Lyrics-Based Music Band and Genre Topic Similarity Analysis

Dr. Dimitrios Geromichalos
RiskDataScience GmbH, 82031 Grünwald, Germany
riskdatascience@web.de

Abstract

Based on hundreds of thousands of song lyrics from thousands of bands, Word2Vec models have been trained to quantitatively identify similarities between band texts and terms. Using prominent examples, this demonstrates for the cases studied, that music bands can be assigned to a similarity network solely on the basis of their song lyrics, which also corresponds to their musical style. Furthermore, using exemplary words, it is demonstrated that semantic term networks vary strongly from genre to genre. In addition, the semantic similarity matrices were studied using network analysis methods. As it turned out, term and band text networks differ significantly. While the former resemble random networks, the latter partly exhibit powerlaw behavior. Both also exhibit threshold-dependent regimes.

1. Introduction

The number and variety of music bands has virtually exploded since the advent of Rock n' Roll and Pop, and now represents a vast cultural treasure. While initially the number of bands increased, from the 1970s onwards there has been a splitting up into different genres such as Hard Rock, Metal, Hip-Hop, Electronic and Techno [1]. Especially in the Metal genre, additional sub-genres such as Melodic, Death, Power, Industrial or Thrash Metal with further sub-sub-genres have formed afterwards [2]. Both music styles and target groups now differ considerably [3]. However, no band or genre exists in a vacuum, but is always integrated into a social community, even if it is sometimes more or less independent [3]. In order to analyze such contexts, network theory is a powerful tool for analyzing associations of any kind [4].

The question remains how the links in the network can be defined and determined. For this purpose, the song lyrics of the bands were used in this paper, which are now available in large quantities. The motivation for this is that lyrics ultimately reflect the “world of thought” [5] of the bands or the genre and can thus be used as a basis for similarity analyses. To determine similarities in content, the methods of Natural Language Processing, specifically the Word2Vec model, were used [6]. In simplified terms, this allows terms to be assigned to vectors based on their statistic usage in texts, which reflect their meaning [7]. Ultimately, this allows both term networks and networks of similar bands to be created.

This was done for several well-known bands of different genres. It turned out that the networks formed on the basis of the song lyrics also correspond to the musical styles of the bands. Furthermore, it could be graphically demonstrated that some bands play a rather special role, while others are strongly integrated in a semantic community of very similar bands. Furthermore, the extent to which the “thought worlds” of individual genres differ was investigated. For this purpose, Word2Vec models were explicitly trained for different genres. Based on these models, semantic term networks were then formed for selected terms and these were compared graphically. As has been shown, these networks differ considerably from each other.

For further analysis, the band network and a selected term network were examined using network analysis methods. This provides numerous metrics with which characteristics of the networks can be quantified and also offers explanations for the formation of networks. Since similarity matrices were initially available, these were decomposed into a large number of “similarity slices” and sub-networks were generated and examined from these. As shown, band networks in large similarity ranges exhibit powerlaw behavior, indicating a self-attachment mechanism, i.e., a (partial) orientation towards successful bands. Beyond that, however, there is also an apparently chance-driven area, which could be due to external reasons. The term network under investigation, on the other hand, exhibits a completely different behavior and resembles a random network. Thus, terms apparently do not emerge through self-attachment. However, the random network has a different character with very large similarities than with smaller similarities.

2. Methodology

2.1 Word2Vec

Word2Vec is a method originating from Natural Language Processing, in which a neural network learns to map terms to (e.g. 300dimensional) vectors based on a text corpus. The terms can be single words as well as groups of words (ngrams).

The mapping is done on the basis of the occurrence of the terms in the texts, taking associations into account. Thus, synonyms are assigned to similar vectors. Based on this, it is possible to perform term arithmetic by adding and subtracting vectors, for example. Furthermore, similarities can be determined quantitatively by calculating cosine similarities.

Since texts ultimately consist of ngrams, this method can also be used to calculate cosine similarities of entire texts. [6, 7]

2.2 Network Analysis

Network Science offers a variety of methods to analyze different networks. Here, according to [4], the following key figures were used, which are relatively easy to determine and can be presented illustratively:

- Mean degree, i.e. the mean number of links for each node.
- Degree standard deviation, i.e. the standard deviation of the number of links. This metric is useful in determining whether a random network is present.
- Size of the biggest cluster (i.e. connected area) in terms of nodes.
- Number of clusters (i.e. connected areas).
- Diameter, i.e. the length of the biggest path connecting two nodes.
- Clustering coefficient, a metric that indicates the average interconnectedness of the nodes' neighbors in the total network.
- Degree correlation coefficient, a metric that indicates the average interconnectedness of nodes with same number of links.
- Powerlaw coefficient, a statistical metric that assumes a powerlaw distribution and is estimated by fitting. For theoretical reasons, only values above 2 are meaningful and only values smaller than just above 3 can be reliably determined. If the powerlaw coefficient is below 3, structural effects can lead to negative degree correlation coefficients (see above).

The size and evolution of the key figures are closely related to the topology and formation of the corresponding network. One of the achievements of network science is to provide categorizations and explanations for different types of networks. Two networks that are of particular importance here are the following:

- Random Networks: These networks are created by a random mechanism. They are characterized by a low standard deviation of the node degrees. Moreover, cluster growth starts abruptly after a certain threshold. The cluster coefficient also scales linearly with the average number of node degrees. In reality, such networks hardly occur, but in the present case they seem to describe term networks well.
- Self-Similar Networks: These networks are created by a self-attachment mechanism, i.e. the probability of further links increases with the number of links. These networks have a much larger standard deviation than would be expected in the random case and still exhibit disconnected clusters even with a large link count. Moreover, they have a powerlaw distribution, with exponents that can range from 2 to over 3 (larger exponents can occur, but are hard to detect reliably in practice). Scale-free networks are widespread in reality (e.g., the Internet, social media), and in the present case they also appear to some extent in the band similarities.

In addition, there are hybrid forms and numerous other models that describe the richness of different network topologies.

3. Analysis

3.1 Data Preparation

The main database used was the processed database “Song lyrics from 79 musical genres” available on Kaggle [8] and this was enriched with additional lyrics obtained via Genius [9]. From these only the English songs were selected. In total, 205,105 songs from 3,155 bands/artists remained.

For further processing, the songs were aggregated per band/artist and cleaned from special characters, etc. In addition, stopword were removed and all words were converted to lower characters.

3.2 Model Training

Word2Vec models were trained from the cleaned data based on the most frequently occurring words (1grams) and 2- and 3-word combinations (2grams and 3grams). The 1, 2 and 3grams will be referred to as “terms” hereafter.

In total, the following Word2Vec models were trained:

- A model based on all lyrics containing 44,548 terms, hereinafter referred to as “big model”.
- A model based on Pop lyrics containing 31,880 terms, hereinafter referred to as “Pop model”.
- A model based on Metal lyrics containing 18,309 terms, hereinafter referred to as “Metal model”.
- A model based on Metal lyrics containing 3,021 terms, hereinafter referred to as “small model”.

The model size was adjusted by using only terms with a minimum occurrence of 75 (for the small model) or 10 (for the other models).

3.3 Similarity Slices

Since the models assign terms to 300dimensional vectors, it is possible to calculate quantitative similarities between two terms by means of cosine similarity. Analogously, it is also possible to determine similarities between texts.

Pairwise comparisons were thus used to determine similarity matrices for both terms and bands (“texts”). Since the similarity matrix of the terms also contained negative values, it was scaled so that all its values were between 0 and 1. This was not necessary for the similarity matrix for the texts (i.e. bands). The principal diagonal of the similarity matrices was initially 1, since each term and text is perfectly similar to itself. For further analysis, the main diagonal was set to 0.

In order to investigate the similarity matrices using (binary) network analysis and at the same time consider their richness resulting from the different similarity values, the following procedure was used. First, the similarity matrix was decomposed into a large number of “similarity slices”, i.e. matrices which have the value 1 if the similarity of the original similarity matrix is between a threshold T as lower limit and T + 1% as upper limit, and which have the value 0 otherwise. Thus, for each given similarity as a threshold, there exists a similarity slice matrix that contains only the values 0 and 1. Each of these sub-matrices can be used as a starting point for a binary sub-net.

3.4 Network Analysis

Based on the similarity slice matrices per similarity, binary undirected networks were formed by using a loop. The network science metrics described in 2.2 could then be obtained for each of these sub-networks.

Since each similarity slice corresponds in principle to a different network, neighboring similarity slices (i.e., similarity slices with similar similarities) are alike in network characteristics. This makes it possible to study the dependence of certain variables on others (e.g., the dependence of the clustering coefficient on the number of degrees).

4. Results and Discussion

In the following, similarity networks are first described as examples for individual bands and terms. Subsequently, the regularities underlying the respective networks are examined by means of network analysis.

4.1 Exemplary Similarities

For this purpose, a small number of catchy bands and terms were selected and their local neighborhood was determined with regard to similar bands or terms. This neighborhood was then represented graphically as a network, whereby similarities between the neighbors were also taken into account.

4.1.1 Band Similarities

With the examined bands/artists it is noticeable that they can be assigned to their musical genre without additional information only on the basis of their lyrics! In particular, the following bands/artists were considered (for more information about the individual bands/artists please refer to the corresponding Wikipedia pages):

- Metallica (Figure 1, left): the Metal band is embedded into a net of Metal bands like Judas Priest and Avenged Sevenfold, but also Punk bands like Misfits.
- Michael Jackson (Figure 1, right): the nearest neighbors are Pop artists like Tina Turner, Justin Bieber and Spice Girls. However, it is noticeable that the neighbors are very similar to each other, which speaks for strong similarities in the genre, which, however, do not apply to Michael Jackson.
- Public Enemy (Figure 2, left): the Hip-Hop band is embedded into a net of Hip-Hop bands like Fugees, Nas and Jay-Z.
- Queen (Figure 2, right): here, the connection to classic bands/artists like Chaka Khan and The Jesus and Mary Chain is striking. The connection to Freddie Mercury is also not surprising.
- Slayer (Figure 3, left): the Thrash Metal band is connected to other Thrash/Death metal bands like Sepultura, Testament and Obituary.
- Tool (Figure 3, right): The closest neighbors of this Alternative Metal band are also Metal bands like Pantera, Rage and Avenged Sevenfold. However, the similarities are small, which is due to the originality of Tool.

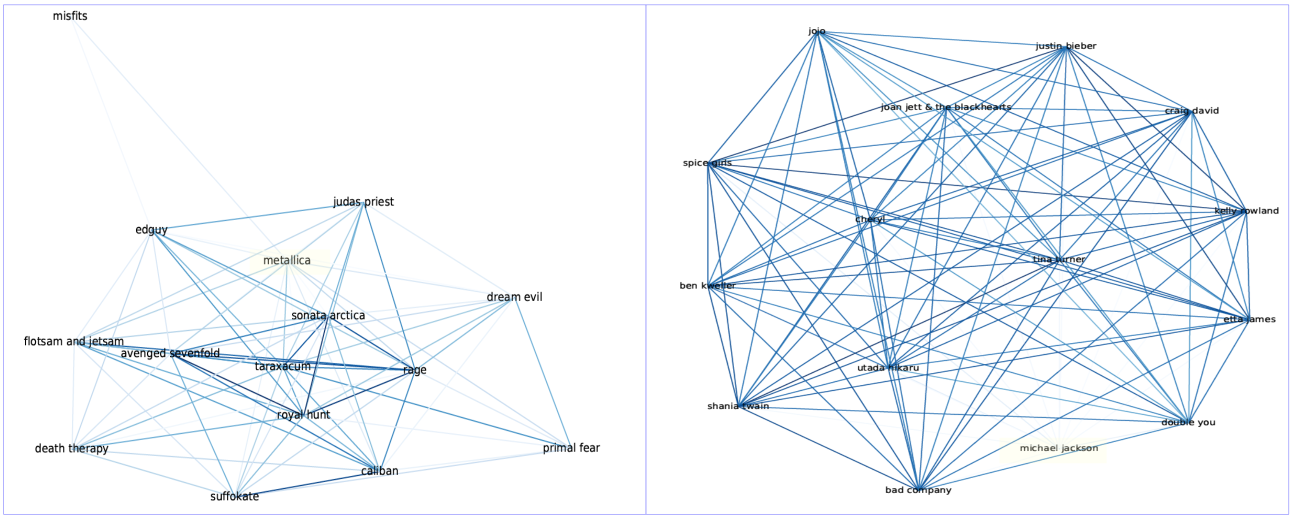

Figure 1: local lyrics-based similarity network for Metallica (left) and Michael Jackson (right)

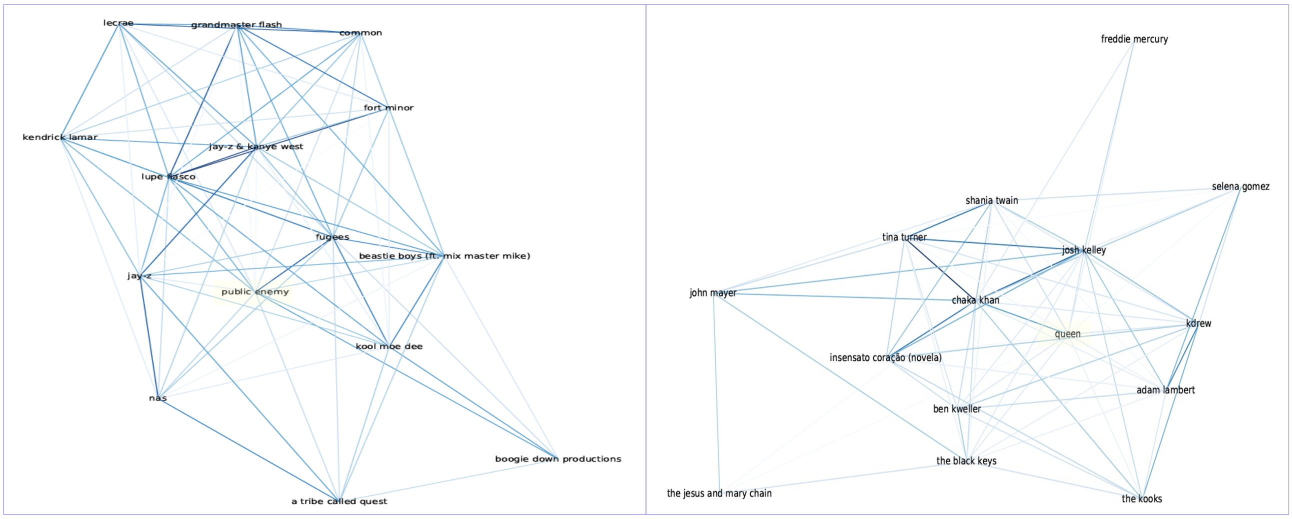

Figure 2: local lyrics-based similarity network for Public Enemy (left) and Queen (right)

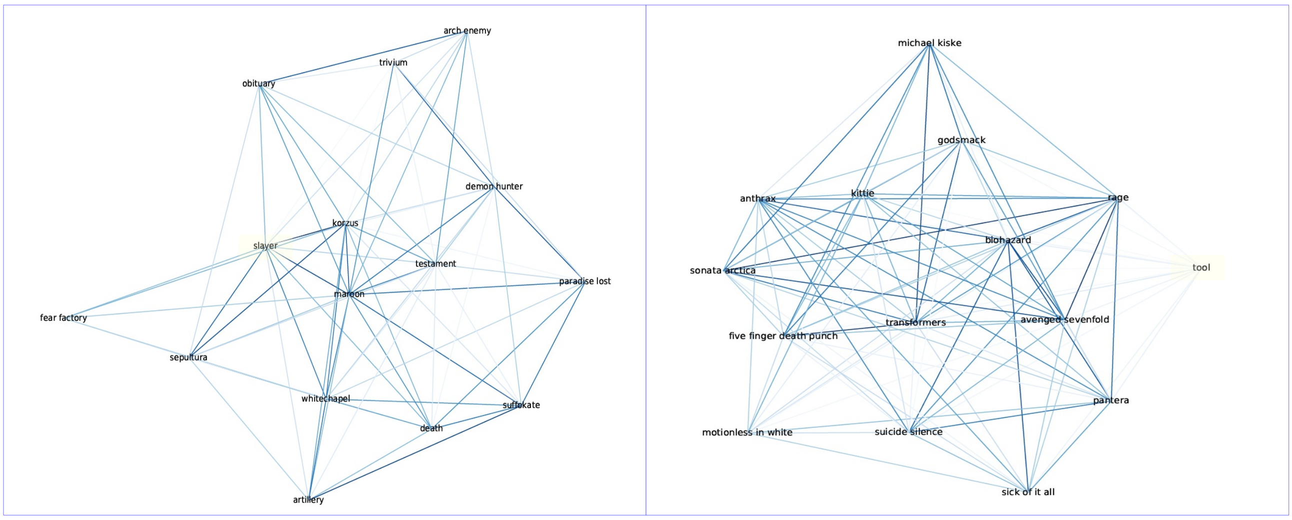

Figure 3: local lyrics-based similarity network for Slayer (left) and Tool (right)

4.1.2 Term Similarities

For the analysis of the term networks, meaningful terms were selected and their local neighborhood of particularly similar terms was determined. In doing so, networks were created and thus the dependencies of the neighbors among each other were also taken into account and graphically displayed. The analyses were always performed with the Word2Vec models “Pop Model” and “Metal Model”, which were trained with Pop and Metal texts, respectively.

In particular, the following terms were examined:

- fight (Figure 4): this term does not have a central role in pop and is mainly associated with commonplace terms such as “right”, “life”, “cause” and “time”, which are much more strongly linked to each other. In metal, however, the concept is quite different: here it is quite central and linked to words like “battle”, “strong”, “win”, but also “die”.
- money (Figure 5): this term also does not play a central role in pop and is associated with commonplace terms and, at most, with terms such as “cash” and “buy”. In Metal, this is further elaborated with terms such as “credit” and “business”; however, money is also associated here with negative terms suggesting violence and corruption.
- sun (Figure 6): “sun” is associated in both Pop and Metal with astronomical and meteorological terms such as “moon”, “stars”, “night” and “wind”. Moreover, in Pop the term is metaphorically

associated with concepts such as “love”, “heart” and “face”. In metal, on the other hand, it is associated with terms like “shadows” and “darkness”, but also with “hope”.

All in all, the fundamentally different orientation of the Pop and Metal genres is already apparent on the basis of a few keywords. While Pop deals with emotional and general topics and is often conceived as radio entertainment, Metal lyrics are often more profound and critical or more negative.

For textual analysis in general, this means that terms must always be considered contextually and are rarely likely to have a general meaning. Therefore, it seems to make sense to consider text corpora separately and to analyze them according to specific associations.

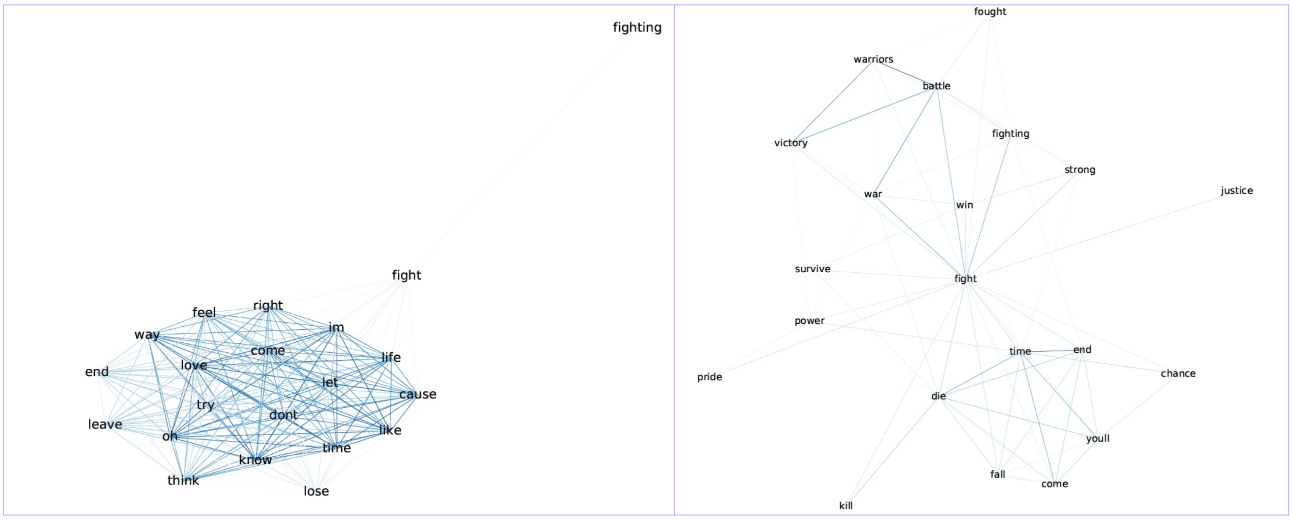

Figure 4: local similarity network for term “fight” in the genres Pop (left) and Metal (right)

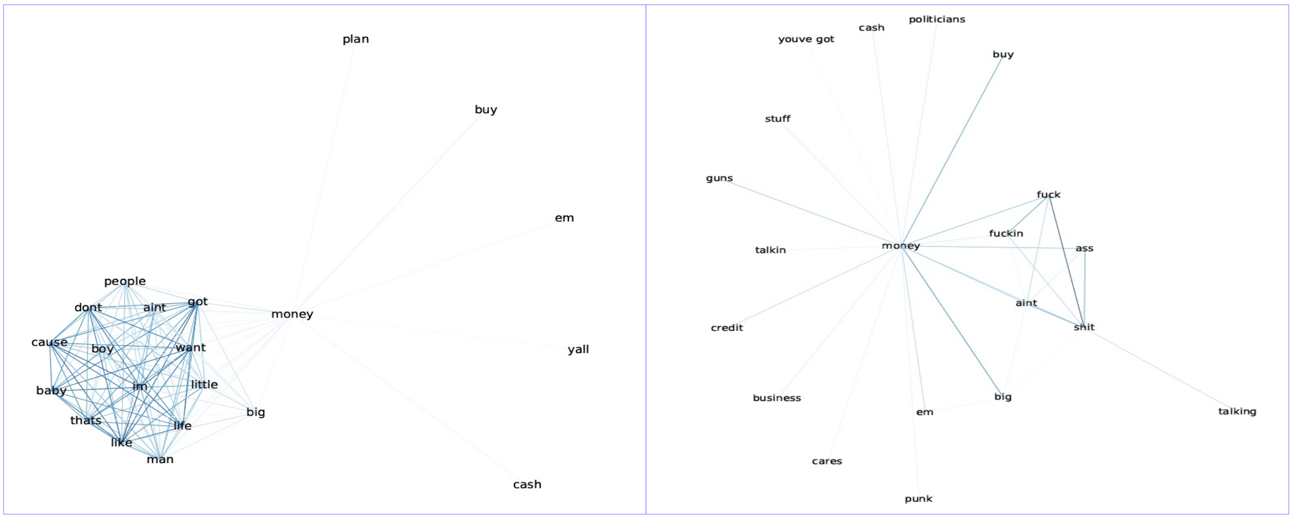

Figure 5: local similarity network for term “money” in the genres Pop (left) and Metal (right)

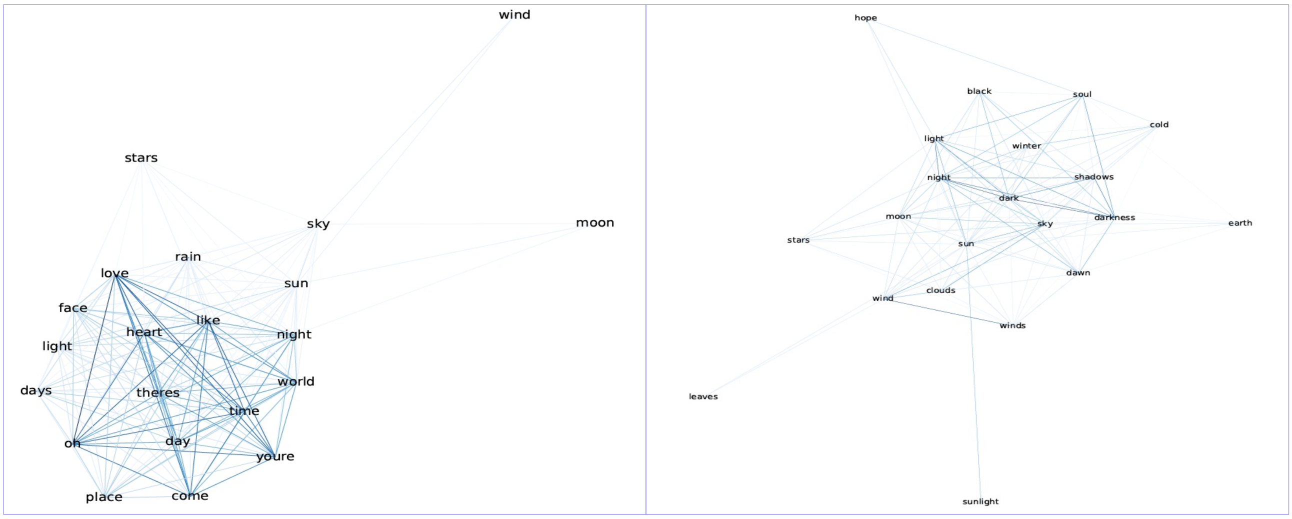

Figure 6: local similarity network for term "sun" in the genres Pop (left) and Metal (right)

4.2 Network Analysis

Network analysis was based on Similarity Slice Networks and was performed for both band similarities (i.e., text similarities) and term similarities of a selected term network. As it turned out, the two networks are fundamentally different from each other.

4.1.1 Band Analysis

Based on the band similarity matrix, similarity slice networks were determined and thus similarity-dependent network metrics were calculated. These show a non-trivial course, which points to several formation mechanisms (Figure 7, Figure 8).

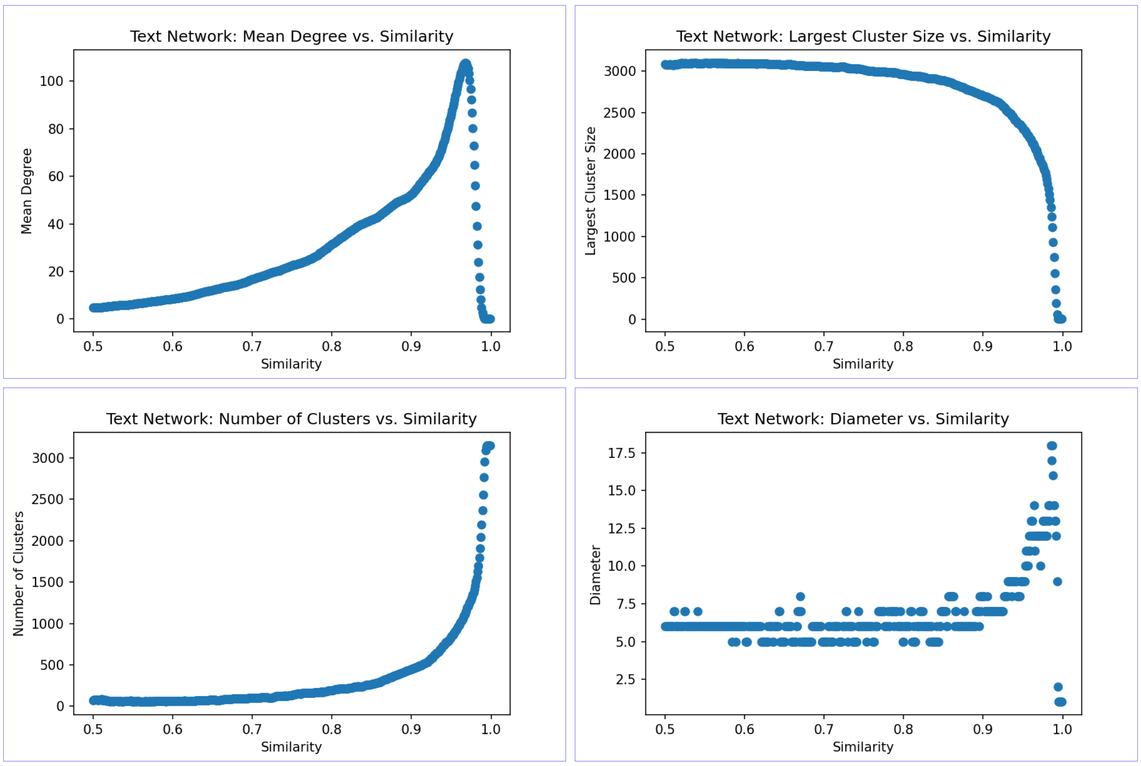

Figure 7: band similarity slice network analysis: mean degree (upper left), largest cluster size (upper right), number of clusters (lower left), diameter (lower right)

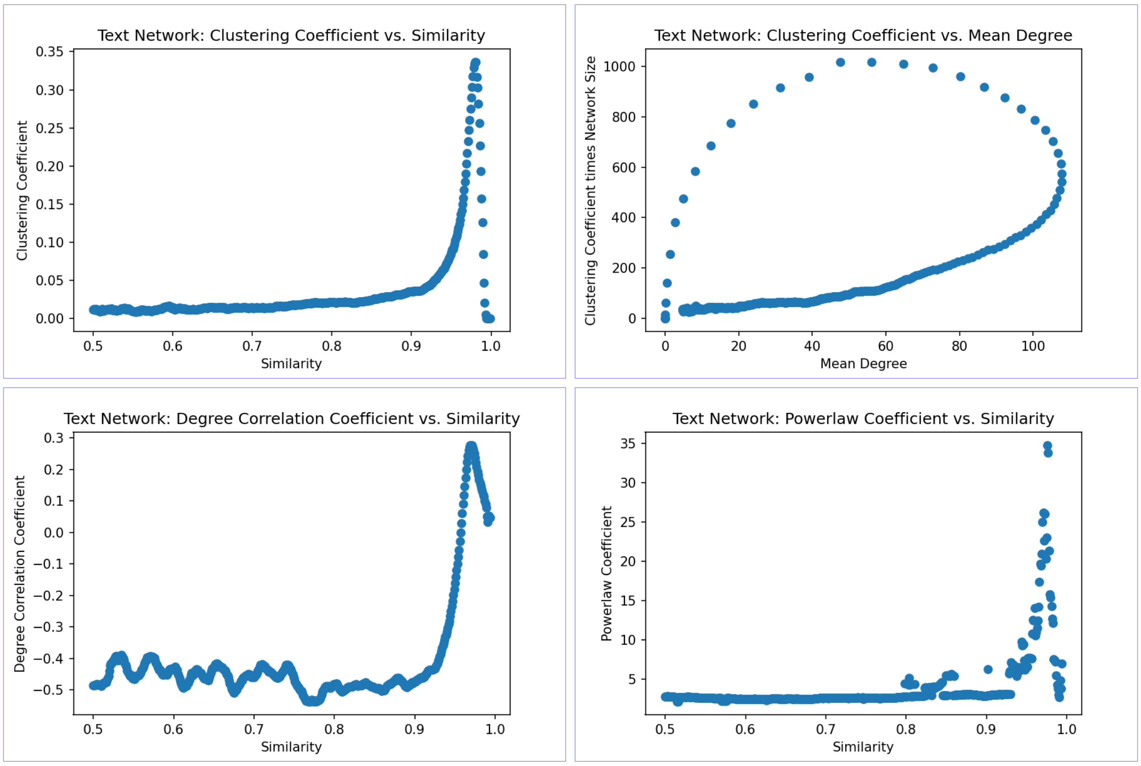

Figure 8: band similarity slice network analysis: clustering coefficient (upper left), clustering coefficient vs. mean degree (upper right), degree correlation coefficient (lower left), powerlaw coefficient (lower right)

This is particularly evident from the fact that networks with relatively low similarities (from 50% to 82%) exhibit powerlaw behavior with a coefficient between 2.5 and 3.0. This even leads to a structural cut-off, i.e. a negative degree correlation coefficient. Thus, bands definitely tend to follow successful examples to a greater or lesser extent in their lyrics.

However, this behavior breaks down at higher similarities, and the powerlaw coefficient leaves the region beyond 3. Nevertheless, a random network does not occur, since the observed variance of the degrees is up to a factor of 10 larger than that expected for random networks. Moreover, the size of the largest cluster decreases despite high mean node degrees. Obviously, in this region of high similarities, the self-attachment mechanism is superimposed by another external mechanism. This mechanism could be due to general social or musical trends. In addition, the style of music could also condition similar lyrics for psychological [10] or musical reasons [11]. However, another reason could be the songwriters behind the scenes, some of which are shared by similar bands [12].

4.1.2 Term Analysis

The reduced Work2Vec model “small model”, trained with metal lyrics, was used as the basis for analysis. The reason for this was that it only contains 3,021 words and could therefore be analyzed without difficulty.

Based on this similarity matrix, similarity slice networks were determined and thus similarity-dependent network metrics were calculated. The progression strongly indicates the presence of random networks (Figure 9, Figure 10).

Thus, the size of the largest cluster decreases rapidly (and the number of clusters increases accordingly) after the mean node degree falls below a critical threshold. With a mean degree of 1.2, about half of all nodes are in the largest cluster. Also, the magnitude of the standard deviation of the node degree is consistently in the range expected for random networks.

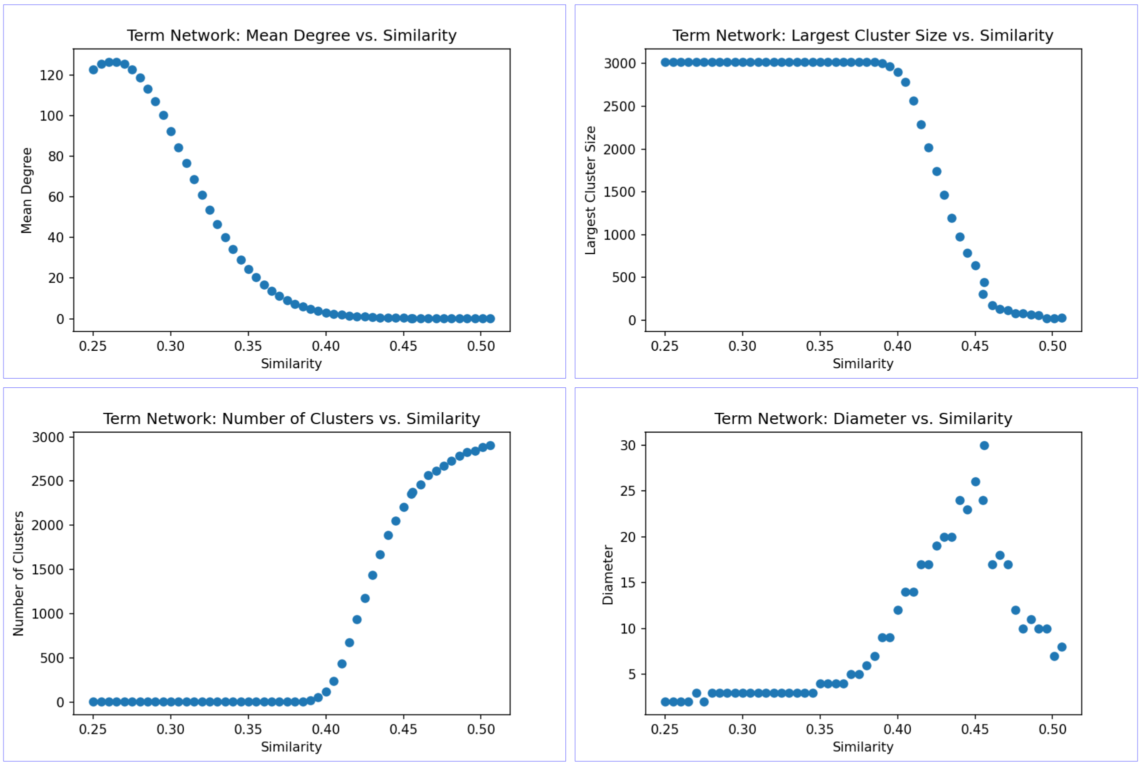

Figure 9: term similarity slice network analysis: mean degree (upper left), largest cluster size (upper right), number of clusters (lower left), diameter (lower right)

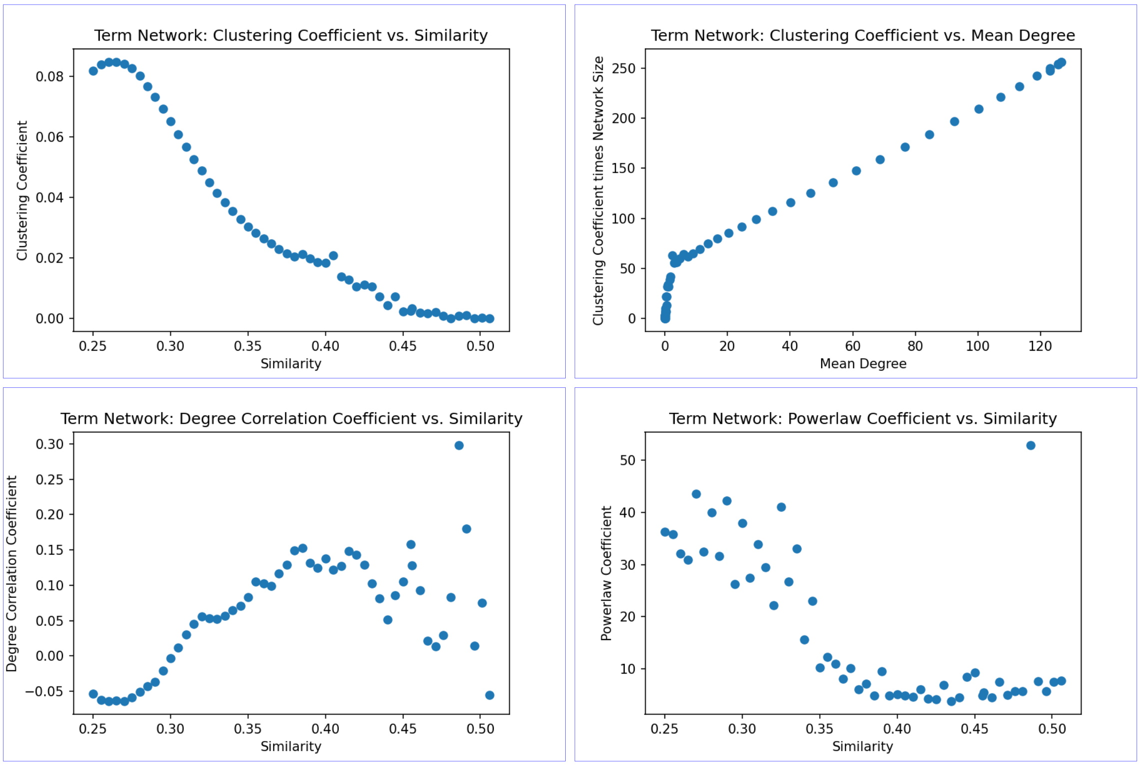

Figure 10: term similarity slice network analysis: clustering coefficient (upper left), clustering coefficient vs. mean degree (upper right), degree correlation coefficient (lower left), powerlaw coefficient (lower right)

However, a difference to a "pure" random network arises if one examines the dependence of the cluster coefficient on the mean node degree. For such a network, a linear course is to be expected, and indeed this is also observed (Figure 10, top right). However, the course shows a kink at a mean degree of 50 (while already at a mean degree of 10 all nodes are located in the largest cluster). It seems that for high similarities the clustering of the neighbors increases with increasing number of links more than expected by pure chance. It almost seems as if "thought patches" of highly interchangeable very similar concepts exist. The occurrence of random networks is highly unusual and also seems to contradict observations regarding semantic networks [13]. However, it should be noted that a purely statistical approach was used here to generate term similarities in a narrow context and the networks are therefore emergent. Overall, then, the issue is how knowledge and associations are represented and in what way this is analyzed by machines [14]. From this point of view, the impression can be gained that terms have no primary properties and merely reflect their underlying facts in an indirect, subjective and accidental way.

5. Summary and Outlook

In summary, Natural Language Processing methods combined with network science are powerful methods for analyzing the similarities of both lyrics and terms.

Using music lyrics, it was shown that similarities can be identified and vividly represented, revealing connections pictorially. Further analysis also revealed that text and term similarities are based on very different mechanisms. While text similarities are rather socially and externally conditioned, psychological and ideological reasons play an important role in term similarities. Furthermore, it was shown that the respective context is of enormous importance when conceptual similarities are analyzed.

The described methods and results open up a multitude of possible applications in a wide variety of fields, such as literature, legal texts, medicine, or news. The context dependency must always be taken into account.

6. References

[1] M. Mauch, R. MacCallum, M. Levy, A. Leroi, "The evolution of popular music: USA 1960-2010", *Royal Society Open Science*, 2. 10.1098/rsos.150081, 2015

[2] B. Hillier, "Considering Genre in Metal Music", *Metal Music Studies*, Volume 6, Number 1, pp. 5-26, 2020

[3] O. Lizardo, "The mutual specification of genres and audiences: Reflective two-mode centralities in person-to-culture data", *Poetics*, Volume 68, 2018, pp. 52-71

[4] A.-L. Barabási, Network Science, Cambridge University Press, 2016

[5] A. Usman, A. Mahmud, A. Daud, S. Dahlan, I. Ternate, "LANGUAGE AS A REFLECTION OF THINKING", *Edukasi*, 19(1):1-12, 2022

[6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, "Distributed Representations of Words and Phrases and their Compositionality", arXiv:1310.4546 [cs.CL], 2013

[7] L. Ma, Y. Zhang, "Using Word2Vec to process big text data", 2015 IEEE International Conference on Big Data (Big Data), pp. 2895-2897, doi: 10.1109/BigData.2015.7364114, 2015

[8] A. Neisse, Song lyrics from 79 musical genres, https://www.kaggle.com/datasets/neisse/scrapped-lyrics-from-6-genres/metadata

[9] https://genius.com/

[10] M. Tizón, F. Gómez, "The Influence of Musical Style in Perceived Emotion", Revista Electronica Complutense de Investigacion Musical, 17. 109-131. 10.5209/reciem.65311, 2020

[11] R. Johnson, D. Huron, L. Collister, "Music and Lyrics Interactions and their Influence on Recognition of Sung Words: An Investigation of Word Frequency, Rhyme, Metric Stress, Vocal Timbre, Melisma, and Repetition Priming", Empirical Musicology Review, 9. 10.18061/emr.v9i1.3729, 2014

[12] S. Qu, , J. Xiao, "The Making of Singer-songwriters: Exploring the Authorship and Ethos of Contemporary Folk Music in Mainland China", Journal of Popular Music Studies, 32. 78-105. 10.1525/jpms.2020.32.1.78, 2020

[13] M. Steyvers, J.B. Tenenbaum, "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth", Cogn Sci. 2005 Jan 2;29(1):41-78. doi: 10.1207/s15516709cog2901_3. PMID: 21702767, 2005

[14] T.K. Landauer, S.T. Dumais, "A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge", Psychological Review, 104, 211-240, 1997