Appendix C

I present here a series of 'model cards' describing the metrics included in the EmDraCor vectors. Each card includes a descriptive name of the feature, the related variable as found in the Python scripts, its source, and an explanation on how it was calculated – especially important when the value was not directly computed by the DraCor metrics service and called via API, but rather calculated autonomously on the basis of other scholars' work. Metrics are grouped according to the four textual aspects they cover: network, cast and speech, size, plot.

Network metrics

Metric Name	**Variable Name**	**Source**	**Explanation**
Size	Size	DraCor API	Represents the total number of nodes in the network.
Average Clustering Coefficient	averageClustering	DraCor API	Calculated as the mean of the local clustering coefficients, which indicates the extent to which nodes cluster or form tightly interconnected groups.

Metric Name	Variable Name	Source	Explanation
Density	density	DraCor API	Represents the ratio of actual edges to possible edges, reflecting the overall interconnectivity of the network.
Average Degree	averageDegree	DraCor API	Represents the average number of edges any node has, measuring the overall connectivity.
Maximum Degree	maxDegree	DraCor API	Represents the degree of the character with the highest number of connections.
Average Betweenness	Betweenness	Computed on data from the DraCor API	Betweenness measures the extent to which a node in a network lies on the shortest paths between other nodes, indicating its importance in facilitating communication or information flow.
Average Closeness	closeness	Computed on data from the DraCor API	Closeness quantifies how close a node is to all other nodes in a graph, with nodes having higher closeness centrality being more central in terms of overall network accessibility. It is calculated as the reciprocal of the sum of the shortest path distances from that node to all other nodes in the network.

Metric Name	Variable Name	Source	Explanation
Average Eigenvector Centrality	eigenvector	Computed on data from the DraCor API	Eigenvector centrality assesses the importance of a node based on both its direct connections and the importance of the nodes it is connected to, with nodes being more central if they are connected to other highly central nodes.
Number Of Edges	numEdges	DraCor API	Represents the total number of edges in the network.
Connected Components	numConnectedComponents	DraCor API	Represents the number of independent subgraphs (connected components) within the network.
Ratio, Average Degree To Maximum Degree	avg_max_degree_ratio	Derived from Szemes and Vida (2024), computed on DraCor API data	"This measure therefore gives the proportion of the average degree number relative to the maximum degree number, and in a sense resembles density" (Szemes and Vida, 2024, 172).
Ratio, Maximum Degree To Number Of Characters	max_degree_num_char_ratio	Derived from Szemes and Vida (2024), computed on DraCor API data	"The ratio shows what percentage of the entire cast the node with the highest degree interacts with" (Szemes and Vida, 2024, 172).

Metric Name	Variable Name	Source	Explanation
Weighted Degree Distribution	low/medium/high-WeightedDegree	Derived from Szemes and Vida (2024), computed on DraCor API data	Weighted degree quantifies the importance or strength of the connections a node has by assigning a weight to each edge. While Szemes and Vida (2024) used *k*-means clustering to identify different groups in the distribution, I preferred a simpler solution, i.e. counting how many characters have a weighted degree falling in the first, second, or third tertile (= low, medium, high WD) and then converting it to a share.
Protagonism	Protagonism	Derived from Algee-Hewitt (2017), computed on DraCor API data	A variation of eigenvector centrality, computed as follows: the ratio between the Gini score (a common measure for inequality, here describing the relationship between the core and the peripheral cast in a play) and the percentage of characters in the top quartile of the eigenvector distribution.

Metric Name	Variable Name	Source	Explanation
Mediated-ness	Mediatedness	Derived from Algee-Hewitt (2017), computed on DraCor API data	A variation of between-ness, refers to the max-imum normalised betweenness among the characters.

Cast and speech

Metric Name	Variable Name	Source	Explanation
Average Characters Per Scene	averageCharin-Scene	Derived from Szemes and Vida (2024), computed on DraCor API data	Computed by dividing the number of charac-ters in one scene by the total number of scenes.
Average Length Of Character Speech	averageChar-Speech	Derived from Szemes and Vida (2024), computed on DraCor API data	Average number of words spoken by a character per line and/or sentence. Only characters who speak more than 10 times are considered.

Metric Name	Variable Name	Source	Explanation
Speech Intensity Distribution	lowSpeech, mediumSpeech, highSpeech	Derived from Szemes and Vida (2024), computed on DraCor API data	Computed following the same logic of weighted degree distribution (word count + division in tertiles), shows the share of characters which speak 'much', 'little', or 'a moderate amount' in the text. Szemes and Vida excluded collective characters from the count, but I reintroduced them since I believe them, in most instances in my corpus, to be fully fledged dramatic characters (i.e. not only 'background noise').
Number Of Speakers	numOfSpeakers	DraCor API	Number of speaking characters in a play (marked with a *@who* attribute inside any <sp> element).
Number Of Speakers By Gender	numOfSpeakersFemale, numOfSpeakersMale, numOfSpeakersUnknown	DraCor API	Number of speaking characters which have a given *@sex* attribute in the <castList> element.
Number Of Collective Speakers	numPersonGroups	DraCor API	Number of speaking characters marked by the element <personGrp> in the <particDesc>.

Size

Metric Name	Variable Name	Source	Explanation
Number Of Acts	numOfActs	DraCor API	Number of elements whose attribute *@type* is "act" (within the <body> element in the TEI markup).
Number Of Segments	numOfSegments	DraCor API	Number of elements of all types, within the <body> element in the TEI markup, including acts, scenes, configurations, and other types of textual segments. Segments inside the <front> or <back> elements, such as prefaces or epilogues, are not counted.
Word Count, Whole Text	wordCountText	DraCor API	Total number of words (tokens) in the entire text.
Word Count, Spoken Text	wordCountSp	DraCor API	Total number of words (tokens) within any <sp> element (i.e. within any speech act).
Word Count, Stage Directions	wordCountStage	DraCor API	Total number of words (tokens) within any <stage> element (i.e. within any stage direction).

Metric Name	Variable Name	Source	Explanation
Number Of Prose Lines	numOfP	DraCor API	Total number of paragraphs in prose (<p>) in the play.
Number Of Verse Lines	numOfL	DraCor API	Total number of lines of verse (<l>) in the play.

Plot dynamics

Metric Name	Variable Name	Source	Explanation
All-In Index	all_in_index	Derived from Trilcke et al. (2017), computed on DraCor API data	Identifies the point in time at which all characters have occurred at least once in a drama, expressed as a percentage (position of the segment where it happens / total number of segments, e.g.: all-in at segment 41/42: 98%).
Final Scene Size	final_scene_size	Derived from Trilcke et al. (2017), computed on DraCor API data	Shows the percentage of all characters of a drama which appear on stage in the last scene.

Metric Name	Variable Name	Source	Explanation
Drama Change Rate	drama_change_rate	Derived from Trilcke et al. (2017), computed on DraCor API data	Based on a modified Levenshtein distance, it takes into account only insertions (+1) and deletions (-1) of characters between scenes (see Fischer et al., 2017). I first compute the change rate between segments through the ratio between the number of edits and the number of individual speakers in any pair of scenes. The mean of all segment change rates is then assumed as the overall drama change rate.