# Clustering analysis of metered Arabic poetry compositions

Abdelmalek Berkani
*Information Management Institute*
*University of Neuchâtel*
Neuchâtel, Switzerland
abdelmalek.berkani@unine.ch

Adrian Holzer
*Information Management Institute*
*University of Neuchâtel*
Neuchâtel, Switzerland
adrian.holzer@unine.ch

***Abstract*—This article focuses on identifying the defining features of a metered Arabic poetic structure. In addition to its research and educational advantages, emphasizing the comprehensibility of the structure, even without language proficiency, enables individuals to grasp its core essence. This understanding opens up opportunities for exploring comparative possibilities and establishing connections with poetic structures in different languages.**

**To achieve this goal, this article presents a novel clustering approach based on the six following poem metrics (*Diversity*, *Evenness*, *Variability*, *Repetition*, *Pattern Potential* and *Pattern Usage*) and utilizing two distinct algorithms. We discovered that the most significant poem metrics are *Repetition*, *Variability* and *Pattern Usage* and two attributes are sufficient to characterize the poetry instead of six.**

***Keywords*—Arabic poetry, Poetry Clustering, Poetry variability, Poetry complexity, Order and Complexity, Arabic poetry meters**

## I. INTRODUCTION

Poetry analysis has long been a subject of research and inquiry. Previous studies have primarily focused on exploring the style, semantic, and figurative language aspects of poetry [1]. Undoubtedly, poets express emotions, imagination, style, and sentiments through their words. These elements directly contribute to the textual content of the poem, but the structural aspect also plays a significant role in poetry aesthetics [2]. Similar to music, the lyrical essence of a song cannot be fully appreciated without considering the musical composition that accompanies it. However, it is worth noting that music can exist without lyrics, whereas poetry invariably requires a text.

This realization prompted us to recognize the importance of the poetic structure as a crucial element that warrants in-depth examination. We investigate metered Arabic poetry structure, aiming to identify distinctive traits that contribute to the unique character of each poem.

This study is situated within the exploring of the notions of order and complexity as they pertain to Arabic poetry. Building upon the findings of previous research (i.e., [3], [4]), we understand order to encompass all aspects of structure and organization, whereas complexity is associated with the quantity and variety of information. Consequently, the number of structural elements contributes to its complexity, while the arrangement of these elements influences its order.

In the context of metered Arabic poetry, specific structural elements are particularly relevant when considering complexity. These elements include used meters and patterns throughout the poem while the arrangement of these structural elements contributes to the overall sense of order. By thoroughly analyzing these components, we aim to extract and evaluate poem metrics related to order and complexity for understanding their impact on poems clustering.

Understanding the intricate interplay between order and complexity in Arabic poetry is important both for research and education purposes. It allows to understand the poetic structures and to explore the potential of their comparison. The structure itself does not necessitate a complete understanding of the language, which paves the path for poetic insights into other languages while considering their unique characteristics and demands concerning meter and patterns. This article aims at contributing to this goal by addressing the following research question: What poem metrics differentiate the structure of a metered Arabic poem from another?

To address this question, this article is structured as follows: Section II gives an overview of metered Arabic poetry, Section III presents previous work related to this research, Section IV describes extracted metrics and Section V details conducted clustering experiments. Then, Sections VI and VII summarize clustering options and insights and finally, Section VIII wraps up with a conclusion.

## II. BACKGROUND

Syllables, patterns and meters are the structural elements of metered Arabic poetry. A syllable is represented by the letter (*L*) for long, (*S*) for short and (*A*) in case of absence of vocalization. Syllabic quantity ($Q$) is the base metric to calculate the weight of any sequence of syllables. A long syllable (*L*) is twice as long as a short syllable (*S*) [5]. The absence of vocalization (*A*) is assumed to have no weight. Thus, the syllable quantity is computed as $Q = 2n_L + 1n_S + 0n_A$. For example the syllable quantity of *LSLA* is 5.

A pattern refers to a named sequence of syllables, e.g., the pattern *Faa'ilaan* is composed by the 4 syllables *LSLA*. The meter is a set of ordered patterns in a verse. In most poems, the verses adhere to one consistent meter. The occurrence of multiple meters within a single poem is uncommon. The meter is related to an entire verse, and the number of pattern positions in the verse depends on the meter, ranging from 2 to 8, as shown in Table I. Since a pattern may exist in more than one meter at one or more positions, any change in verse patterns arrangement may have an impact on the verse's structure. Meters are assigned unique names and serve to establish the rhythm of the verse. There are 16 root meters and 29 different variants in metered Arabic poetry based on 43 distinct patterns. A variant that retains the same number of patterns as the root meter is considered as *Complete (C)* while one that does not is considered as *Partial (P)* [6].

TABLE I
AVAILABLE POSITIONS AND PATTERNS PER METER VARIANT

| Meter | Pos., Pat. | Meter | Pos., Pat. |
|---|---|---|---|
| Alṭṭawil C. | 8, 6 | Almutadaarak C. | 8, 6 |
| Albasiṭ C. | 8, 8 | Almutadaarak P.1 | 6, 7 |
| Albasiṭ P.1 | 6, 11 | Almutadaarak P.2 | 4, 7 |
| Albasiṭ P.2 | 6, 9 | Almutaqaarib C. | 8, 5 |
| Alhazaj C. | 4, 6 | Almutaqaarib P. | 6, 5 |
| Alkaamil C. | 6, 10 | Alrrajaz C. | 6, 13 |
| Alkaamil P. | 4, 9 | Alrrajaz P.1 | 4, 13 |
| Alkhafif C. | 6, 10 | Alrrajaz P.2 | 3, 8 |
| Alkhafif P. | 4, 7 | Alrrajaz P.3 | 2, 6 |
| Almadid C. | 6, 10 | Alrramal C. | 6, 12 |
| Almadid P. | 4, 7 | Alrramal P. | 4, 12 |
| Almuḍaari' C. | 4, 4 | Alssari' C. | 6, 11 |
| Almujtath C. | 4, 7 | Alwaafir C. | 6, 4 |
| Almunsariḥ C. | 6, 9 | Alwaafir P. | 4, 4 |
| Almuqtaḍab C. | 4, 5 | | |

## III. RELATED WORK

### A. Classification of non Arabic poetry

Previous research demonstrated that concrete features such as rhyme and alliteration, as well as abstract features like positive emotions and psychological well-being, can be utilized to distinguish if a particular poem was written in the style of professional poets or amateur poets [7]. The research [8] used orthographic, syntactic, phonemic and lexical features related to word types to classify Bangla poems into four categories: "Devotional", "Love", "Nature", and "Nationalism". The study compares the performance of Super Vector Machine (SVM) and Naive Bayes algorithms and finds that SVM, trained on both lexical and stylistic features, outperforms the other method. Similarly, the article [9] comes to a close conclusion when categorizing Punjabi poems into categories: "Nature and Festival", "Linguistic and Patriotic", "Relation and Romantic" and "Philosophy and Spiritual" using different lexical, syntactic and semantic features. In the subject-based poetry classification research [10], poems are classified in categories such "Love", "Nature", "Religion", "Living", "Arts & Sciences" using term frequency and probabilistic models (Latent Dirichlet Allocation). The SVM classifier gives an overall accuracy ranging between 0.6 and 0.8. In emotion classification field, [11] used deep learning algorithms to categorize a dataset of English poems into emotional states such "Love", "Joy", "Hope", "Sadness", "Anger" etc. Experimental results obtained by using the C-BiLSTM model achieves an overall accuracy of 88%. Some researchers [12], [13] direct their attention towards binary classification in order to differentiate between metered and non-metered poems or to distinguish poetry from prose.

Whereas the research discussed above focuses on supervised learning, other research investigated unsupervised machine learning classification techniques or statistical approaches when poems categories are unknown. For instance, a study analyzed a collection of American poems and computed weighted metrics based on different features [14]: orthographic, syntactic and phonetic. The findings suggest that these features are better suited for clustering and distinguishing between different poetry styles than traditional word-occurrence features. Another study [15] uses a clustering approach for mining T.S. Eliot's poem *"The Waste Land"* using a slightly modified version of K-means algorithm and achieves better results than baseline approaches. A recent paper also suggests a clustering approach for mining emotional patterns in Romanian poetry using lexicon-based emotion features [16]. That study employs a hierarchical clustering algorithm that yields 50 clusters, which are validated by a Silhouette Index of 0.79. Rather than using machine learning techniques, the research [17] analyzed the authorship of the 21-poem collection *"The Passionate Pilgrim"*, associated with William Shakespeare using three statistical techniques: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Vector Space Method (VSM), along with four neurolinguistic-based features. The study reveals that only 15 poems are authored by Shakespeare.

### B. Classification of Arabic poetry

The Arabic poetry categorization by era conducted by researchers [18] involved using multiple classifiers on a vast collection of poems ranging from the 5th to 15th century. The best-performing model for this task was the Multinomial Naive Bayes with a word tokenizer and no stop words, achieving an accuracy score of 70.21%. For the same purpose of classifying poems by era, the study [19] employed deep learning models and discovered that polarity categorization between "Modern" and "Non Modern" poems achieved a high score of over 91% due to significant vocabulary differences. To classify Arabic poetry based on emotions, study [20] categorized poems into four classes: "Elegy", "Love", "Pride", and "Satire". The study compared four algorithms: Naive Bayes, SVM, Voting Feature Intervals (VFI), and Hyperpipes, and found that Hyperpipes with non-stemmed and non-rooted features achieved the highest precision of 79%. Similarly, [21] used three classifiers, namely SVM, Linear Support Vector Classification (SVC), and Naive Bayes to categorize modern Arabic poems into four classes: "Love", "Islamic", "Social" and "Political". The study found that SVC and Naive Bayes outperformed the SVM algorithm. In the same topic, the research [22] investigates the use of deep learning techniques for categorizing Arabic poems into emotional categories, such as "Joy", "Sadness", and "Love". After training several models, the study found that AraBERT, an Arabic variant of Bidirectional Encoder Representations from Transformers (BERT), achieved the highest performance. In the field of authorship recognition, the research [23] used lexical, syntactic, and semantic features extracted from poems written by 73 poets and achieve an overall accuracy of 98.63% when using SVM classifier on all features.

This literature review shows that researchers have used various techniques to classify poems using mainly textual features. Some of them have explored poetry structure. The majority of this work has focused on English and other languages. This is particularly noteworthy given that Arabic is one of the most widely spoken languages in the world, yet there has been little research investigating its poetry. The exploration of poetry structure for clustering purposes aims to bridge this gap and expand the natural language processing by incorporating analysis of metered Arabic poetry. To the best of our knowledge, this article represents the first attempt to cluster metered Arabic poetry using the notions of order and complexity.

## IV. Poem structure metrics

As shown in Table II, we selected 6 metrics of order and complexity. We discarded other metrics for correlation reasons. For example, we excluded Simpson's indices [24] from the diversity metrics because they are correlated with Shannon's indices. Some syllabic similarity metrics using Gestalt pattern matching [25] are found to be correlated with repetition metrics. Similarly, metrics that gauge variability through the Manhattan distance between verses syllabic quantity are correlated with the variability metric using standard deviation.

TABLE II
POEM METRICS

| Metric | Att. | Type |
|---|---|---|
| Diversity | *DIV* | Complexity |
| Eveness | *EVE* | Complexity |
| Variability | *VAR* | Order |
| Repetition | *REP* | Order |
| Pattern Potential | *POT* | Complexity |
| Pattern Usage | *USG* | Complexity |

### A. Diversity

The Shannon diversity index [26], also known as the Shannon-Wiener index or Shannon entropy, is a measure used to quantify the diversity or richness of a dataset. While it is commonly applied in the context of biodiversity to assess species diversity within an ecosystem, we assume that the index is applicable to the context of metered Arabic Poetry considering the whole poem as a community and patterns as the species inside the community.

$$DIV = -\sum_{i=1}^{P}(p_i \log_2(p_i)) \tag{1}$$

*DIV* is the Shannon entropy or diversity index. $P$ is the total number of patterns in the poem. $p_i$ is the proportion or relative abundance of patterns in the poem. The higher the value of *DIV*, the higher the diversity. A value of $DIV = 0$ indicates a poem uses one pattern only.

### B. Evenness

The Shannon evenness index [26], also known as Equitability index, is a metric that quantifies the evenness or uniformity of different patterns in the poem. This value ranges from 0 to 1 where 1 indicates complete uniformity.

*EVE* is the Shannon evenness index. *DIV* is the Shannon diversity index. $P$ is the total number of patterns in the poem.

$$EVE = \frac{DIV}{\log_2(P)} \tag{2}$$

### C. Variability

This metric is computed using the standard deviation of syllabic quantity ($Q$) of all poem's verses ($V$). If all verses have the same value ($Q$), the variability is 0.

$$VAR = \sqrt{\frac{1}{V}\sum_{i=1}^{V}(Q_i - \bar{Q})^2} \tag{3}$$

### D. Repetition

This metric evaluates the repetition of syllabic sequences in the poem. Its highest value is 1 when all verses have the same syllabic sequence, as the maximum count of repeated syllables matches the number of verses ($V$). Conversely, the lowest value is achieved when each verse has its distinct syllabic sequence, resulting in a maximum count of repeated syllables of 1. The repetition metric is then computed as the reciprocal of the number of verses, represented by the division of 1 by the number of verses.

$$REP = \frac{\text{Maximum Count of repeated syllables}}{V} \tag{4}$$

### E. Pattern Potential

This metric calculates the difference between the patterns accessible to the poet through the meter ($P_a$) of the poem and the patterns actually employed in the poem ($P_u$). It serves as an indicator of the number of unused patterns at the poet's disposal.

$$POT = P_a - P_u \tag{5}$$

### F. Pattern Usage

This metric highlights the usage of few patterns in the poem's composition by assuming that composing a limited number of verses with a diverse range of patterns is generally easier than constructing a large number of verses with only a few patterns.

$$USG = \frac{V \cdot POT}{L \cdot P_a} \tag{6}$$

*USG* is the pattern usage metric, $V$ is the number of verses, *POT* is the pattern potential metric described in equation 5,

$P_a$ is the number of patterns available in the poem's meter, $L$ is the upper limit of verses processed per poem. In this study, $L$ is set to 100. In most cases, *USG* varies between 0 and 1 except when the poet uses more than one meter in the same poem.

## V. Experiments

The experimental process includes three sequential phases. The first involves corpus data exploration. In the second phase, clustering tests are applied using various attributes combinations. In the final phase, clustering options that exhibit the highest clustering validation indices are retained.

### A. Corpus

We used the poetry corpus used by [6] and extended it to 119 poems [27]. The number of verses per poem ranges from 5 to 100. The corpus contains 68 distinct poets and counts 3381 verses. This corpus covers various historical periods, spanning from pre-Islamic times to the present. Additionally, it encompasses all poetic meters. Another advantage is that the text within the corpus is diacritized, enabling the detection of poetic meter with greater accuracy. Poetry datasets and clustering data are available at [27].

TABLE III
Data summary before standardizing

| | *USG* | *POT* | *REP* | *DIV* | *EVE* | *VAR* |
|---|---|---|---|---|---|---|
| **min** | -0.15 | -3 | 0.05 | 0.00 | 0.19 | 0.00 |
| **median** | 0.09 | 3 | 0.20 | 1.22 | 0.85 | 0.86 |
| **mean** | 0.11 | 3.96 | 0.29 | 1.13 | 0.83 | 0.78 |
| **max** | 0.71 | 11 | 1.00 | 1.94 | 1.00 | 2.42 |

The corpus contains some outliers data points that are retained as they could provide insights into unconventional compositions. All data points are centered and scaled.

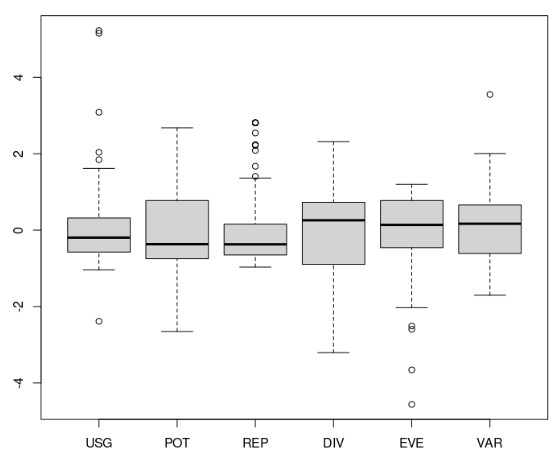


Fig. 1. Scaled data summary

### B. Clustering analysis

The goal of clustering Arabic poetry compositions is to explore the degree of similarity or dissimilarity among poetic structures, aiming to enhance their characterization and categorization beyond the traditional classification based on meter. Clustering, also known as cluster analysis, consists of grouping objects into a number of clusters in a manner that objects in the same groups are similar to each other and dissimilar to those in other groups [28]. Clustering is a fundamental component of exploratory data analysis. Unlike supervised learning, where models are trained to map input to a specific output, clustering is considered as an unsupervised learning approach [29]. The task of clustering is challenging due to the fact that clustering methods will generate clusters even if the data lacks inherent cluster structure. Consequently, it becomes imperative to assess the cluster tendency of the dataset prior to employing any clustering algorithm. Once the tendency is deemed sufficient, the analysis can proceed by clustering the data, subsequently evaluating the quality of the obtained results. Another hurdle arises in attribute selection, as there is no target output available. This necessitates context dependent approaches [30].

In clustering analysis, five commonly used clustering indicators [31] are computed. The Hopkins statistic is used for clustering tendency. The Silhouette coefficient, the Calinski-Harabasz index, the Davies-Bouldin index and the Dunn index are used for cluster validation. The Hopkins statistic [32] is employed to evaluate the clustering tendency of a dataset, aiming to determine if the data distribution is non-random. The Hopkins statistic value ranges from 0, indicating a lack of discernible clustering possibilities, to 1, denoting a substantial capacity for effective classification. Unlike the Hopkins statistic used for assessing clustering tendency, the Silhouette, Calinski-Harabasz, Davies-Bouldin and Dunn indexes are used for clustering validation. The Silhouette coefficient [33] serves as a metric for evaluating the degree of separation between clusters. It quantifies the proximity of each data point within a cluster to the data points in neighboring clusters. Ranging between -1 and 1, a value approaching 1 signifies that a data point is significantly distant from neighboring clusters. A value of 0 indicates that the data point is in close proximity to neighboring clusters. Conversely, a negative value suggests that the data point may have been wrongly assigned to an incorrect cluster. According to [34] and [35], when the Silhouette coefficient exceeds the threshold of 0.50, it indicates a good clustering quality. A value surpassing 0.70 provides strong evidence of excellent clustering quality. The Calinski-Harabasz index [36], also known as the variance ratio criterion, is a clustering evaluation metric for assessing different clustering alternatives. It is a ratio of the between-cluster variance and the within-cluster variance, and it is used to determine the number of clusters that should be used in a clustering solution. The Calinski-Harabasz index ranges from 0 to infinity and should be maximized. The higher the index, the better the solution. The Davies-Bouldin index [37] is a measure used to evaluate the clustering quality. It is based on the average similarity between clusters considering within-cluster dispersion and separation between clusters. It theoretically ranges from 0 to infinity and should be minimized. The lower the index, the better the solution. The Dunn index [38] is another clustering evaluation metric defined as the ratio between the minimum inter-cluster distance and the maximum intra-cluster distance. It theoretically ranges from 0 to infinity ans should be maximized. The higher the index, the better the solution.

In the context of this research and due to the lack of available output attribute, the clustering of all combinations of attributes is considered. Attributes are combined without repetition. Each combination includes at least 2 attributes and up to 6. Two distinct clustering algorithms, namely *hclust* (Agglomerative Hierarchical Clustering) and *K-means* (Partitioning clustering) are applied to each combination. For each algorithm, 8 clustering options are tested by varying the number of clusters from 2 to 9. Consequently, 912 clustering options are tested, as summarized in Table IV.

TABLE IV
ATTRIBUTES COMBINATIONS AND EXPERIMENTS

| Atts. | Comb. | Alg. | Clusters | Exp. |
|---|---|---|---|---|
| 6 | 1 | 2 | 8 | 16 |
| 5 | 6 | 2 | 8 | 96 |
| 4 | 15 | 2 | 8 | 240 |
| 3 | 20 | 2 | 8 | 320 |
| 2 | 15 | 2 | 8 | 240 |
| | **57** | | | **912** |

**Algorithm 1** Clustering experiments

**for** $att = 2$ to $6$ **do**
  $attributes \leftarrow Combinations(att)$
  **for all** $comb$ in $attributes$ **do**
    $h \leftarrow Hopkins(comb)$
    **for all** $alg$ in $(hclust, kmeans)$ **do**
      **for** $k = 2$ to $9$ **do**
        $c \leftarrow Clustering(comb, alg, k)$
        $sil \leftarrow Silhouette(c)$
        $ch \leftarrow CalinskiHarabasz(c)$
        $db \leftarrow DaviesBouldin(c)$
        $dn \leftarrow Dunn(c)$
        $b \leftarrow CountBad(sil)$
        $Result(att, comb, alg, k, h, sil, ch, db, dn, b)$
      **end for**
    **end for**
  **end for**
**end for**

The output of algorithm 1 is a dataset containing 912 clustering results including attributes combination, algorithm, cluster number, and all the five clustering metrics: Hopkins statistic, Silhouette coefficient, Calinski-Harabasz index, Davies-Bouldin index, Dunn index and finally the number of wrongly clustered data points.

### C. Clustering candidates

During this phase, the objective is to identify the most promising clustering options. To achieve this goal, we filtered the clustering results in three steps. The first one is setting a comparison baseline according to the Silhouette threshold set to 0.50. The baseline is the clustering result where the Silhouette index is immediately greater or equal than 0.50 as shown in the Table V.

TABLE V
CLUSTERING BASELINE

| | |
|---|---|
| **Silhouette (sil)** | 0.50 |
| **Calinski-Harabasz (ch)** | 52.38 |
| **Davies-Bouldin (db)** | 0.73 |
| **Dunn (dn)** | 0.10 |
| **bad (b)** | 6.00 |

In the second step, clustering results are filtered according to baseline values of the Silhouette coefficient and the number of misclassified clusters i.e. Silhouette $>= sil$ and wrongly classified clusters $<= b$. 16 clustering results satisfy these conditions.

In the third and final step, in order to ensure a more comprehensive evaluation and selection of the optimal clustering solution, additional conditions are introduced to avoid relying solely on a single validation index, such as the Silhouette coefficient. Among the 16 candidates, it is verified whether at least two other indexes exhibit superior performance compared to the baseline. Specifically, the Calinski-Harabasz and Dunn indexes are expected to exceed the baseline value for maximization, while the Davies-Bouldin index is anticipated to be minimized and thus be lower than the baseline.

TABLE VI
CLUSTERING FINAL CANDIDATES

| alg | att | k | sil | ch | db | dn | b |
|---|---|---|---|---|---|---|---|
| hclust | *POT,REP* | 2 | 0.53 | 62.53 | 0.65 | 0.07 | 5 |
| hclust | *REP,DIV* | 2 | 0.61 | 109.64 | 0.58 | 0.15 | 2 |
| hclust | *REP,EVE* | 2 | 0.59 | 48.72 | 0.38 | 0.18 | 5 |
| hclust | *REP,VAR* | 2 | 0.58 | 111.25 | 0.65 | 0.05 | 3 |
| hclust | *REP,VAR* | 3 | 0.56 | 147.91 | 0.63 | 0.06 | 4 |
| hclust | *USG,EVE* | 2 | 0.72 | 37.21 | 0.25 | 0.38 | 1 |
| hclust | *USG,REP* | 2 | 0.58 | 68.78 | 0.62 | 0.06 | 4 |
| kmeans | *USG,REP* | 3 | 0.58 | 116.57 | 0.70 | 0.08 | 0 |

## VI. RESULTS

The final outcome presented in Table VI, retains 8 clustering options. Figure 2 shows that the majority of poems have a repetition rate spanning from low to moderate. Cluster 2, encompassing nearly 10% of the corpus, presents poems characterized by a high *REP* alongside a *POT* ranging between low to moderate, i.e., poets use either all available patterns or approximately half of them. Other poets in the cluster 1 maintain a low repetition rate even when using few patterns. The sole case when both *REP* and *POT* reach their maximum values is when the poem uses one pattern only.

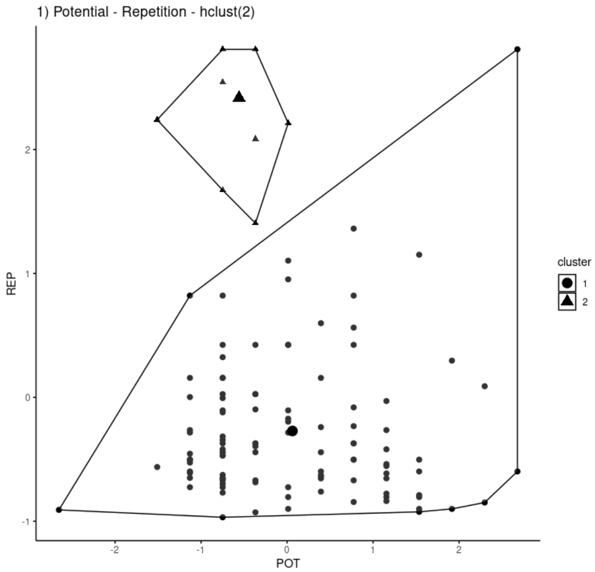


Fig. 2. Clustering option 1

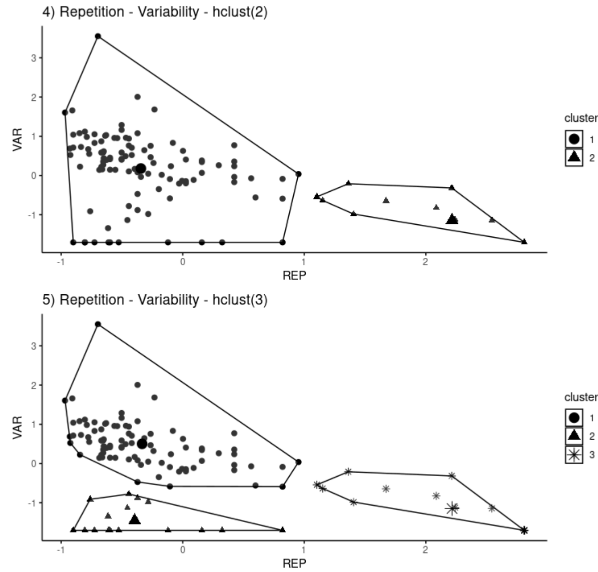


Fig. 4. Clustering options 4 and 5

Option 2, as shown in Figure 3, exhibits a similar distribution pattern (107 in cluster 1, 12 in cluster 2) as the option 1. The composition of cluster 2 is nearly identical. The high repetition in this cluster results in decreased diversity, leading to a higher likelihood of using the same patterns. The more the same patterns are repeated, the more the diversity decreases.

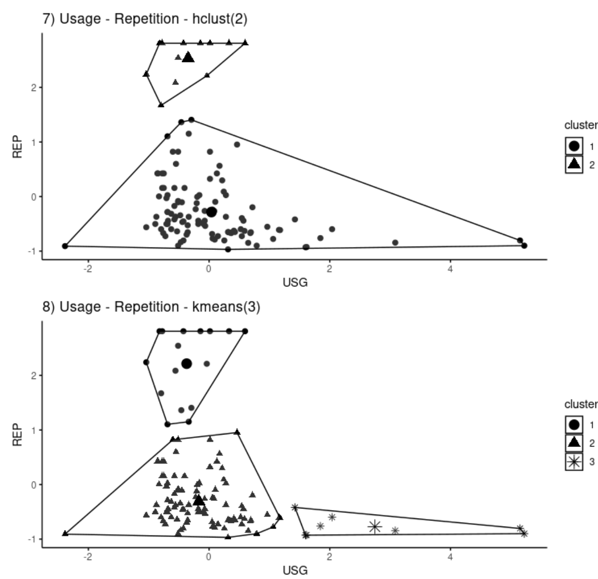


Fig. 5. Clustering options 7 and 8

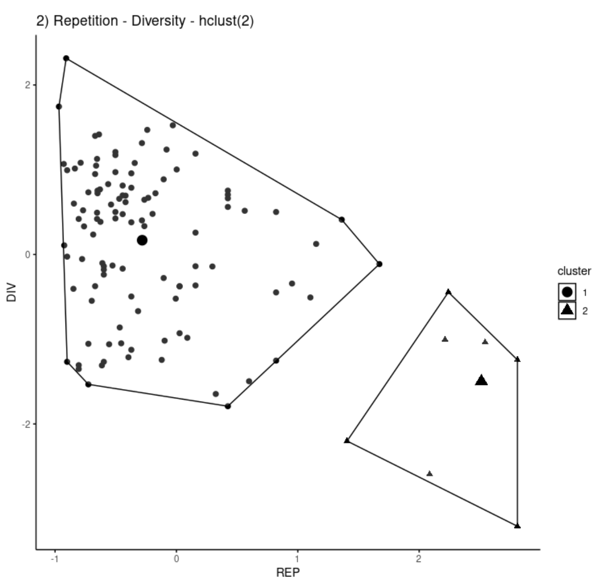


Fig. 3. Clustering option 2

The options 4 and 5, as shown in Figure 4, maintain almost the same cluster where repetition is high as in option 1 and 2. The option 5 shows that variability ranges between low and medium except some rare data points. In the option 5, most of poems have medium variability (86 in cluster 1).

Regarding the clustering using the attributes of pattern usage and repetition, as shown in Figure 5, poems that have a high level of pattern usage maintain low repetition. The option 8 splits data in three clusters: 16 in cluster 1, 95 in cluster 2 and 8 in cluster 3. The latter represents the most difficult compositions in the corpus since poems in this cluster use few patterns to write many verses while keeping a low level of variability and repetition.

We performed an inspection of the clustering characteristics of each option and found that the options 3 and 6 shown in Figure 6 do not provide a clustering meaning in the context of metered Arabic poetry. Both involve the evenness attribute.

## VII. Discussion

As indicated in the results presented in Table VI, all of the selected options propose that only two attributes are necessary

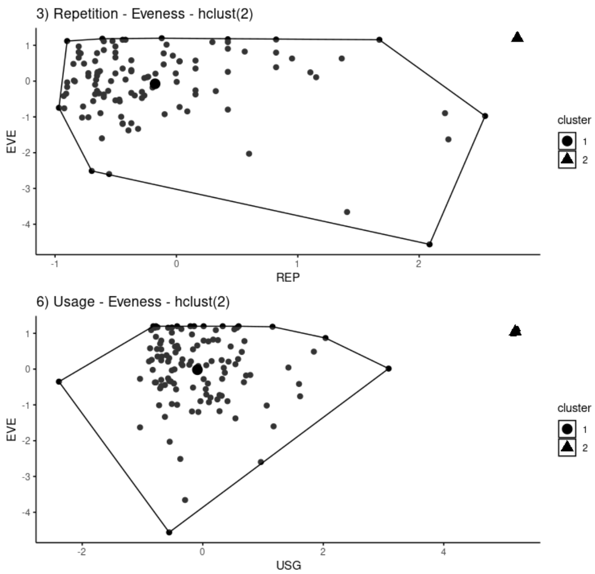


Fig. 6. Clustering options 3 and 6

for clustering the poetry data, in contrast to the initial six attributes. Among the eight options considered, the metric *REP* appears seven times, *USG* appears three times, *VAR* and *EVE* appear twice, *POT* and *DIV* appear once.

In the clusters with highest number of members, the majority of poems exhibit characteristics such as low to moderate pattern usage *USG*, moderate to high diversity *DIV*, low to moderate repetition *REP*, and moderate variability *VAR*. However, poets who look to distinguish their poems from the prevailing compositions belong to smaller groups. These groups demonstrate either a high repetition rate *REP* (options 1, 2, 4, 5, 7, 8), high pattern usage *USG* (option 8) or low variability *VAR* (option 5). As an example of remarkable compositions, the two poets *Antarah ibn Shaddad* and *Labid ibn Rabi'ah* achieved a high pattern usage rate in their poems *Almu'allaqaat* [39], by composing 89 and 88 verses respectively, utilizing a mere 2 patterns out of the available 10 while maintaining the variability rate to 0 by keeping a consistent syllabic quantity across all verses.

Given their prominence in the selected clustering options and their influential role in dividing the data, the metrics *REP*, *VAR*, and *USG* emerge as the most significant factors in the clustering of metered Arabic poetry. The nature of a composition relies then on the poet's skills in effectively merging and harmonizing these three metrics, which may occasionally conflict with one another.

The composition's direction is contingent upon the poet's intended emphasis. The poem's semantics, encompassing its message, emotional impact, and imagery, significantly influence the meter selection. When poets prioritize low variability, they write verses that maintain consistent syllabic quantities across the poem, ensuring a uniform rhythmic flow. Furthermore, by employing a selected set of patterns and incorporating them into multiple verses, they can achieve a heightened pattern usage rate. However, if poets opt to avoid repetitive patterns to introduce structural diversity, composing with a limited number of patterns becomes challenging. Varying their placement poses the risk of deviating from the established meter and potentially altering syllabic quantities, thereby compromising the goal of low variability.

It is therefore preferable to look at these attributes in pairs and not all 3 at the the same time. *USG* with *REP* or *VAR* with *REP* as obtained in option 5 and 8. These two options suggest clustering Arabic poetry in 3 clusters. We assume that by widening the corpus, other clusters may exist. However, metered poetry, unlike free poetry or prose, is rarely highly variable. Consequently, by expanding the scope of the study significantly, we can anticipate the emergence of a number of clusters ranging from 6 to maximum 9 according to the possible combinations of the three mentioned attributes. (*VAR*: Low, Medium), (*REP*: Low, Medium, High), (*USG*: Low, Medium, High).

## VIII. Conclusion

The objective of this research is to identify what differentiates the structure of a metered Arabic poem from another. We conducted clustering analysis of 119 metered Arabic poems, considering six independent variables: *Diversity*, *Evenness*, *Variability*, *Repetition*, *Pattern Potential* and *Pattern Usage*. By exploring all possible combinations of attributes, we used two distinct algorithms, namely *K-means* and *hclust*, while varying the clustering possibilities from 2 to 9. Throughout the process, we conducted clustering tests on 912 attribute combinations, progressively measuring and validating the results. We applied evaluation metrics such as Hopkins statistic, Silhouette coefficient, Calinski-Harabasz index, Davies-Bouldin index, Dunn index and checked the number of misclassified points to refine the selection. This iterative approach allowed to narrow down the clustering options from 912 options to 16, and subsequently to 8, by establishing a threshold for clustering quality. Results show that two attributes are sufficient to characterize the poetry instead of six. This finding highlights the potential significance of a reduced set of attributes in effectively clustering the poetry data.

Regarding the metrics, *Repetition*, *Variability* and *Pattern usage* metrics are found to be the most significant factors in the clustering of metered Arabic poetry. Their consistent presence and notable impact on the clustering process highlight their importance and suggest their crucial role in capturing the distinguishing characteristics of the poetry data.

The maximum number of clusters suggested by the result is 3. However, we formulate the hypothesis that by expanding the corpus extensively, the number of clusters would converge to a range between 6 and 9.

As a potential future research, we suggest enlarging the corpus to validate the hypothesis of number of clusters ranging between 6 and 9 and go further in studying the link between Arabic poetry structure and the notions of order, complexity and aesthetics.

## REFERENCES


[1] Dewi Syafitri and Melisa Marlinton. An analysis of figurative language used in edgar allan poe's poems. *Linguistic, English Education and Art (LEEA) Journal*, 2(1):43–59, 2018.

[2] Christian Obermeier, Winfried Menninghaus, Martin Von Koppenfels, Tim Raettig, Maren Schmidt-Kassow, Sascha Otterbein, and Sonja A Kotz. Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in psychology*, 4:10, 2013.

[3] Eline Van Geert and Johan Wagemans. Order, complexity, and aesthetic appreciation. *Psychology of aesthetics, creativity, and the arts*, 14(2):135, 2020.

[4] Eline Van Geert, Christophe Bossens, and Johan Wagemans. The order & complexity toolbox for aesthetics (octa): A systematic approach to study the relations between order, complexity, and aesthetic appreciation. *Behavior Research Methods*, pages 1–24, 2022.

[5] Ibrahim Anis. *Al Aswat Al Loghawiya (Linguistic sounds)*. Maktabat Al Anglo Masriya, Cairo, 1975.

[6] Abdelmalek Berkani, Adrian Holzer, and Kilian Stoffel. Pattern matching in meter detection of arabic classical poetry. In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–8. IEEE, 2020.

[7] Justine Kao and Dan Jurafsky. A computational analysis of style, affect, and imagery in contemporary poetry. In *Proceedings of the NAACL-HLT 2012 workshop on computational linguistics for literature*, pages 8–17, 2012.

[8] Geetanjali Rakshit, Anupam Ghosh, Pushpak Bhattacharyya, and Gholamreza Haffari. Automated analysis of bangla poetry for classification and poet identification. In *Proceedings of the 12th international conference on natural language processing*, pages 247–253, 2015.

[9] Jasleen Kaur and Jatinderkumar R Saini. Designing punjabi poetry classifiers using machine learning and different textual features. *Int. Arab J. Inf. Technol.*, 17(1):38–44, 2020.

[10] Andres Lou, Diana Inkpen, and Chris Tanasescu. Multilabel subject-based classification of poetry. *Nature*, 2218:30–7, 2015.

[11] Shakeel Ahmad, Muhammad Zubair Asghar, Fahad Mazaed Alotaibi, and Sherafzal Khan. Classification of poetry text into the emotional states using deep learning technique. *IEEE Access*, 8:73865–73878, 2020.

[12] Chris Tanasescu, Bryan Paget, and Diana Inkpen. Automatic classification of poetry by meter and rhyme. In *The Twenty-Ninth International Flairs Conference*, 2016.

[13] Amitha Gopidi and Aniket Alam. Computational analysis of the historical changes in poetry and prose. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 14–22, 2019.

[14] David M Kaplan and David M Blei. A computational approach to style in american poetry. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 553–558. IEEE, 2007.

[15] Julian Brooke, Graeme Hirst, and Adam Hammond. Clustering voices in the waste land. In *Proceedings of the Workshop on Computational Linguistics for Literature*, pages 41–46, 2013.

[16] Mihaiela Lupea, Anamaria Briciu, and Elena Bostenaru. Emotion-based hierarchical clustering of romanian poetry. *Stud. Inform. Control*, 30:109–118, 2021.

[17] David Kernot, Terry Bossomaier, and Roger Bradbury. Stylometric techniques for multiple author clustering. *International Journal of Advanced Computer Science and Applications*, 8(3), 2017.

[18] Mourad Abbas, Mohamed Lichouri, and Ahmed Zeggada. Classification of arabic poems: from the 5th to the 15th the century. In *New Trends in Image Analysis and Processing–ICIAP 2019: ICIAP International Workshops, BioFor, PatReCH, e-BADLE, DeepRetail, and Industrial Session, Trento, Italy, September 9–10, 2019, Revised Selected Papers*, pages 179–186. Springer, 2019.

[19] Mariam Orabi, Hozayfa El Rifai, and Ashraf Elnagar. Classical arabic poetry: Classification based on era. In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–6. IEEE, 2020.

[20] Ouais Alsharif, Deema Alshamaa, and Nada Ghneim. Emotion classification in arabic poetry using machine learning. *International Journal of Computer Applications*, 65(16), 2013.

[21] Munef Abdullah Ahmed, Raed Abdulkareem Hasan, Ahmed Hussein Ali, and Mostafa Abdulghafoor Mohammed. The classification of the modern arabic poetry using machine learning. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(5):2667–2674, 2019.

[22] Sakib Shahriar, Noora Al Roken, and Imran Zualkernan. Classification of arabic poetry emotions using deep learning. *Computers*, 12(5):89, 2023.

[23] A-F Ahmed, Ramdani Mohamed, and Bellafkih Mostafa. Machine learning for authorship attribution in arabic poetry. *Int. J. Future Comput. Commun*, 6(2):42–46, 2017.

[24] Edward H Simpson. Measurement of diversity. *nature*, 163(4148):688–688, 1949.

[25] PE Black. Ratcliff/obershelp pattern recognition. http://www.nist.gov/dads/HTML/ratcliffObershelp.html/, 2021. Accessed: 2022-04-06.

[26] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[27] Abdelmalek Berkani. Arabic poetry analysis datasets. https://doi.org/10.5281/zenodo.8256825/, 2023. Publisher: Zenodo, Version 1.0.

[28] Michael Steinbach, Levent Ertöz, and Vipin Kumar. The challenges of clustering high dimensional data. *New directions in statistical physics: econophysics, bioinformatics, and pattern recognition*, pages 273–309, 2004.

[29] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[30] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.

[31] Olatz Arbelaitz, Ibai Gurrutxaga, Javier Muguerza, Jesús M Pérez, and Iñigo Perona. An extensive comparative study of cluster validity indices. *Pattern recognition*, 46(1):243–256, 2013.

[32] Brian Hopkins and John Gordon Skellam. A new method for determining the type of distribution of plant individuals. *Annals of Botany*, 18(2):213–227, 1954.

[33] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[34] Edwin S Dalmaijer, Camilla L Nord, and Duncan E Astle. Statistical power for cluster analysis. *BMC bioinformatics*, 23(1):1–28, 2022.

[35] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.

[36] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[37] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, 1(2):224–227, 1979.

[38] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.

[39] Al-Zawzani. *Sharh al-mu'allaqat al-ashr (Explanation of the ten odes)*. Dar Maktabat Al Hayat, Beirut, 1983.