CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (2) (2026), Pages 2515–2531
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Application of a Heterogeneous Graph for Analyzing Workforce Competency Networks

Nasya Lailatur Romadona* and Muhammad Husnul Khuluq

Department of Mathematics, Faculty of Mathematics and Natural Sciences, State University of Surabaya, Surabaya, Indonesia

Abstract

Recruitment in outsourcing companies requires effective matching between candidate competencies and job requirements. This study constructs and analyzes a heterogeneous candidate–competency–position graph using real recruitment data from an outsourcing company to examine the structural roles of competencies and their community organization. The recruitment database comprised 28,289 records collected from October 2025 to August 2026, with 875 unique candidates across five selected positions: Fitter, Maintenance, Mechanical Technician, Electrical Technician, and Welder. Using stratified, criterion-based purposive sampling with equal allocation, 20 candidates were selected from each position, resulting in 100 candidates. The resulting graph contained 115 vertices and 665 edges, consisting of 634 candidate–competency relationships and 31 competency–position relationships. Normalized betweenness centrality ranked Technical Drawing Reading highest ($C_B = 0.197077$), followed by Welding (0.185315) and Teamwork (0.141790). Louvain community detection with resolution 1.0, `weight=None`, and random seed 38 identified five communities with modularity $Q = 0.079237$. A 30-run stability analysis produced three to six communities, with a mean pairwise ARI of 0.137730, indicating variation across random initializations. The findings are interpreted as exploratory structural patterns rather than statistically representative community structures.

Keywords: Betweenness Centrality; Heterogeneous Graph; Louvain Community Detection; Workforce Competency

Copyright © 2026 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

1. Introduction

The workforce recruitment process involves matching candidates with the requirements of a specific job position [1]. This process is increasingly complex because a position typically requires a combination of complementary skills, while the vast number of potential candidate-position pairings makes finding a suitable match particularly challenging [1, 2]. The digitalization of the economy has expanded the availability of data on job vacancies, curricula vitae, and job seekers across various online platforms. These data contain information on skills and competencies that can complement conventional administrative statistics and facilitate a deeper analysis of the balance between skill supply and demand, including the identification of skill mismatches and shortages of workers with specific expertise [3, 4]. Consequently, an analytical approach

*Corresponding author. E-mail: nasya.23041@mhs.unesa.ac.id

Received: September 09, 2026 Revised: October 16, 2026 Accepted: October 27, 2026
DOI: https://doi.org/10.18860/cauchy.v11i2.46006

is required that goes beyond analyzing skills in isolation to also consider the interrelationships between the various elements of the recruitment process.

A graph-theoretic approach can be used to analyze relationships among skills by representing skills as vertices and their co-occurrence in job postings as edges. Liu et al. analyzed 65 million job postings in the UK using community detection, as well as closeness and betweenness centrality, to map the national skill graph structure [5]. Maghsoudi applied four centrality measures—degree, betweenness, closeness, and eigenvector centrality—to a skill graph constructed from computer science job postings [6]. Meanwhile, Waters and Shutters analyzed skill co-location networks across metropolitan areas using weighted degree, closeness, and betweenness centrality, and applied Louvain community detection to examine their community structures [7]. Beyond skill-level analysis, research in the recruitment domain has begun using heterogeneous graphs, which allow multiple types of entities and relationships to be represented simultaneously. Behar et al. constructed a heterogeneous recruiting graph linking job information with a candidate knowledge graph to support job recommendations [2]. Subsequently, Behar et al. developed TIMBRE (Temporal Integrated Model for Better Recommendations), a heterogeneous recruitment graph comprising eleven vertex types derived from real application data, and used it to recommend candidates with Graph Neural Networks [8]. Frazzetto et al. also constructed a candidate graph based on similarities in curricula vitae, including skills, industry sectors, education, and languages, and used Graph Neural Networks to predict candidates' progression through different stages of the job selection process [9].

To address this research gap, this study employs betweenness centrality and Louvain community detection, which provide distinct yet complementary structural insights. Betweenness centrality is used to identify vertices that hold pivotal positions as connectors or bridges, based on the extent to which they participate in the shortest paths between other pairs of vertices. Research by Zhang and Chen indicates that modularity-based community detection methods, particularly the Louvain method, can be applied to heterogeneous networks even when the number of communities is not known in advance [10]. The choice of community detection algorithm must align with the research question and the nature of the graph, given that each method entails a different approach and set of characteristics [11]. In this study, the Louvain method was selected because it utilizes a modularity optimization approach capable of effectively identifying community structures in large-scale graphs [12]. However, the Louvain method has a known limitation called the resolution limit, which can make it difficult to detect small communities [13].

Applying both methods together allows this study not only to identify competencies that occupy structurally central positions, but also to examine how these competencies are organized within community structures. To address the identified gap, this study models the recruitment process as a heterogeneous graph with three vertex types (candidates, competencies, and job positions) using internal recruitment data from an outsourcing company. Unlike approaches that focus mainly on skill co-occurrence or candidate-job recommendation, the proposed framework jointly examines normalized betweenness centrality, Louvain community structure, the candidate-position match score $F(c, p)$, and the formal breadth of competency requirements. The candidate-position application relationship is deliberately excluded from the graph topology, so that candidate-position suitability is represented indirectly through competency relationships. This combination offers a structural perspective on workforce competencies that complements formal position requirements. The novelty of this study lies in combining structural centrality, community detection, competency-position alignment, and formal competency breadth within the same heterogeneous recruitment graph, using actual internal recruitment records. This integration makes it possible to examine whether competencies that are formally required across many positions also occupy structurally central roles in candidate competency profiles. The analysis provides an empirical observation that formal breadth and network centrality do not necessarily coincide, offering a complementary structural perspective that is not obtained from formal

competency requirements alone.

2. Methods

The recruitment database comprised 28,289 records collected between October 2025 and August 2026. From this database, 875 unique candidates were identified across five selected job positions: 126 Fitter candidates, 357 Maintenance candidates, 65 Mechanical Technician candidates, 39 Electrical Technician candidates, and 288 Welder candidates. Instead of random sampling, this study employed stratified, criterion-based purposive sampling with equal allocation across strata [14].

The five positions were deliberately chosen because their competency requirements overlap, which allows them to form a connected candidate–competency–position network. The remaining records were then grouped into strata according to the position each candidate had applied for. A record was deemed eligible when the candidate (1) had applied for one of the five selected positions, (2) had sufficient information in the *Experience* and/or *Skills* fields to identify at least one competency, and (3) held at least one mandatory competency for that position, meaning a match score of $F(c, p) > 0$. Candidates who held none of the mandatory competencies for the position they applied for ($F(c, p) = 0$) were removed.

The original database included some candidates who had submitted more than one application. The final sample was therefore screened for duplicates, and none were found among candidate IDs, candidate-position combinations, or candidate-competency combinations. In addition, every selected record had complete competency information. In each position stratum, the researcher manually chose 20 candidates who had applied for that position and whose competencies were relevant to its mandatory requirements [14]. The analytical sample therefore consisted of 100 candidates.

The competency information was derived from the free-text Experience and Skills fields recorded for each candidate. Because candidates expressed similar abilities using different words or phrases, the information from these two fields was manually interpreted and grouped into ten standardized competency categories used in the graph analysis: Teamwork, Occupational Health and Safety (OHS), Welding, Electrical Fundamentals, Troubleshooting, Machine Maintenance, Machine Operation, Technical Drawing Reading, Use of Measuring Tools, and AutoCAD. The grouping was based on semantic similarity rather than exact wording. For example, expressions describing the ability to work with a team were mapped to Teamwork, while expressions describing the ability to read technical drawings, use measuring instruments, perform welding, operate machines, conduct machine maintenance, perform troubleshooting, understand electrical fundamentals, or use AutoCAD were mapped to the corresponding standardized competency. A candidate was considered to possess a competency when the corresponding meaning was identified in either the Experience or Skills field. The standardized competency information was then used to construct the candidate–competency relationships in the heterogeneous graph.

Equal allocation was used so that all five positions would be represented comparably in the heterogeneous graph. This prevents any one position from dominating the network structure and keeps the graph manageable for structural analysis and interpretation. Nevertheless, because the allocation was intentionally balanced instead of proportional, and because candidates were chosen manually instead of randomly, the network topology of this sample may not mirror that of the full recruitment population. The results should thus be read as structural characteristics of the balanced sample, not as statistically representative patterns of the whole database. Lastly, although the applied position guided both candidate selection and stratification, the candidate–position relationship was not included as a direct edge in the heterogeneous graph. Candidates and positions are connected only indirectly through competency vertices.

Formal breadth denotes the number of job positions whose mandatory competency requirements include a given competency. A heterogeneous graph $G = (V, E)$ features more than one type of vertex and/or relation; V represents the set of vertices, and E represents the set of edges

Table 1: Mandatory competencies by position

Competency	Fitter	Electrical Technician	Mechanical Technician	Maintenance	Welder	Positions requiring
Machine Operation						0
Machine Maintenance			✓	✓		2
Troubleshooting		✓	✓	✓		3
Teamwork	✓	✓	✓	✓	✓	5
Use of Measuring Tools	✓	✓	✓	✓	✓	5
OHS	✓	✓	✓	✓	✓	5
Electrical Fundamentals		✓	✓	✓		3
Welding	✓		✓	✓	✓	4
AutoCAD			✓			1
Technical Drawing Reading	✓	✓			✓	3
Number of competencies	5	6	8	7	5	

indicating relationships between vertices, including relationships between different vertex types [15, 16]. This structure is suitable for modeling problems involving multiple types of entities and relationships that occur simultaneously, such as in recommendation systems and job matching [8, 17]. The set of vertices V consists of three types of entities: candidates, competencies, and job positions, while the set of edges E comprises the relationships between candidates and competencies, and between competencies and job positions, where the latter are defined by the mandatory competencies listed in Table 1. For instance, a candidate might possess welding skills, whereas a welder requires welding skills; thus, competencies serve as a link connecting candidates to suitable jobs. This method enables relationships between different entities to be represented within a single graph structure. The vertex set is defined in Eq. (1), which specifies the three types of vertices included in the graph: candidates, competencies, and job positions.

$$V = V_c \cup V_k \cup V_p \tag{1}$$

where:

- V_c denotes the set of candidate vertices;
- V_k denotes the set of competency vertices;
- V_p denotes the set of job-position vertices.

In the constructed graph, the vertex sets consist of 100 candidates, 10 competencies, and 5 job positions, respectively. Therefore, the total number of vertices is calculated as follows:

$$|V| = |V_c| + |V_k| + |V_p| = 100 + 10 + 5 = 115.$$

Meanwhile, the set of edges E connects these entities and encompasses two types of relationships: candidate-competency relationships, indicating that a candidate possesses a specific competency, and competency-position relationships, indicating that a job position requires a specific competency [18]. The edge set E contains no direct relationships between candidates and the positions they applied for. Although the recruitment database records the position each candidate officially applied for, this information was used only to select candidates and to define the reference position p in Eq. (2). The application relationship was deliberately not represented as an edge, so that the graph reflects candidate suitability based on competency relationships rather than on the initial application. Candidate-position suitability is thus captured indirectly through candidate-competency and competency-position relationships.

The graph was constructed in Python using the NetworkX library as an undirected, unweighted simple graph, so no weights were assigned to its edges. Under this simple-graph structure, multiple occurrences of the same vertex pair were collapsed into a single edge. This one graph served as the basis for the visualization, the betweenness centrality analysis, and the Louvain community detection. The resulting graph comprised 115 vertices (100 candidates, 10 competencies, and

5 job positions) and 665 edges, of which 634 were candidate-competency edges and 31 were competency-position edges.

A candidate's suitability for a position is calculated based on the degree to which the candidate possesses the competencies required by that position. This formula was developed specifically for this study to demonstrate how well mandatory competencies are met in a manner that is simple, quantifiable, and interpretable, yielding values ranging from 0 to 1. The candidate-position match score is calculated using Eq. (2):

$$F(c,p) = \frac{M(c,p)}{|R_p|} \tag{2}$$

Here, p denotes the actual job position applied for by candidate c as recorded in the recruitment database. It is used as the reference position for calculating the match score and is not inferred from the graph. The terms are defined as follows:

- $F(c,p)$ denotes the match score of candidate c for position p, ranging from 0 to 1. A score of 0 indicates that candidate c possesses none of the mandatory competencies, whereas a score of 1 indicates that candidate c possesses all of them;
- $M(c,p)$ denotes the number of mandatory competencies for position p possessed by candidate c;
- R_p denotes the set of mandatory competencies required for position p, and $|R_p|$ denotes the number of competencies in that set.

Eq. (2) is used to calculate the match score of each candidate for the position they applied for. The resulting scores are used to screen eligible candidates based on the criterion $F(c,p) > 0$, summarize the match-score distribution in Table 2, and compare match-score statistics across positions in Table 3. This competency intersection ratio approach aligns with the suitability ratio method used in skill-extraction-based candidate-job matching systems [19].

Additionally, betweenness centrality is used to identify capabilities that serve as crucial connectors within a graph. This metric indicates how frequently a vertex lies on the shortest path connecting two other vertices. Consequently, competencies with high betweenness centrality values can be viewed as bridging competencies that connect various groups of competencies and distinct positions [20]. The normalized betweenness centrality is calculated using Eq. (3).

$$C_B(v) = \frac{2}{(n-1)(n-2)} \sum_{s<t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

where:

- $C_B(v)$ denotes the normalized betweenness centrality of vertex v;
- σ_{st} denotes the total number of shortest paths connecting vertices s and t;
- $\sigma_{st}(v)$ denotes the number of shortest paths connecting vertices s and t that pass through vertex v, where $v \neq s$ and $v \neq t$;
- n denotes the total number of vertices in the graph.

Betweenness centrality was calculated using the `betweenness_centrality` function from the NetworkX library on the same undirected and unweighted graph, with normalization enabled (`normalized=True`). The resulting values correspond to the normalized betweenness centrality defined in Eq. (3). These values were used to rank all ten competency vertices, produce the results presented in Table 4, and establish the descriptive classification of Bridge and Specific competencies. A competency with a high betweenness centrality value can be viewed as one that potentially serves as a strategic link between various groups of competencies and positions. This approach aligns with the concept of bridge skills in job and competency graphs [21].

For descriptive classification, the ten competency vertices were ranked according to their normalized betweenness centrality values. The five highest-ranked competencies were categorized

as Bridge competencies, whereas the remaining five were categorized as Specific competencies. No absolute betweenness-centrality threshold was used; the classification was based solely on the rank of the ten competency vertices, with ranks 1–5 classified as Bridge and ranks 6–10 as Specific. This classification is relative to the ten competencies included in the analysis and is intended solely for descriptive interpretation rather than as an externally validated competency taxonomy.

Community detection in this study was performed using the Louvain algorithm implemented in NetworkX version 3.4.2. The algorithm was applied to the same undirected and unweighted graph with a resolution parameter of 1.0 and `weight=None`. A fixed random seed of 38 was used for the primary analysis to ensure reproducibility of the reported partition. The Louvain algorithm groups vertices that share relatively strong connections through a modularity optimization process [12]. Its application aims to identify communities of competencies and positions that exhibit similar connectivity patterns. Consequently, the naturally formed structure of job and competency groups within the graph can be identified. The modularity value Q is calculated using Eq. (4).

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{4}$$

where:

- Q denotes the modularity value of the graph; higher values generally indicate stronger community structure;
- A_{ij} denotes the adjacency-matrix element for vertices i and j, equal to 1 if an edge connects them and 0 otherwise;
- k_i and k_j denote the degrees of vertices i and j, respectively, and m denotes the total number of edges;
- c_i and c_j denote the communities assigned to vertices i and j;
- $\delta(c_i, c_j)$ denotes the Kronecker delta, equal to 1 if $c_i = c_j$ and 0 otherwise.

Eq. (4) is used to evaluate the quality of the community partition generated by the Louvain algorithm through the modularity value Q. The modularity value of the primary partition obtained using random seed 38 is reported in Section 3.3, while the modularity results across 30 random seeds are summarized in Table 5. The modularity measure was introduced by Newman and Girvan in 2004 [22]. It was subsequently used as the optimization objective in the Louvain algorithm [12]. This algorithm operates iteratively by moving vertices from one community to another to increase the Q value. Subsequently, the formed communities are merged and represented as new vertices in the next iteration. This process continues until there is no further significant increase in modularity. The application of the Louvain algorithm to job skill graphs has also been employed in previous research to group jobs and competencies into communities that are strongly interconnected internally yet relatively distinct from other communities [21].

2.1. Louvain Stability Analysis

A stability analysis was performed to examine whether the detected community structure depends on random initialization. Louvain was run 30 times on the same graph with a resolution parameter of 1.0, using fixed random seeds 1–29 and 38, where seed 38 produced the primary partition. The number of communities and the modularity value were recorded for every run. Partition similarity was then quantified with the Adjusted Rand Index (ARI), both across all pairs of the 30 partitions and between each partition and the primary partition (seed 38).

2.2. Sampling Robustness Analysis

To further assess the sensitivity of the network results to candidate composition, a sampling-robustness analysis was performed using 30 alternative balanced samples. Each alternative

sample contained 20 candidates from each of the five applied-position strata, resulting in 100 candidates per sample. Candidate pools were formed after applying the same match-score criterion, $F(c, p) > 0$, used in the primary study. The resulting candidate pools consisted of 80 Fitter candidates, 292 Maintenance candidates, 55 Mechanical Technician candidates, 36 Electrical Technician candidates, and 282 Welder candidates.

The 30 alternative samples were generated using fixed sampling seeds from 101 to 130. A candidate could appear in different alternative samples but could not be selected more than once within the same sample. For each alternative sample, the same ten standardized competency categories, semantic mapping rules, graph construction procedure, and 31 mandatory competency-position relationships used in the primary analysis were retained. The graph remained an undirected and unweighted heterogeneous graph without direct candidate-position edges.

For each alternative sample, normalized betweenness centrality was recalculated and compared with that of the primary graph using the overlap of the five highest-ranked competencies and Spearman rank correlation. The rank position of Teamwork was also recorded. Louvain community detection was performed using the same settings as the primary analysis, namely NetworkX version 3.4.2, resolution 1.0, `weight=None`, and random seed 38. Because candidate vertices differed across alternative samples, community similarity with the primary graph was evaluated only on the 15 shared non-candidate vertices, consisting of the ten competency vertices and five job-position vertices, using the Adjusted Rand Index (ARI). The number of detected communities was also recorded for each alternative sample.

3. Results and Discussion

The heterogeneous graph constructed from the company's internal recruitment data comprises 115 vertices: 100 candidate vertices, 10 competency vertices, and 5 job-position vertices. The graph contains 665 edges, consisting of 634 candidate–competency relationships and 31 competency–position relationships. The 100 candidates are evenly distributed across the five applied-position strata, namely Fitter, Maintenance, Mechanical Technician, Electrical Technician, and Welder, with 20 candidates selected within each stratum. As shown in Figure 1, job-position vertices are arranged at the top, competency vertices occupy the middle rows, and candidate vertices are aligned along the bottom. The connecting edges represent candidate–competency relationships and mandatory competency–position relationships.

3.1. Candidate-Position Match Score Analysis

A match score $F(c, p)$ is determined for each candidate based on a comparison between the mandatory competencies they possess and the total mandatory competencies required for the position applied for. Mandatory competency requirements vary by position: the Mechanical Technician role has the highest requirement, with 8 competencies, followed by Maintenance with 7, Electrical Technician with 6, and Fitter and Welder with 5 competencies each. Based on the score distribution in Table 2, 54 out of 100 candidates achieved an F score of 0.8 or higher, while 20 candidates scored below 0.6.

Table 2: Distribution of Match Scores $F(c, p)$

Score Range	Number of Candidates
0.0–0.2	0
0.2–0.4	1
0.4–0.6	19
0.6–0.8	26
0.8–1.0	54

Table 2 shows that 54% of candidates obtained scores of 0.8–1.0, while 20% scored below 0.6. The overall mean match score was 0.754. The variation in match scores across positions is presented in Table 3.

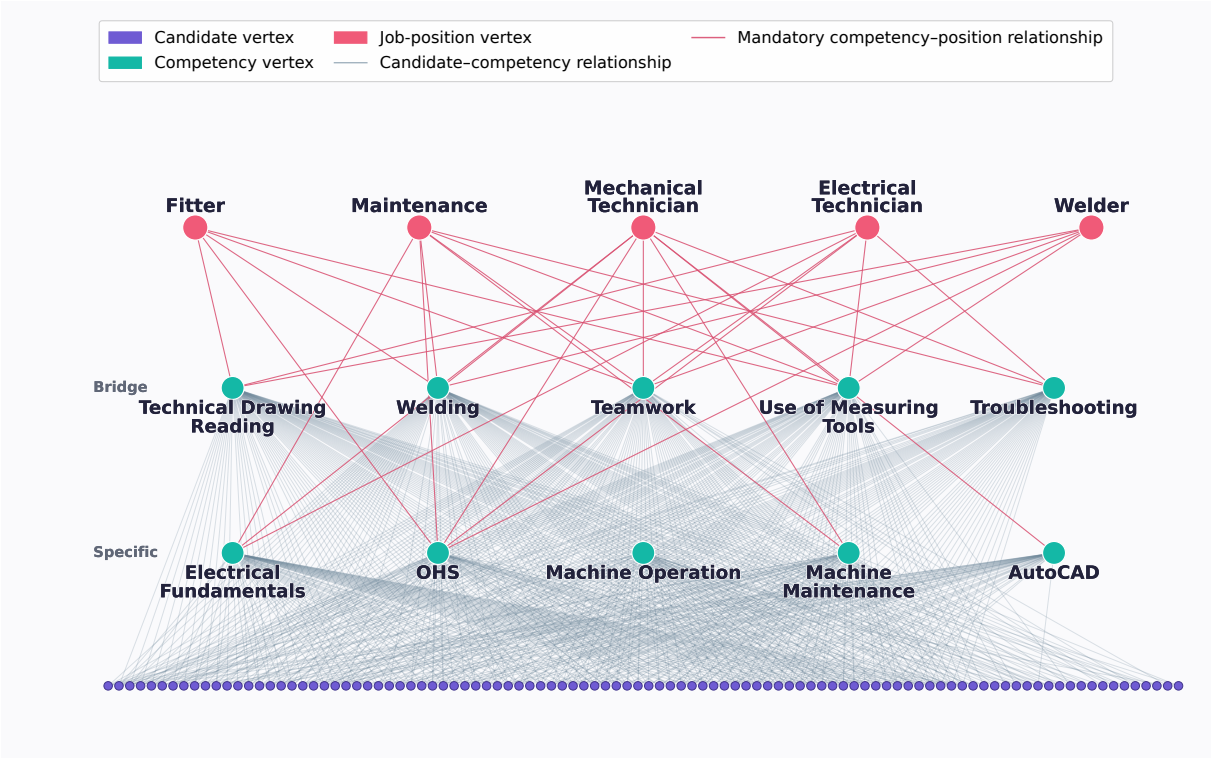


Fig. 1: Heterogeneous Workforce Competency Graph

Table 3: Match Score Statistics $F(c, p)$ by Position

Position	Number of Candidates	Average F	Min	Max
Maintenance	20	0.879	0.571	1.000
Electrical Technician	20	0.875	0.333	1.000
Welder	20	0.840	0.400	1.000
Mechanical Technician	20	0.725	0.500	0.875
Fitter	20	0.450	0.400	0.600

The results show that Maintenance had the highest average match score of 0.879, followed by Electrical Technician at 0.875 and Welder at 0.840. Mechanical Technician had an average score of 0.725, while Fitter had the lowest average score at 0.450. These differences reflect variation in the extent to which the selected candidates possessed the mandatory competencies of their respective positions. The candidate alignment patterns for the five positions are presented in Figures 2–6.

3.2. Betweenness Centrality Analysis

The betweenness centrality calculation results show that Technical Drawing Reading has the highest normalized betweenness centrality value ($C_B = 0.197077$), followed by Welding (0.185315), Teamwork (0.141790), and Use of Measuring Tools (0.095099), as presented in Table 4. Technical Drawing Reading has the highest normalized betweenness centrality, indicating that it occupies an important bridging position between different parts of the heterogeneous network. Welding and Teamwork follow as the second- and third-ranked competencies, respectively. The high centrality of these competencies suggests that they participate in shortest paths connecting diverse candidate competency profiles and position requirements. In contrast, AutoCAD has the lowest betweenness centrality value (0.024269), indicating more limited structural connectivity in the observed network. The complete ranking is shown in Table 4.

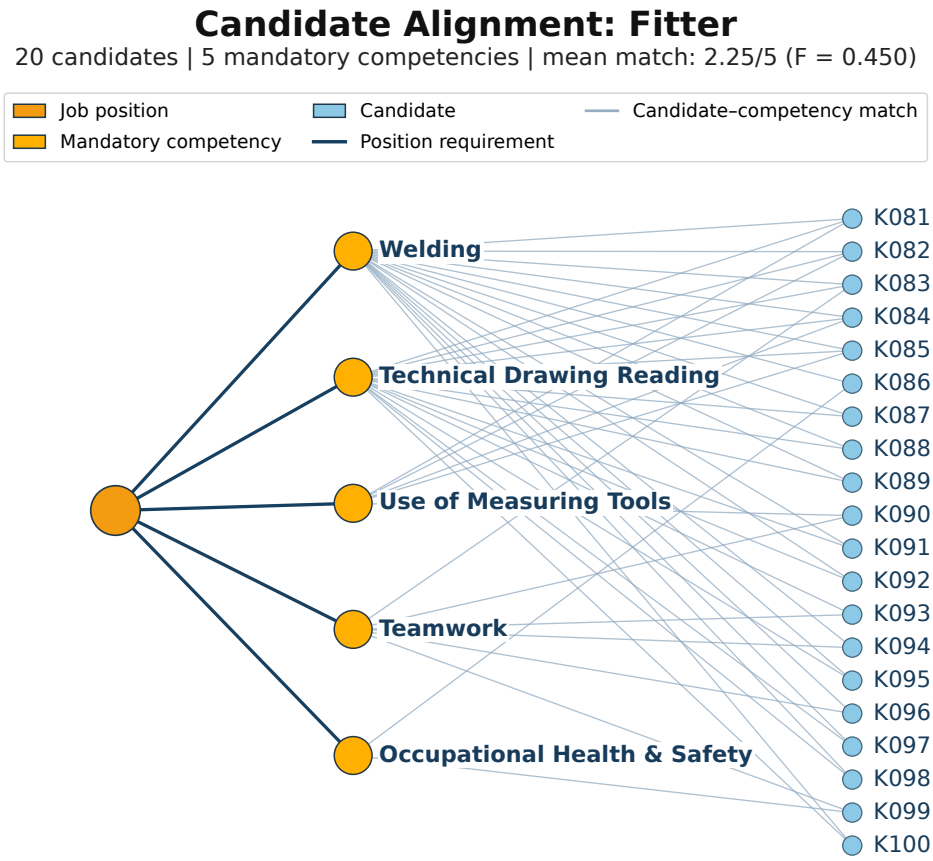


Fig. 2: Candidate alignment with mandatory competencies for Fitter.

Table 4: Competencies with the Highest Betweenness Centrality

Rank	Competency	Betweenness Centrality
1	Technical Drawing Reading	0.197077
2	Welding	0.185315
3	Teamwork	0.141790
4	Use of Measuring Tools	0.095099
5	Troubleshooting	0.077063
6	Electrical Fundamentals	0.069133
7	OHS (Occupational Health and Safety)	0.057647
8	Machine Operation	0.034395
9	Machine Maintenance	0.033441
10	AutoCAD	0.024269

3.3. Louvain Community Detection

Running the Louvain algorithm on the heterogeneous workforce competency graph with a fixed random seed of 38 produced five communities with a modularity of $Q = 0.079237$. The analysis used the same undirected, unweighted graph of 115 vertices and 665 edges. The resulting partition is shown in Figure 7.

A modularity of $Q = 0.079237$ indicates relatively weak separation among the detected communities. The community structure is therefore best interpreted as an exploratory partition rather than as evidence of sharply separated competency groups. This low modularity is consistent with the overlapping connectivity among candidate competency profiles, competencies, and job positions. The communities in the seed-38 partition are described below in terms of the competency and job-position vertices:

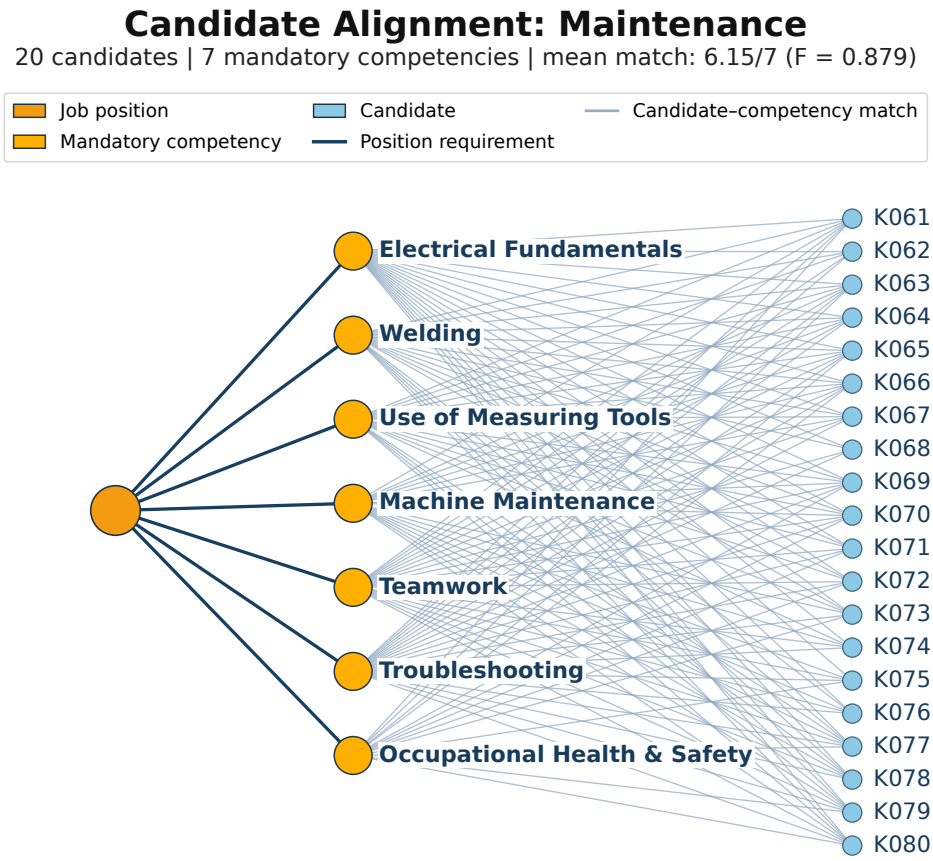


Fig. 3: Candidate alignment with mandatory competencies for Maintenance.

1. Community K0 contains Teamwork and Occupational Health and Safety (OHS), together with the Fitter and Welder position vertices.
2. Community K1 contains Technical Drawing Reading and Welding.
3. Community K2 contains Electrical Fundamentals, Machine Maintenance, Troubleshooting, and Use of Measuring Tools, together with the Electrical Technician, Mechanical Technician, and Maintenance position vertices.
4. Community K3 contains Machine Operation as its only competency vertex.
5. Community K4 contains AutoCAD as its only competency vertex.

Technical Drawing Reading has the highest normalized betweenness centrality and belongs to Community K1, while Welding also belongs to K1. Teamwork belongs to Community K0 together with OHS and the Fitter and Welder position vertices. This illustrates that betweenness centrality and community membership capture different structural properties. Betweenness centrality identifies vertices that frequently lie on shortest paths between other vertices, whereas Louvain groups vertices according to their connectivity patterns. A competency can therefore have high structural centrality without being distributed across multiple communities.

To assess the robustness of the partition, a Louvain stability analysis was run with 30 different random seeds, keeping the graph, resolution parameter, and weighting scheme fixed. Three communities appeared in 1 of the 30 runs, four communities in 16 runs, five communities in 12 runs, and six communities in 1 run. Modularity ranged from 0.071103 to 0.088973, with a mean of 0.078088 and a standard deviation of 0.004690. Seed 38 produced a modularity of 0.079237, but it was not the maximum among the tested initializations. Nevertheless, the variation in the number of communities and modularity demonstrates sensitivity to random initialization.

Partition similarity was also assessed with the Adjusted Rand Index (ARI). Across all pairs

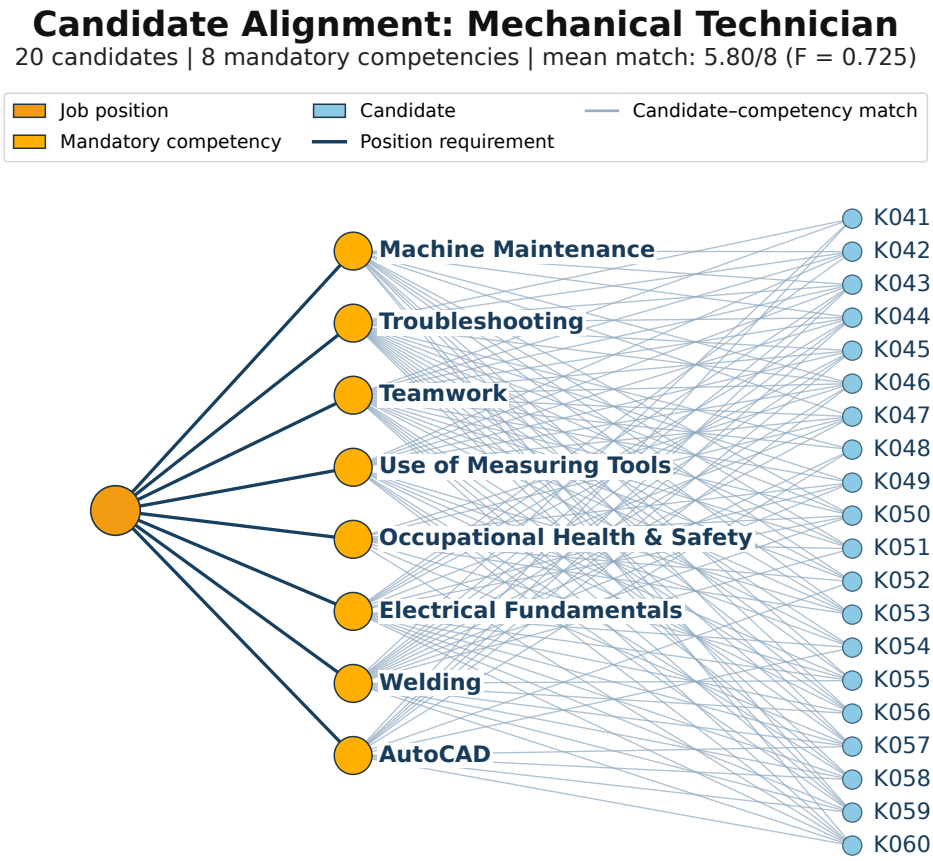


Fig. 4: Candidate alignment with mandatory competencies for Mechanical Technician.

of the 30 partitions, the mean ARI was 0.137730, with values between −0.144487 and 0.654378. For the 29 runs other than the reference seed 38, the ARI against the seed-38 partition ranged from −0.017857 to 0.654378, with a median of 0.232804 and a mean of 0.229715. These values indicate that the exact community partition varies across random initializations. The seed-38 structure is therefore used as a reproducible reference partition for interpretation, while its specific community membership should not be treated as a uniquely stable solution. The 30-run Louvain stability results are summarized in Table 5.

Table 5: Louvain Stability Analysis Across 30 Random Seeds

Measure	Value
Number of runs	30
3 communities	1/30
4 communities	16/30
5 communities	12/30
6 communities	1/30
Mean modularity	0.078088
Standard deviation of modularity	0.004690
Minimum modularity	0.071103
Maximum modularity	0.088973
Mean pairwise ARI	0.137730
Minimum pairwise ARI	-0.144487
Maximum pairwise ARI	0.654378

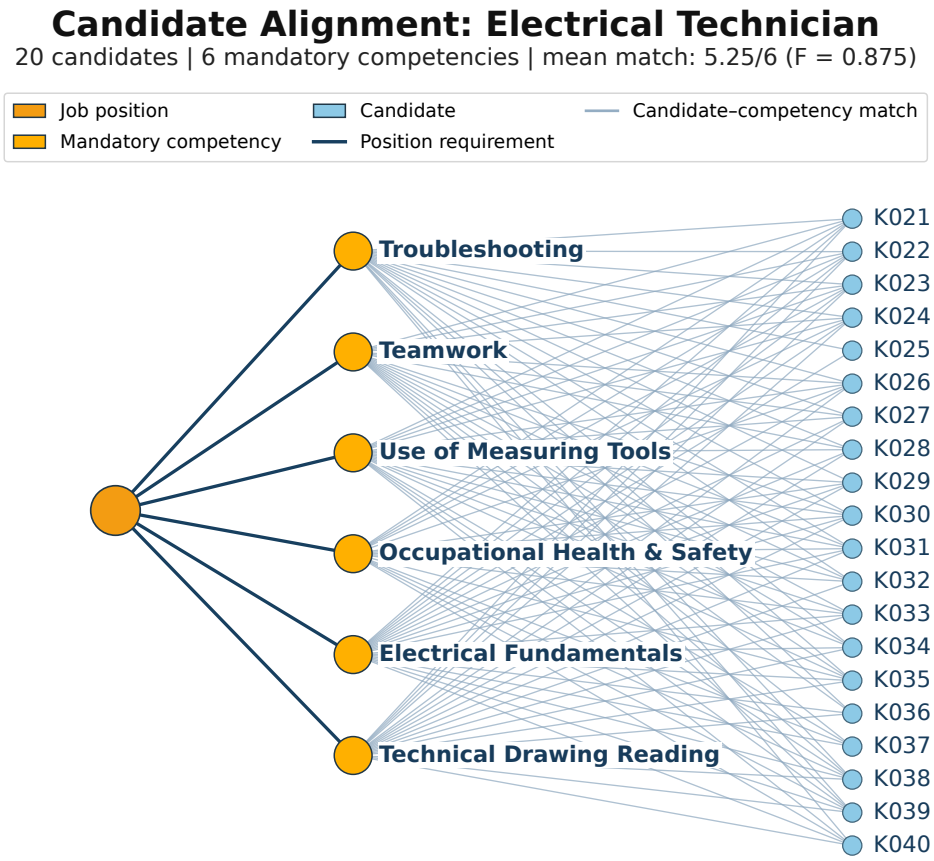


Fig. 5: Candidate alignment with mandatory competencies for Electrical Technician.

3.4. Sampling Robustness Results

The sampling-robustness analysis used 30 alternative balanced samples, each consisting of 100 candidates drawn as 20 candidates from each of the five applied-position strata. The alternative samples were generated from the candidate pools after applying the same match-score criterion $F(c, p) > 0$ used in the primary study. The candidate pools contained 80 Fitter candidates, 292 Maintenance candidates, 55 Mechanical Technician candidates, 36 Electrical Technician candidates, and 282 Welder candidates.

This analysis assessed whether the observed competency centrality rankings and community structure were sensitive to changes in candidate composition, while the graph construction and competency–position requirements were held fixed. Across the 30 alternative samples, the mean overlap between the five highest-ranked competencies and the primary top-five ranking was 2.067. The mean Spearman rank correlation between the normalized betweenness centrality rankings of the alternative samples and that of the primary graph was $\rho = 0.227071$. Teamwork remained among the upper-ranked competencies, with its rank ranging from 4th to 7th across the alternative samples.

The community structure also varied across alternative candidate compositions. The number of Louvain communities ranged from 3 to 7 across the 30 alternative samples, compared with five communities in the primary graph. When evaluated only on the 15 competency and job-position vertices shared across graphs, the mean ARI relative to the primary partition was 0.179801. These results indicate that both the centrality ranking and community organization are sensitive to candidate composition, although some competencies remained among the relatively higher-ranked competencies across alternative samples. The results of the sampling-robustness analysis across the 30 alternative samples are summarized in Table 6.

The relationship between the structural roles of competencies, their community membership,

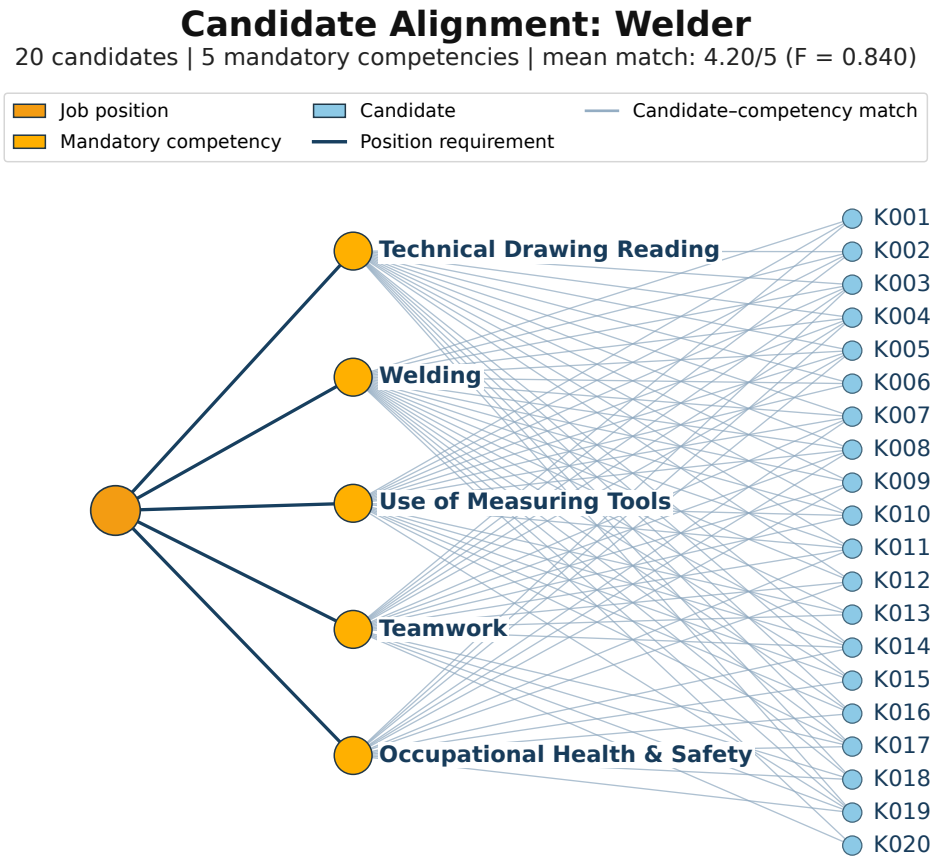


Fig. 6: Candidate alignment with mandatory competencies for Welder.

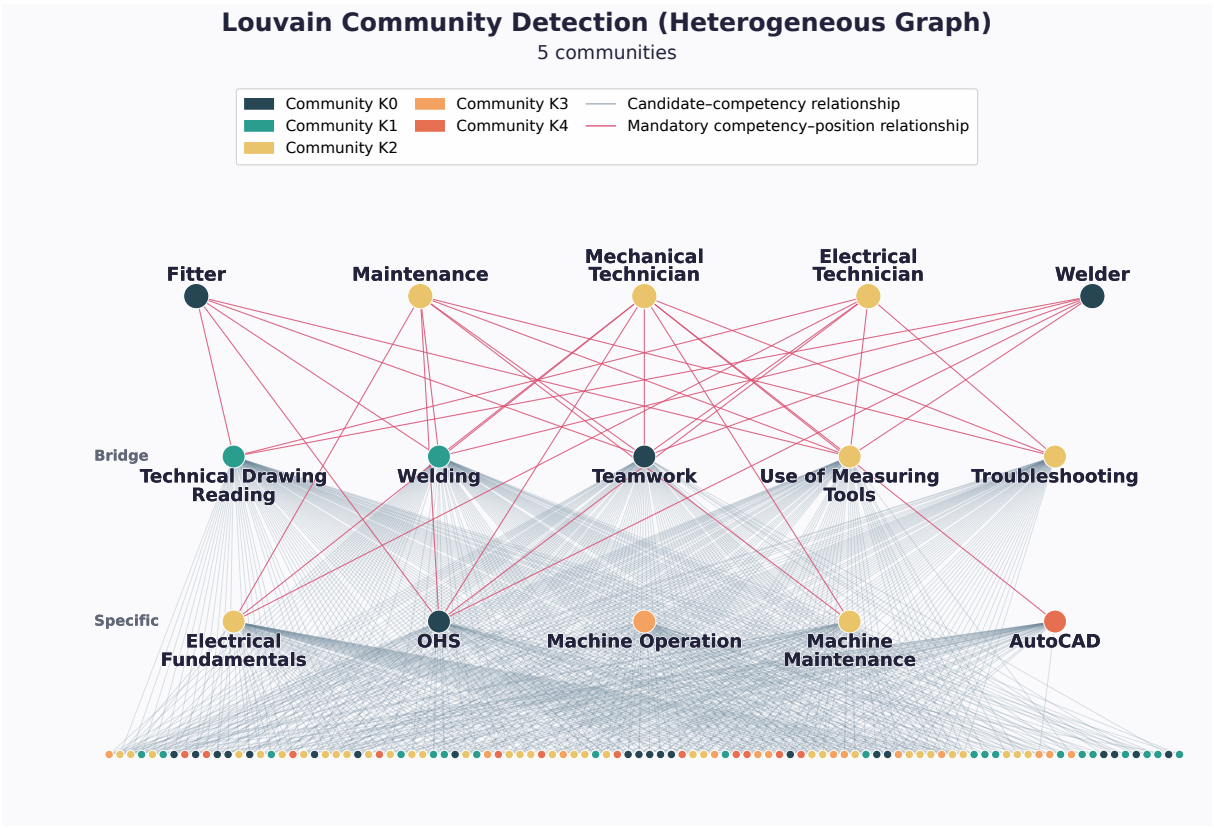


Fig. 7: Louvain Community Detection Results on the Heterogeneous Workforce Competency Graph

Table 6: Sampling Robustness Analysis Across 30 Alternative Samples

Measure	Value
Number of alternative samples	30
Candidates per sample	100
Candidates per position	20
Mean top-5 overlap with primary	2.067
Mean Spearman ρ with primary	0.227071
Teamwork rank range	4–7
Community count range	3–7
Mean ARI vs. primary (15 shared vertices)	0.179801

and their formal breadth was examined by combining the normalized betweenness centrality results in Table 4, the Louvain community assignments, and the formal breadth of each competency. Formal breadth denotes the number of job positions whose mandatory competency requirements include a given competency. The combined results are presented in Table 7.

Table 7: Structural Position and Formal Breadth of Competencies

Competency	BC	Rank	Community	Community Composition	Positions Requiring	Category
Technical Drawing Reading	0.197077	1	K1	Technical Drawing/Welding	3/5	Bridge
Welding	0.185315	2	K1	Technical Drawing/Welding	4/5	Bridge
Teamwork	0.141790	3	K0	Fitter/Welder	5/5	Bridge
Use of Measuring Tools	0.095099	4	K2	Electrical/Maintenance/Mechanical	5/5	Bridge
Troubleshooting	0.077063	5	K2	Electrical/Maintenance/Mechanical	3/5	Bridge
Electrical Fundamentals	0.069133	6	K2	Electrical/Maintenance/Mechanical	3/5	Specific
OHS (Occupational Health and Safety)	0.057647	7	K0	Fitter/Welder	5/5	Specific
Machine Operation	0.034395	8	K3	Machine Operation	0/5	Specific
Machine Maintenance	0.033441	9	K2	Electrical/Maintenance/Mechanical	2/5	Specific
AutoCAD	0.024269	10	K4	AutoCAD	1/5	Specific

The results show that structural centrality and formal breadth do not necessarily coincide. Teamwork and Use of Measuring Tools are formally required for all five positions, with Teamwork ranking third ($C_B = 0.141790$) and Use of Measuring Tools ranking fourth ($C_B = 0.095099$), whereas OHS, despite also having a formal breadth of 5/5, ranks seventh ($C_B = 0.057647$). Welding is formally required for four of the five positions and ranks second, while Technical Drawing Reading is formally required for three positions but has the highest betweenness centrality. Machine Operation is not a mandatory requirement for any of the five selected positions, giving it a formal breadth of 0/5, yet it retains a non-zero betweenness centrality of 0.034395 because it appears in candidate competency profiles. AutoCAD has the lowest centrality and is formally required for only one position. These findings indicate that structural importance is influenced not only by formal competency breadth but also by how competencies are connected through candidate profiles in the network.

In the seed-38 partition, the five Bridge competencies are distributed across Communities K0, K1, and K2, rather than being concentrated in a single community. The presence of separate communities for Machine Operation and AutoCAD further indicates that some competencies form structurally isolated parts of the observed network. Because the stability analysis shows variation across random initializations, however, these community-level patterns should be treated as exploratory rather than as fixed organizational domains.

These findings provide a structural perspective on competency requirements, but their practical interpretation should remain cautious. The results should not be used directly to assign workers to job rotations or to design training programs, particularly because the study relies on a purposively selected, equally allocated sample of 100 candidates from a single recruitment

database, and the Louvain partition is sensitive to random initialization. The analysis is also limited to competency-based relationships: the position each candidate actually applied for was recorded and used as the reference position p in computing $F(c, p)$, but the candidate–position application relationship was intentionally left out of the graph as an edge. Future studies could compare the competency-based alignment score $F(c, p)$ with the positions candidates actually applied for, to examine how formal competency requirements, observed competency structures, and application behavior relate to one another.

4. Conclusion

This study constructed a heterogeneous candidate–competency–position graph from internal recruitment data to examine the structural role of workforce competencies. The final graph comprised 115 vertices and 665 edges, consisting of 100 candidates, 10 competencies, and 5 job positions. Normalized betweenness centrality identified Technical Drawing Reading as the most structurally central competency ($C_B = 0.197077$), followed by Welding (0.185315) and Teamwork (0.141790). The results demonstrate that structurally central competencies do not necessarily correspond directly to the breadth of their formal requirements across job positions.

The Louvain analysis with random seed 38 produced five communities with a modularity of $Q = 0.079237$. The stability analysis and sampling-robustness analysis showed variation in both community structure and competency rankings across different initializations and candidate compositions. Therefore, the detected community structure and centrality patterns should be interpreted as exploratory characteristics of the balanced sample rather than as uniquely stable or statistically representative structures. The findings are limited by the purposively selected, equally allocated sample of 100 candidates from a single recruitment database, the sensitivity of the network results to candidate composition and Louvain initialization, and the exclusion of direct candidate–position application relationships from the graph topology. Future research could extend the analysis using larger and more diverse datasets, alternative sampling designs, weighted competency relationships, and comparisons between competency-based alignment and actual application behavior.

CRediT Authorship Contribution Statement

Nasya Lailatur Romadona: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, & Writing–Original Draft Preparation. **Muhammad Husnul Khuluq:** Writing–Review and Editing, Methodology, & Supervision.

Declaration of Generative AI and AI-assisted technologies

Generative AI and AI-assisted technologies were used in the writing of this manuscript to improve language, grammar, and writing structure, as well as to assist in creating and reviewing the data analysis workflow. All results, interpretations, and final statements have been reviewed and verified by the authors, who bear full responsibility for the content of the manuscript.

Declaration of Competing Interest

The authors declare that there is no conflict of interest in this study.

Ethics Approval and Consent to Participate

This study has obtained permission from the relevant company to use internal recruitment data for research purposes. The data were processed with strict confidentiality, ensuring that no personal information capable of identifying candidates was included. Ethical approval was not

required for this study, as it utilized internal recruitment data for which company authorization had already been secured.

Funding and Acknowledgments

This research received no external funding.

Data and Code Availability

The data used in this study were obtained from an outsourcing company with permission for research purposes. Because the company and its recruitment data are confidential, the dataset cannot be made publicly available.

References

[1] Chen Yang, Yupeng Hou, Yang Song, Tao Zhang, Ji-Rong Wen, and Wayne Xin Zhao. "Modeling two-way selection preference for person-job fit". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 2022, pp. 102–112. DOI: 10.1145/3523227.3546752.

[2] Eric Behar, Julien Romero, Amel Bouzeghoub, and Katarzyna Wegrzyn-Wolska. "Tackling Cold Start for Job Recommendation with Heterogeneous Graphs". In: *Proceedings of the 3rd Workshop on Recommender Systems for Human Resources (RecSys in HR 2023)*. Vol. 3490. CEUR Workshop Proceedings. 2023. https://ceur-ws.org/Vol-3490/RecSysHR2023-paper_9.pdf.

[3] Yuchen Guo, Christina Langer, Fabio Mercorio, and Francesco Trentini. "Skills mismatch, automation, and training: Evidence from 17 European countries using survey data and online job ads". In: *EconPol Forum* 23.5 (2022), pp. 11–15. https://www.ifo.de/en/econpol/publications/2022/article-journal/skills-mismatch-automation-and-training.

[4] Nik Dawson, Marian-Andrei Rizoiu, Benjamin Johnston, and Mary-Anne Williams. "Adaptively selecting occupations to detect skill shortages from online job ads". In: *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. 2019, pp. 1637–1643. DOI: 10.1109/BigData47090.2019.9005967.

[5] Zhaolu Liu, Jonathan M Clarke, Bertha Rohenkohl, and Mauricio Barahona. "Patterns of co-occurrent skills in UK job adverts". In: *PLOS Complex Systems* 2.2, e0000028 (2025). DOI: 10.1371/journal.pcsy.0000028.

[6] Mehrdad Maghsoudi. "Uncovering the skillsets required in computer science jobs using social network analysis". In: *Education and Information Technologies* 29.10 (2024), pp. 12759–12780. DOI: 10.1007/s10639-023-12304-4.

[7] Keith Waters and Shade T. Shutters. "Impacts of Skill Centrality on Regional Economic Productivity and Occupational Income". In: *Complexity* 2022.1, 5820050 (2022). DOI: 10.1155/2022/5820050.

[8] Eric Behar, Julien Romero, Amel Bouzeghoub, and Katarzyna Wegrzyn-Wolska. "TIMBRE: Efficient job recommendation on heterogeneous graphs for professional recruiters". In: *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE. 2024, pp. 8–15. DOI: 10.1109/WI-IAT62293.2024.00010.

[9] Paolo Frazzetto, Muhammad Uzair Ul Haq, Flavia Fabris, and Alessandro Sperduti. "From text to talent: A pipeline for extracting insights from candidate profiles". In: *arXiv preprint arXiv:2503.17438* (2025). DOI: 10.48550/arXiv.2503.17438. https://arxiv.org/abs/2503.17438.

[10] Jingfei Zhang and Yuguo Chen. "Modularity based community detection in heterogeneous networks". In: *Statistica Sinica* 30.2 (2020), pp. 601–629. DOI: 10.5705/ss.202017.0399.

[11] Santo Fortunato. "Community detection in graphs". In: *Physics Reports* 486.3–5 (2010), pp. 75–174. DOI: 10.1016/j.physrep.2009.11.002.

[12] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008 (2008). DOI: 10.1088/1742-5468/2008/10/P10008.

[13] Santo Fortunato and Marc Barthélemy. "Resolution limit in community detection". In: *Proceedings of the National Academy of Sciences* 104.1 (2007), pp. 36–41. DOI: 10.1073/pnas.0605965104.

[14] Steve Campbell, Melanie Greenwood, Sarah Prior, Toniele Shearer, Kerrie Walkem, Sarah Young, Danielle Bywaters, and Kim Walker. "Purposive sampling: Complex or simple? Research case examples". In: *Journal of Research in Nursing* 25.8 (2020), pp. 652–661. DOI: 10.1177/1744987120927206.

[15] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and Philip S. Yu. "Heterogeneous Information Network Embedding for Recommendation". In: *IEEE Transactions on Knowledge and Data Engineering* 31.2 (2019), pp. 357–370. DOI: 10.1109/TKDE.2018.2833443.

[16] Yizhou Sun and Jiawei Han. "Mining heterogeneous information networks: a structural analysis approach". In: *ACM SIGKDD Explorations Newsletter* 14.2 (2012), pp. 20–28. DOI: 10.1145/2481244.2481248.

[17] Fu Li, Guangsheng Ma, Feier Chen, Qiuyun Lyu, Zhen Wang, and Jian Zhang. "Enhanced enterprise-student matching with meta-path based graph neural network". In: *Journal of King Saud University - Computer and Information Sciences* 36.6, 102116 (2024). DOI: 10.1016/j.jksuci.2024.102116.

[18] K. Aark Tejha and K. Srinivas Rao. "Skill-Wise Resume Matching Using NLP and Machine Learning". In: *International Journal of Research Publication and Reviews* 6.8 (Aug. 2025), pp. 4365–4368. DOI: 10.55248/gengpi.6.0825.3126.

[19] Shiyao Wu. "JobHam-place with smart recommend job options and candidate filtering options". In: *arXiv preprint arXiv:2303.17930* (2023). DOI: 10.48550/arXiv.2303.17930. https://arxiv.org/abs/2303.17930.

[20] Linton C. Freeman. "A set of measures of centrality based on betweenness". In: *Sociometry* 40.1 (1977), pp. 35–41. DOI: 10.2307/3033543.

[21] Ahmed Dawoud, Sondos Samir, and Mahmoud Mohamed. "Graph-Based Analysis of AI-Driven Labor Market Transitions: Evidence from 10,000 Egyptian Jobs and Policy Implications". In: *arXiv preprint arXiv:2601.06129* (2026). DOI: 10.48550/arXiv.2601.06129. https://arxiv.org/abs/2601.06129.

[22] M. E. J. Newman and M. Girvan. "Finding and evaluating community structure in networks". In: *Physical Review E* 69.2, 026113 (2004). DOI: 10.1103/PhysRevE.69.026113.