '19*, page 1198–1209, New York, NY, USA, 2019.
580 Association for Computing Machinery.
- 581 [30] Jiahui Wen, Jingwei Ma, Yiliu Feng, and Mingyang Zhong. Hybrid attentive answer selection in cqa with deep users modelling. In *32nd
582 AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 2556–2563. AAAI Press, 2018.
- 583 [31] Zhou Zhao, Hanqing Lu, Vincent W Zheng, Deng Cai, Xiaofei He, and Yueting Zhuang. Community-based question answering via
584 asymmetric multi-faceted ranking network learning. In *AAAI*, volume 17, pages 3532–3539, 2017.
- 585 [32] Zan-Xia Jin, Bo-Wen Zhang, Fang Zhou, Jingyan Qin, and Xu-Cheng Yin. Ranking via partial ordering for answer selection. *Information
586 Sciences*, 538:358–371, 2020.
- 587 [33] Xiaoqiang Zhou, Baotian Hu, Qingcai Chen, and Xiaolong Wang. Recurrent convolutional neural network for answer selection in community
588 question answering. *Neurocomputing*, 274:8 – 18, 2018. Query Understanding.
- 589 [34] Shahar Harel, Sefi Albo, Eugene Agichtein, and Kira Radinsky. Learning novelty-aware ranking of answers to complex questions. In *The
590 World Wide Web Conference, WWW '19*, page 2799–2805, New York, NY, USA, 2019. Association for Computing Machinery.
- 591 [35] Zejun Sun, Yanan Sun, Xinfeng Chang, Qiming Wang, Xuetong Yan, Zhongqiang Pan, and Zong ping Li. Community detection based on
592 the matthew effect. *Knowledge-Based Systems*, 205:106256, 2020.
- 593 [36] Ming Li, Xiuzhi Lu, Lisheng Chen, and Jun Wang. Knowledge map construction for question and answer archives. *Expert Systems with
594 Applications*, 141:112923, 2020.

- 595 [37] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. No country for old members: User
596 lifecycle and linguistic change in online communities. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW*
597 '13, pages 307–318, New York, NY, USA, 2013. ACM.
- 598 [38] Chaogang Fu. Tracking user-role evolution via topic modeling in community question answering. *Information Processing & Management*,
599 56(6):102075, 2019.
- 600 [39] Bei Yan and Lian Jian. Beyond reciprocity: The bystander effect of knowledge response in online knowledge communities. *Computers in*
601 *Human Behavior*, 76:9–18, 2017.
- 602 [40] Yuxiang (Chris) Zhao, Zhouying Liu, Si Chen, and Qinghua Zhu. From free to fee: Exploring the factors that influence the askers' switching
603 behavior on online q&a platforms. *Proceedings of the Association for Information Science and Technology*, 56(1):517–520, 2019.
- 604 [41] Alexander Espina and Alejandro Figueroa. Why was this asked? automatically recognizing multiple motivations behind community question-
605 answering questions. *Expert Systems with Applications*, 80:126 – 135, 2017.
- 606 [42] Chencheng Fang and Jiantong Zhang. Users' continued participation behavior in social q&a communities: A motivation perspective.
607 *Computers in Human Behavior*, 92:87–109, 2019.
- 608 [43] Langtao Chen, Aaron Baird, and Detmar Straub. Why do participants continue to contribute? evaluation of usefulness voting and commenting
609 motivational affordances within an online knowledge community. *Decision Support Systems*, 118:21–32, 2019.
- 610 [44] Mohamed Bouguessa and Lotfi Ben Romdhane. Identifying authorities in online communities. *ACM Trans. Intell. Syst. Technol.*, 6(3):30:1–
611 30:23, April 2015.
- 612 [45] J. Atkinson and A. Maurelia. Redundancy-based trust in question-answering systems. *Computer*, 50(01):58–65, jan 2017.
- 613 [46] Pradeep K. Roy, Jyoti P. Singh, Abdullah M. Baabdullah, Hatice Kizgin, and Nripendra P. Rana. Identifying reputation collectors in
614 community question answering (cqa) sites: Exploring the dark side of social media. *International Journal of Information Management*,
615 42:25 – 35, 2018.
- 616 [47] Dorra Attiaoui, Arnaud Martin, and Boutheina Ben Yaghlane. Belief measure of expertise for experts detection in question answering
617 communities: case study stack overflow. *Procedia Computer Science*, 112:622 – 631, 2017. Knowledge-Based and Intelligent Information
618 & Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France.
- 619 [48] Mahmood Neshati, Zohreh Fallahnejad, and Hamid Beigy. On dynamicity of expert finding in community question answering. *Information*
620 *Processing & Management*, 53(5):1026 – 1042, 2017.
- 621 [49] Chaogang Fu. User intimacy model for question recommendation in community question answering. *Knowledge-Based Systems*, 188:104844,
622 2020.
- 623 [50] Piero Molino, Luca Maria Aiello, and Pasquale Lops. Social question answering: Textual, user, and network features for best answer
624 prediction. *ACM Trans. Inf. Syst.*, 35(1):4:1–4:40, September 2016.
- 625 [51] Grégoire Burel, Paul Mulholland, Yulan He, and Harith Alani. Predicting answering behaviour in online question answering communities.
626 In *Proceedings of the 26th ACM Conference on Hypertext & #38; Social Media*, HT '15, pages 201–210, New York, NY, USA, 2015. ACM.
- 627 [52] Konstantinos Pelechrinis, Vladimir Zadorozhny, Velin Kounev, Vladimir Oleshchuk, Mohd Anwar, and Yiling Lin. Automatic evaluation of
628 information provider reliability and expertise. *World Wide Web*, 18(1):33–72, Jan 2015.
- 629 [53] Jiahui Wen, Hongkui Tu, Xiaohui Cheng, Renquan Xie, and Wei Yin. Joint modeling of users, questions and answers for answer selection
630 in cqa. *Expert Systems with Applications*, 118:563 – 572, 2019.
- 631 [54] Imrul Kayes, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, and Francesco Bonchi. The social world of content abusers in
632 community question answering. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, pages 570–580,
633 Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.
- 634 [55] William L Hamilton, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. Loyalty in online communities.
635 In *Eleventh International AAAI Conference on Web and Social Media*, pages 540–543, 2017.
- 636 [56] Zhe Liu and Bernard J. Jansen. Identifying and predicting the desire to help in social question and answering. *Information Processing &*
637 *Management*, 53(2):490 – 504, 2017.
- 638 [57] Yan Zhang, Mingli Zhang, Nuan Luo, Yu Wang, and Tao Niu. Understanding the formation mechanism of high-quality knowledge in social
639 question and answer communities: A knowledge co-creation perspective. *International Journal of Information Management*, 48:72–84, 2019.
- 640 [58] Tao Guan, Le Wang, Jiahua Jin, and Xiaolong Song. Knowledge contribution behavior in online q&a communities: An empirical investigation.
641 *Computers in Human Behavior*, 81:137 – 147, 2018.
- 642 [59] Onur Kucuktunc, B. Barla Cambazoglu, Ingmar Weber, and Hakan Ferhatosmanoglu. A large-scale sentiment analysis for yahoo! answers.
643 In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM '12*, pages 633–642, New York, NY,
644 USA, 2012. ACM.
- 645 [60] Alejandro Figueroa. Male or female: What traits characterize questions prompted by each gender in community question answering? *Expert*
646 *Systems with Applications*, 90:405 – 413, 2017.
- 647 [61] P. Morrison and E. Murphy-Hill. Is programming knowledge related to age? an exploration of stack overflow. In *2013 10th Working*
648 *Conference on Mining Software Repositories (MSR)*, pages 69–72, May 2013.
- 649 [62] Alejandro Figueroa and Mohan Timilsina. What identifies different age cohorts in yahoo! answers? *Knowledge-Based Systems*, 228:107278,
650 2021.
- 651 [63] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and
652 applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.
- 653 [64] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM*
654 *SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710, 2014.
- 655 [65] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international*
656 *conference on Knowledge discovery and data mining*, pages 1225–1234, 2016.

- 657 [66] Li Liu, William K Cheung, Xin Li, and Lejian Liao. Aligning users across social networks using network embedding. In *Ijcai*, pages
658 1774–1780, 2016.
- 659 [67] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international
660 conference on Knowledge discovery and data mining*, pages 855–864, 2016.
- 661 [68] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph
662 neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020.
- 663 [69] Mohan Timilsina, Declan Patrick Mc Kernan, Haixuan Yang, and Mathieu D’Aquin. Synergy between embedding and protein functional
664 association networks for drug label prediction using harmonic function. *IEEE/ACM Transactions on Computational Biology and
665 Bioinformatics*, 2020.
- 666 [70] Predrag Radivojac, Wyatt T Clark, Tal Ronnen Oron, Alexandra M Schnoes, Tobias Wittkop, Artem Sokolov, Kiley Graim, Christopher Funk,
667 Karin Verspoor, Asa Ben-Hur, et al. A large-scale evaluation of computational protein function prediction. *Nature methods*, 10(3):221–227,
668 2013.
- 669 [71] Lars Backstrom and Jure Leskovec. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the
670 fourth ACM international conference on Web search and data mining*, pages 635–644, 2011.
- 671 [72] Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. Asymmetric transitivity preserving graph embedding. In *Proceedings of
672 the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1105–1114, 2016.
- 673 [73] Chang Zhou, Yuqiong Liu, Xiaofei Liu, Zhongyi Liu, and Jun Gao. Scalable graph embedding for asymmetric proximity. In *Proceedings of
674 the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
- 675 [74] Anton Tsitsulin, Davide Mottin, Panagiotis Karras, and Emmanuel Müller. Verse: Versatile graph embeddings from similarity measures. In
676 *Proceedings of the 2018 world wide web conference*, pages 539–548, 2018.
- 677 [75] Renchi Yang, Jieming Shi, Xiaokui Xiao, Yin Yang, and Sourav S Bhowmick. Homogeneous network embedding for massive graphs via
678 reweighted personalized pagerank. *arXiv preprint arXiv:1906.06826*, 2019.
- 679 [76] Zexi Huang, Arlei Silva, and Ambuj Singh. A broader picture of random-walk based graph embedding. In *Proceedings of the 27th ACM
680 SIGKDD conference on knowledge discovery & data mining*, pages 685–695, 2021.
- 681 [77] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*,
682 2016.
- 683 [78] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
- 684 [79] Hongwei Wang, Jia Wang, Jialin Wang, Miao Zhao, Weinan Zhang, Fuzheng Zhang, Xing Xie, and Minyi Guo. Graphgan: Graph
685 representation learning with generative adversarial nets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
- 686 [80] Quanyu Dai, Qiang Li, Jian Tang, and Dan Wang. Adversarial network embedding. In *Proceedings of the AAAI Conference on Artificial
687 Intelligence*, volume 32, 2018.
- 688 [81] Quanyu Dai, Xiao Shen, Liang Zhang, Qiang Li, and Dan Wang. Adversarial training methods for network embedding. In *The World Wide
689 Web Conference*, pages 329–339, 2019.
- 690 [82] Amr Ahmed, Nino Shervashidze, Shravan Narayanamurthy, Vanja Josifovski, and Alexander J Smola. Distributed large-scale natural graph
691 factorization. In *Proceedings of the 22nd international conference on World Wide Web*, pages 37–48, 2013.
- 692 [83] Mohan Timilsina, Meera Tandan, Mathieu d’Aquin, and Haixuan Yang. Discovering links between side effects and drugs using a diffusion
693 based method. *Scientific reports*, 9(1):10436, 2019.
- 694 [84] Shaosheng Cao, Wei Lu, and Qiongfai Xu. Grarep: Learning graph representations with global structural information. In *Proceedings of
695 the 24th ACM international conference on information and knowledge management*, pages 891–900, 2015.
- 696 [85] Jie Zhang, Yuxiao Dong, Yan Wang, Jie Tang, and Ming Ding. Prone: Fast and scalable network representation learning. In *IJCAI*, volume 19,
697 pages 4278–4284, 2019.
- 698 [86] Xin Liu, Tsuyoshi Murata, Kyoung-Sook Kim, Chatchawan Kotarasu, and Chenyi Zhuang. A general view for network embedding as matrix
699 factorization. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 375–383, 2019.
- 700 [87] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah,
701 Michal Kosinski, David Stillwell, Martin EP Seligman, et al. Personality, gender, and age in the language of social media: The open-
702 vocabulary approach. *PLoS one*, 8(9):e73791, 2013.
- 703 [88] Francisco Rangel, Paolo Rosso, Moshe Koppel, Efstathios Stamatatos, and Giacomo Inches. Overview of the author profiling task at pan
704 2013. In *CLEF Conference on Multilingual and Multimodal Information Access Evaluation*, pages 352–365. CELCT, 2013.
- 705 [89] Francisco Rangel, Paolo Rosso, Irina Chugur, Martin Potthast, Martin Trenkmann, Benno Stein, Ben Verhoeven, and Walter Daelemans.
706 Overview of the 2nd author profiling task at pan 2014. In *CLEF 2014 Evaluation Labs and Workshop Working Notes Papers, Sheffield, UK,
707 2014*, pages 1–30, 2014.
- 708 [90] Francisco Rangel, Paolo Rosso, Ben Verhoeven, Walter Daelemans, Martin Potthast, and Benno Stein. Overview of the 4th author profiling
709 task at pan 2016: cross-genre evaluations. In *Working Notes Papers of the CLEF 2016 Evaluation Labs. CEUR Workshop Proceedings/Balog,
710 Krisztian [edit.]; et al.*, pages 750–784, 2016.
- 711 [91] Francisco Manuel Rangel Pardo, Fabio Celli, Paolo Rosso, Martin Potthast, Benno Stein, and Walter Daelemans. Overview of the 3rd author
712 profiling task at pan 2015. In *CLEF 2015 Evaluation Labs and Workshop Working Notes Papers*, pages 1–8, 2015.
- 713 [92] Roy Christopher Bayot and Teresa Gonçalves. Author profiling using svms and word embedding averages. In *CLEF*, 2016.
- 714 [93] A. Culotta, N. Kumar, and Jennifer Cutler. Predicting the demographics of twitter users from website traffic data. In *AAAI*, 2015.
- 715 [94] Aron Culotta, Nirmal Kumar Ravi, and Jennifer Cutler. Predicting twitter user demographics using distant supervision from website traffic
716 data. *Journal of Artificial Intelligence Research*, 55:389–408, 2016.
- 717 [95] Negin Ghasemi, Ramin Fatourechi, and Saeedeh Momtazi. User embedding for expert finding in community question answering. *ACM
718 Trans. Knowl. Discov. Data*, 15(4), March 2021.

- 719 [96] Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng. Expert finding for question answering via graph regularized matrix completion. *IEEE*
720 *Transactions on Knowledge and Data Engineering*, 27(4):993–1004, 2015.
- 721 [97] Alejandro Figueroa, Carlos Gómez-Pantoja, and Günter Neumann. Integrating heterogeneous sources for predicting question temporal
722 anchors across yahoo! answers. *Information Fusion*, 50:112 – 125, 2019.
- 723 [98] Bolun Chen, Fenfen Li, Senbo Chen, Ronglin Hu, and Ling Chen. Link prediction based on non-negative matrix factorization. *PloS one*,
724 12(8):e0182968, 2017.
- 725 [99] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae). *Geoscientific Model Development*
726 *Discussions*, 7(1):1525–1534, 2014.
- 727 [100] Mala Saraswat, Anil Dubey, Satyam Naidu, Rohit Vashisht, and Abhishek Singh. Web-based movie recommender system. In *Ambient*
728 *Communications and Computer Systems*, pages 291–301. Springer, 2020.
- 729 [101] Supriyo Ahmed, Ripon K Chakraborty, Daryl L Essam, and Weiping Ding. Poly-linear regression with augmented long short term memory
730 neural network: Predicting time series data. *Information Sciences*, 2022.
- 731 [102] Francis Galton. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and*
732 *Ireland*, 15:246–263, 1886.
- 733 [103] Corinna Cortes and Vladimir Vapnik. Support-vector networks. In *Machine Learning*, pages 273–297, 1995.
- 734 [104] Marvin HJ Gruber. *Improving efficiency by shrinkage: the James-Stein and ridge regression estimators*. Routledge, 2017.
- 735 [105] Robert Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B*
736 *(Statistical Methodology)*, 73(3):273–282, 2011.
- 737 [106] Evelyn Fix and Joseph Lawson Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International*
738 *Statistical Review/Revue Internationale de Statistique*, 57(3):238–247, 1989.
- 739 [107] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1,
740 pages 278–282. IEEE, 1995.
- 741 [108] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy*
742 *of sciences*, 79(8):2554–2558, 1982.
- 743 [109] Kenneth Jonsson, Josef Kittler, YP Li, and Jiri Matas. Support vector machines for face authentication. *Image and Vision Computing*,
744 20(5-6):369–375, 2002.
- 745 [110] T. Joachims. Training linear SVMs in linear time. In *ACM SIGKDD International Conference On Knowledge Discovery and Data Mining*
746 *(KDD)*, pages 217–226, 2006.
- 747 [111] B. Strauss, W. Strauss, and N. Howe. *Generations: The History of America’s Future, 1584 to 2069*. Morrow, 1991.
- 748 [112] Minghui Hu, Jet Heng Chion, Ponnuthurai Nagarathnam Suganthan, and Rakesh Kumar Katuwal. Ensemble deep random vector functional
749 link neural network for regression. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022.