features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Message features (M)					
Topic	M_TopicM1, M_TopicM2, ..., M_TopicM19		A measure of the likelihood that the message belongs to a specific topic in the 19-topic scheme. (1-19: arts&culture, business&entrepreneurs, celebrity&pop_culture, diaries&daily_life, family, fashion&style, film_tv&video, fitness&health, food&dining, gaming, learning&educational, music, news&social_concern, other_hobbies, relationships, science&technology, sports, travel&adventure, youth&student_life)	a float within the range [0.0, 1.0] which represents the likelihood	https://huggingface.co/cardiffnlp/tweet-topic-21-multi
	M_TopicMMain		The most likely topic with the highest likelihood in the 19-topic scheme.	an integer within [0,18], each integer represents a topic.	
	M_TopicMNum		The number of identified topics with likelihood > 0.5 in the 19-topic scheme.	an integer within [0,19]	
	M_TopicG1, M_TopicG2, ..., M_TopicG6		A measure of the likelihood that the message belongs to a specific topic in the 6-topic scheme. (1-6: arts&culture, business&entrepreneurs, pop_culture, daily_life, sports&gaming, science&technology)	a float within the range [0.0, 1.0] which represents the likelihood	https://huggingface.co/cardiffnlp/tweet-topic-21-single
	M_TopicGMain		The most likely topic with the highest likelihood in the 6-topic scheme.	an integer within [0,5], each integer represents a topic.	
	M_TopicGNum		The number of identified topics with likelihood > 0.5 in the 6-topic scheme.	an integer within [0,6]	
	M_TopicLDA1, M_TopicLDA2, ..., M_TopicLDA10	Y	A measure of the likelihood that the message belongs to one of the 10 topics identified by the LDA model.	a float within the range [0.0, 1.0] which represents the likelihood	sklearn.decomposition.LatentDirichletAllocation: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html
	M_CharNum		The number of characters of the message.	an integer>0	len(text)
	M_WordNum		The number of words of the message.	an integer within [1, text_len]	len(text.split())
	M_Grammer1		Scores for grammatical and spelling correctness on the level of words.	a float within the range [0.0, 1.0], where 1.0 means totally correct.	language-tool-python: https://pypi.org/project/language-tool-python/
	M_Grammer2		Scores for grammatical and spelling correctness of the whole message.	0 or 1, where 0 means the text has grammatical or spelling errors and 1 means there are none.	

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Language	M_Polarity		The orientation of the expressed tone of the message.	a float within the range [-1.0, 1.0], -1 means the text expresses a negative tone and 1 means the text expresses a positive tone.	TextBlob, Simplified Text Processing: https://textblob.readthedocs.io/en/dev/
	M_Subjectivity		A measure of the amount of personal opinion and factual information contained in the message. The higher subjectivity means that the message contains more personal opinion than factual information.	a float within the range [0.0, 1.0] where 0.0 means extremely objective and 1.0 means extremely subjective.	
	M_Irony		A measure of how ironic the message is.	a float within the range [0.0, 1.0] where 0.0 means not ironic and 1.0 means extremely ironic.	Twitter-roBERTa-base: https://huggingface.co/cardiffnlp
	M_Offensive		A measure of how offensive the message is.	a float within the range [0.0, 1.0] where 0.0 means not offensive and 1.0 means extremely offensive.	
	M_Emoji		The most likely emoji that describes the text content	an integer within [0, 19], each integer represents an emoji	
	M_Masculinity		Whether the language style of the text is masculine.	0 or 1, where 0 means feminine and 1 means masculine.	Logistic Regression model (sklearn: TfidfVectorizer, LogisticRegression) trained on public data source: https://www.kaggle.com/datasets/crowdflower/twitter-user-gender-classification?datasetId=409&sortBy=voteCount
Readability	M_Readability1, M_Readability2, ..., M_Readability6		Measures of how difficult the message in English is to understand. Each metrix has a different focus. (1-6: Kincaid, ARI, Coleman-Liau, FleschReadingEase, GunningFogIndex, SMOGIndex.)	a float which can indicate the number of years of education(the U.S. grade level) required to understand this text, relevant when the value > 10.	https://github.com/andreasvc/readability/ DaleChallIndex uses a list of 3000 words and any word not on that list to be difficult. complex_words does not include proper nouns, familiar jargon, or compound words. Do not include common suffixes (such as -es, -ed, or -ing) as a syllable;
	M_Readability7		7: LIX, 8: RIX. LIX is a measure on how difficult to read the message. RIX is an imporved version.	a float>0	
	M_Readability8				
	M_Readability9		DaleChallIndex: A measure that provides a numeric gauge of the comprehension difficulty that readers come upon when reading the message.		
	M_Readability10		complex_words: Count of words in the message that consisting of three or more syllables.	an integer>=0	
	M_Readability11		complex_words_DaleChallIndex: Count of words in the message that are not on the list used in DaleChallIndex.		
	M_Sentiment1		Negative sentiment score of the message.	a float within the range [0.0, 1.0],	NLTK-VADER's Sentiment Intensity

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Sentiment	M_Sentiment2		Neutral sentiment score of the message.	where 0.0 represents no sentiment expressed and 1.0 means an extreme sentiment expressed.	Analyzer: https://www.nltk.org/index.html
	M_Sentiment3		Positive sentiment score of the message.		
	M_Sentiment4		Compound value of M_sentiment1, M_sentiment2, and M_sentiment3	a float between -1 (extreme negative) and +1 (extreme positive)	https://github.com/cjhutto/vaderSentiment#about-the-scoring
	M_SentimentMain		The most likely sentiment label according to M_sentiment4	an integer within [0, 2], each integer represents a sentiment (neutral, negative, positive).	
Emotion	M_Emotion1, M_Emotion2, ..., M_Emotion7		The probability that the message express a specific type of emotion. (1-7: anger, joy, fear, disgust, surprise, sad, others)	a float within the range [0.0, 1.0]	A Transformer-based library for SocialNLP tasks: https://github.com/pysentimiento/pysentimiento
	M_EmotionMain		The most likely emotion expressed by the message.	an integer within [0, 6], each integer represents an emotion.	
Hate-speech	M_Hate1		A measure of aggressiveness of the message.	a float within the range [0.0, 1.0], where 0.0 represents no hate-speech detected and 1.0 means an extreme hate-speech detected.	A Transformer-based library for SocialNLP tasks: https://github.com/pysentimiento/pysentimiento
	M_Hate2		A measure of hatefulness of the message.		
	M_Hate3		A measure of whether the message is targeting an individual (or entity). 0 - not targeting; 1 - targeting an individual.		
	M_HsNum		the number of the above types of hate speech (above given thresholds) detected in the text	an integer within [0,3]	
Others	M_hashtag		The label of the hashtag containing in the message.	an integer within [0,13], each integer represents a hashtag.	/
	M_S_id		The id of the sender of the message.	array-like of shape (n_samples,)	LabelEncoder() from sklearn.preprocessing
	M_R_id		The id of the recipient of the message.		
User features (U)					
	U_R_AccountAge		The number of days since the recipient's account was registered (until a given cutoff date).	a float >0	01/09/2022-creation time
	U_R_FollowerNum		The number of the recipient's followers.	an intenger>=0	Obtained directly from raw data
	U_R_FolloweeNum		The number of the recipient's followings.	an intenger>=0	
	U_R_TweetNum		the number of total tweet messages that the recipient has posted since registration.	an intenger>=0	
	U_R_ListedNum		The number of lists the recipient is included in.	an intenger>=0	
	U_R_SpreadActivity	Y	The ratio of the number of the recipient's messages to the maximal messages number among all users.	a float with in [0,1]	U_R_TweetNum/max (U_R_TweetNum, U_S_TweetNum)
	U_R_FollowerNumDay		The number of followers normalized by the age of this account.	a float>=0	U_R_FollowerNum/U_R_AccountAge
	U_R_FolloweeNumDay		The number of followings normalized by the age of this account.	a float>=0	U_R_FolloweeNum/U_R_AccountAge

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Profile	U_R_TweetNumDay		The number of total tweet messages normalized by the age of this account.	a float>=0	U_R_TweetNum/U_R_AccountAge
	U_R_ListedNumDay		The number of lists that the user is included in normalized by the age of this account.	a float>=0	U_R_ListedNum/U_R_AccountAge
	U_R_ProfileVerified		Whether the recipient is verified.	[0,1], 1 represents this user is verified by X	Obtained directly from raw data
	U_R_ProfileUrl		Whether a URL is shown in the recipient's profile	[0,1], 1 represents there is a URL shown in the user profile	
	U_S_AccountAge		The number of days since the sender's account was registered (until a given cutoff date).	a float >0	01/09/2022-creation time
	U_S_FollowerNum		The number of the sender's followers	an intenger>=0	Obtained directly from raw data
	U_S_FolloweeNum		The number of the sender's followings	an intenger>=0	
	U_S_TweetNum		The number of total tweet messages that the sender has posted since registration.	an intenger>=0	
	U_S_ListedNum		The number of lists the sender is included in.	an intenger>=0	
	U_S_SpreadActivity		The ratio of the number of the sender's messages to the maximal messages number among all users.	a float with in [0,1]	U_S_TweetNum/max (U_R_TweetNum, U_S_TweetNum)
	U_S_FollowerNumDay		The number of followers normalized by the age of this account.	a float>=0	U_S_FollowerNum/U_S_AccountAge
	U_S_FolloweeNumDay		The number of followings normalized by the age of this account.	a float>=0	U_S_FolloweeNum/U_S_AccountAge
	U_S_TweetNumDay		The number of total tweet messages normalized by the age of this account.	a float>=0	U_S_TweetNum/U_S_AccountAge
	U_S_ListedNumDay		The number of lists that the sender is included in normalized by the age of this account.	a float>=0	U_S_ListedNum/U_S_AccountAge
	U_S_ProfileVerified		Whether the sender is verified.	[0,1], 1 represents this user is verified by X	Obtained directly from raw data
	U_S_ProfileUrl		Whether a URL is shown in the sender's profile	[0,1], 1 represents there is a URL shown in the user profile	
Network	U_R_LeaderRank	Y	A value to measure the influence of the recipient. A variation of PageRank.	a float with in [0,1]	NetworkX: https://networkx.org/
	U_R_Indegree	Y	The number of users who had ever followed and interacted with the recipient.	an intenger>=0	Obtained directly from raw data
	U_R_FollowS		Whether the recipient is following the sender.	[0,1], 1 represents the recipient is following the sender	
	U_S_LeaderRank		A value to measure the influence of the sender. A variation of PageRank.	a float with in [0,1]	NetworkX: https://networkx.org/

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
	U_S_Indegree		The number of users who had ever followed and interacted with the sender.	an intenger>=0	Obtained directly from raw data
	U_S_FollowR		Whether the sender is following the recipient	[0,1], 1 represents the sender is following the recipient	
Historical Message content features (HM)					
We first calculate the feature value for each historical tweet of the user and then take the average of numerical values and the mode of category values.					
Historical Message features are designed samely for both the recipient (HM_R) and the sender (HM_S).					
Historical Message content features for the recipient (HM_R)					
Topic	HM_R_TopicM1, HM_R_TopicM2, ..., HM_R_TopicM19		The average of the likelihoods that the recipient's historical messages belong to a specific topic in the 19-topic scheme. (1-19: arts&culture, business&entrepreneurs, celebrity&pop_culture, diaries&daily_life, family, fashion&style, filM_tv&video, fitness&health, food&dining, gaming, learning&educational, music, news&social_concern, other_hobbies, relationships, science&technology, sports, travel&adventure, youth&student_life)	a float within the range [0.0, 1.0] which represents the likelihood	https://huggingface.co/cardiffnlp/tweet-topic-21-multi
	HM_R_TopicMMain		The topic in the 19-topic scheme that appears most frequently in the recipient's historical messages.	an integer within [0,18], each integer represents a topic.	
	HM_R_TopicMNum		The average number of identified topics with likeihood > 0.5 in the 19-topic sechme in the recipient's historical messages.	an integer within [0,19]	
	HM_R_TopicG1, HM_R_TopicG2, ..., HM_R_TopicG6,		The average of the likelihoods that the recipient's historical message belong to a specific topic in the 6-topic scheme. (1-6: arts&culture, business&entrepreneurs, pop_culture, daily_life, sports&gaming, science&technology)	a float within the range [0.0, 1.0] which represents the likelihood	https://huggingface.co/cardiffnlp/tweet-topic-21-single
	HM_R_TopicGMain		The topic in the 6-topic scheme that appears most frequently in the recipient's historical messages.	an integer within [0,5], each integer represents a topic.	
	HM_R_TopicGNum		The average number of identified topics with likeihood > 0.5 in the 6-topic sechme in the recipient's historical messages.	an integer within [0,6]	
	HM_R_TopicLDA1, HM_R_TopicLDA2, ..., HM_R_TopicLDA10	Y	The average of the likelihoods that the recipient's historical messages belong to a one of the 10 topics identified by the LDA model..	a float within the range [0.0, 1.0] which represents the likelihood	sklearn.decomposition.LatentDirichletAllocation: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html
	HM_R_TopicMatched	Y	The cosine distance between the reposted message's topic vector and the recipient’s topic preference, based on the LDA model.	a float within the range [0.0, 1.0]	sklearn.metrics.pairwise.cosine_distances

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Language	HM_R_CharNum		The average number of characters of the recipient's historical messages.	an integer>0	len(text)
	HM_R_WordNum		The average number of words of the recipient's historical messages.	an integer within [1, text_len]	len(text.split())
	HM_R_Grammer1		Average scores for grammatical and spelling correctness of the recipient's historical messages. The minimum checking unit is a word.	a float within the range [0.0, 1.0], where 1.0 means totally correct.	language-tool-python: https://pypi.org/project/language-tool-python/
	HM_R_Grammer2		Average scores for grammatical and spelling correctness of the recipient's historical messages.	0 or 1, where 0 means the text has grammatical or spelling errors and 1 means there are none.	
	HM_R_Polarity		The average orientation of the expressed tone of the recipient's historical messages.	a float within the range [-1.0, 1.0], -1 means the text expresses a negative tone and 1 means the text expresses a positive tone.	TextBlob, Simplified Text Processing: https://textblob.readthedocs.io/en/dev/
	HM_R_Subjectivity		Average of the amount of personal opinion and factual information contained in the recipient's historical tweet messages. The higher subjectivity means that the text contains personal opinion rather than factual information.	a float within the range [0.0, 1.0] where 0.0 means extremely objective and 1.0 means extremely subjective.	
	HM_R_Irony		Average of the measure of how ironic the recipient's historical tweet messages are.	a float within the range [0.0, 1.0] where 0.0 means not ironic and 1.0 means extremely ironic.	Twitter-roBERTa-base: https://huggingface.co/cardiffnlp
	HM_R_Offensive		Average of the measure of how offensive the recipient's historical tweet messages are.	a float within the range [0.0, 1.0] where 0.0 means not offensive and 1.0 means extremely offensive.	
	HM_R_Emoji		The most likely emoji that describes the recipient's historical tweet messages.	an integer within [0, 19], each integer represents an emoji	
	HM_R_MasculinityPer		Percentage of the recipient's historical tweet messages of which the language style is masculine.	a float within the range [0.0, 1.0]	Logistic Regression model(sklearn: TfidfVectorizer, LogisticRegression) trained on public data source: https://www.kaggle.com/datasets/crowdflower/twitter-user-gender-classification?datasetId=409&sortBy=voteCount
	HM_R_Readability1, HM_R_Readability2, ..., HM_R_Readability6		Average of the measure of how difficult the recipient's historical tweet messages in English are to understand. Each metrix has a different focus. (1-6: Kincaid, ARI, Coleman-Liau, FleschReadingEase, GunningFogIndex, SMOGIndex.)	a float which can indicate the number of years of education(the U.S. grade level) required to understand this text, relevant when the value > 10.	https://github.com/andreasvc/readability/ DaleChallIndex uses a list of 3000 words and any word not on that list to
	HM_R_Readability7		Average of 7: LIX, 8: RIX. LIX is a measure on how difficult to read a		

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Readability	HM_R_Readability8		message. RIX is an imporved version.	a float>0	words and any word not on that list to be difficult. complex_words does not include proper nouns, familiar jargon, or compound words. Do not include common suffixes (such as -es, -ed, or -ing) as a syllable;
	HM_R_Readability9		Avegage of DaleChallIndex: A measure that provides a numeric gauge of the comprehension difficulty that readers come upon when reading a message.		
	HM_R_Readability10		complex_words: Average count of words consisting of three or more syllables in the recipient's historical tweet messages	an integer>=0	
	HM_R_Readability11		complex_words_DaleChallIndex: Average count of words that not on the list used in DaleChallIndex in the recipient's historical tweet messages		
Sentiment	HM_R_Sentiment1		Average negative sentiment score of the recipient's historical tweet messages	a float within the range [0.0, 1.0], where 0.0 represents no sentiment expressed and 1.0 means an extreme sentiment expressed.	NLTK-VADER's Sentiment Intensity Analyzer: https://www.nltk.org/index.html
	HM_R_Sentiment2		Average neutral sentiment score of the recipient's historical tweet messages		
	HM_R_Sentiment3		Average positive sentiment score of the recipient's historical tweet messages		
	HM_R_Sentiment4		Compound value of of the HM_R_sentiment1, HM_R_sentiment2, and HM_R_sentiment3	a float between -1 (extreme negative) and +1 (extreme positive)	
	HM_R_SentimentMain		The most likely sentiment label according to HM_R_sentiment4	an integer within [0, 2], each integer represents a sentiment (neutral, negative, positive).	https://github.com/cjhutto/vaderSentiment#about-the-scoring
Emotion	HM_R_Emotion1, HM_R_Emotion2, ..., HM_R_Emotion7		The average probability that the recipient's historical tweet messages express a specific type of emotions. (1-7: anger, joy, fear, disgust, surprise, sad, others)	a float within the range [0.0, 1.0]	A Transformer-based library for SocialNLP tasks: https://github.com/pysentimiento/pysentimiento
	HM_R_EmotionMain		The most likely emotion expressed by the recipient's historical tweet messages	an integer within [0, 6], each integer represents an emotion.	
Hate-speech	HM_R_Hate1		The average of the measure of aggressiveness of the recipient's historical messages.	a float within the range [0.0, 1.0], where 0.0 represents no hate-speech detected and 1.0 means an extreme hate-speech detected.	A Transformer-based library for SocialNLP tasks: https://github.com/pysentimiento/pysentimiento
	HM_R_Hate2		The average of the measure of hatefulness of the recipient's historical messages.		
	HM_R_Hate3		The average of the level in which the recipient's historical messages are targeted to an individual or entity. 0 - not targeting; 1 - targeting an individual.		

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
	HM_R_HsNum		The average number of types of hate speech detected in the recipient's historical messages.	an integer within [0,3]	
Historical Message content features for the recipient (HM_S)					
Topic	HM_S_TopicM1, HM_S_TopicM2, ..., HM_S_TopicM19		The average of the likelihoods that the recipient's historical messages belong to a specific topic in the 19-topic scheme. (1-19: arts&culture, business&entrepreneurs, celebrity&pop_culture, diaries&daily_life, family, fashion&style, filM_tv&video, fitness&health, food&dining, gaming, learning&educational, music, news&social_concern, other_hobbies, relationships, science&technology, sports, travel&adventure, youth&student_life)	a float within the range [0.0, 1.0] which represents the likelihood	https://huggingface.co/cardiffnlp/tweet-topic-21-multi
	HM_S_TopicMMain		The topic in the 19-topic scheme that appears most frequently in the recipient's historical messages.	an integer within [0,18], each integer represents a topic.	
	HM_S_TopicMNum		The average number of identified topics with likeihood > 0.5 in the 19-topic sechme in the recipient's historical messages.	an integer within [0,19]	
	HM_S_TopicG1, HM_S_TopicG2, ..., HM_S_TopicG6,		The average of the likelihoods that the recipient's historical message belong to a specific topic in the 6-topic scheme. (1-6: arts&culture, business&entrepreneurs, pop_culture, daily_life, sports&gaming, science&technology)	a float within the range [0.0, 1.0] which represents the likelihood	https://huggingface.co/cardiffnlp/tweet-topic-21-single
	HM_S_TopicGMain		The topic in the 6-topic scheme that appears most frequently in the recipient's historical messages.	an integer within [0,5], each integer represents a topic.	
	HM_S_TopicGNum		The average number of identified topics with likeihood > 0.5 in the 6-topic sechme in the recipient's historical messages.	an integer within [0,6]	
	HM_S_TopicLDA1, HM_S_TopicLDA2, ..., HM_S_TopicLDA10		The average of the likelihoods that the recipient's historical messages belong to a one of the 10 topics identified by the LDA model..	a float within the range [0.0, 1.0] which represents the likelihood	sklearn.decomposition. LatentDirichletAllocation: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html
	HM_S_TopicMatched		The cosine distance between the reposted message's topic vector and the recipient's topic preference, based on the LDA model.	a float within the range [0.0, 1.0]	sklearn.metrics.pairwise. cosine_distances
	HM_S_CharNum		The average number of characters of the recipient's historical messages.	an integer>0	len(text)
	HM_S_WordNum		The average number of words of the recipient's historical messages.	an integer within [1, text_len]	len(text.split())
	HM_S_Grammer1		Average scores for grammatical and spelling correctness of the recipient's historical messages. The minimum checking unit is a word.	a float within the range [0.0, 1.0], where 1.0 means totally correct.	language-tool-python: https://pypi

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Language	HM_S_Grammer2		Average scores for grammatical and spelling correctness of the recipient's historical messages.	0 or 1, where 0 means the text has grammatical or spelling errors and 1 means there are none.	language-tool-python: https://pypi.org/project/language-tool-python/
	HM_S_Polarity		The average orientation of the expressed tone of the recipient's historical messages.	a float within the range [-1.0, 1.0], -1 means the text expresses a negative tone and 1 means the text expresses a positive tone.	TextBlob, Simplified Text Processing: https://textblob.readthedocs.io/en/dev/
	HM_S_Subjectivity		Average of the amount of personal opinion and factual information contained in the recipient's historical tweet messages. The higher subjectivity means that the text contains personal opinion rather than factual information.	a float within the range [0.0, 1.0] where 0.0 means extremely objective and 1.0 means extremely subjective.	
	HM_S_Irony		Average of the measure of how ironic the recipient's historical tweet messages are.	a float within the range [0.0, 1.0] where 0.0 means not ironic and 1.0 means extremely ironic.	Twitter-roBERTa-base: https://huggingface.co/cardiffnlp
	HM_S_Offensive		Average of the measure of how offensive the recipient's historical tweet messages are.	a float within the range [0.0, 1.0] where 0.0 means not offensive and 1.0 means extremely offensive.	
	HM_S_Emoji		The most likely emoji that describes the recipient's historical tweet messages.	an integer within [0, 19], each integer represents an emoji	
	HM_S_MasculinityPer		Percentage of the recipient's historical tweet messages of which the language style is masculine.	a float within the range [0.0, 1.0]	Logistic Regression model(sklearn: TfidfVectorizer, LogisticRegression) trained on public data source: https://www.kaggle.com/datasets/crowdflower/twitter-user-gender-classification?datasetId=409&sortBy=voteCount
Readability	HM_S_Readability1, HM_S_Readability2, ..., HM_S_Readability6		Average of the measure of how difficult the recipient's historical tweet messages in English are to understand. Each metrix has a different focus. (1-6: Kincaid, ARI, Coleman-Liau, FleschReadingEase, GunningFogIndex, SMOGIndex.)	a float which can indicate the number of years of education(the U.S. grade level) required to understand this text, relevant when the value > 10.	https://github.com/andreasvc/readability/ DaleChallIndex uses a list of 3000 words and any word not on that list to be difficult. complex_words does not include proper nouns, familiar jargon, or compound words. Do not include
	HM_S_Readability7		Average of 7: LIX, 8: RIX. LIX is a measure on how difficult to read a message. RIX is an imporved version.	a float>0	
	HM_S_Readability8				
	HM_S_Readability9		Avegage of DaleChallIndex: A measure that provides a numeric gauge of the comprehension difficulty that readers come upon when reading a message.		

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
	HM_S_Readability10		complex_words: Average count of words consisting of three or more syllables in the recipient's historical tweet messages	an integer>=0	common suffixes (such as -es, -ed, or -ing) as a syllable;
	HM_S_Readability11		complex_words_DaleChallIndex: Average count of words that not on the list used in DaleChallIndex in the recipient's historical tweet messages		
Sentiment	HM_S_Sentiment1		Average negative sentiment score of the recipient's historical tweet messages	a float within the range [0.0, 1.0], where 0.0 represents no sentiment expressed and 1.0 means an extreme sentiment expressed.	NLTK-VADER's Sentiment Intensity Analyzer: https://www.nltk.org/index.html
	HM_S_Sentiment2		Average neutral sentiment score of the recipient's historical tweet messages		
	HM_S_Sentiment3		Average positive sentiment score of the recipient's historical tweet messages		
	HM_S_Sentiment4		Compound value of of the HM_S_sentiment1, HM_S_sentiment2, and HM_S_sentiment3	a float between -1 (extreme negative) and +1 (extreme positive)	
	HM_S_SentimentMain		The most likely sentiment label according to HM_S_sentiment4	an integer within [0, 2], each integer represents a sentiment (neutral, negative, positive).	https://github.com/cjhutto/vaderSentiment#about-the-scoring
Emotion	HM_S_Emotion1, HM_S_Emotion2, ..., HM_S_Emotion7		The average probability that the recipient's historical tweet messages express a specific type of emotions. (1-7: anger, joy, fear, disgust, surprise, sad, others)	a float within the range [0.0, 1.0]	A Transformer-based library for SocialNLP tasks: https://github.com/pysentimiento/pysentimiento
	HM_S_EmotionMain		The most likely emotion expressed by the recipient's historical tweet messages	an integer within [0, 6], each integer represents an emotion.	
Hate-speech	HM_S_Hate1		The average of the measure of aggressiveness of the recipient's historical messages.	a float within the range [0.0, 1.0], where 0.0 represents no hate-speech detected and 1.0 means an extreme hate-speech detected.	A Transformer-based library for SocialNLP tasks: https://github.com/pysentimiento/pysentimiento
	HM_S_Hate2		The average of the measure of hatefulness of the recipient's historical messages.		
	HM_S_Hate3		The average of the level in which the recipient's historical messages are targeted to an individual or entity. 0 - not targeting; 1 - targeting an individual.		
	HM_S_HsNum		The average number of types of hate speech detected in the recipient's historical messages.	an integer within [0,3]	
A shared feature between both the sender and the recipient.					

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Topic similarity	HM_SR_TopicSimilarity	Y	The cosine distance between the recipient and the sender's topic preference, based on the LDA model.	a float within the range [0.0, 1.0]	sklearn.metrics.pairwise. cosine_distances
Historical User Activity features (HU)					
Activity	HU_R_TweetNum		The number of the recipient's historical tweet messages.	an integer within [0,50]	Obtained directly from raw data
	HU_R_TweetPercent		Percentage of original tweets/retweets/quotes/replies in the recipient's historical tweet messages.	a float within the range [0.0, 1.0]	# of historical original tweets/retweets/quotes/replies/repostings / # of historical tweet messages
	HU_R_RetweetPercent				
	HU_R_QuotePercent				
	HU_R_ReplyPercent				
	HU_R_InteractiveProb		Percentage of repostings (retweets, quotes, and replies) in the recipient's historical messages.	a float within the range [0.0, 1.0]	
	HU_R_AverageInterval		The average time interval (days) between the posting of recipient's historical tweet messages.	a float >0	(creation time of the latest historical tweet - creation time of the oldest historical tweet) / # of historical tweet messages
	HU_R_RepostLatency		Latency (days) from the creation time of the reposted messages to the creation time of the corresponding repost.	a float >0	creation time of the repost - creation time of the tweet
	HU_R_MentionS		The number of times that recipient mentioned sender in recipient's historical tweet messages.	an integer within [0,50]	Obtained directly from raw data.
	HU_R_MentionSPer		Percentage of the number of messages in which recipient mentioned sender in the number of recipient's historical tweet messages.	a float within the range [0.0, 1.0]	HU_R_MentionS/ # of historical tweet messages
	HU_S_TweetNum		The number of the sender's historical tweet messages.	an integer within [0,50]	Obtained directly from raw data
	HU_S_TweetPercent		Percentage of original tweets/retweets/quotes/replies in the sender's historical tweet messages.	a float within the range [0.0, 1.0]	# of historical original tweets/retweets/quotes/replies / # of historical tweet messages
	HU_S_RetweetPercent				
	HU_S_QuotePercent				
	HU_S_ReplyPercent				
	HU_S_InteractiveProb	Y	Percentage of repostings (retweets, quotes, and replies) in the sender's historical messages.	a float within the range [0.0, 1.0]	/
	HU_S_AverageInterval		The average time interval (days) between the posting of sender's historical tweet messages.	a float >0	(creation time of the latest historical tweet - creation time of the oldest historical tweet) / # of historical tweet messages
	HU_S_MentionR		The number of messages in which the sender mentioned the recipient in the sender's historical tweet messages.	an integer within [0,50]	Obtained directly from raw data.
	HU_S_MentionRPer		Percentage of the number of messages in which sender mentioned recipient in sender's historical tweet messages.	a float within the range [0.0, 1.0]	HU_S_MentionR / # of sender's historical tweet messages

features table

Feature Type & Sub-type	Feature Name	Used in TORS	Description	Value Range	Calculation Method/Reference
Popularity	HU_R_RetweetedRate		Rate of the number of times that the recipient's historical tweet messages were retweeted/quoted/replied/liked (by any user) to the number of the recipient's historical tweet messages.	a float within the range [0.0, 1.0]	sum(the number of retweets/quotes/repiles/likes that historical tweet messages received) / # of historical tweet messages
	HU_R_QuotedRate				
	HU_R_RepliedRate				
	HU_R_LikedRate				
	HU_S_RetweetedRate		Rate of the number of times that the sender's historical tweet messages were retweeted/quoted/replied/liked (by any user) to the number of the sender's historical tweet messages.	a float within the range [0.0, 1.0]	sum(the number of retweets/quotes/repiles/likes that historical tweet messages received) / # of historical tweet messages
	HU_S_QuotedRate				
	HU_S_RepliedRate				
	HU_S_LikedRate				
Some shared features between both the sender and the recipient.					
Activities under topics	HU_SR_TORS1, HU_SR_TORS2, ..., HU_SR_TORS10	Y	A multi-dimensional vector to depict users' Topic-Oriented Relationship Strength under different topics.	a float within the range [0.0, 1.0]	sklearn.metrics.pairwise.cosine_distances: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html
	HU_SR_PathWidth	Y	The cosine distance between the message's topic vector and the Topic-Oriented Relationship Strength between sender and recipient.	a float within the range [0.0, 1.0]	