Tabelle 4.2: Übersicht über die verschiedenen Aufrufmöglichkeiten und Parameter der entwickelten Applikation

Programmaufruf für kompletten Prozess (Grapherstellung + Auswertung)	
python3 ___init___.py graph_save_calculate plain_file encoded_file threshold <optionen>	
Programmaufruf für Ähnlichkeitsgrapherstellung	
python3 ___init___.py graph_save plain_file encoded_file threshold <optionen>	
Programmaufruf für Berechnung der Präzisionswerte aus dem Graph	
python3 ___init___.py graph_load pickle_file <optionen>	
Positionsgebundene Argumente	
plain_file	String: Pfad zur Klartext-CSV-Datei
encoded_file	String: Pfad zur CSV-Datei mit den enkodierten Daten
pickle_file	String: Pfad zur erstellten Pickle-Graph-Datei
threshold	Float: Schwellwert für die Berechnung des Ähnlichkeitsgraphen
Optionale Argumente (graph_load und graph_save_calculate)	
--min_comp_size	Int: Minimale Anzahl der Komponentengröße Default: 0 (graph_load), sonst 3
--results_path	String: Speicherpfad für die Resultate
--lsh_size_node_matching	Int: Größe des Vektors für Hamming-LSH beim Node-Matching
--lsh_count_node_matching	Int: Anzahl der Vektoren für Hamming-LSH beim Node-Matching
--node_matching_tech	String: Technik für das Node-Matching (mögliche Werte: **shm**, smm, mwm)
--weight_list	List<Float>: Gewichte (für NF) für die Berechnung der Embedding-Ähnlichkeit zwischen Node-Features und -Embeddings, Default: 0.9, 0.8, ..., 0.1
--graphwave_sg_lib	Boolean: Wenn gesetzt, dann wird für GraphWave die StellarGraph-Implementierung verwendet (ohne Kantengewichte)
--hp_config_file	String: Dateiname (ohne .py) für die im config-Ordner liegende Konfigurationsdatei für das Hyperparametertuning
--scaler	String: Skalierungstechnik für die Node-Features und -Embeddings (minmaxscaler oder **standardscaler**)
--num_top_pairs	List<Int>: Mengen an Top-Matches, die jeweils für die Präzisionsberechnung betrachtet werden sollen

weiter auf der nächsten Seite

Tabelle 4.2: Übersicht über die verschiedenen Aufrufmöglichkeiten und Parameter der entwickelten Applikation fortgesetzt

--node_matching_threshold	Float: Schwellwert für die Cosinus-Ähnlichkeit beim bipartiten Graphen im Node-Matching-Schritt
--vidanage_weights	List<Float> (Länge 3): Gewichte für die Neuberechnung der endgültigen Ähnlichkeit beim bipartiten Graphen für Cosinus-Ähnlichkeit, Ähnlichkeits- und Grad-Konfidenz **(0.6, 0.3, 0.1)**
Optionale Argumente (graph_save und graph_save_calculate)	
--graph_path	String: Speicherpfad für die Pickle-Datei mit dem berechneten StellarGraph sowie den echten Matches
--remove_frac_plain	Float: relativer Anteil an Records, der vom Klartextsatz entfernt wird
--remove_frac_encoded	Float: relativer Anteil an Records, der vom enkodierten Satz entfernt wird
--record_count	Int: Anzahl der Records, die vom Datensatz betrachtet werden
--node_features	String: Konfiguration bzgl. der zu verwendenden Knotenmerkmale (fast, egonet1, egonet2, **all**)
--node_count	Boolean: Wenn gesetzt, dann wird die Knoten-anzahl als Knotenmerkmal mitverwendet
--nf_scaled	String: Wenn gesetzt (standardscaler oder minmaxscaler), dann werden die Knotenmerkmale (für die Node-Embedding-Techniken) der beiden Graphen getrennt skaliert
--padding	Boolean: Wenn gesetzt, dann wird davon ausgegangen, dass die enkodierten Daten auf Basis von Padding berechnet wurden
--lsh_size_blocking	Int: Größe des Vektors für Hamming-LSH beim Blocking für den Ähnlichkeitsgraphen
--lsh_count_blocking	Int: Anzahl der Vektoren für Hamming-LSH beim Blocking für den Ähnlichkeitsgraphen
--qgram_attributes	List<String>: Spaltenname der Attribute, für die die Q-Gramme berechnet werden sollen
--encoded_attr	String: Spaltenname für das Attribut, wo sich der enkodierte Bloom-Filter befindet
--min_comp_size	Int: Minimale Anzahl der Komponentengröße, Default: 3