Reputation Deflation Through Dynamic Expertise Assessment in Online Labor Markets

Marios kokkodis
Boston College
kokkodis@bc.edu

ABSTRACT

Current reputation systems in online labor markets (e.g., Freelancer, PeoplePerHour) experience three major shortcomings: (1) reputation inflation (i.e., reputation scores are inflated to "above average" values) (2) reputation attribution (i.e., attribution of reputation scores to individual skills is unfeasible) and (3) reputation staticity (i.e., reputation scores are uniformly averaged over time). These shortcomings render online reputation systems uninformative, and sometimes even misleading. This work proposes a data-driven approach that deflates reputation scores by solving the problems of reputation attribution and saticity. The deflating process starts with projecting any random set of skills to a set of competency dimensions. For each competency dimension, a Hidden Markov Model estimates a contractor's current (but latent) competency-specific expertise. Aggregation of competency-specific estimates provides expertise predictions for any given set of required skills. Empirical analysis on 61,330 completed tasks from a major online labor market shows that the resulting estimates are deflated and they better predict contractor performance. These results suggest a series of implications for online (labor) markets and their users.

CCS CONCEPTS

• **Information systems → Data analytics**;

KEYWORDS

Reputation deflation, Expertise assessment, Digital markets, Online markets, HMM, Word embeddings

ACM Reference Format:
Marios kokkodis. 2019. Reputation Deflation Through Dynamic Expertise Assessment in Online Labor Markets. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3308558.3313479

1 INTRODUCTION

Online labor markets such as Peopleperhour and Freelancer connect buyers with contractors around the globe to accomplish a diverse set of tasks. These tasks span multiple categories, including web development, graphic design, accounting and sales and marketing. On par with other online platforms, online labor markets have experienced an accelerated growth in the past decade [43]. A typical transaction in these markets starts with a buyer posting a job description. Contractors who are looking for opportunities submit their job applications to relevant jobs. The buyer then chooses to hire one (or more) contractor(s) from the pool of applicants. Once hired, the contractors complete the task and receive the prearranged compensations.

To identify the best candidate for each opening, buyers assess a series of observed and latent characteristics of the available contractors. Observed characteristics include the contractors' skills, work histories, and certifications. Latent characteristics include the contractors' true knowledge and abilities. The existence of latent characteristics, the heterogeneity that appears in the observed ones [26], and the interactions between the two create an uncertain environment of information asymmetry [4].

To reduce information asymmetry and help buyers make better-informed hiring decisions, online labor markets have developed reputation systems. Contractors get rated for the tasks they complete, and these ratings become part of their online resumes. Buyers then consider the contractors' past performance as a signal for future performance. In theory, such reputation mechanisms should inform and facilitate buyers' choices. In practice however, current reputation systems experience three major shortcomings:

- Reputation inflation: In online labor markets, raters feel pressure to rate "above average" [14]. At the same time, contractors who receive low ratings are unable to get hired. As a result, they leave the market and look for alternatives, or rejoin with different credentials [20]. These two synergic forces result in a highly inflated rating distribution [14, 19].
- Reputation attribution: Current reputation systems in online labor markets provide a uni-dimensional accumulated reputation score for each contractor [41]. Because these markets support heterogeneous distributions of skills and qualifications [26], accurate attribution of accumulated reputation scores to individual skills is unfeasible. For instance, consider a contractor who completes a task that requires `java`, `SQL` and `python` and receives a feedback score 9/10. Does this score capture the contractor's performance on `java`, on `SQL`, on `python`, or on any combinations of those skills?
- Reputation staticity: Over time, contractors evolve, as they gain expertise on skills that they repeatedly use. Current reputation systems however uniformly average received ratings to provide an aggregate score, thereby assuming that the latent expertise of a contractor does not change over time [27].

The end result of these shortcomings is that current reputation scores are often uninformative and sometimes even misleading.

This work proposes a data-driven framework that directly solves reputation attribution and reputation staticity, and, indirectly, deflates the reputation distribution. The framework first incorporates

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313479

word embeddings to project skills into a set of competency dimensions. For each one of these competency dimensions a Hidden Markov Model (HMM) uses observed signals to dynamically estimate a contractor's competency-specific level of expertise. These estimates are a solution to reputation attribution, as they map expertise to specific skillsets. Because the HMMs allow contractors to dynamically evolve to various expertise states, they provide *current* estimates of expertise (solution to reputation staticity). Aggregation of competency-specific expertise estimates provides expertise predictions for any given set of required skills (HMM-W2V reputation).

In theory, compared to current reputation scores in online labor markets, HMM-W2V reputation should be closer to the true expertise of contractors on a given set of skills. Since these markets attract millions of contractors of various abilities, HMM-W2V reputation is likely to form a deflated, Paretian distribution [38]. Empirical evaluation of the HMM framework on 61,330 completed tasks from a major online labor market shows that, compared to current reputation scores, HMM-W2V reputation is (1) deflated, and (2) a better predictor of contractor performance on a any given set of skills.

This work is the first to propose a data-driven solution that dynamically estimates contractors' expertise in online labor markets. Competency-specific estimates provide deflated scores for the dynamic expertise of contractors on any given set of skills. These estimates can improve an online labor market's transaction efficacy through better and more relevant candidate recommendations [30]. Better-informed buyers are likely to make more successful hiring decisions (through accurate opening-specific expertise estimates) and as a result continue using the platform. At the same time, the market can better understand the demand and supply of experienced contractors in every competency dimension, and create any necessary interventions. Finally, the proposed framework generalizes beyond online labor markets, and it is directly applicable to other markets that suffer from reputation inflation, such as Amazon, Yelp, and TripAdvisor.

2 RESEARCH CONTEXT

Digital markets signal the quality of their services (or products) through online reputation systems. The strong economic impact of these systems on product selection and quality assessment has been repeatedly verified in multiple contexts [1, 3, 12, 13, 23, 29, 44]. Next, we discuss the importance of reputation signals in the context of an online labor market.

2.1 Reputation in online labor markets

Buyers place a very significant weight on contractors' reputation when making hiring decisions [45], and they are willing to trade-off reputation and price to accept higher bids posted by more reputable contractors [36]. This trade-off is also observed on the contractor side, where the contractor's past performance has a significant impact on future wages [7]. Reputation can be a significant predictor of success for fixed-price contracts [32], and it can substantially improve contractors' subsequent employment outcomes [39]. Furthermore, reputation can capture latent qualities of a contractor that transfer across multiple task categories that require a diverse set of skills [27]. Overall, these works show that reputation drives both hiring decisions and subsequent task outcomes.

Table 1: Data summary

	Obs.	Min	Mean	Median	Max	StD
Hourly wage	61,330	3	13.6	11.1	219	10.6
Skills per user	671,340	1	9.6	9	61	5.9
Skills per opening	61,330	1	2.4	2	21	1.7
Applications	671,340					
Hires	61,330					
Openings	61,330					
Skills ($\mathbb{R}$)	215					
Unique skillsets	17,449					

Contractors come from 182 countries. The dataset spans 4 years.

2.2 Reputation inflation

Response bias, i.e., who chooses to rate a service, drives the overall rating of a service more so than the actual service characteristics [35]. One type of response bias is acquisition bias, which recognizes that buyers typically choose services that they expect to like [18]. This predisposition of buyers introduces a bias in the sample of users who are able to rate a service, which partially explains the observation that overall ratings tend to be positively inflated on most digital platforms [19].

Reputation inflation is present in online labor markets as well [14]. Even state-of-the art bilateral feedback systems [8] suffer from inflation [14]. However, acquisition bias might not be the only driving force of this inflation. One reason for this is "Customer death" [20]: users that receive low feedback scores leave the marketplace since they cannot get hired, and they either create new accounts and rejoin, or join alternative marketplaces to use. An alternative explanation is that raters likely feel pressure to assign good ratings due to inefficiencies of current reputation systems [14]. The proposed dynamic framework in this work tackles these inefficiencies from a design science perspective: it uses information from multiple observed signals in an online labor market to create a deflated, informative reputation.

2.3 Reputation as a signal of expertise

In theory, reputation systems resolve information asymmetries by providing an accurate estimate of the service (or product) quality [12]. In the context of online labor markets, service quality is bound to the expertise of a contractor on a given set of skills. Hence, ideally, reputation scores in online labor markets should signal a contractor's level of expertise. In practice however, due to reputation inflation reputation attribution and reputation staticity, current reputation scores fail to accurate assess a contractor's expertise.

The focal framework of this study relies on accurate skillset-specific expertise assessment. Expertise is latent, and its assessment is a very hard task due to lack of observable signals [26]. Item Response Theory [16] is the classic test-based approach for assessing

expertise on a given set of skills. Other approaches use observational data to identify experts (in online communities) through analyses of networks [21, 40, 46] and static user profiles [6].

In the context of an online labor market there are multiple observed signals that draw a more complete picture of a contractor's expertise. Such signals include the total number of completed jobs and the contractor's premium (hourly compensation). Furthermore, through observing contractors over multiple years, online labor markets become aware of dynamic changes in user profiles (e.g., new skill acquisitions, wage fluctuations, etc.). The proposed framework incorporates these observed signals to estimate the up-to-date expertise of a contractor on any given set of skills.

3 REPUTATION INFLATION IN AN ONLINE LABOR MARKET

The focal data forms a snapshot of 671,340 job applications that led to 61,330 completed tasks by 17,510 contractors from a major online labor market, `GigWork` (pseudonym). `GigWork` supports a diverse set of tasks, including software and web development, writing and translation, sales and marketing, data science, etc. The dataset spans four years, and it considers only contractors who have completed two or more tasks. For these contractors the dataset includes complete information, including any new skill acquisition, and the complete set of their job applications.

Table 1 summarizes the dataset. Overall, the contractor profiles and job openings span across 215 skills. The contractors come from 182 different countries, and they complete tasks for which they earn between $3 and $219 per hour.

On `GigWork`, buyers rate contractors after the completion of a task with a score $Y \in \{0, 1/9, 2/9, \ldots, 9/9\}$. The platform supports a state of the art reputation system: The buyer and the contractor each gets two weeks to leave their feedback score – however this is a double blind process: neither party learns its own rating before leaving a rating for the other party. Figure 1 shows the resulting accumulated feedback distribution on `GigWork`. The mean of this distribution is 0.96 and the median 0.98. Simply put, most of the contractors appear to have an almost perfect reputation. Hence, it is fair to say that `GigWork` suffers from reputation inflation.

At the same time, the heterogeneity in terms of skills and qualifications on the platform in combination with the fact that feedback scores are assigned uniformly (reputation attribution) adds to the noise of the inflated distribution. Figure 2 shows that skills accumulate vastly different feedback scores: from `translation`, with median 1 and mean 0.91, to `twitter-marketing` with median 7/9 and mean 0.74. This variation suggests that there is a need for skill-specific reputation.

4 PROBLEM FORMULATION AND METHODOLOGY

Independent contractors in an online labor market are on the lookout for suitable opportunities. When they find good matches, contractors submit job applications and (some of them) eventually get hired. Once hired, they complete the assigned tasks and receive a pre-arranged compensation along with the respective performance feedback scores. Due to inefficiencies of current reputation systems,

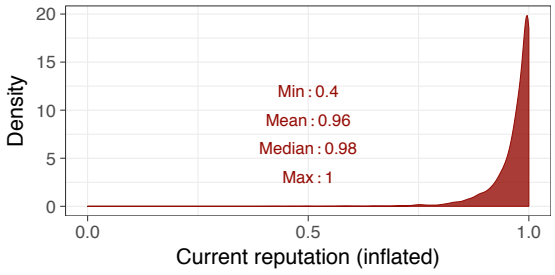

Figure 1: Reputation inflation on `GigWork`. Accumulated contractor feedback scores are skewed to the right. The mean is 0.96 and the median 0.98.

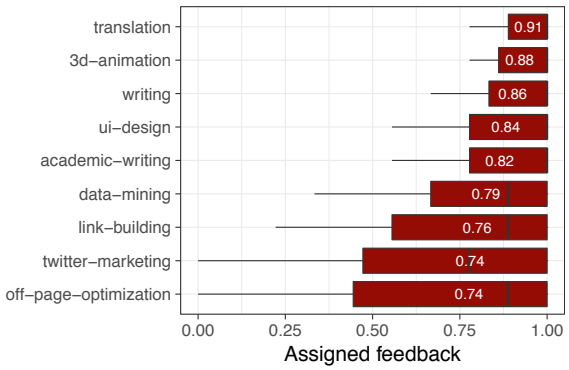

Figure 2: Boxplots of assigned feedback scores for a subset of skills. Each boxplot includes the mean feedback score for the respective skill.

performance feedback scores are inflated (Figure 1), and have high variability across different skills (Figure 2).

Problem definition: *Given a contractor's history on the market, our goal is to estimate a score that captures the current, deflated and true contractor's expertise on any given set of skills.*

The proposed framework (`HMM-W2V-framework`) consists of three components. The first component focuses on decomposing skills into competency dimensions that allows the framework to generalize and accommodate any arbitrary number of skills. The second component presents a dynamic model that combines multiple signals to estimate a contractor's current, competency-specific expertise. The final component aggregates competency-specific predictions to get a current estimate of the contractor's expertise on any arbitrary set of skills. Figure 3 draws the interconnections of these three components, which we describe in detail next.

4.1 Component A: Skills decomposition

Buyers can request contractors with any arbitrary combination of skills. This creates a space of tens of thousands of available combinations of unique skillsets (Table 1). Theoretically, we could directly estimate the expertise of a contractor on any observed

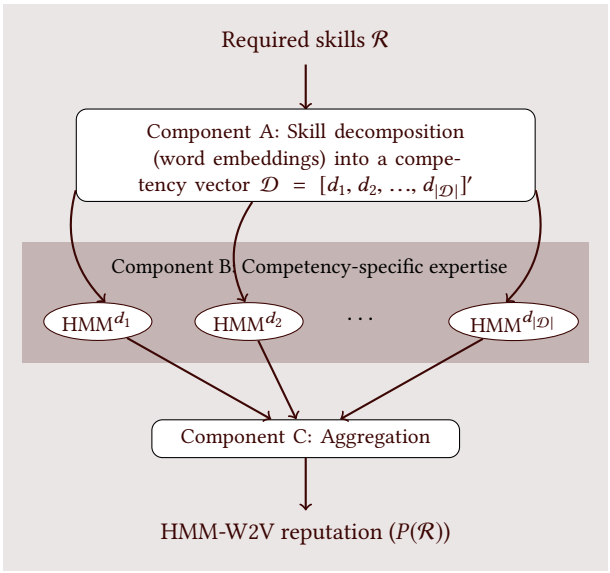

Figure 3: The three components of the `HMM-W2V-framework`. Component A maps any set of skills to competency dimensions through word embeddings. Component B provides dynamic models (HMM) that make competency-specific and up-to-date expertise estimates. Component C aggregates these estimates to provide a skillset-specific expertise assessment.

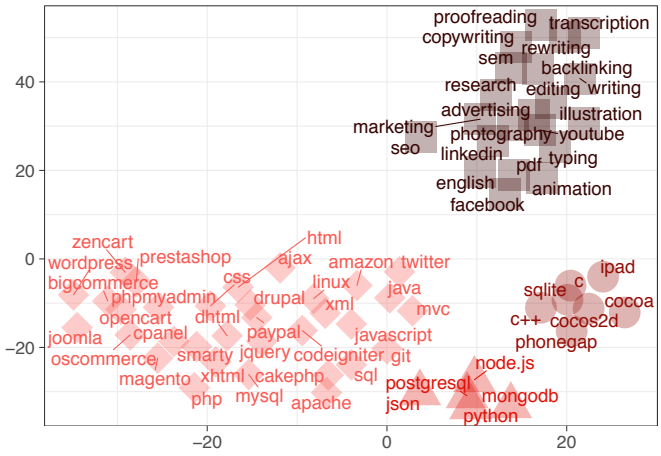

Figure 4: Visualization of mapped skills through W2V in a reduced, two dimensional space through stochastic neighbor embedding [17]. Contextually similar skills appear close to each other (e.g., `c++, c, sqlite`), while contextually dissimilar skills appear far away from each other (e.g., `mongodb` vs. `proofreading`).

skillset. There are three major drawbacks of this approach: first, independent of the size of the data analyzed, considering only skillset-specific observations will result in very sparse training datasets. Second, expertise is transferable [25, 27] and many skillsets are highly correlated. Third, an entrance of new skills would require retraining of the complete framework from scratch.

To overcome these shortcomings we use a distributed representation of words model (word embeddings–W2V) [2, 33] that projects individual skills into a set of competency dimensions. W2V embed words from a vocabulary into a lower dimensional space (i.e., competency dimensions), in which semantically similar words appear close to each other, while semantically dissimilar words appear far away from each other [33]. In the context of an online labor market, a "skill" maps to a "word," and a "skillset" to a "document." Based on this representation, W2V projects contextually similar skills close to each other in the $|\mathcal{D}|$-dimensional space of competencies. (The actual number of competency dimensions $|\mathcal{D}|$ is an input parameter.)

Figure 4 shows how a randomly selected subset of skills maps into a reduced two-dimensional space. The plot reveals hidden contextual similarities between skills; For instance it shows that buyers who request `c++` usually also request `sqlite`, while buyers who request `python` usually also request `mongodb`. The plot hence indicates that `python` is contextually closer to `mongodb` than `sqlite`.

The `HMM-W2V-framework` maps any observed skillset $\mathcal{R}$ into a $|\mathcal{D}|$-dimensional vector space of competencies (Figure 3). To do so, it aggregates competency-specific scores of each skill in a given skillset. Specifically, a skillset $\mathcal{R} = \{R_1, R_2, \ldots, R_{|\mathcal{R}|}\}$, $R_k \in \mathbb{R}$ maps to an aggregated representation as follows:

$$w^d_{\mathcal{R}} = \sum_{R \in \mathcal{R}} \text{W2V}^d(R), \quad \forall d \in \mathcal{D}\,, \tag{1}$$

where $w^d_{\mathcal{R}}$ is the $d-$competency score of skillset $\mathcal{R}$, and $\text{W2V}^d(R)$ is the W2V score for skill R in dimension d. (Alternatively, we could use a distributed memory model (document embeddings, D2V) [31]. In practice, W2V outperformed D2V in this context.)

4.2 Component B: Competency-specific expertise

At any given time (t), each participating contractor has a level of expertise in each available competency dimension $d \in \mathcal{D}$. This expertise is latent (unobserved). However, for each completed task with required skills $\mathcal{R}$ we observe a feedback score (Y_t) that the contractor receives, which maps into competency-specific scores $Y^d_t = w^d_{\mathcal{R}} Y_t, \forall d \in \mathcal{D}$. These scores ($Y^d_t$) form up-to-date proxies of the contractor's expertise in each competency dimension $d \in \mathcal{D}$.

In addition to latent, a contractor's expertise dynamically evolves with time. Contractors gain (or lose) expertise in any competency dimension through learning new skills, improving their knowledge in others, or by completely abandoning skills that the market deems obsolete. The `HMM-W2V-framework` formulates this evolution through a structured Hidden Markov Model (HMM [15, 28]). At any point in time t, contractors operate from a latent, competency-specific state, which determines their propensity to perform with score Y^d_t. Depending on these performance scores on completed tasks, contractors stochastically transition to new latent states. The framework assumes a set of $\mathcal{S}^d$ latent states that describe K^d different levels of expertise for each competency d, $\mathcal{S}^d$ =

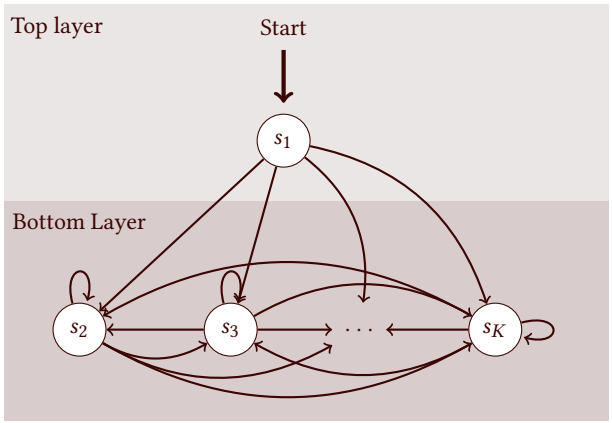

Figure 5: Two-layered structure of each competency-specific HMM. The top layer includes an initial state s_1, which clusters all contractors when they first join the platform. Once contractors start completing tasks, they emit observations and transition to an appropriate state in the bottom layer.

$\{s_1, s_2, \ldots, s_{K^d}\}$. The rest of the paper assumes that the unit of time (t) is a new completed task.

HMM structure: Every new contractor who joins the platform has a certain level of expertise across the competency dimensions $d \in \mathcal{D}$. As the contractor completes new tasks on the platform, we observe signals of the contractor's latent expertise. To capture this behavior, the `HMM-W2V-framework` imposes a two-layer structure: In the first layer, it assume only an initial latent state s_1, where all new contractors land. This state makes an average initial estimate of a contractor's competency-specific expertise. (Alternatively, contractors could land stochastically to one of the second layer states. This would add noise to the estimates, and as a result it could potentially hurt the performance of the framework.) Once the contractors complete their first job and emit an observation (i.e., $Y_1^d, \quad \forall d \in \mathcal{D}$), they stochastically transition to one of the $K^d - 1$ states of the second layer. Figure 5 presents the structure with the two layers and derives the possible transitions from each state.

To define an HMM for a given competency dimension d, we need (1) a vector of initial state probabilities $\boldsymbol{\pi}^d$, (2) a transition matrix T^d that stores the transition probabilities between states, and (3) an emission matrix E^d that describes the state-specific probability distributions for observations Y_t^d. Since every new contractor lands in state s_1, the initial probability vector of the HMM is the following:

$$\boldsymbol{\pi}^d = [1, 0, 0, ..., 0]' \ . \tag{2}$$

A contractor's history provides multiple observable signals that drive transitions to new expertise states (e.g., received feedback scores in previous tasks, total wages received, hiring rates, number of completed tasks). Such historical attributes define a vector $\boldsymbol{Z}_t^d$, which directly affects the transition probability matrix T^d. Formally, we define the transition probability of a given contractor for a given dimension to move from state s_k to state s_l at time t as follows:

$$\lambda^{d,s_k s_l}_{\boldsymbol{\gamma}^d_{kl} \boldsymbol{Z}^d_{t-1}} = \Pr(S_t^d = s_l | S_{t-1}^d = s_k; \boldsymbol{\gamma}_{kl}^d, \boldsymbol{Z}_{t-1}^d) = \text{softmax}(\boldsymbol{\gamma}_{kl}^d \boldsymbol{Z}_{t-1}^d) \ . \tag{3}$$

In the previous Equation, $\boldsymbol{\gamma}_{kl}^d$ is the vector of coefficients of state s_k that define the weights of $\boldsymbol{Z}_{t-1}^d$ in estimating the transition probability to state s_l.

Because of the HMM structure (Figure 5), the transition matrix is not completely filled with elements of Equation 3. Specifically, at time t, the transition matrix has the following form:

$$T^d(\Gamma^d, \boldsymbol{Z}_{t-1}^d) = \begin{bmatrix} 0 & \lambda^{d,s_1 s_2}_{\boldsymbol{\gamma}^d_{12} \boldsymbol{Z}^d_{t-1}} & \cdots & \lambda^{s_1 s_{K^d}}_{\boldsymbol{\gamma}_{1K^d} \boldsymbol{Z}_{t-1}} \\ 0 & \lambda^{d,s_2 s_2}_{\boldsymbol{\gamma}^d_{22} \boldsymbol{Z}^d_{t-1}} & \cdots & \lambda^{d,s_2 s_{K^d}}_{\boldsymbol{\gamma}^d_{2K^d} \boldsymbol{Z}^d_{t-1}} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \lambda^{d,s_{K^d} s_2}_{\boldsymbol{\gamma}^d_{K2} \boldsymbol{Z}^d_{t-1}} & \vdots & \lambda^{s_{K^d} s_{K^d}}_{\boldsymbol{\gamma}^d_{K^d K^d} \boldsymbol{Z}^d_{t-1}} \end{bmatrix} , \tag{4}$$

where $\Gamma^d = [\boldsymbol{\gamma}_{11}^d, \boldsymbol{\gamma}_{12}^d, \ldots, \boldsymbol{\gamma}_{K^d K^d}^d]'$.

Similarly, current, opening-specific observed characteristics (e.g., bidding price or whether or not a contractor has been invited to apply) drive the observed emissions of the HMM (matrix E^d). These characteristics form a vector $\boldsymbol{X}_t^d$ which affects the elements of the emission matrix. Formally, the conditional probability of observing Y_t^d given the current state of the contractor S_t^d is:

$$\Pr(Y_t^d | S_t^d = s_k; \boldsymbol{\theta}_k^d, \boldsymbol{X}_t^d) = f^d(\boldsymbol{\theta}_k^d \boldsymbol{X}_t^d) \ , \tag{5}$$

where $f^d(.)$ is a continuous probability distribution (e.g., Gaussian), and $\boldsymbol{\theta}_k^d$ is the parameter vector of the continuous distribution for state k and competency d. The vector of complete parameter vectors $\boldsymbol{\Theta}^d$ (for all states $s_k \in \mathcal{S}^d$) is as follows:

$$\boldsymbol{\Theta}^d = [\boldsymbol{\theta}_1^d, \boldsymbol{\theta}_2^d, ..., \boldsymbol{\theta}_{K^d}^d]' . \tag{6}$$

HMM identification: Given the structure of the HMM the focus turns on estimating the parameter vectors $\boldsymbol{\Theta}^d, \boldsymbol{\Gamma}^d$. To do so, we maximize the conditional probability of the set of observations given the HMM. (For simplicity, in the following analysis we drop the superscript d. However, keep in mind that this estimation process happens independently for each dimension $d \in \mathcal{D}$.)

Let us assume that we have the following sequence of M observations for a given contractor i:

$$\boldsymbol{Y}_i = Y_{i1}, Y_{i2}, \ldots, Y_{iM} \ . \tag{7}$$

These observations correspond to a sequence of input vectors:

$$\boldsymbol{X}_{1:M} = \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M \ . \tag{8}$$

Furthermore, let us assume that $\boldsymbol{Y}_i$ is the result of a sequence of latent states, $\boldsymbol{S}_i$:

$$\boldsymbol{S}_i = S_{i1}, S_{i2}, \ldots, S_{iM} \ , \tag{9}$$

where $S_{im} \in \mathcal{S}$. This sequence of states is affected by the sequence of historic vectors $\boldsymbol{Z}_{1:M-1}$:

$$\boldsymbol{Z}_{1:M-1} = \boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_{M-1} \ . \tag{10}$$

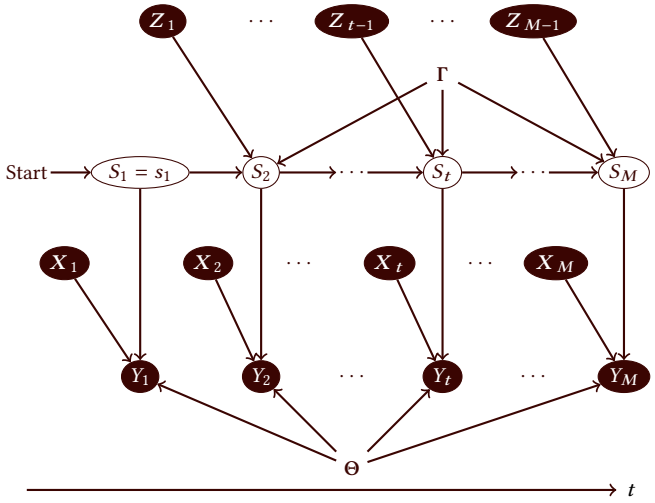

Figure 6: Temporal evolution of the HMM. The structure of the latent state sequence S_i, the observed sequence of outcomes Y_i, the parameter vectors Θ, Γ, and the sequences of input vectors $X_{1:M}, Z_{1:M-1}$ for a given contractor i. For better readability we have dropped the contractor subscript i and the competency superscript d. As with traditional probabilistic graphical models, latent states are in clear ellipses, and observed features in shaded ones.

Figure 6 shows these sequences along with their interactions. Based on the structure of the graph, we get the conditional likelihood of observing sequence $\boldsymbol{Y}_i$:

$$\Pr(\boldsymbol{Y}_i|\boldsymbol{S}_i;\boldsymbol{\Theta},\boldsymbol{X}_{1:M}) = \prod_{t=1}^{M} \Pr(Y_{it}|S_{it};\boldsymbol{\Theta},\boldsymbol{X}_t)\ , \tag{11}$$

where Equation 5 estimates the right hand side. From Figure 6, the conditional probability of observing the sequence $\boldsymbol{S}_i$ is:

$$\Pr(\boldsymbol{S}_i|\boldsymbol{\Gamma},\boldsymbol{Z}_{1:M-1}) = \boldsymbol{\pi}(S_1) \prod_{t=2}^{M} \Pr(S_{it}|S_{it-1};\boldsymbol{\Gamma},\boldsymbol{Z}_{t-1})\ , \tag{12}$$

where $\boldsymbol{\pi}(S_1)$ is the the prior probability of being at state S_1. Equation 3 estimates these transition probabilities. Since the structure of the HMM (Figure 5) imposes that every new contractor lands in state s_1 ($\boldsymbol{\pi}(S_1 = s_1) = 1$), the previous equation becomes:

$$\Pr(\boldsymbol{S}_i|\boldsymbol{\Gamma},\boldsymbol{Z}_{1:M-1}) = \prod_{t=2}^{M} \Pr(S_{it}|S_{it-1};\boldsymbol{\Gamma},\boldsymbol{Z}_{t-1})\ . \tag{13}$$

Table 2: Descriptive statistics for the attributes in vectors X, Z, and the outcome variable Y_{it}.

		Mean	Median	StD	Min	Max
	Observed outcome (Y_{it})	0.86	1	0.22	0	1
Vector Z_t	Current reputation (FB_{it})	0.96	0.98	0.05	0.4	1
	Total money earned ($)	202	44	883	3	56,433
	Completed jobs	3.14	2	2.17	2	39
	Hiring rate	0.20	0.09	0.26	0.002	1
Vector X_t	Invited	0.3	0	0.46	0	1
	Hourly bid ($)	11.93	8.89	12.31	3	400
	Buyer's reputation	0.93	1	0.13	0	1
	Rehires	0.02	0	0.1	0	1

Based on this analysis and the graph in Figure 6, the likelihood of this sequence of observations for contractor i is as follows:

$$\begin{aligned} l(\boldsymbol{Y}_i;\boldsymbol{\Theta},\boldsymbol{\Gamma}) &= \Pr(\boldsymbol{Y}_i|\boldsymbol{\Theta},\boldsymbol{\Gamma},\boldsymbol{X}_{1:M},\boldsymbol{Z}_{1:M-1}) \\ &= \sum_{\forall \boldsymbol{S}_i} \Pr(\boldsymbol{Y}_i,\boldsymbol{S}_i|\boldsymbol{\Theta},\boldsymbol{\Gamma},\boldsymbol{X}_{1:M},\boldsymbol{Z}_{1:M-1}) \\ &\overset{\text{Figure 6}}{=} \sum_{\forall \boldsymbol{S}_i} \Pr(\boldsymbol{Y}_i|\boldsymbol{S}_i;\boldsymbol{\Theta},\boldsymbol{X}_{1:M}) \Pr(\boldsymbol{S}_i|\boldsymbol{\Gamma},\boldsymbol{Z}_{1:M-1}) \\ &= Pr(Y_{i1}|S_{i1};\boldsymbol{\Theta},\boldsymbol{X}_1) \\ &\times \sum_{\forall \boldsymbol{S}_i} \prod_{t=2}^{M} \Pr(Y_{it}|S_{it};\boldsymbol{\Theta},\boldsymbol{X}_t) \\ &\times \Pr(S_{it}|S_{it-1};\boldsymbol{\Gamma},\boldsymbol{Z}_{t-1}), \end{aligned} \tag{14}$$

where we used the structure of the HMM to decompose the joint probability of $\Pr(\boldsymbol{Y}_i,\boldsymbol{S}_i|\boldsymbol{\Theta},\boldsymbol{\Gamma},\boldsymbol{X}_{1:M},\boldsymbol{Z}_{1:M-1})$. Then, the complete likelihood for a dataset with N contractors is:

$$L(\boldsymbol{\Theta},\boldsymbol{\Gamma}) = \prod_{i=1}^{N} l(\boldsymbol{Y_i};\boldsymbol{\Theta},\boldsymbol{\Gamma})\ . \tag{15}$$

Maximization of this likelihood estimates the parameters $\boldsymbol{\Theta}, \boldsymbol{\Gamma}$. We do this numerically through the limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [9]. (In practice we minimize the negative log-likelihood.)

4.3 Component C: Aggregation

This process happens independently for each competency dimension $d \in \mathcal{D}$. As a result, for a given contractor i, each HMM estimates a current (t) competency-specific expertise (p_{it}^d). To predict the expertise of a contractor for a given opening with a set of required skills $\mathcal{R}$ the `HMM-W2V-framework` aggregates the available competency-specific estimates. Let us define a vector with all the competency-specific weights for the opening at hand (Equation 1),

$$\boldsymbol{w}_{\mathcal{R}} = [w_{\mathcal{R}}^1, w_{\mathcal{R}}^2, ..., w_{\mathcal{R}}^{|\mathcal{D}|}]'\ , \tag{16}$$

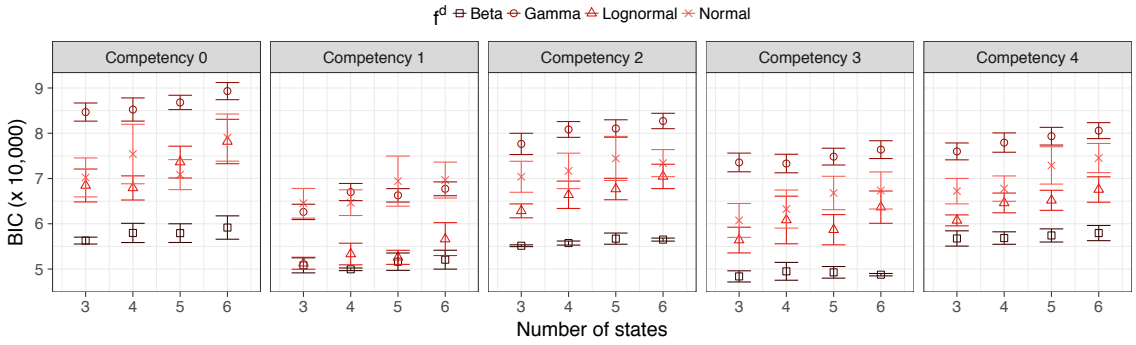

Figure 7: Number of states selection and choice of function f^d, for the five competency dimensions $d \in \mathcal{D}$. Error bars represent 95% confidence intervals.

and a vector with the competency-specific expertise estimates,

$$\boldsymbol{p}_{it} = [p_{it}^1, p_{it}^2, ..., p_{it}^{|\mathcal{D}|}]' \ . \tag{17}$$

The aggregated expertise for skillset $\mathcal{R}$ is then the inner product of the two:

$$P_{it}(\mathcal{R}) = \langle \boldsymbol{w}_{\mathcal{R}} \cdot \boldsymbol{p}_{it} \rangle \ . \tag{18}$$

5 IMPLEMENTATION DETAILS AND RESULTS

Next, we discuss the implementation details and the performance of the `HMM-W2V-framework`.

5.1 Features

Figure 6 shows how vectors $\mathbf{Z}_t$ and $\mathbf{X}_t$ affect the transitions and the emissions of the HMM respectively. What variables compose these two vectors?

Transition probabilities to new states are driven by the framework's current expertise estimate of any given contractor. Even though actual expertise is latent, online labor markets provide observed signals that are correlated with expertise. Such signals form vector $\mathbf{Z}_t$, and include the accumulated feedback score, the total amount of received wages, the hiring rate of the contractor, and the total number of completed jobs. For each competency $d \in \mathcal{D}$, the `HMM-W2V-framework` customizes these signals by weighing them with $w_{\mathcal{R}}^d$. (Note that these signals are estimated at time $t-1$, i.e., at the completion of task $t-1$.)

Similarly, emission probabilities are affected by other, opening-specific characteristics that are not directly related to the latent expertise of the worker (feature vector $\mathbf{X}_t$). Such characteristics are whether or not the contractor has been invited to apply to the specific opening, the bidding price of the contractor, the number of times that a contractor has previously worked with the buyer at hand, and the focal employer's reputation.

Table 2 shows the descriptive statistics of the variables that form vectors $\boldsymbol{X}_t, \boldsymbol{Z}_t$, as well as the outcome variable Y_t.

5.2 Modeling choices and parameter estimation

For the purposes of this study, we pick the number of competency dimensions to be $|\mathcal{D}| = 5$. (In practice, engineers can experiment with various competency dimensions – however, to keep the analysis focused, we only present results for $|\mathcal{D}| = 5$.) We split the dataset into training (66%) and test sets (34%) based on contractors (i.e., the sequence of outcomes of each contractor can only be in either the test or the training set, but not in both). We use the training set to estimate the parameters of the framework and to choose appropriate functions. We then use the test set to evaluate the performance of the framework.

Function f^d (Equation 5) and the number of states (K^d) for each competency are fundamental choices for the implementation of the framework. To make the best possible choices, we estimate the configurations that yield the smallest Bayesian information criterion (BIC) scores in the training set [37]. In particular, the framework considers the following continuous probability distributions for modeling function f^d:

$$f^d \in \{ \text{Gaussian, Gamma, Beta, LogNormal} \} \ \forall d \in \mathcal{D} \ , \tag{19}$$

and the following set of number of states:

$$K^d \in \{2, 3, ..., 6\}, \forall d \in \mathcal{D} \ . \tag{20}$$

For each combination in $\{K^d \times f^d\}$, we estimate the the set of parameters $\boldsymbol{\Theta}^d, \boldsymbol{\Gamma}^d$ that maximize the likelihood of Equation 15. Because the optimization process depends on the initialization of $\boldsymbol{\Theta}^d, \boldsymbol{\Gamma}^d$, it is prone to stuck in local maxima. To increase the likelihood of reaching a potential global maximum, we conduct a search of 1,000 random initializations for each combination in $\{K^d \times f^d\}$. Finally, for each competency dimension $d \in \mathcal{D}$ we pick the configuration that yields the lowest BIC score.

Figure 7 shows the BIC scores for a series of configurations, for each one of the five competencies we consider. In terms of the emission function f^d, Beta seems to perform statistically significantly better than the other three choices, for competencies 0,2,3,4. In competency 1, Beta performs on par with the Lognormal distribution. In terms of number of states selection, none of the available choices is statistically significantly better than the rest (given f^d =Beta). The configurations that yielded the minimum BIC scores are:

- Competency-0: $K^0 = 6$ states, f^0 =Beta emissions.
- Competency-1: $K^1 = 3$ states, f^1 =Beta emissions.
- Competency-2: $K^2 = 4$ states, f^2 =Beta emissions.
- Competency-3: $K^3 = 6$ states, f^3 =Beta emissions.

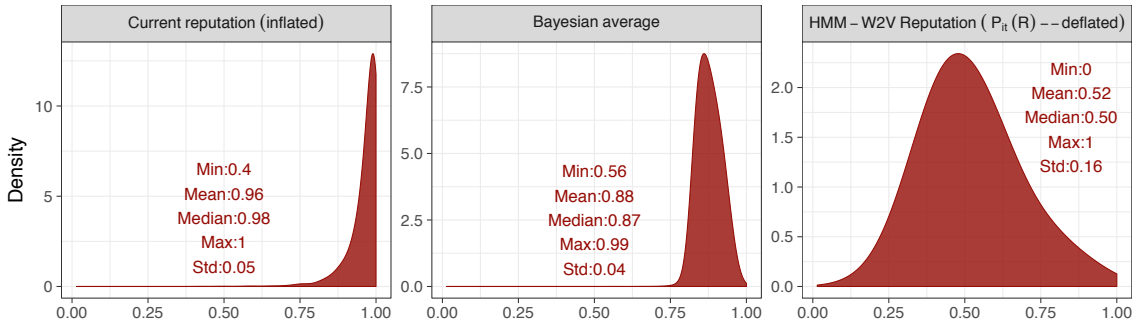

Figure 8: Reputation deflation: Distributions of the inflated current reputation (left), Bayesian average (middle), and HMM-W2V reputation (right).

Table 3: Regression analysis of the `HMM-W2V-framework` and the two baselines.

DV: Observed rating	(A1)	(A2)	(A3)	(A4)	(A5)	(A6)	(A7)	(A8)
Current reputation	-0.004	-0.097***					-0.017*	-0.108***
	(0.01)	(0.02)					(0.01)	(0.02)
Current reputation (squared)		0.121***						0.124***
		(0.03)						(0.03)
Bayesian average			0.047	-0.090			0.070	0.302
			(0.04)	(0.36)			(0.05)	(0.36)
Bayesian average (squared)				0.084				-0.169
				(0.22)				(0.22)
HMM-W2V reputation					0.175***	0.561***	0.177***	0.567***
					(0.02)	(0.11)	(0.02)	(0.11)
HMM-W2V reputation (squared)						-0.313***		-0.315***
						(0.08)		(0.08)
R^2	0.000	0.001	0.000	0.000	0.010	0.010	0.010	0.012

Robust standard errors in parentheses. The constant term is estimated but omitted from the table. ($^{***}p-$value < 0.001, $^{**}p-$value < 0.01, $^{*}p-$value < 0.05).

- Competency-4: $K^4 = 4$ states, f^4 =Beta emissions.

After estimating the parameters and choosing the states k^d and functions f^d, the `HMM-W2V-framework` is ready to estimate the latent expertise of the contractors in the test set.

5.3 Results

Because contractors' levels of expertise are latent, the evaluation of the `HMM-W2V-framework` is not straightforward. Recall that the main goal of this study is to estimate scores that (1) are not inflated (like the current reputation scores), and (2) they capture the up-to-date expertise of contractors on any arbitrary set of skills. Testing whether the `HMM-W2V-framework` deflates reputation is straightforward through estimating and plotting the resulting distribution of $P_{it}(\mathcal{R})$. To evaluate whether the information of HMM-W2V reputation is more accurate than current reputation, we use as ground truth the observed outcomes of each task in the test set.

Baselines: Two baseline algorithms form a benchmark for the HMM-W2V reputation. The first one, is the current accumulated feedback score of `GigWork`. The second baseline is a Bayesian average of current reputation, which is frequently used for ranking items with star ratings [10, 34]. The Bayesian average for contractor i at time τ is as follows:

$$\text{Bayesian average:} BA_{i\tau} = \frac{h_{\text{tasks}} h_{\text{fb}} + \sum_{t=1}^{\tau-1} Y_{it}}{h_{\text{tasks}} + (\tau - 1)}, \qquad (21)$$

where h_{tasks} and h_{fb} are priors for the number of completed tasks per contractor and the contractor outcomes respectively. We set these priors to the mean values of the training set.

Reputation deflation: Figure 8 compares the distributions of the two baselines and the HMM-W2V reputation. Visually, the two baseline distributions are shifted to the right, while the focal HMM-W2V reputation distribution is almost centered at 0.5. While the two baselines take values in [0.4,1] (current reputation) and [0.56.0.99]

Table 4: Predictive performance of the `HMM-W2V-framework` compared to the two baselines.

	Predictive analysis		Ranking correlations		Mutual Information
	MAE	MSE	Kendall τ	Spearman's rank	
Current reputation	0.17 ± 0.007	0.061 ± 0.001	0.07^{***}	0.05^{***}	0.006
Bayesian average	0.18 ± 0.007	0.061 ± 0.001	0.01^{**}	0.01^{**}	0.006
HMM-W2V reputation	0.14 ± 0.003	0.057 ± 0.001	0.12^{***}	0.09^{***}	0.016
Improvement over baselines	16%-17%	7%	71%-1100%	80% - 800%	166%

(95% Confidence intervals for MAE and MSE. $^{***}p-$value < 0.001, $^{**}p-$value < 0.01.)

(Bayesian average), HMM-W2V reputation deflates these scores to [0,1]. The mean of current reputation is 0.96, while Bayesian average brings this mean down to 0.88. However, the mean of HMM-W2V reputation is down to 0.52. Finally, while the two baselines have low variation (Standard deviation around 0.05), the deflated distribution clearly ranks contractors in the available space, with a standard deviation (0.16) 3.2 times higher than the baselines.

HMM-W2V reputation accuracy: The deflated distribution would have had very little value if the deflated HMM-W2V reputation was not a better (than the baselines) representation of the latent contractor expertise. Four metrics test this:

- Regression analysis: Does HMM-W2V reputation better explain the observed outcomes Y_{it} than the two baselines?
- Predictive analysis: Does HMM-W2V reputation better predict the observed outcomes Y_{it} than the two baselines?
- Ranking correlation: Does HMM-W2V reputation rank contractors better than the two baselines?
- Mutual information: Does HMM-W2V reputation and the observed outcomes Y_{it} yield higher mutual information than the two baselines?

For the regression analysis we estimate the following specification:

$$Y_{it} = c_1 + \boldsymbol{c_2}\text{poly}(P_{it}(\mathcal{R}), 2) + \boldsymbol{c_3}\text{poly}(FB_{it}, 2) + \boldsymbol{c_4}\text{poly}(BA_{it}, 2)\,, \quad (22)$$

where poly(x,2) estimates the first and second order of variable x. Table 3 shows the results. These models use the complete test set to estimate the necessary parameters. Columns (A1,A2) show models that consider only current reputation scores (FB_{it}). Columns (A3,A4) show models that consider only Bayesian average scores (BA_{it}). Columns (A5,A6) show models that consider only HMM-W2V reputation scores ($P_{it}(\mathcal{R})$). The comparison shows that HMM-W2V reputation has higher explanatory power than the two baselines, as it yields more than 10 times greater R^2, while coefficients $\boldsymbol{c_2}$ are consistently statistically significant ($p < 0.001$). Models (A7,A8) show how the six variables behave when we estimate the complete specification of Equation 22. Again, the coefficients of HMM-W2V reputation remain large and statistically significant ($p < 0.001$).

Models A2, A4 and A6 in Table 3 test the predictive performance of the three reputation scores. Table 4 shows the 10-fold cross validation Mean Absolute Error (MAE) and Mean Squared Error (MSE) for the three models. The last row shows the percentage improvement of the HMM-W2V reputation over the two baselines. For both metrics HMM-W2V reputation outperforms the two baselines, showing an improvement (decrease in errors) between 7% and 17%.

A correct expertise assessment should rank contractors according to their likelihood of performing good in any given opening with required skills $\mathcal{R}$. Two ranking correlation coefficients (Kendall τ [22], Spearman's rank [42]) test the ordinal associations between expertise estimates and observed outcomes. Table 4 shows the clear superiority of HMM-W2V reputation over the two baselines.

Finally, mutual information measures the dependence between two variables [11]. Observed outcomes should be dependent to an accurate expertise estimate. Table 4 shows that observed outcomes are 166% more dependent to HMM-W2V reputation compared to the two baselines.

Overall, the empirical analysis in this section shows a clear superiority of the HMM-W2V reputation over the two baselines, both in terms of deflating reputation scores and estimating the latent contractor expertise (accuracy). This promising performance of the `HMM-W2V-framework` suggests important implications for online (labor) markets and their users. We discuss these next.

6 IMPLICATIONS

Deflated reputation scores help (1) contractors to differentiate, (2) buyers to make better-informed and faster (reduced search cost [5]) decisions, and (3) the market to build more accurate recommendation algorithms but also to better understand the supply distributions across latent competencies. When contractors can be accurately differentiated, the quality of the supply side of the market naturally increases. High-quality contractors are more likely to keep participating, when low-quality contractors could potentially give space to newcomers. When buyers make better-informed (and faster) decisions that lead to better outcomes become more likely to keep participating in the market. At the same time, markets can improve the performance of their recommendation algorithms by using HMM-W2V reputation as an additional feature. Better recommendations imply higher income and higher transaction efficacy [24, 30]. Finally, through studying the resulting HMM-W2V reputation, markets can get a better picture of the distribution of skills and abilities, and intervene where they think appropriate (e.g., through targeted advertising on skills that there is a deficit).

Besides the direct implications to online labor markets, the `HMM-W2V-framework` can be adjusted and implemented in any online market that has similar rating distributions for products or services. Amazon for instance could deploy the proposed framework to deflate product ratings. In this scenario, the dimensions could be latent product characteristics, and the required skills could be the preferences of the buyer across these latent product characteristics. Similarly, reputation platforms such as Yelp and TripAdvisor could use this framework, as they could decompose venue features to latent dimensions and ask for user preferences across these dimensions.

In conclusion, we presented a framework that deflates reputation in online labor markets through solving reputation attribution and reputation staticity. Analysis of real hiring decisions from an online labor market showed a clean superiority of the proposed framework over current reputation and a Bayesian average baseline.

ACKNOWLEDGMENTS

The author thanks professors Robert Fichman, Panagiotis G. Ipeirotis, Konstantinos Pelechrinis and Sam Ransbotham for their constructive comments and guidance.

REFERENCES

[1] Panagiotis Adamopoulos. 2013. What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. In *International Conference on Information Systems*. AIS.
[2] Panagiotis Adamopoulos, Anindya Ghose, and Vilma Todri. 2018. The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms. *Information Systems Research* 29, 3 (2018), 612–640.
[3] Panagiotis Adamopoulos and Alexander Tuzhilin. 2015. The business value of recommendations: A privacy-preserving econometric analysis. In *International Conference on Information Systems*. AIS.
[4] George A Akerlof. 1970. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics* (1970), 488–500.
[5] J Yannis Bakos. 1997. Reducing buyer search costs: Implications for electronic marketplaces. *Management Science* 43, 12 (1997), 1676–1692.
[6] Krisztian Balog and Maarten De Rijke. 2007. Determining expert profiles (with an application to expert finding). In *Proceedings of the 20th International Joint Conference on Artifical intelligence*. Morgan Kaufmann Publishers Inc., 2657–2662.
[7] Rajiv D Banker and Iny Hwang. 2008. Importance of measures of past performance: Empirical evidence on quality of e-service providers. *Contemporary Accounting Research* 25, 2 (2008), 307–337.
[8] Gary Bolton, Ben Greiner, and Axel Ockenfels. 2013. Engineering trust: reciprocity in the production of reputation information. *Management Science* 59, 2 (2013), 265–285.
[9] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16, 5 (1995), 1190–1208.
[10] Mung Chiang. 2012. *Networked Life: 20 Questions and Answers*. Cambridge University Press.
[11] Thomas M Cover and Joy A Thomas. 2012. *Elements of information theory*. John Wiley & Sons.
[12] Chrysanthos Dellarocas. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science* 49, 10 (2003), 1407–1424.
[13] Chrysanthos Dellarocas, Guodong Gao, and Ritu Narayan. 2010. Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products? *Journal of Management Information Systems* 27, 2 (Oct. 2010), 127–158.
[14] Apostolos Filippas, John J. Horton, and Joseph Golden. 2018. Reputation Inflation. *Working paper* (2018).
[15] Anindya Ghose and Vilma Todri. 2016. Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly* 40, 4 (2016), 889–910.
[16] R.K. Hambleton, H. Swaminathan, and H.J. Rogers. 1991. *Fundamentals of Item Response Theory*. SAGE Publications. http://books.google.com/books?id=cmJU9SHCzecC
[17] Geoffrey E Hinton and Sam T Roweis. 2003. Stochastic neighbor embedding. In *Advances in neural information processing systems*. 857–864.
[18] Nan Hu, Paul A Pavlou, and Jie Zhang. 2017. On self-selection biases in online product reviews. *MIS Quarterly* 41, 2 (2017), 449–471.
[19] N. Hu, J. Zhang, and P.A. Pavlou. 2009. Overcoming the J-shaped distribution of product reviews. *Commun. ACM* 52, 10 (2009), 144–147.
[20] Kinshuk Jerath, Peter S Fader, and Bruce GS Hardie. 2011. New perspectives on customer "death" using a generalization of the Pareto/NBD model. *Marketing Science* 30, 5 (2011), 866–880.
[21] Pawel Jurczyk and Eugene Agichtein. 2007. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the sixteenth ACM Conference on Information and Knowledge Management*. ACM, 919–922.
[22] Maurice G Kendall. 1938. A new measure of rank correlation. *Biometrika* 30, 1/2 (1938), 81–93.
[23] Marios Kokkodis. 2012. Learning from positive and unlabeled Amazon reviews: Towards identifying trustworthy reviewers. In *Proceedings of the 21st International Conference on World Wide Web (WWW)*. ACM, 545–546.
[24] Marios Kokkodis. 2018. Dynamic Recommendations for Sequential Hiring Decisions in Online Labor Markets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 453–461.
[25] Marios Kokkodis and Panagiotis G. Ipeirotis. 2013. Have you done anything like that?: predicting performance using inter-category reputation. In *Proceedings of the sixth ACM International Conference on Web Search and Data Mining (WSDM)*. ACM, 435–444.
[26] Marios Kokkodis and Panagiotis G. Ipeirotis. 2014. The utility of skills in online labor markets. *Proceedings of the AIS International Conference on Information Systems* (2014).
[27] Marios Kokkodis and Panagiotis G. Ipeirotis. 2016. Reputation transferability in online labor markets. *Management Science* 62, 6 (2016), 1687–1706.
[28] Marios Kokkodis and Theodoros Lappas. 2016. Realizing the Activation Potential of Online Communities. In *International Conference on Information Systems*.
[29] Marios Kokkodis and Theodoros Lappas. 2016. The Relationship Between Disclosing Purchase Information and Reputation Systems in Electronic Markets. In *International Conference on Information Systems*.
[30] Marios Kokkodis, Panagiotis Papadimitriou, and Panagiotis G. Ipeirotis. 2015. Hiring behavior models for online labor markets. In *Proceedings of the eighth ACM International Conference on Web Search and Data Mining (WSDM)*. ACM.
[31] Q Le and T Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*. 1188–1196.
[32] Mingfeng Lin, Yong Liu, and Siva Viswanathan. 2018. Effectiveness of Reputation in Contracting for Customized Production: Evidence from Online Labor Markets. *Management Science* 64, 1 (2018), 345–359.
[33] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[34] Evan Miller. 2014. Ranking items with star ratings. http://www.evanmiller.org/ranking-items-with-star-ratings.html. Accessed: 2018-10-24.
[35] Wendy W Moe and David A Schweidel. 2012. Online product opinions: Incidence, evaluation, and evolution. *Marketing Science* 31, 3 (2012), 372–386.
[36] Antonio Moreno and Christian Terwiesch. 2014. Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research* 25, 4 (2014), 865–886.
[37] Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. The MIT Press.
[38] Ernest O'Boyle Jr and Herman Aguinis. 2012. The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology* 65, 1 (2012), 79–119.
[39] Amanda Pallais. 2014. Inefficient hiring in entry-level labor markets. *The American Economic Review* 104, 11 (2014), 3565–3599.
[40] Konstantinos Pelechrinis, Vladimir Zadorozhny, Velin Kounev, Vladimir Oleshchuk, Mohd Anwar, and Yiling Lin. 2015. Automatic evaluation of information provider reliability and expertise. *World Wide Web* 18, 1 (2015), 33–72.
[41] PeoplePerHour. 2018. Hire expert freelancers. https://www.peopleperhour.com/hire-freelancers. Accessed: 2018-10-22.
[42] Charles Spearman. 1904. The proof and measurement of association between two things. *The American journal of psychology* 15, 1 (1904), 72–101.
[43] Statista. 2018. Quarterly share of e-commerce sales of total U.S. retail sales from 1st quarter 2010 to 2nd quarter 2018. https://www.statista.com/statistics/187439/share-of-e-commerce-sales-in-total-us-retail-sales-in-2010/. Accessed: 2018-10-09.
[44] Vilma Todri and Panagiotis Adamopoulos. 2014. Social commerce: An empirical examination of the antecedents and consequences of commerce in social network platforms. In *International Conference on Information Systems*.
[45] Hema Yoganarasimhan. 2013. The value of reputation in an online freelance marketplace. *Marketing Science* 32, 6 (2013), 860–891.
[46] Jun Zhang, Mark S Ackerman, and Lada Adamic. 2007. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th International Conference on World Wide Web*. ACM, 221–230.