IBIMA Publishing
Communications of the IBIMA
https://ibimapublishing.com/articles/CIBIMA/2025/149149/
Vol. 2025 (2025), Article ID 149149, 15 pages, ISSN: 1943-7765
https://doi.org/10.5171/2025.149149
Research Article

Data-Driven Learning Path Generation from Job Postings Using Bayesian Networks with Mutual Information and NO TEARS Structure Learning

Florin STOICA, Dana SIMIAN, Laura Florentina STOICA and Elena-Cristina RAULEA

Lucian Blaga University of Sibiu, Sibiu, Romania

ORCID ID: 0000-0002-9073-0781; ORCID ID: 0000-0002-5210-1810
ORCID ID: 0000-0002-4758-6606; ORCID ID: 0009-0001-5522-7128

Correspondence should be addressed to: Florin STOICA; florin.stoica@ulbsibiu.ro

Received date:2 June 2025; Accepted date:17 November 2025; Published date: 18 December 2025

Academic Editor: Octavian Dospinescu

Abstract

In the context of a rapidly changing labor market, determining the skills demanded by employers and supporting learners in acquiring them effectively presents a considerable challenge. Within the framework of the ENTEEF (Fostering Entrepreneurship through Freelancing) project, this paper proposes an approach based on data-driven methods for identifying competency relationships and generating effective learning paths. The ENTEEF project targets the alignment of educational outcomes with industry expectations by fostering the development of freelancing-relevant competencies among students and other beneficiaries. To address this need, we present a methodology based on Bayesian Networks (BNs) to model the relationships among skills extracted from ~30,000 job postings and to derive data-driven learning sequences from the resulting structures. In our model, every skill is encoded as a node within the BN, and directed edges denote statistically significant relationships, either dependencies or frequent co-occurrences, identified through patterns observed in job market data. We employ two weighting schemes for the BN's arcs, namely the Raw Mutual Information (RMI) score and the Normalized Mutual Information (NMI) score, to measure the strength of the relationships between skills. We also incorporate the NO TEARS framework, a differentiable, continuous optimization method for learning Directed Acyclic Graph (DAG) structures. For each model, pruned networks, skill-centric subgraphs, and Jaccard similarity analyses are used to assess structural coherence and to extract potential learning paths. Results show that while RMI and NMI yield highly similar structures, NO TEARS produces a more globally optimized DAG with stronger, more selective edges. Across methods, the derived learning paths align with realistic upskilling trajectories relevant to freelancing careers. By integrating MI-based weighting with a modern continuous optimization approach, this study offers a scalable, market-aligned framework for generating personalized learning paths grounded in real labor-market data.

Keywords: Bayesian Networks, learning paths, mutual information, NO TEARS.

Cite this Article as: Florin STOICA, Dana SIMIAN, Laura Florentina STOICA and Elena-Cristina RAULEA (2025)," Data-Driven Learning Path Generation from Job Postings Using Bayesian Networks with Mutual Information and NO TEARS Structure Learning", Communications of the IBIMA, Vol. 2025 (2025), Article ID 149149, https://doi.org/10.5171/2025.149149

Introduction

Aligning educational programs with the rapidly evolving demands of the job market is crucial for improving employability and lifelong learning. Learners often struggle to decide what to learn next to reach their career goals, especially in fields like freelancing where required competencies are multifaceted. The ENTEEF project, under the Erasmus+ programme, aims to address this challenge by improving entrepreneurship competences among students and other target groups, preparing them to work as freelancers, and promoting lifelong learning and micro-credentials (ENTEEF, 2025). ENTEEF's approach includes thorough analysis of the freelancer job market to identify key skills, a Competency Assessment Tool (CAT) to determine an individual's skill gaps, and a suite of 12 Massive Open Online Courses (MOOCs) to help fill those gaps. Through a competency gap test and targeted MOOCs, learners receive personalized learning paths that enable them to acquire the skills needed for successful freelancing careers.

However, manually constructing these learning paths from labor market data is labor-intensive. Job postings contain rich information about skills sought by employers, and mining this data can reveal which skills are most in-demand and how they relate to each other. If we can automatically infer a structured map of skills from job offer data, we can generate data-driven learning pathways guiding learners from their current competencies to those required by their target jobs. Recent research underscores the value of aligning learning content with job market needs: for example, the study (Carroll and Schlippe, 2023) achieved over 94% accuracy in identifying job-market skills within course materials using AI and found that showing such alignment to learners improved their motivation and provided valuable career insights. This exemplifies how connecting education to real job requirements can inspire and direct learners.

In this paper, we propose a method to generate recommended learning paths based on skills extracted from job offers. Our approach uses a Bayesian Network to model the relationships among skills, treating each skill as a node and learning a directed acyclic graph from data. The context for this work is the ENTEEF project's goal of guiding learners (prospective freelancers) to acquire competencies demanded by the market. By mining ~30,000 job postings for skills, we create a probabilistic model of how these skills co-occur or depend on one another. We then use information-theoretic measures - specifically the mutual information score and the normalized mutual information score - to weight the connections (arcs) in the network, and we additionally incorporate the NO TEARS algorithm as an alternative, differentiable framework for learning a DAG structure. The resulting weighted Bayesian Networks serve as the foundation for constructing learning paths: a higher weight on an edge suggests a stronger relation that can be interpreted as a recommended transition or prerequisite link between skills.

Therefore, this study is grounded in the hypothesis that "*A Bayesian Network learned from large-scale job postings, when weighted using mutual information measures or using the NO TEARS algorithm for continuous DAG optimization, can effectively model skill relationships and generate learning paths that reflect realistic upskilling trajectories aligned with job market demands*". By extracting and structuring these relationships from collected data, we aim to provide a data-driven foundation for guiding learners in acquiring market-relevant competencies, particularly in freelancing contexts.

This paper represents a substantially extended and enhanced version of our earlier conference contribution presented at the 45th International Business Information Management Association (IBIMA) Conference (Stoica et al., 2025). While the conference paper introduced a Bayesian Network approach weighted using mutual information to derive learning paths from job postings, the present journal version significantly extends that work by incorporating additional structure-learning techniques, deeper comparative analyses, and

expanded experimental validation. In particular, this article introduces the NO TEARS continuous optimization framework for learning directed acyclic graphs, enabling a more global and differentiable modeling of skill dependencies, and provides an extensive comparative evaluation against the original mutual-information-based models.

The paper is organized as follows. In the next section, we review related work. In the third section, we detail our methodology including data processing, Bayesian Network learning, the two arc-weighting schemes, structure learning for BN using NO TEARS, and generating learning paths. The fourth section presents aspects related to the construction of models and their comparative evaluation. Fifth section offers a discussion of the findings, practical implications, and limitations of our study. Finally, are presented some conclusions and opportunities for future investigations in the last section.

Related Work

There is a growing body of research focused on extracting skill requirements from job postings using Natural Language Processing (NLP). Skill extraction is considered a core task in computational job market analysis. Early approaches relied on keyword matching or statistical language models, but recent methods leverage deep learning. For instance, Zhang et al. (2022) introduce the SKILLSPAN dataset for skill extraction and demonstrate state-of-the-art results using BERT-based models on job posting text. Their work is among the first to apply advanced language models to identify both hard and soft skills from job ads, highlighting the feasibility of automatically obtaining structured skill data. Such techniques provide the initial input (a set of skills per job offer) for our problem.

Bayesian Networks have been widely used to model knowledge and learning in educational technology. BNs provide a principled way to represent probabilistic relationships among a set of variables (e.g. skills or concepts) and have been used extensively as student models in intelligent tutoring systems. By encoding dependencies between knowledge components, BNs can infer a learner's mastery level or suggest next learning steps. For example, the study presented in (Culbertson, 2016) reviews numerous assessment systems that employ BNs to diagnose student understanding. In the context of learning path generation, Shen et al. (2020) proposed using a Bayesian network-based association rule algorithm to discover optimal learning paths among microlearning units. Their study created navigation paths for learners by finding correlations among course units, which is conceptually similar to our goal of linking skills. These works validate the idea of using network structures to represent learning sequences.

Beyond Bayesian networks, researchers have explored various techniques for recommending or generating personalized learning paths. A recent systematic review by Rahayu et al. (2023) indicates that many learning path recommender systems rely on ontology or knowledge-based representations. In such systems, relationships between concepts (or skills) are explicitly modeled (e.g., prerequisites in a domain ontology), and algorithms then search these graphs for an optimal path tailored to the learner's profile. Our approach aligns with this knowledge-based trend, but instead of manually crafted ontologies it uses a data-driven BN learned from job data.

Our prior work (Stoica et al., 2025) demonstrated the feasibility of generating learning paths from job postings using Bayesian Networks whose structures were learned via greedy hill-climbing and whose edges were weighted post-hoc using mutual information measures. While effective, such approaches rely on discrete structure search and pairwise dependency measures. In contrast, the present study substantially extends this line of research by integrating the NO TEARS framework (Zheng et al., 2018), a differentiable, continuous optimization method for learning directed acyclic graph structures. Unlike mutual-information-based weighting applied after structure discovery, NO TEARS jointly learns both structure and edge weights under a global acyclicity constraint. This enables a more coherent representation of skill dependencies and allows for a principled comparison

between probabilistic, information-theoretic, and continuous optimization approaches for learning path generation. To the best of our knowledge, this is among the first studies to apply NO TEARS to labor-market-driven skill dependency modeling and learning path construction.

Methodology

This paper extends our previous work (Stoica et al., 2025), in which an initial Bayesian Network - based framework for learning path generation from job postings was introduced. Parts of the baseline methodology and selected illustrative figures are reused for continuity, while the present work substantially expands the modeling, evaluation, and methodological depth.

Our methodology consists of four main steps: (1) Data Collection and Skill Extraction, (2) Bayesian Network Structure Learning, (3) Arc Weighting using Mutual Information and NO TEARS, and (4) Learning Path Generation. Below we describe each step in detail, including key algorithms to illustrate the implementation.

Dataset and Skill Extraction

We gathered a dataset of approximately 30,000 job offers, each accompanied by a list of skills required for that position. These job postings were collected from Upwork, a top-tier platform for freelancers and clients worldwide, using a custom web scraper developed in Python by members of the ENTEEF project team.

Ethical principles were consistently applied during the entire data collection process. The dataset was compiled exclusively from publicly accessible information, with all client-identifiable details removed to ensure privacy. Data selection was conducted without the use of filters, relying on Upwork's default sorting to retrieve the most recent and active job postings. The resulting dataset contains job postings featuring project titles, required skills, budget structures, client information, and geographic details related to the job offers.

The job postings were processed through an NLP pipeline to extract skill keywords, using techniques similar to those described in the literature (Rahayu et al., 2023).

We normalized skill names (e.g., handling synonyms and variations) to ensure consistency. By averaging the frequency of skill mentions in job postings across both continental and global levels, a ranking of 60 in-demand skills was generated. We will focus on these 60 most requested skills in the following. The result is a binary jobs × skills matrix M of size 30,000 × 60. An entry $M[i, j] = 1$ indicates that job posting i mentions skill j, and 0 otherwise. Each job posting thus provides a data point linking a combination of skills that employers expect together.

Learning a Bayesian Network Structure using Hill-Climbing Search

Using the processed dataset, we learn a Bayesian Network structure that captures the probabilistic dependencies between skills. We treat each skill as a binary variable (present/absent in a job posting). Structure learning is performed with a Hill-Climbing Search (HCS) algorithm, a common greedy search method for BN structure discovery (Ankan and Textor, 2024; Dubois et al., 2008). The HCS algorithm starts with an empty network and iteratively adds, removes, or reverses edges to maximize a scoring function (such as the Bayesian Information Criterion or K2 score). In our implementation, we utilized the HillClimbSearch class from the pgmpy library in Python with a BIC scoring metric (Ankan and Textor, 2024).

This search yields a directed acyclic graph (DAG) G where nodes are skills and directed edges suggest a dependency (potentially interpreted as a prerequisite or strong association). For example, the algorithm might learn that an edge goes from skill A (e.g., *"HTML"*) to skill B (*"CSS"*), indicating that HTML knowledge often accompanies or precedes CSS in job requirements - a hint that learning HTML might be a prerequisite to learning CSS.

The learned BN structure encodes which skills tend to appear together in job descriptions and the directionality that best fits the data (note: the direction of an edge in a learned BN does not *strictly* imply pedagogical prerequisite, but we hypothesize it often aligns with a plausible learning order). The structure acts as a skill graph from which we can derive candidate learning paths.

Arc Weighting with Mutual Information

While the BN structure defines the qualitative relationships (which pairs of skills are connected), we next quantify the strength of each connection by computing weights for each arc. We use two different weighting schemes based on mutual information (MI): (a) the raw mutual information score (RMI) and (b) the normalized mutual information score (NMI).

The raw mutual information is an information-theoretic measure of the dependency between two variables. In our context, it measures how much knowing one skill's presence in a job posting reduces uncertainty about the presence of another skill. Formally, for two skill variables X and Y, the raw mutual information $RMI(X;Y)$ is defined as:

$$RMI(X;Y) = \sum_{x\in\{0,1\}} \sum_{y\in\{0,1\}} P(X=x, Y=y)\, log \frac{P(X=x, Y=y)}{P(X=x)P(Y=y)}$$

where $P(X=x)$ is the probability of outcome x.

We treat the job dataset as empirical observations to estimate these probabilities for each pair of connected skills. The higher the RMI, the more information one skill gives about the other, meaning they strongly co-occur (either both present or both absent more often than chance). However, RMI is unbounded on the upper end and tends to grow with the entropy of the variables (for instance, very frequent or very rare skills can influence the magnitude of RMI).

To enable comparison across different skill pairs, we also compute the normalized mutual information (NMI), which scales the mutual information to a standardized range [0,1].

The NMI is given by (Sklearn, 2025):

$$NMI(X;Y) = \frac{RMI(X;Y)}{\sqrt{H(X)H(Y)}},$$

where $H(\bullet)$ denotes entropy and is defined as follows:

$$H(X) = -\sum_{x\in\{0,1\}} P(X=x)\ log\, P(X=x)$$

Considering that the logarithm is in base 2, entropy is measured in bits.

This normalization accounts for the overall occurrence rates of skills X and Y, yielding 0 when skills are independent and 1 when knowing one perfectly predicts the other. Intuitively, NMI tells us the *fraction of maximum possible information* that X and Y share, thus providing a comparable "connection strength" on a zero-to-one scale.

We computed both metrics for every directed edge ($A \rightarrow B$) in the learned BN. Using these metrics, we assign two sets of weights to the network's edges. For example, if skill A and skill B are connected in the BN, we might find $RMI(A;B) = 0.85$ bits and $NMI(A;B) = 0.42$. A higher weight (closer to 1 in NMI, or a larger RMI value) implies a stronger coupling of those skills in job ads.

A pruning threshold was applied (RMI < 0.009; NMI < 0.03) to eliminate weak dependencies from the initial network. The resulting reduced network retains only the stronger connections, for clearer path generation, and the summary statistics of their RMI/NMI scores are presented in Table 1 alongside those of the full network.

Table 1: Summary statistics of RMI, NMI and NO TEARS weights in the Bayesian Networks

Statistic	RMI Full Network	RMI Pruned Network threshold: 0.009	NMI Full Network	NMI Pruned Network threshold: 0.03	NO TEARS Full Network	NO TEARS Pruned Network threshold: 0.16
Count	210	85	210	96	3782	88
Mean	0.016	0.035	0.084	0.169	0.01	0.353
Median	0.006	0.027	0.024	0.136	0	0.317
Std Dev	0.022	0.023	0.112	0.117	0.06	0.151
Min	0	0.009	0	0.036	-0.129	0.166
Max	0.102	0.102	0.492	0.492	0.749	0.749

Structure Learning for Bayesian Networks Using NO TEARS

The NO TEARS (Non-combinatorial Optimization via Trace Exponential and Augmented Lagrangian for Structure Learning) algorithm (Zheng et al., 2018) is a continuous optimization method for learning the structure of Directed Acyclic Graphs (DAGs). Unlike traditional BN structure learning methods, which treat the problem as a combinatorial search over graph space (e.g., Hill-Climbing), NO TEARS reformulates DAG learning into a differentiable optimization problem. This makes it possible to use gradient-based optimization to learn graph structures efficiently and at scale.
NO TEARS optimizes a weighted adjacency matrix $W \in \mathbb{R}^{d \times d}$, where:

- $W_{ij} \neq 0$ means there is an edge $i \rightarrow j$
- If $W_{ij} = 0$, then no edge connects i and j
- The learned matrix must encode a DAG (also known as Bayesian network)

NO TEARS learns how skills jointly predict each other in job postings. Strong positive coefficients W_{ij} indicate "Skill *i* frequently precedes or implies skill *j*". Resulting graph is a data-driven skill dependency DAG.

The algorithm defines a smooth function $h(W)$ such that $h(W) = 0$ if and only if W represents a DAG. The function is:

$$h(W) = \text{trace}(e^{W \circ W}) - d$$

For a weighted adjacency matrix W, if we square each element (the elementwise square or Hadamard product):

$$M = W \circ W$$

M^i encodes paths of length *i* and the matrix exponential of a square matrix M:

$$e^M = \exp(M) = I + M + \frac{1}{2!}M^2 + \frac{1}{3!}M^3 + \cdots$$

has contributions from all possible path lengths.

The trace of the square matrix $e^M \in \mathbb{R}^{d \times d}$ is the sum of the elements on the main diagonal. The trace is important in NO TEARS because it captures a scalar summary of the

diagonal of the matrix exponential, which encodes cycle information.

Thus, acyclicity becomes a differentiable equality constraint:

$$\text{trace}(e^{W \circ W}) = d$$

which holds only when the learned structure represents a directed acyclic graph (only the identity matrix contributes to the trace).

Using the NO TEARS structure-learning approach, we generate a weighted DAG from the job-posting data, then prune the resulting network by removing all arcs with weights below the 0.16 threshold. The descriptive statistics for both the full and the reduced networks are also summarized in Table 1.

Generating Learning Paths

The final step is to utilize the weighted Bayesian network to suggest learning paths. A learning path in this context is a sequence of skills $[S_1 \rightarrow S_2 \rightarrow \ldots \rightarrow S_k]$ that a learner could follow, where each transition $S_i \rightarrow S_{i+1}$ is an edge in the BN indicating a strong relationship. To generate a path for a given target job or role, we proceed as follows:

- **Identify Target Skills**: From the job role of interest (or job postings of that role), determine the set of key skills required. For example, a "Data Scientist" role might require {Python, Machine Learning, Data Visualization, Statistics}.
- **Subgraph Extraction**: Extract the subgraph of the BN that contains these target skills, and any skills directly or indirectly connected to them (this subgraph represents the domain of relevant competencies).
- **Path Search**: Within this subgraph, find paths that connect a learner's current skills to the target skills. We assume the learner's current skills (e.g., those they already possess or have mastered) are known via the competency assessment test (CAT). Starting from a current skill node, we perform a forward search through the network toward a target skill node. We prioritize moves along edges with higher weights (indicating stronger skill association). This can be implemented as a weighted graph search (e.g., Dijkstra's algorithm if interpreting 1-*weight* as a distance (for NMI method), or simply greedily following the highest-weight edge).

Example: If a learner knows HTML but needs to learn React (a JavaScript framework) for a job, the BN might contain a path HTML → CSS → JavaScript → React. Each arrow is supported by strong mutual information scores (indicating these skills frequently co-occur in jobs). The suggested learning path would then be to start with HTML (already known), then learn CSS, then JavaScript, and finally React, in that order. This path aligns with both job data (many web development postings list those skills together) and pedagogical logic (each skill builds on the previous).

The outcome is a recommended sequence of skills (and by extension, relevant MOOCs teaching those skills) personalized to the learner's goals and gaps. In ENTEEF's framework, once such a path is identified, the platform can present the learner with the specific MOOCs corresponding to each skill in the sequence. The Competency Assessment Tool ensures that the learner skips skills they have already mastered, focusing on the ones they lack, while the Bayesian network-derived structure ensures the order of learning is sensible and supported by real-world demand.

By using raw mutual information, normalized mutual information and NO TEARS for arc weighting, we can experiment with different path generation criteria. The raw mutual info score weighting might favor edges that involve generally highly demanded skills (since those contribute more bits of information), whereas normalized mutual info score might highlight niche but strongly linked skill pairs (by controlling for base

frequency). For the skill-dependency modeling presented in this paper, NO TEARS offers an appealing alternative to Hill-Climbing Search, producing sparser and more stable skill graphs and capturing global dependency patterns beyond pairwise mutual information effects.

Model Development and Evaluation

The initial Bayesian Network structure obtained with HillClimbSearch is shown in Figure 1. By filtering out very weak connections, defined as edges with an NMI below a threshold, we obtained the pruned network presented in Figure 2.

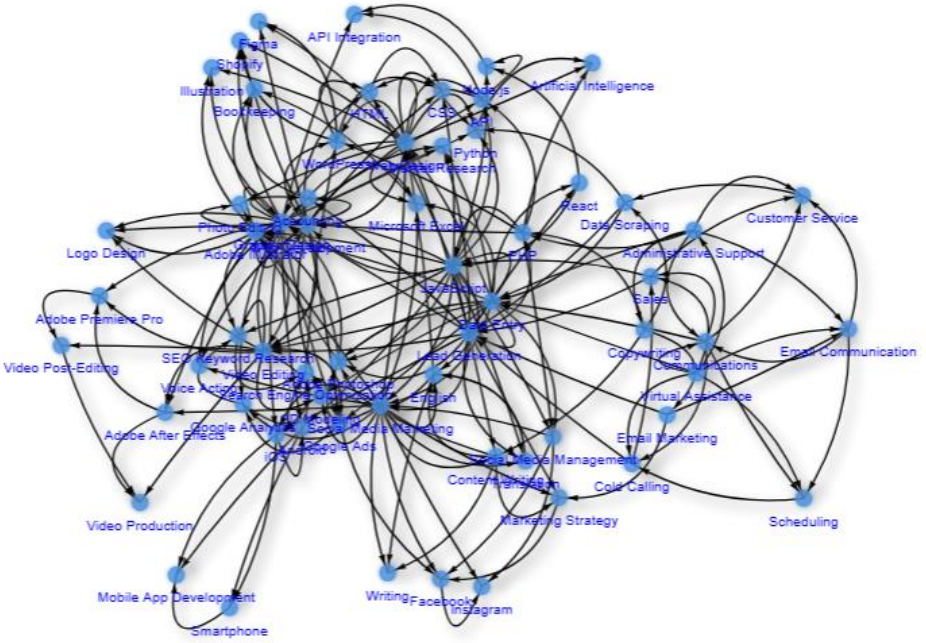


Fig 1. The initial learned Bayesian Network structure (Stoica et al., 2025)

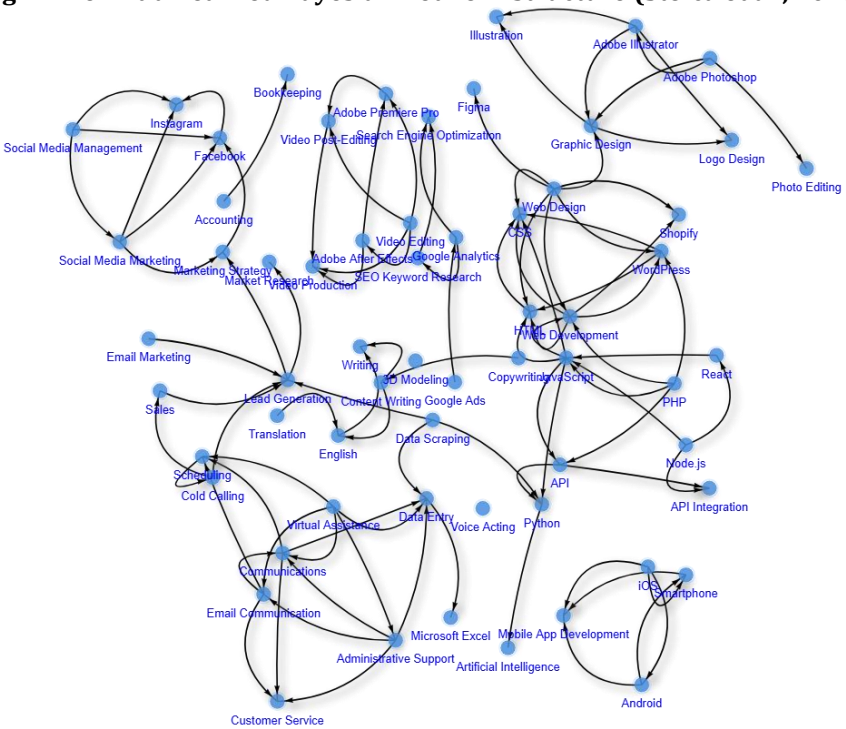


Fig 2. The pruned network, using NMI with threshold 0.03 (Stoica et al., 2025)

Figure 3 presents the pruned network obtained using the NO TEARS framework for BN structure discovery, retaining from the initial network only those arcs whose weights exceed the threshold of 0.16.

To compare the models produced by the three approaches (RMI, NMI and NO TEARS), we extracted the subgraphs for each skill from the respective models, centering each subgraph around the corresponding skill. In an empirical evaluation of the models generated for the most requested skills, we observed that the two Bayesian network weighting methods (RMI and NMI), as well as NO TEARS, produce relatively similar results. Figure 4 shows that, for the Graphic Design skill, the NMI and RMI subgraphs are identical, while the NO TEARS subgraph includes one fewer node.

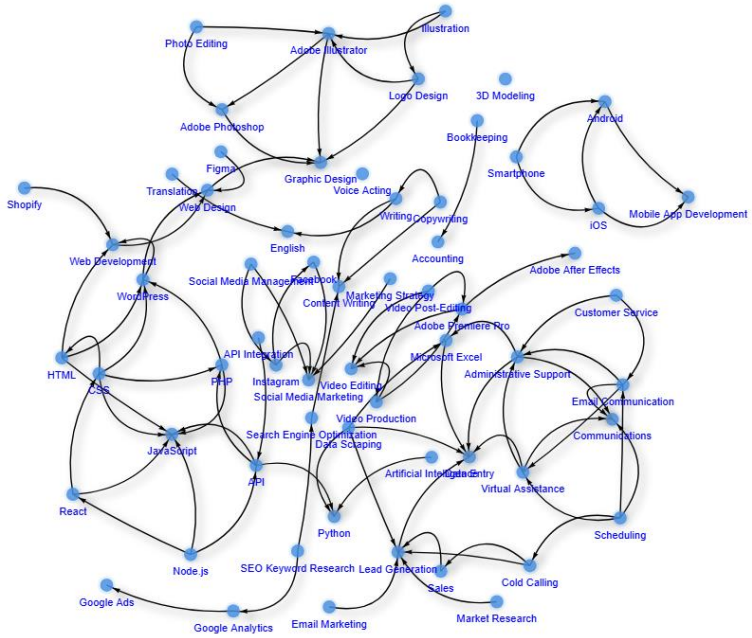
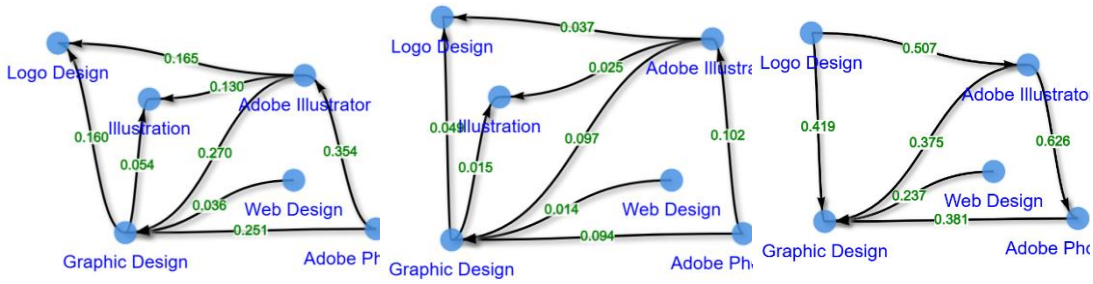


Fig 3. The pruned network, using NO TEARS with threshold 0.16



a) Arc weighting with NMI method (Stoica et al., 2025)

b) Arc weighting with RMI method (Stoica et al., 2025)

c) Arc weighting with NO TEARS method

Fig 4. The extracted subgraphs for Graphic Design skill

However, Figures 5, 6 and 7 illustrate that, for Web Development, the models are slightly different.

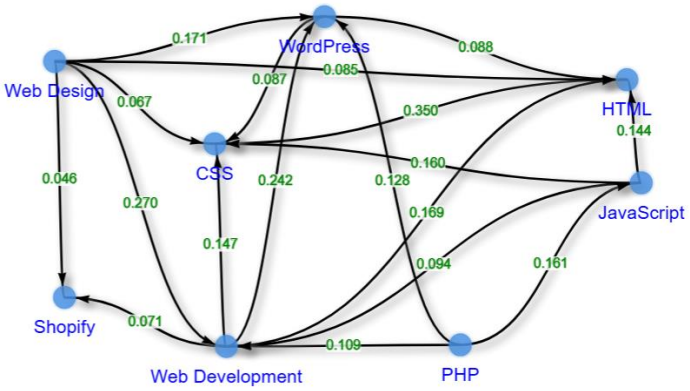


Fig 5. Extracted BN subgraph, based on NMI weights, for the Web Development skill (Stoica et al., 2025)

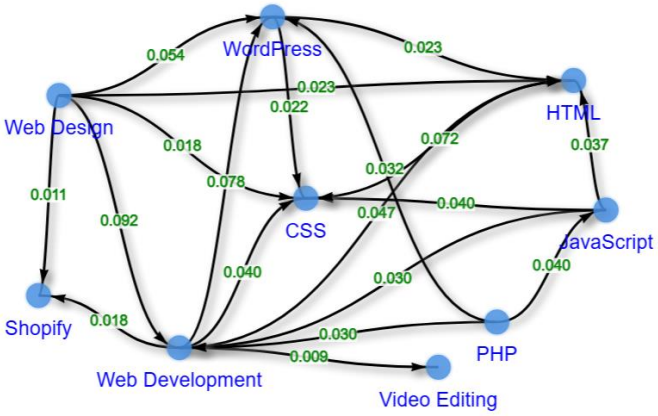


Fig 6. Extracted BN subgraph, based on RMI weights, for the Web Development skill (Stoica et al., 2025)

For an automated comparative evaluation, we measured the similarity of all subgraphs using the Jaccard similarity metric focused on the node sets (node perspective).

Given a graph $G = (V, E)$ and two nodes u and v, the Jaccard similarity of their neighborhoods is defined as follows:

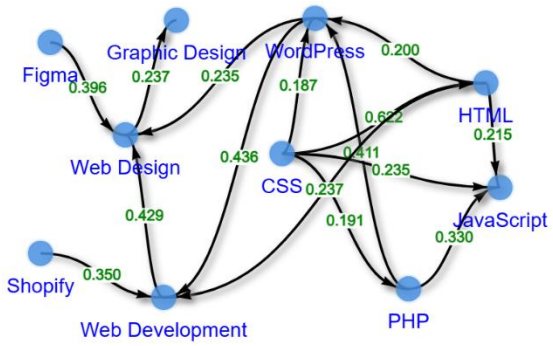


Fig 7. Extracted BN subgraph, based on NO TEARS weights, for the Web Development skill

$$J(u,v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

where $N(u)$ is the set of nodes adjacent to u.

Figures 8, 9, and 10 present only the skills that display meaningful Jaccard similarity values (i.e., different from 1) in the pairwise comparisons between NO TEARS, NMI, and RMI.

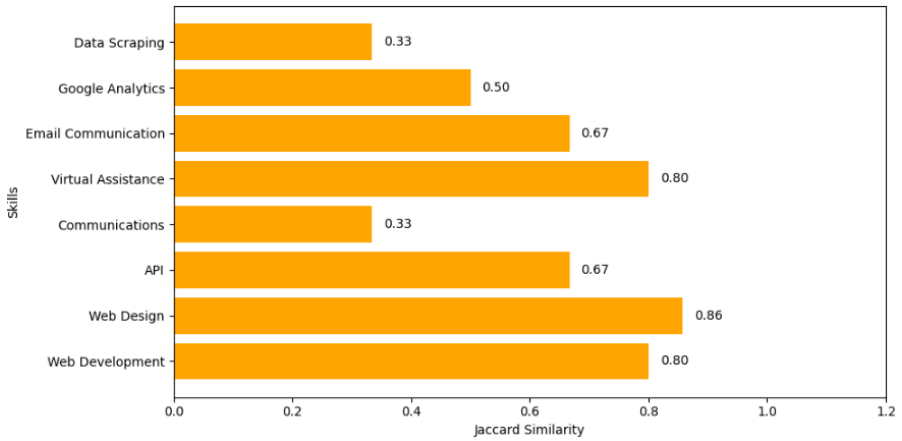


Fig 8. Jaccard similarity differences between RMI and NMI for the skill set (Stoica et al., 2025)

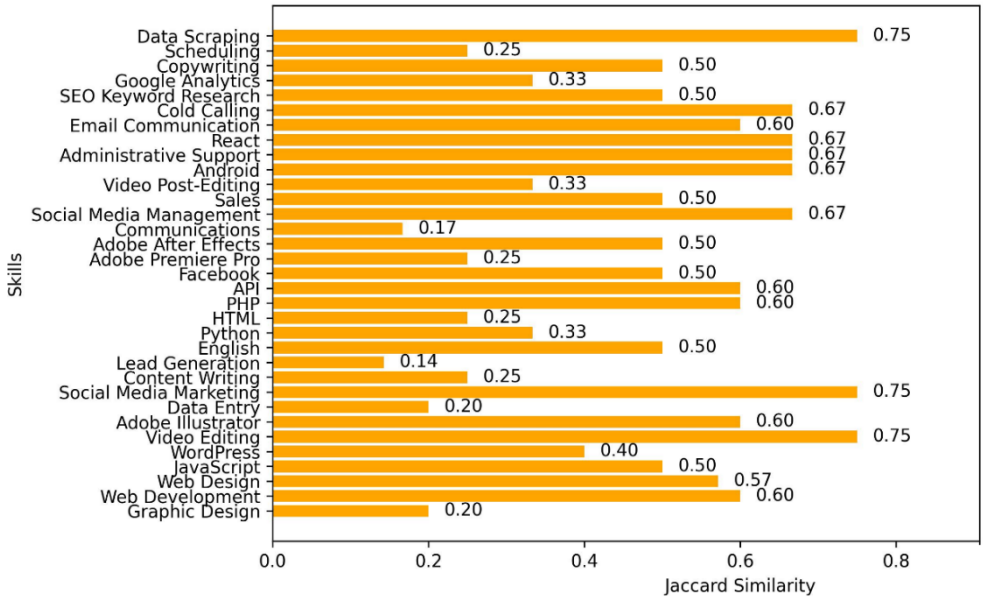


Fig 9. Jaccard similarity differences between NO TEARS and NMI for the skill set

Figure 11 presents the subgraphs extracted from the Bayesian Networks learned from job postings using the three analyzed methods (RMI, NMI, and NO TEARS). Each subgraph is centered on the five most requested competencies (Graphic Design, Web Development, Web Design, JavaScript, and Adobe Photoshop) as target skills and includes only the skills involved in the potential learning path.

Within these subgraphs, learning-path generation entails finding a path from the learner's existing competencies to the desired target competencies.

Discussion

Based on the ENTEEF project's goals and our data-driven modeling approach, we hypothesize that a Bayesian Network learned from large-scale job postings, when weighted using mutual information metrics or NO TEARS, can effectively generate learning paths that reflect real-world skill progression and employer expectations. This hypothesis guides our methodological design and evaluation, aiming to contribute a scalable framework for personalized upskilling grounded in labor market data.

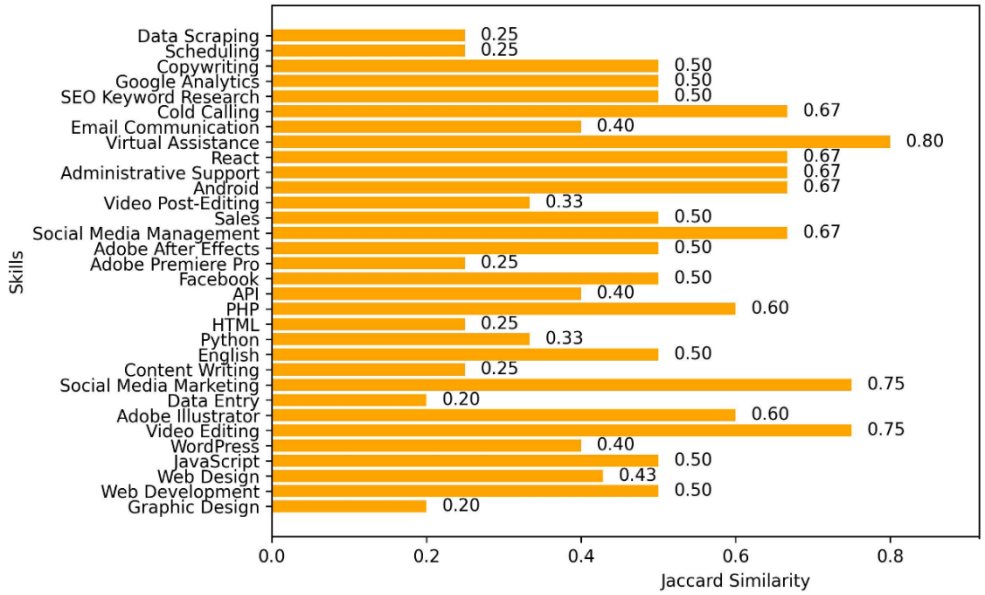
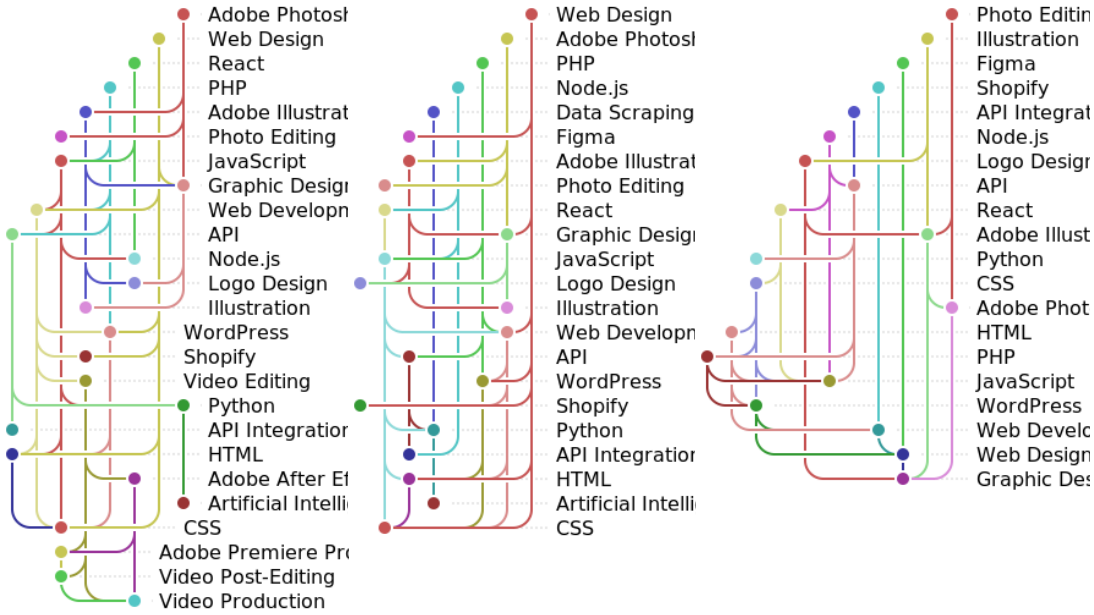


Fig 10. Jaccard similarity differences between NO TEARS and RMI for the skill set

To validate this hypothesis, we provide:

1. *Empirical support from real job data*
   We constructed the Bayesian Network using a large dataset (~30,000 job postings) covering diverse freelancing roles. The structure learned reflects co-occurrence and dependency relationships between skills actually required in the market.
2. *Weighting and comparative evaluation*
   We employed two distinct arc-weighting methods (Raw Mutual Information and Normalized Mutual Information) as well as NO TEARS, a structure learning algorithm that jointly estimates both the graph structure and the edge weights, to quantify the strength of skill relationships. We compared the resulting networks using Jaccard similarity and visual subgraph analysis for multiple key skills.
3. *Practical alignment with learning sequences*
   The learning paths generated align with both the BN structure and logical pedagogical progression, supporting the validity of the approach.
4. *Contribution to knowledge*
   Prior work in the learning path generation often relies on expert-defined ontologies or student learning data. Our contribution is a data-driven, scalable method that infers skill pathways from the job market itself, providing a novel bridge between labor analytics and educational technology.



a) Arc weighting with NMI method

b) Arc weighting with RMI method

c) Arc weighting with NO TEARS method

Fig 11. The extracted subgraphs for the five most requested skills (Graphic Design, Web Development, Web Design, JavaScript, Adobe Photoshop)

The results of the Bayesian Network modeling support our research hypothesis. Furthermore, the ability to extract clear subgraphs and measure node similarity confirms the structural consistency of our model, validating the hypothesis that data-driven BN structures can serve as effective frameworks for learning path generation.

This validation demonstrates the practical and theoretical soundness of our methodology and its contribution to knowledge in both learning path generation and data-driven curriculum design.

Next, we will present some limitations of our study:

- The learning paths are inferred from job data but not yet validated through actual learner outcomes (e.g., whether following a path leads to better job placement or skill mastery).
- The edges in the Bayesian Network are based on statistical dependencies, not necessarily *true pedagogical prerequisites*.
- The system has not yet been tested in real-world learning scenarios, so user experience and motivation effects are still to be studied.
- The pruning thresholds for the RMI, NMI, and NO TEARS models were set empirically. Different thresholds may lead to different path suggestions, and this may affect consistency.

However, our approach has several important advantages and implications, including:

- Our method enables educational institutions to align learning paths with *real-world job market needs*, especially in fast-changing domains like freelancing and tech.
- Learners can follow targeted learning sequences based on actual labor demand, improving the efficiency and relevance of lifelong learning programs.
- Because the Bayesian Network is learned from job postings, the model can be updated regularly to reflect emerging skills and trends.
- While applied here to freelancing, this methodology can be adapted to any domain with sufficient job data, such as

healthcare, cybersecurity, or digital marketing.

In summary, our study demonstrates the feasibility and potential of using Bayesian Networks to derive learning paths directly from labor market data. While the approach offers strong implications for personalized and demand-driven education, its effectiveness depends on the quality of skill extraction and the interpretability of inferred dependencies. These limitations highlight the need for future validation through real learner outcomes and expert review.

Conclusions and Future Work

Our research presents a novel approach to generating learning paths by combining skill extraction from job postings with Bayesian Network modeling. In the ENTEEF project context, this approach automates the linkage between labor market requirements and educational content, paving the way for data-informed upskilling pathways. Early results indicate that mutual-information-based weighting and the NO TEARS structure-learning approach both provide meaningful measures of skill relatedness, and the resulting learning paths align with logical prerequisite structures.

Using NO TEARS to learn relationships between skills extracted from job data offers several benefits. Edges and their strengths (weights) are learned simultaneously, unlike HCS where edges are discrete decisions and weights must be computed afterward. NO TEARS enables scalable gradient-based optimization for datasets with many skills. Also, RMI/NMI edge weights depend on pairwise frequencies; NO TEARS uses global loss minimization, capturing more coherent structures.

Although NO TEARS offers a globally optimized DAG structure, mutual-information based methods (RMI/NMI) also retain several advantages. RMI/NMI provide a direct and transparent measure of skill co-occurrence that is easy to interpret. They capture nonlinear dependencies and rare but meaningful associations that may be smoothed out by the NO TEARS framework. Furthermore, RMI/NMI are computationally simpler, scale more easily to large skill sets, and are often preferred in benchmarking and visualization contexts. For these reasons, RMI/NMI remain valuable complementary approaches to NO TEARS.

Comparing the learning paths derived from NO TEARS with those produced using NMI/RMI weighted Bayesian networks helps to assess the relative strengths of these methods in modeling realistic upskilling paths based on labor market data.

We will compare learning paths obtained under each weighting scheme in future evaluations to see which better guides learners to acquire comprehensive, job-relevant skill sets. Future work will involve validating these generated paths with expert educators and measuring learner outcomes when following the recommended paths, as well as integrating the approach into the ENTEEF platform for real-world use.

Acknowledgements

Co-funded by the European Union (Project no 2024-1-PL01-KA220-HED-000248152). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them. We gratefully acknowledge the contributions of the ENTEEF team members: Bartłomiej Balsamski, Jakub Kanclerz, Mukhammad Andri Setiawan, Ahmad Fathan Hidayatullah and Ratih Dianingtyas Kurnia for their valuable work in the primary data collection and preparation phase of this research. Their efforts in gathering and organizing job posting data provided a critical foundation for the analysis and modeling presented in this study.

References

- Ankan, A. and Textor J. (2024), 'pgmpy: A Python Toolkit for Bayesian Networks', *Journal of Machine Learning Research*, 25(265), 1-8.

- Carroll, D. and Schlippe, T. (2023), 'Connecting Learning Material and the Demand of the Job Market Using Artificial Intelligence', *Artificial Intelligence in Education Technologies: New Development and Innovative Practices*, pp. 282-298.
- Culbertson, M. J. (2016). 'Bayesian Networks in Educational Assessment: The State of the Field', *Applied Psychological Measurement*, 40(1), 3–21.
- Dubois, D., Prade, H. and Smets, P. (2008), 'A definition of subjective possibility', *International Journal of Approximate Reasoning*, 48(2), 352-364. [Online], [Retrieved November 17, 2025], Available: https://doi.org/10.1016/j.ijar.2007.01.005
- ENTEEF Project Fostering Entrepreneurship through Freelancing (2025), Erasmus+ Programme Project Website. [Online], [Retrieved November 17, 2025], Available: https://enteef.uek.krakow.pl
- Rahayu, N.W., Ferdiana, R. and Kusumawardani, S.S. (2023), 'A systematic review of learning path recommender systems', *Education and Information Technologies*, 28 (6), 7437-7460. [Online], [Retrieved November 17, 2025], Available: https://doi.org/10.1007/s10639-022-11460-3
- Shen, H., Liu T. and Zhang Y. (2020). 'Discovery of Learning Path Based on Bayesian Network Association Rule Algorithm', *International Journal of Distance Education Technologies*, 18(1), 117-130.
- Sklearn Metrics Normalized Mutual Info Score - Scikit-learn 1.7 documentation (2025). [Online], [Retrieved November 17, 2025], Available: https://scikit-learn.org/1.7/modules/generated/sklearn.metrics.normalized_mutual_info_score.html
- Stoica, F., Simian, D., Stoica, L.F. and Raulea, E.C. (2025), 'Generating Learning Paths from Job Postings via Bayesian Networks', Proceedings of the 45th International Business Information Management Association (IBIMA) Computer Science Conference, ISBN: 979-8-9867719-7-7, 25-26 June 2025, Cordoba, Spain, 458 – 466.
- Zhang, M., Jensen K. N., Sonniks S. D. and Plank, B. (2022), 'SkillSpan: Hard and Soft Skill Extraction from English Job Postings', Proceedings of NAACL 2022 (Association for Computational Linguistics), 4962-4984.
- Zheng, X., Aragam, B., Ravikumar, P. and Xing, E. P. (2018), 'DAGs with NO TEARS: Continuous Optimization for Structure Learning', *Advances in Neural Information Processing Systems* (NeurIPS 2018), 12 pages.