Disciples of the Heinous Path: Social network structure and genre hierarchy in Heavy Metal

Maarten van Hees
Media Technology MSc Program, Leiden University
vanheesmaarten@hotmail.com
August 2018

Thesis supervisors:
Peter van der Putten, Leiden University
Maarten H. Lamers, Leiden University

Abstract

Organization of social networks has been studied extensively viewed through the lens of inter-personal relations. However, the aim of this exploratory study is to find if there are different, domain-specific factors in play influencing the structure of the social network. As a case study, an extensive data-set from *Encyclopaedia Metallum*, a Heavy Metal band archive, was used to test the hypothesis whether or not the social network structure is influenced by a hierarchy of genres. Attempts were made to re-categorize bands according to genre and to counter the fluid nature of the data, after which a Word2Vec analysis and hierarchical clustering was performed, the latter using the number of shared artists between bands as distance metric. A range of histograms where generated to gain insight into the distribution of genres, along with a colored dendrogram to see how genre distribution converges. While the Word2Vec results do show some interesting results with regards to stylistic clustering, the results from the hierarchical clustering remain inconclusive. A number of issues have been identified to address in future research.

Keywords: Social network analysis, Heavy Metal, community detection, hierarchical clustering, genre hierarchy, word2vec

0 Preface

You may be asking yourself why I named this paper the way I did. It's a fair question. It seems to be a pretty over-the-top title. However, if there is something Heavy Metal has taught me, it's that being over-the-top is actually quite important. In fact, it is essential. There is not a more theatrical style of music, be it the funny kind of theatrical or a more sinister variety, not a more ostentatious style of music and not a more extreme style of music than Heavy Metal. It would be remiss of me not to get involved with the theatrical absurdity of it. Metal is among some of most extreme examples of human creative behavior, which makes it very interesting. But apart from that: I think everyone can agree that being a bit *silly* from time to time is just a lot of fun. Metal certainly offers ample opportunity for that, as it tends to *amplify* pretty much every activity, however mundane an activity it may be. You can try this yourself. Next time you are driving to work or doing the dishes, play *Blood and Thunder* by Mastodon and see what happens.

So, what's with the interest in metal? Well, the short answer is this: I really like metal. I'm a *metalhead*. I wear woeful band shirts, buy tapes and enjoy listening to hairy dudes abusing perfectly serviceable musical instruments while some other guy wearing camo-shorts is gurgling Listerine. Or at least, that's what it sounds like. Anyway, the slightly longer answer would be: There are tons of interesting things about metal to wrap your head around. For example, there's all these strange cult-like movements around with semi-religious undertones, which is quite interesting. Also, metal is absolutely everywhere. In fact, there are at least two *brutal death metal* (which is a subgenre of *death metal*, by the way) bands from North Korea, a country not typically known for its tolerance towards alternative creative outlets, or creative outlets in general. Also, Antarctica hosts a band that composes something called (in their words) 'ambient cosmic extreme funeral post-drone metal' which sounds all kinds of pleasant and interesting. This band records their music at the South Pole, which is pretty *metal*, all things considered. Another thing which interests me lies in the fact is that it is a world-wide phenomenon. More particularly, the aspect of metal which is world-wide: the underground side of it. Normally, an underground music scene is something very regional, be it from a city or a province. However, there seems to be something about metal which transcends the typical hallmarks of an underground scene. Perhaps there is something uniquely *human* about metal in a way, as metal music is often about dealing with, channeling or unleashing your inner demons.

The thesis lying in front of you, or displayed in front you, is the culmination of the fruits of my labor during the Media Technology MSc program, at Leiden University. I would like to thank a number of people, starting with all my classmates from 2014's Media Technology vintage. It has been a blast, guys. Also, I would like in particular thank Dr M. H. Lamers for successfully channeling my sometimes unfocused and obstinate self, Dr P.W.H. van der Putten for helping me out with the data mining and analysis issues and Dr M.J. van Duijn for setting it all in motion. It is fair to say I wasn't banking on *this* being the end result when I started this journey. At times it has been infuriating, frustrating and even soul-destroying, but incredibly worthwhile in the end. I hope you will enjoy reading this thesis as much as I enjoyed writing it.

Renkum, July 2018 Maarten van Hees

1 Introduction

Social network analysis has been studied extensively through the lens of inter-personal relations, while few studies focus on possible domain specific reasons why a network has a certain structure. In order to gain a different perspective on the structure of social networks, it was decided conduct an explorative study into these possible effects. As a case study, a large dataset was acquired from the metal band archive website *Encyclopaedia Metallum*, which contains bands, albums, artists that are interlinked where there are shared artists, thus containing a social network of metal bands. This dataset was used to investigate whether or not the genre hierarchy contained within the genre of Heavy Metal has any bearing on the structure of the social network. As far as we know, this is the first time an explorative study such as this one is attempted with this dataset.

The genre of Heavy Metal has always been one rife with controversy, ignominy and notoriety **[12]**. A lot of this notoriety was based around the genre's properties that sets it apart from 'traditional' society. Its detractors would (and often times, will) argue that taking part in the Heavy Metal subculture would lead to depression, nihilism and even suicide, the latter having been studied extensively **[14][18][58]**. However, there is some evidence it is actually country music **[13]** or rap music **[57]** which is more suitable to receive the epithets often hurled at Heavy Metal. While there have certainly been some very strong cases for its detractors, mainly the absolute descent into said depression, nihilism and suicide that characterized the early Norwegian Black Metal scene **[52]**. While this has been the subject of many documentaries **[53][54][55]**, it has evolved into a fairly standard-fair counter-culture and very much went the way of the punk scene. Sure enough, while Heavy Metal's excesses are more extreme than with any other genre, its bands run the full political gamut. Ranging from the radical leftist, social justice warrior side of things to the openly fascist side of things. The pronounced extremism present in Heavy Metal, availability of data and a general interest in data mining problems and Heavy Metal where the main reasons behind the chosen case-study.

The paper is organized as follows. Firstly, an outline of preliminary and related works is given to provide context for this study (section 2), followed by a formal description of the problem statement, the dataset that was used and methodology in section 3, which also describes the experimental setup. Section 4 deals with the results of the experiment, and provides some discussion, followed by section 5, which provides conclusions and future work.

2 Related works

This section will first outline available previous literature focusing in the field of Heavy Metal. Next, an overview of existing literature on collaborative networks arising from music and their surrounding communities will be given. After that an overview will be given about social network analysis in general, after which the focus will shift to community detection and hierarchical clustering, followed by a brief overview of different label classification approaches.

2.1 Heavy Metal subculture

Heavy metal subculture has been the subject in a number of studies in a wide range of fields. Robert L Gross notes in his 1990 paper **[17]** that while the genre of Heavy Metal has always been controversial since its inception in the late 1960s, the debate regarding Heavy Metal subculture reached a fevered pitch in the late 80s and early 90s with events such as the Tipper Gore fronted PMRC's effort to censure Heavy Metal music because of its supposed effect on teenage suicide and depression and the church-burnings by the Norwegians of the early Black Metal scene. Consequently, a number of studies have been done focusing on said suicide and depression as an effect of the Heavy Metal subculture **[18][19][14]** whose findings aren't consensual at all. Other studies focus on behavioral aspects **[20]**, importance of geography within collaborative networks **[21]**, sense of community **[22]**, the usage of masks and shadows **[23][24]**, merchandising and creative practices **[25][26][27]**, philosophy **[28]**, religion in general **[29][31]** and Biblical exegesis in particular **[30]**. The field of social network analysis in conjunction with Heavy Metal has been relatively unexplored, this will be explored in the next sections.

2.2 Social networks in music

There have been a few studies using the type of collaborative networks that arise from music genres. A study done by Pablo M Gleiser en Leon Danon **[1]** focused on the community structure in Jazz. They used a dataset of 128 jazz bands from 1912 to 1940 and found quantitative correlations to an email based social network with regard to typical community degree-distribution. The resulting network also shows correlations between racial segregation, recording location and the community structure.

Teemu Makkonen investigated the Finnish heavy metal scene, exploring the collaboration between music genre heavyweights, evolution of their collaboration networks and whether or not geographical location is important in the development of this collaboration network. **[16]** While earlier literature **[1]** suggests that the presence of a local music scene is vital for the development of a band's success, Makkonen's findings suggest otherwise, namely that successful bands are more likely to communicate via global pipelines, rendering the geographical aspect of a collaboration network as less important than expected.

2.3 Social network analysis

The study done by M Girvan and M.E.J. Newman **[2]** notes that a number of characteristics appear to be shared by social (and biological) networks. For example, social networks tend be scale-free networks, meaning that their degree distribution tend to follow a power law. The network's degree distribution is a measure of the connectedness, i.e. how much connections a vertex has to other vertices, of each vertex and how that connectedness is distributed in the network. Where said degree-distribution follows a power law, meaning that relatively few nodes dominate a certain network **[3]**, are tend to be called scale-free networks. This type of network is opposed by the exponential, or random, network **[4][5]**, where degree distribution follows a Poisson distribution where lambda (the mean occurrence) peaks strongly at the typical connectedness.

Another thing common to social networks is the small-world effect. A small-world network is one where a given node doesn't neighbor most other nodes, but the neighbors of said node are typically neighbors of one another. A well-known family of small-world networks where initially postulated by D Watts and S Strogatz **[6]**. Any of the small-world network's nodes can be reached from any other given node with a small number of steps. This small-world property tends to coalesce with the scale-free property, i.e. the tendency of networks to be dominated by very strong clusters.

2.4 Clustering and community detection methods

Since this prime focus of this study is to find genre hierarchy in data by means of the social network contained within, this section will provide a few notable methods of analyzing a social network. Each method will be viewed in light of the study and where its drawbacks may lie. The community detection methods are included since they offer a way to split up a social network in manageable chunks. The assumption here is that these communities will share fairly similar genres, or at least have a genre distribution that leans heavily to one or two genres.

2.4.1 Hierarchical clustering

One of the techniques typically used with networks that contain clusters of densely connected communities, but where the clusters themselves aren't as densely connected, is the hierarchical clustering method **[8]**. Hierarchical clustering is a method whereby for each vertex pair o,p a weight w is calculated. This weight can be calculated in any number of ways. One way to do it is to count the number of vertexes that stand between o and p on its shortest path and take the higher amount of vertices as a higher weight. After calculating said weights, one takes all vertices and start adding edges between them according to the weights of the pairs. As edges are added, clusters of vertices will appear. These are taken as communities. Its advantages are that it is an explicit procedure, has a clear interpretation and has many public implementations available. There are a number of drawbacks with this method however. There are many kinds of weighting criteria, it can be very difficult to find the correct weighting criteria for the network analyzed. Also the amount of subsets used in the clusters is arbitrary, it can be difficult to get the correct resolution for the network used. The biggest drawback however is in the way the algorithm operates. Once a vertex has been assigned to a grouping, that grouping can't be undone at a later stage. This has a tendency to isolate peripheral vertices that have a low connectedness but still should rightly belong to some community, it can lead to false positives and poorly defined community clusters **[2]**.

Hierarchical clustering methods come in 2 varieties: top-down, also referred to as *divisive,* and bottom-up, also referred to as *agglomerative*.

2.4.2 Edge betweenness method

One way of dealing with the disadvantages of the hierarchical clustering method was proposed by Girvan and Newman **[2]**. Instead of focusing on the edges that are most central to communities (i.e. most 'well-traveled' edges) Girvan's and Newman's method focusses on the opposite, the edges that are most 'between' communities. Simply put, it works by progressively removing the least central edges, thus isolating dense clusters with high degree distribution, centered on vertices with a high degree of connectedness. This vertex betweenness was originally proposed by Freeman **[9]** as a measure of the influence a vertex has in these clusters. This is measured by counting the amount of paths pass through a certain vertex. The Girvan-Newman algorithm is a divisive (top-down) hierarchical clustering method.

2.4.3 Louvain method

The Louvain method **[10]** is aimed at optimizing modularity, a measure of the density of links between vertices inside communities as opposed to the typical density of links between vertices. The method describes a greedy optimization algorithm that takes a bottom-up approach. Firstly, small, local communities are identified by optimizing modularity locally. These small, local communities are then aggregated into a new network in which the local communities are single nodes. These 2 steps are run until modularity is maximally optimized, producing a hierarchy of communities. The algorithm is very fast and can therefore be used on large networks. It has been combined with other types of modularity optimization, such as leader detection. **[15]** There are a number of drawbacks, not just with the Louvain method, but with modularity optimization in general, as Good, de Montjoye and Clauset outline **[11]**. First of all, it is not certain that the found optimal partitions coincide with the most intuitive partition. This is referred to as the resolution-limit problem **[12]** and is driven by the assumption of methods like these that the network analyzed is in fact a type of random network, which typically has a different degree distribution then a typical social network.

2.5 Label classification

In conjunction with the social network approach, different, more traditional approaches exists to find structure in labels. All of the techniques considered revolve around *word embedding*, which are a number of techniques aimed at mapping words to vectors, making them easier to work with in a machine learning environment. This section will give a brief overview of two notable ones.

2.5.1 LSA

LSA stands for latent semantic analysis refers to a technique to extract *concepts* from semantic text **[63]**. This works by analyzing a corpus of documents for relationships between the terms contained within these documents. This process maps out what concept is related to whichever other concepts. Using a technique called SVD (singlevalue decomposition), which is a mathematical process by which the number of rows (dimensions) in the final matrix is reduced while retaining the overall structure and variation, one can lower the dimensionality of the dataset. In this technique, concepts relate to each other by means of a vector distance. Because this technique maps words to vectors, one can use distance measures such as a cosine distance.

2.5.2 Word2Vec

Word2Vec **[59]**, created by a team of researchers led by Thomas Mikolov, is another technique for word embedding. Using a two layer neural network, Word2Vec maps context of words using a corpus of sentences. The result is a vector space containing vectors representing words. The vectors that are placed closer to each other are more similar, i.e. will appear together more regularly than vectors placed further apart. This style of notation is perfect for visualization, as one can simply use a t-SNE **[60]** style plot to visualize similarity within this vector space. This is much like the LSA technique described earlier. The difference with LSA lies in how similarity is defined. With LSA this is done using counting, where similar concepts across documents have a similar count. Word2Vec *predicts* similarity by means of a neural network. This means that Word2Vec lends itself more for non-semantic content, where someone just wants to map out similarity without worrying too much about semantics

3 Problem statement and methodology

This section will first provide a description of the problem statement and general approach, followed by the used data source, method of data collection, interpretation and resulting data.

3.1 Problem statement

In this study we have chosen to taken an exploratory approach **[32]**. There is no fixed question as of yet, we are simply curious what we may encounter with the dataset which will be described shortly. However, there is one matter that will be explored explicitly, which is whether or not the social network contained within Heavy Metal is organized by its genre. Two things will be attempted in order to gain more insight into possible answers. Firstly, a Word2Vec analysis will be performed using a t-SNE plot in which the similarity across genres can be inspected. Then a hierarchical clustering method will be applied to the band collaboration graph, disregarding the node's genres, after which a genre hierarchy will be compared to the resulting social network's clustering. More on methodology later in this section.

3.2 Encyclopedia Metallum

The dataset used in this study was obtained from *Encyclopedia Metallum* **[7]**, a website devoted to cataloguing Heavy Metal bands, albums and individual artists. EM's content is entirely generated by its community, which outlie a number of challenges. EM contains, at time of writing, 121,856 bands with associated artists, album releases, album reviews, song lyrics and more. EM has been around since 2002 and has a very active user base **[33]** and has been used sparingly in previous studies, such as a study done by Ingeborg H Aarsand on the usefulness of representing natural utopias in music as a response to environmental crises **[34]**, the influence of Heavy Metal in times of social instability **[35]** and a study on collaborative networks of the Finnish metal scene **[16]**.

3.3 Data Collection

The data contained in *Encyclopedia Metallum* is structured alphabetically, regardless of any differentiating property. For technical reasons such as a lack of bandwidth and processing power and time constraints, it was decided to include 50% of the bands within each lettered list, going from A to Z and ending with a special section containing bands with names written in non-Latin characters. Using a custom-built website scraper the data was obtained over the course of several months, interpreted and saved to a relational database for future use.

3.4 Resulting data

The collected data was saved into a relational database. Figure 1 shows the general structure of the database.

Figure 1: structure of the resulting relational database. Main entites are colored orange, seconday entities green, link tables blue.

As is visible in figure 1, the dataset has been structured in a way that closely mimics *Encyclopedia Metallum*'s format. Roughly speaking, 3 main entities are identified: Band, Album and Artist. A Band can contain Artist via two routes, past member or current member. Hence the two link tables that stand between Band and Artist. Album contains a single type of link to Artist. This is because auxiliary staff for an Album are not necessarily part of the Band, such as album cover artists, recording engineers and so on, but these entities are listed as Artist entities. This is where the Role entity comes into play. The Role entity serves to handle the possibility that an Artist can have different roles in different bands. For example, Artist Z can be drummer of Band A and the vocalist for Band B, but may also have mastered some Album for Band C. Finally, the Song entity contains information about an album song and, if available, song lyrics. The RelatedLink entities contain whatever links for Artists and Bands have been provided, such as links to webpages, social media and so on.

As previously stated, 50% of the EM's total dataset has been obtained, for reasons also previously stated. The total amount of obtained entities are shown in Table 1.

Entity name	Entity count
Band	63,306
Artist	288,872
Album	178,355
Song	1,273,940

Table 1: Amount of the main entities in the current dataset. Song was included because of its standalone nature. Role and the other link tables are therefore not included.

3.4.1 Genre family tree

In order for the one explicit question to be answered, a genre family tree had to be obtained. A genre family tree is a hierarchical structure that sheds some insight into what 'main' genre leads to what 'subgenre'. For example, death metal comes from trash metal, trash metal comes from heavy metal and so on. While fierce discussion about the structure of this family tree remains **[56]**, the hierarchy with regards to main genres, such as the examples listed before, of Metal are widely accepted.

There are a number of such genre family trees available online, but for reasons of availability of data and ease of access, the genre family tree by *Bound By Metal* **[36]** was chosen. This family tree is very extensive and even lists non-metal genres that have influenced the development of certain metal subgenres. Such as *Techno*'s strong influence in the *Neue Deutsche Härte* movement, regional traditional *Folk* music in the development of *Folk Metal*, and *Blues* in the development of *Doom Metal*. Obtaining the data from this genre family tree was done using browser development tools, UltraEdit and a program that parsed the data into a usable hierarchy, which is shown in appendix 1**.**

The parsed genre hierarchy doesn't contain non-metal genres. The exclusion of the non-metal genres was done partly manually, partly computational. The computational part involved the dataset from *Encyclopaedia Metallum*, where the genre hierarchy's tags where compared to the genres present in the dataset. Those that weren't present in both where excluded. In order to account for ambiguous naming a manual rewrite was performed. For example, *raw* black metal in the dataset and *war* metal in the genre hierarchy are the same, albeit obscure, subgenre. The decision to rewrite genres was partly done by personal reckoning and partly done after performing Google searches and reading up on forum topics on the matter. As stated before, discussion about the structure of the genre hierarchy and nature of genres, names, bands and so on are widely spread. Thus it became matter of reading up on these discussions and gauging the consensus.

3.4.2 Normalizing the dataset by genre

Some genres in this family tree are not present in the data from *Enclyclopedia Metallum*, such as bands from the aforementioned *Neue Deutsche Härte* movement. In order to make the genre hierarchy comply more closely with the dataset, a kind of normalization had to take place.

First of all, due to the way *Encyclopaedia Metallum* was set up, anyone can type out the band's genre when submitting a band. Which leads to some pretty creative genres and makes certain bands extremely hard to define by a 'main' genre, as shown in Table 2.

Band	**Genre according to Encyclopaedia Metallum**
Rotting Christ	Grindcore, Black Metal (early), Gothic Metal (mid), Melodic Black Metal (later)
In Flames	Melodic Death Metal (early), Melodic Groove Metal/Alternative Rock/Metalcore (later)
Meshuggah	Technical Groove/Trash Metal (early), Djent (later)
Darkthrone	Death Metal (early), Black Metal (mid), Black/Heavy/Speed Metal (later)

Table 2: some examples of creative genre labeling.

This, of course, is not illogical at all. Bands are rarely pinned down by a single genre and will often times try something else when composing music. All this data is submitted by volunteers, who may or may not have a different reading to a certain genre than the next person. However, this makes is extremely hard to parse a band according to its genre. Excluding these bands from the final dataset has its risks; this might skew the final dataset, as the bands that have these very detailed genre definitions tend to be the bigger and more influential bands, as is the case with the four bands listed above.

In order to make a bit of sense of band genres such as the ones listed above a normalization method was developed. After having considered several string similarity algorithm (e.g. Levenshtein distance **[46]**, Sørensen–Dice coefficient **[47]** and cosine distance, with genres represented as vectors), it was decided to create a custom method based on string exclusion. For each band the genre was stripped of all 'non-genre' terms, meaning things like '(early)' , '(later)', slashes, backslashes, comma's, and finally, the word 'metal'. This creates a list of terms which can be used to 'score' a band according to its supposed genre(s). So in the case of the Greek band *Rotting Christ* this would yield the terms *Grindcore*, *Black*, *Gothic* and *Melodic Black*, which means it scores 0.25 on all terms, meaning the maximum score of a genre is 1.0. These scores can then be used to place the band (in the Euclidian sense) somewhere on the genre hierarchy. Lastly, a distinction was made between main genres and subgenres by counting the amount of terms in the isolated genre and ranking them according to that. This ensured that subgenres like *Melodic Black*, *Atmospheric Black* and *Depressive Black* all followed the main genre of *Black.* This method yielded an ordered list of genres.

Figure 2: Ordered list of genres, abbreviated.

Because bands score on 1 or more of these genres, some overlap and ambiguity is still present. Bands that have a genre of *Black/Death* would score 0.5 on *Black* and 0.5 on *Death* while, arguably, the band's genre should be *Blackened Death Metal*. Then again, this particular subgenre is distinct enough to be in its own category, but doesn't take into account a band's evolution from one genre to the next, as could be the case with bands that have the aforementioned *Black/Death* genre.

3.5 Methodology

The first target of this study was to find out whether or not the domain specific genre hierarchy is present in the data itself. In order to gain a better understanding of the data, a word embedding technique was performed. Next, a more social-network oriented approach was taken to determine whether or not the structure of the social network follows the genre-hierarchy. This section will mainly deal with the methodology concerning these questions. Also a couple of other considerations regarding different questions that have arisen from this study will be given.

3.5.1 Word embedding with Word2Vec

As stated before, the search for the genre hierarchy in the data has two approach vectors. Firstly, by using Word2Vec, just by looking at co-occurrence of band genre we might be able to verify or falsify the notion that bands, regardless of the social network, *mix* genres in a particular way: namely by means of this genre hierarchy. If this is the case, it might be an argument that the genre-hierarchy is a different kind of product from the data, not something that arises from the social network contained in the data. Word2Vec is part of the Pyhton3 package *Gensim* **[61]**. This package allows for easy implementation and result inspection. The t-SNE plot was generated with a *scikit-learn* **[62]** module. The reason why we went with Word2Vec is two-fold. While LSA focuses more on presenting important concepts within the corpus of documents, Word2Vec allows for predicting per term which terms are closely related to each term. This makes visualization easier and facilitates in building a map of related terms, one of the targets of this study. The second reason lies in the data. Since LSA works with concepts, its target data is in fact semantic text. The data we use for the word embedding isn't semantic text, but a list of genres, so having a method that maps similarity by co-occurrence seems much more logical.

3.5.2 Giant Component

To facilitate the completion of the hierarchical clustering method it was decided to use the biggest component in the social network, the so-called *giant component* **[42]**. Nodes in the giant component (or in any component with $N > 1$) have a degree of $D > 1$, meaning all nodes within the component are connected and can be reached from whichever node in the component. Since the distance metric, which will be described shortly, works with shared artists (the edge between the bands), using the giant component for the clustering method has the added benefit of making the termination criteria of the clustering method fairly straightforward $C = 1$, where C is the list of clusters the hierarchical clustering method works on. The giant component was obtained using a bog-standard BFS algorithm implemented using a FIFO queue **[44]**, which yielded a list of Band entities which where all connected which each other one way or another via shared artists. This list of Band entities was then compiled into a same-size list of clusters where all Band entities have their own cluster.

3.5.3 Floyd-Warshall algorithm

Using the cluster C obtained from the BFS algorithm, the initial distances had to be calculated before clustering could take place. This was done using an implementation of the Floyd-Warshall **[38]** algorithm, which conveniently yields a table with the shortest path length between each pair of nodes in the provided dataset. Because this algorithm runs in $\theta(V^3)$ and is therefore very time intensive, it was parallelized after considering findings from a comparative study focusing on the parallelization of the Floyd-Warshall algorithm **[37]**. This yielded a performance increase of roughly 250% on the main test computer, which is equipped with a run-of-the-mill quad-core CPU. There are Floyd-Warshall implementations using the GPU **[39][40]**, which can provide more powerful parallelized options, but because of the added complexity and time constraints this wasn't pursued further.

3.5.4 Distance metric and clustering method

The hierarchical clustering method needs a distance metric to determine which cluster is closest to which cluster **[41]**. Normally, with a graph that exists in Euclidian space, one could simply use the Euclidian distance between the clusters as a measure of distance, i.e. utilizing a 'centroid' that determines that Euclidian center of a cluster. However, this social network doesn't exist in Euclidian space, thus the non-Euclidian centroid, called a 'clustroid' was used. A clustroid is simply the most 'central' node within a cluster, given a certain criteria. In the case of this study, the most central node was defined as the node that has the lowest average distance to all the other nodes in the cluster, which would, intuitively, be at the center of a cluster of nodes.

The distance between clusters was based on edge weight, which in turn is defined by shared artists. If Band A and Band B have an artist in common, there is an edge between those 2 bands. The edge-weight was determined by the number of shared artists, i.e. 4 shared artists would mean a lower edge weight than as would be with 1 shared artist. Also, edge-weight was influenced by the status of the artist, if the artist was still part of the band, it would yield a lower edge-weight than if the artist is no longer part of the band.

The clustering method works by first assigning each band to its own cluster, resulting in a list of clusters C. Then, for each cluster C^i the closest cluster C^j was obtained using the distance table yielded by the Floyd-Warshall algorithm and the distance metric described above. After iterating through C a pair of clusters with the shortest distance

C^i and C^j where aggregated into a new cluster C^{i+j} which was then added to list C. Finally clusters C^i and C^j where removed from C. This process continues until there is but 1 cluster remaining in list C. After the clustering method runs to completion, for each cluster a histogram was generated to show the distribution of genres within the cluster.

4 Results and discussion

This section will provide some results obtained from the Word2Vec analysis and the hierarchical clustering. It will also provide some insights into the data, and the social network contained within, using Gephi.

4.1 Word embedding

The Word2Vec method takes in a list of documents from which the co-occurrence vector space is constructed. This list of documents is a band-wise list of genres. In order to circumvent the sometimes messy annotation and general muddiness of the user-generated data, the data was put through the normalization method described in section 3.4.2. This yielded a list of 28007 band-wise genre co-occurrence documents as illustrated below.

```
"Black, Sludge, Doom"
"Stoner, Doom"
"Black, Death"
"Black, Brutal Death, Grindcore"
"Melodic Black, Death"
"Heavy, Power"
```

Figure 2: First six band-wise genre co-occurrence documents

Using this kind of data with the Word2Vec algorithm isolates each term in each document. For example, the first document contains three terms: *Black, Sludge* and *Doom*. Running this through the neural network links these two terms together, i.e. it will generate vector representations which will lie close to each other. If this particular combination shows up later in the corpus, the vectors will get closer and closer, strengthening the similarity.

The Word2Vec model was implemented using the Python *gensim* package, ignoring all documents with less than 75 occurrences. After experimenting with the number of minimum occurrences, 75 turned out to give a reasonable balance in resolution and provided a readable t-SNE plot which is displayed below. The clusters, displayed as colored circles, have been added manually while inspecting the plot.

Figure 3: t-SNE plot of Word2Vec results

There's a couple of interesting things to note in this plot. Broadly speaking, the plot clusters heavily on style of genre. The blue cluster contains all the *ambient* genres, which is a slow, melodic style of metal. The only two outliers in this genre are *Funeral Doom* and *Depressive Black*, but these should arguably be part of the *ambient* genre group as these two genres focus heavily on atmospherics and have a similar style in lyrics. The second cluster worth noting is the red cluster. These are, roughly speaking, the extreme styles of *Death metal*. *Goregrind*, *Grindcore* and *Brutal Death Metal* are often used interchangeably as they are fairly similar genres. *Deathcore* is very much a spin-off and, arguably, a simplification of the *Brutal Death Metal* genre. The green cluster features all of the genres emerging from the fusion of *Punk* and *Trash*, of which the genre *Crossover* is the most pure mixture of *Punk* and *Trash*. One genre that, at first glance, shouldn't be part of this cluster is the *RAC* genre, which stands for *Red-Anarchist* and is a subgenre of *Black Metal*, which doesn't share much similiarity with the typical p*unk* and *trash* genres. However, a case can be made for its inclusion, as *RAC* is very leftist politically oriented, very much like the wider *punk* movement. The final two clusters are fairly notable in much the same way. The purple cluster are the *pagan* styles of metal, which focus heavily on folklore, mythology and traditional folk music elements. The yellow cluster are the *progressive* styles of metal, generally speaking. This is probably the most nebulous cluster, although it is pretty interesting that *shred, neo-classical* and *technical* are clustered close together along with *progressive rock* and *symphonic*. This makes sense, these are all off-shoots from the initial Heavy Metal style of music, incorporating more instruments (*symphonic*), experimenting with song structure (*progressive*) and extreme technicality (*shred*, *technical* and *neo-classical*).

As stated before, this analysis simply looks at genre co-occurrence, band by band. The clusters found in the plot do point to a trend in the data. Bands tends to not much mix their genres around much and when they do, it tends to be variations on a certain theme, such as the political themes of the *punk* genres, the explicit violence and gore of the extreme *Death Metal* genres and the heavy focus on ambience and depressive lyrics of the *ambient* genres. This does seem to point towards an organization in genre choice. However, apart from the few notable clusters, there isn't much linking these clusters together, but there is certainly an organization to it.

4.2 Social network

The hierarchical clustering method was ran with 3 subsets of the full dataset, all bands from Canada, Belgium and Colombia. As stated before, the analysis was run on the giant component of each country's band collaboration graph. The reason for using these 3 countries was for their limited network size and the difference in locality which may or may not have bearing on the properties of the social network. Table 3 shows the size of C for each country. First, a couple of experiments were conducted with Gephi. These are very explorative in nature and serve to make the data more clear and provide a number of new possible research avenues. Section 4.2 deals with the actual question as stated in section 3.

Country	**Size of *C***
Belgium	362
Canada	891
Colombia	462

Table 3: Sizes of C for experimental datasets.

4.2.1 Gephi

To gain some insight into the type of network contained within the dataset, how the network looks and to verify the dataset, a few experiments using Gephi **[43]** where conducted on the previously described experimental datasets.

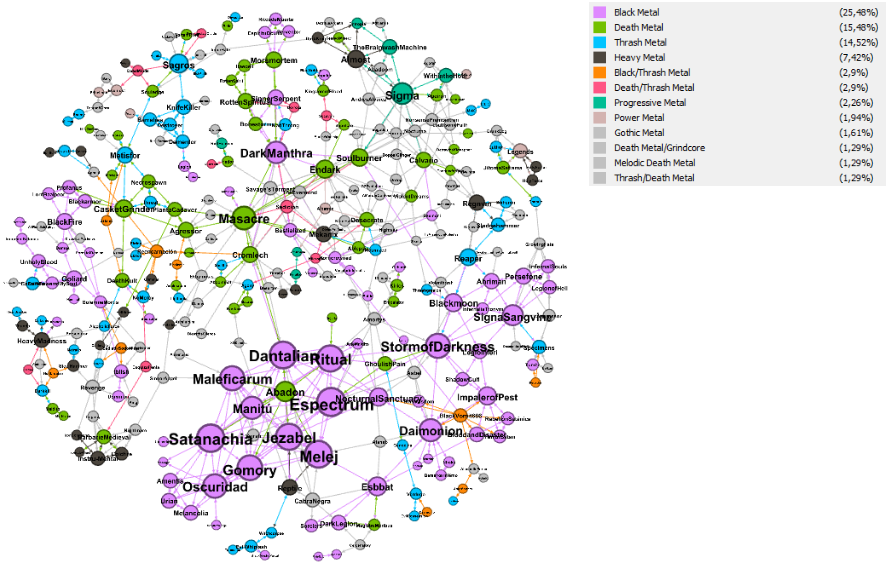


Figure 4: All Colombian bands, colored by genre

The above image shows the social network of all Colombian bands. One thing that immediately stands out is the *black metal* cluster of bands in the lower part of the image. There is a strong *black metal* tradition in Colombia dating back the early 80s which, arguably, was influential in the development of the revered Norwegian scene **[47].** This scene was based at the city of Medellin, which can be seen (at least partly) when the graph is colored with the band location:

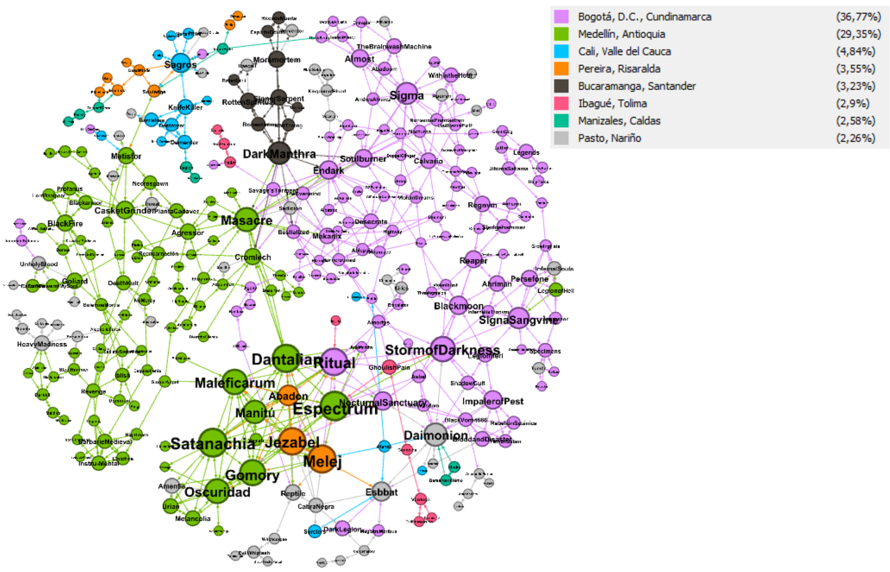


Figure 5: All Colombian bands, colored by location

The bands that are part of the isolated black metal subscene are and have been part of the Medellin "ultra-metal" movement. Interestingly enough, there is a clear split between the 2 most represented localities, Medellin and Bogota. This could be because of the strong scene based around Medellin, and subsequent scenes evolving around that city. Also, since the Medellin scene is relatively unknown outside of Colombia **[47]**, it could be that heavy metal hasn't taken root outside of the major cities or that local communication lines are more important than regional or provincial communication lines, this would oppose the situation in Finland, as Makonnen observed **[16]**.

The Canadian network has a slightly different look. It doesn't have the insular community that Colombia has, but a number of interconnected communities that center around high-degree nodes and the major cities.

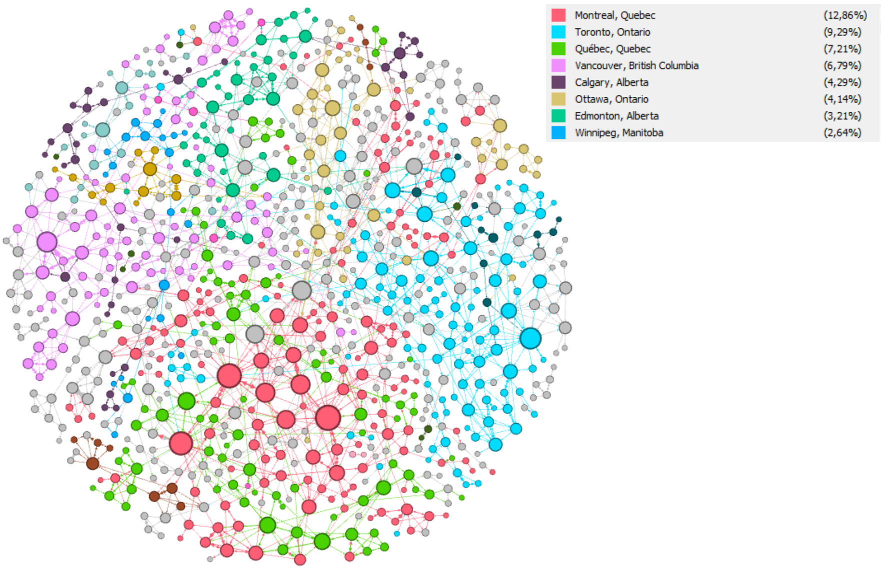


Figure 6: Canadian bands, colored by location

These communities become visible when applying the Louvain method (resolution of 3.0) and coloring according to modularity class.

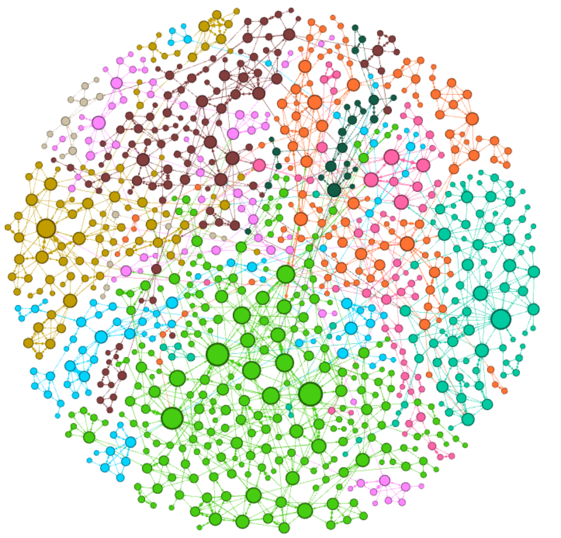


Figure 7: Canadian bands, Louvain method applied.

Interestingly enough, the Montreal and Quebec bands are put into the same community as they are fairly close together and share a language, while the bands from Vancouver and Toronto are in their own communities, being at pretty much opposite ends of the country. This again may oppose Makonnen's findings, although his dataset didn't take into account regional language differences, it might be interesting to investigate the effect language has on the development of these kinds of social networks.

The Colombian Medellin scene has clearly influenced the structure of the network. This Medellin scene was also one that didn't really pass the Colombian borders; it was very much a local scene. This could prove to be the way to find the more extreme, and more ostracized, parts of the metal community. In particular the National-Socialist and Red Anarchist (which combines far-left and environmental philosophies with *black metal*) scenes. Especially the former is a small minority in the wider black metal movement **[49][51]**, even though C Dornbusch and H.P. Killguss' *Unheilige Allianzen* **[48]** suggest that neo-nazism and ethnic paganism is a growing trend in black metal. On the other extreme end of the spectrum, Red Anarchist black metal remains an understudied subject. Especially the Cascadian scene has many bands that subscribe to this ideology **[50]**, and it would be interesting to investigate with regards to the wider Cascadian independence movement, which also revolves heavily around ecological matters.

4.2.2 Hierarchical clustering

The hierarchical clustering method yielded a number of histograms which were used to view the convergence of genre within the network. Below are the histograms showing the genre distribution of the experimental datasets.

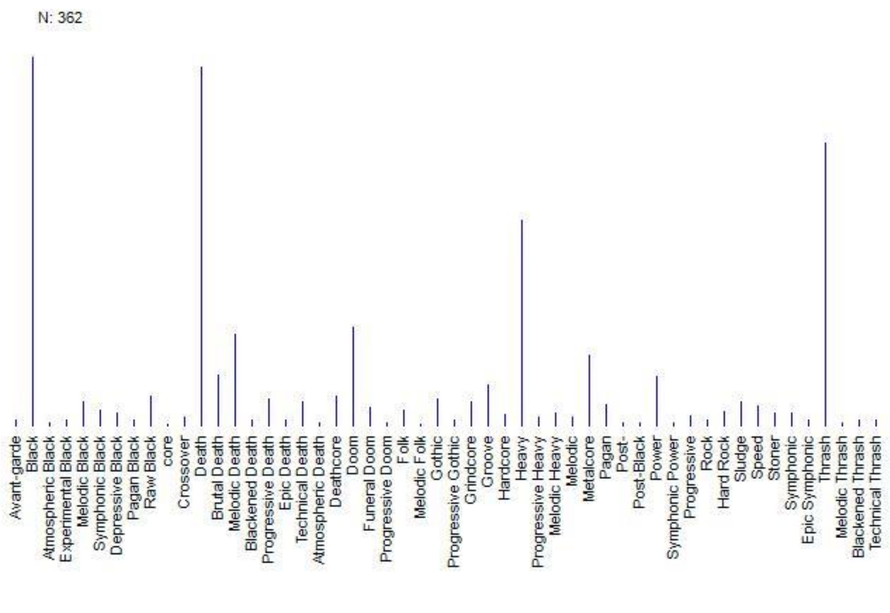


Figure 8: Genre distribution in Belgium

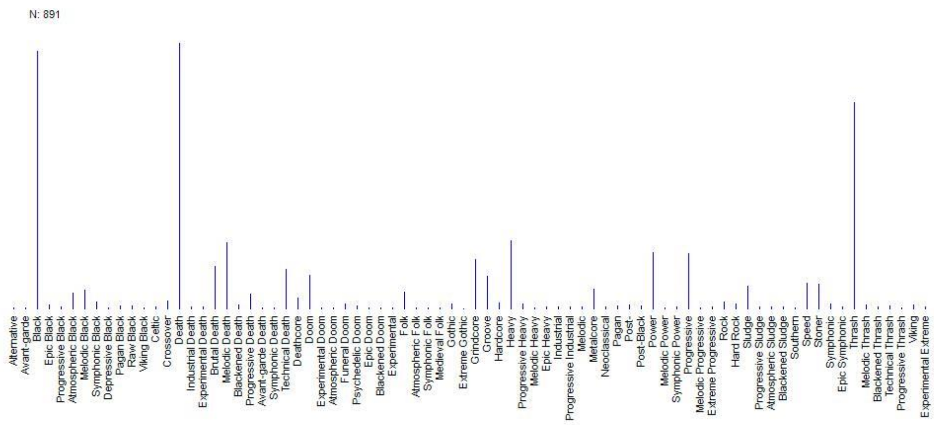


Figure 9: Genre distribution in Canada

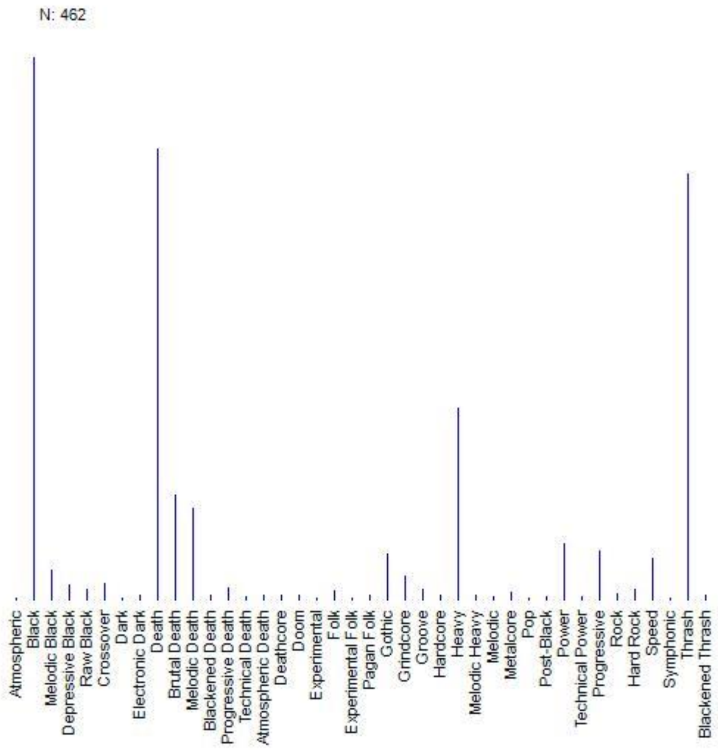


Figure 10: Genre distribution in Colombia

Al 3 countries have a fairly similar genre distribution. The *black*, *death*, *trash* and *heavy* genres are most frequent. Something that stands out is the black metal frequency for Colombia, which is a fair bit higher than Belgium and Canada. This makes sense, as there is a strong black metal scene in Colombia, as discussed before. However, at this stage it is unclear whether or not the actual genre hierarchy is reflected in the structure of the social network. To gain more insights into this, a colored dendrogram was created.

The links of the dendrogram have been colored according to the highest distribution of genre. Each cluster has its own genre distribution and an associated histogram, like the ones shown in figures 8-10. For each cluster, the highest frequency genre was picked and assigned a color.

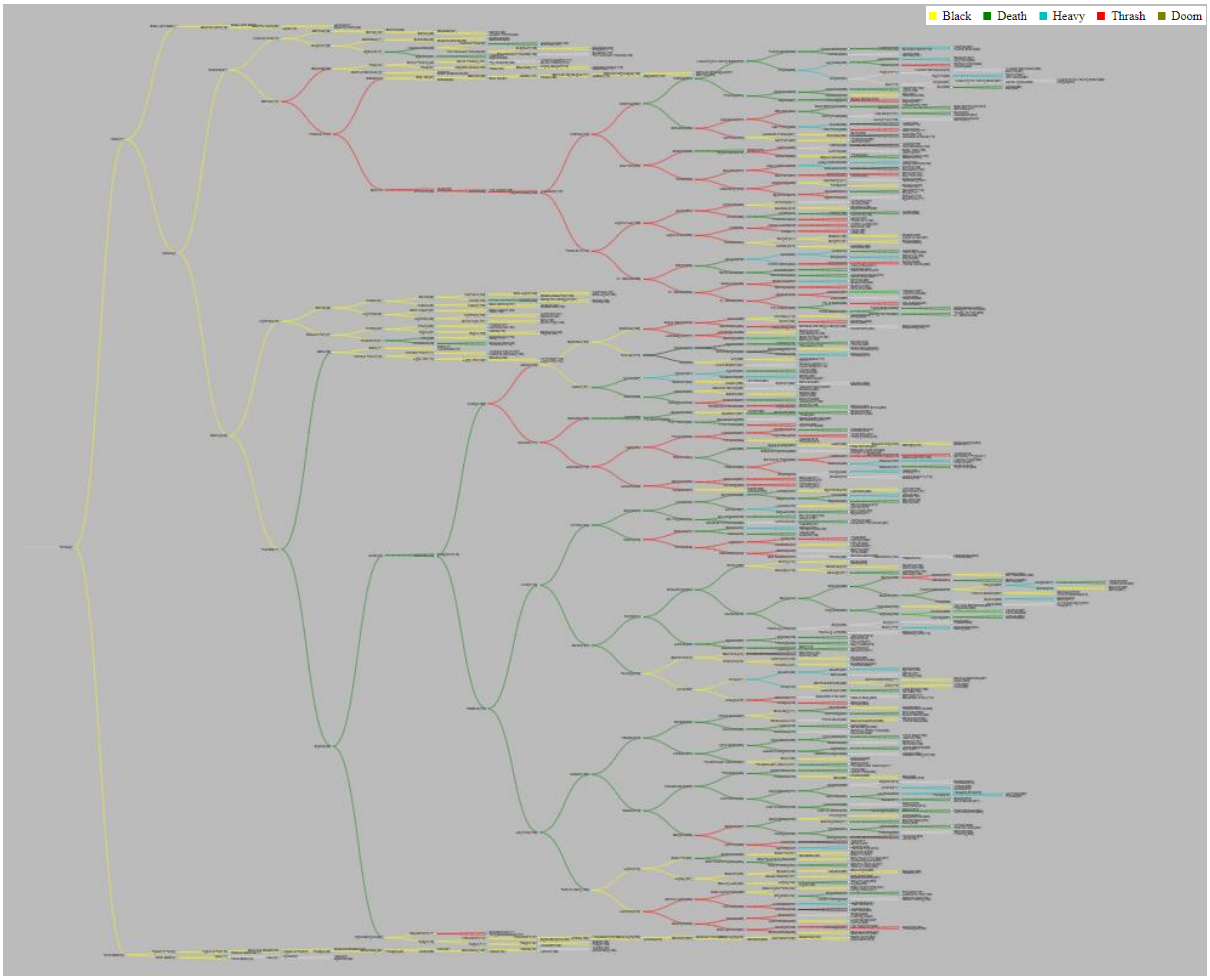


Figure 11: Hierarchically clustered Colombian bands, colored by highest genre distribution

The image above shows the colored dendrogram for the Colombian bands. One thing that stands out is the high density branch in the center of the dendrogram. These are the bands from the aforementioned Medellin scene, contributing a high percentage to the overall black metal frequency. However, there is a big problem here. Because of the distribution of genres, i.e. the vast overrepresentation of the genres black, death, trash and heavy, the dendrogram loses its overview. At some resolution, the overall genre distribution takes shape and from that point onward, moving deeper into the tree, the clusters will have a distribution similar to that of the total tree. It may be more interesting to look at the complete distribution at certain points and balance the data so that it contains an even representation of genres, even though it may prove challenging to retain the social network structure when balancing the data.

At the local resolution there is some more visible variation in genres, although there isn't much to suggest the structure follows the genre hierarchy. The same goes for the

Belgian bands. The comparison with the genre hierarchy was done using visual inspection, unfortunately it doesn't quite look the way we hoped it would. It is unclear if the genre hierarchy is present in the hierarchically clustered data. One of the reasons is simply down to the nature of both hierarchies. The hierarchies as seen in figures 9 and 10 are hierarchies of distribution. This hierarchy is weighted and contains quantitative information with regards to genre distribution. The Genre hierarchy itself as seen in Appendix A is not a distribution, but a plain old hierarchy of tags.

It stands to reason to do the inspection using a discrete computational method. However, it is unclear whether such a method exists or can be modified to work for this particular case. Before a computational method can be used for comparing these 2 types of hierarchies, either a kind of normalization needs to take place or a different method of clustering must be attempted, which brings its own challenges. Section 5.4 will deliberate on this further.

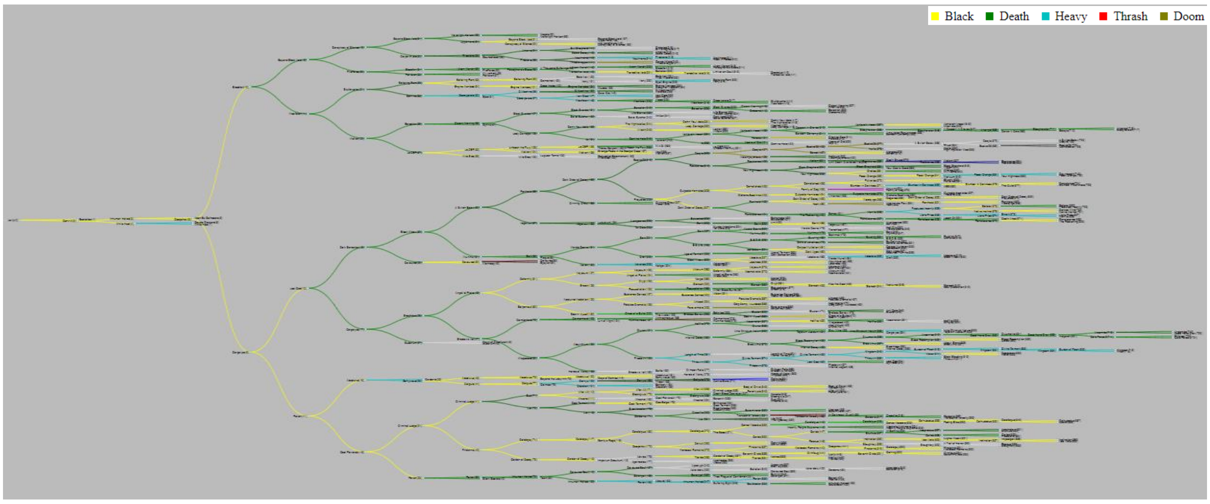


Figure 12: Hierarchically clustered Belgian bands, colored by highest genre distribution.

The dendrogram shown above is a lot less "clustered" than the Colombian dendrogram. Belgium lacks a strong metal scene, something that Colombia does have. This means that the structure of the social network is a lot more connected as there is no isolated scene. This can also be seen when visualizing the network with Gephi.

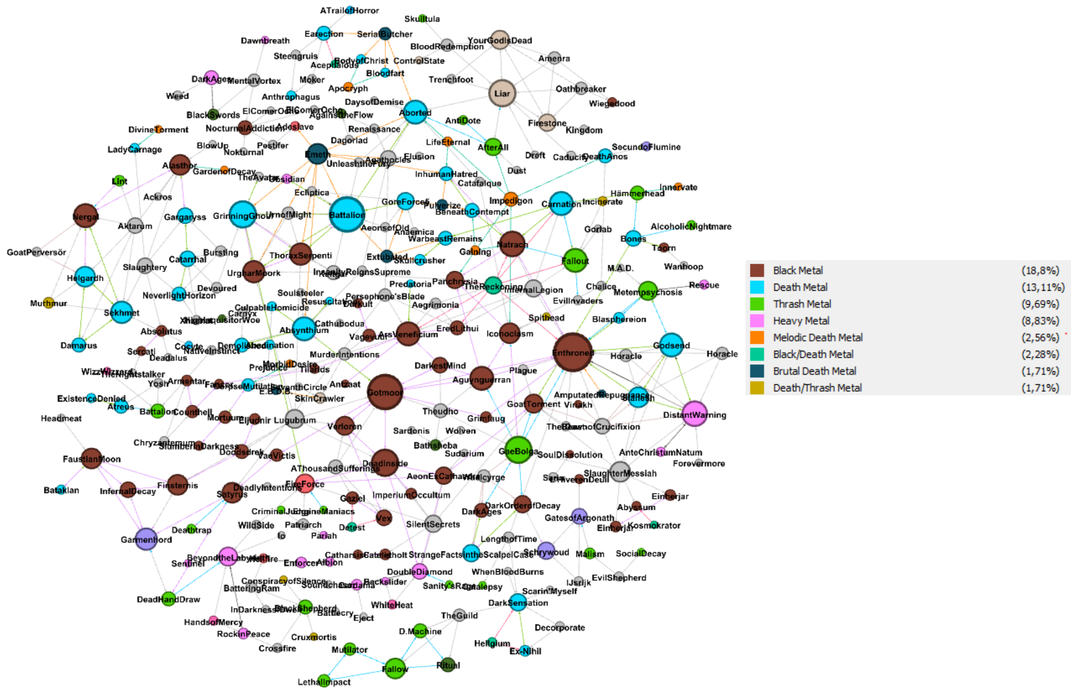


Figure 13: All Belgian bands, colored by genre.

This is where the hierarchical clustering method, and in particular, the coloring, runs into another issue. With networks such as this one, the hierarchical clustering method and coloring doesn't quite work. There are a fair number of trash metal bands in Belgium, but these do not show in the dendrogram. This is probably because these get clustered later along with black and death metal bands, meaning the trash metal frequency never gets higher than the black and death metal frequencies. Also, on the local resolution, there is nothing that suggests a genre hierarchy has bearing on the social network structure.

5 Conclusions and future work

This section will draw some conclusions and reflect on them. The final subsection will provide some directions regarding future work.

5.1 Methodology

In this study we have taken an exploratory approach. Whilst seeming a good fit for this type of study initially, a number of drawbacks have been identified. One of these drawbacks is a problem of *pathfinding.* Normally, one starts with a question, maps out the route, follows it and ends up at, ideally, the envisioned end-goal. However, an exploratory approach works in a slightly different manner. Instead of posing a question and mapping out a route, this approach forces one to keep an open mind when it comes to the approach, yet without an actual end-goal, since the question arises while experimenting with different approaches. This laissez-faire approach is on one hand pretty good for exploring a dataset such as the one used in this study, on the other hand can lead to a premature optimization when it comes to the ideal approach. Also, at times it becomes tricky to keep focus. Normally, one make a plan and stick to it, but with an approach such as this it gets pretty tempting to continuously go back to the drawing table when some result isn't the one you were looking for. Thusly, this exploratory approach has the possibility of causing some serious scope creep, potentially endangering the goal of the initial study.

5.2 The scraper and data

The dataset provided by *Encyclopaedia Metallum* has proved itself to be extensive enough for usage in fairly large scale social network analysis. However, the data itself wasn't useful right off the bat, the meta-data contained within was often simply too fuzzy to use without some kind of normalization. This was very much expected, as this is often the problem with user-generated content. Luckily for us, not all meta-data was fuzzy. Things like country, location and status where pretty easy to parse, but, unfortunately, genre wasn't so easy to parse. The genres themselves found within the dataset where unique, but the way the genres where registered for each band proved to be a big obstacle, as discussed in previous sections. Running any normalization method against this fuzzy data risks obscuring the semantic meaning of the data. But because of the volume of data, it had to be some kind of automatic method. Taking manual samples after the normalization did show good results, but the verification could have possibly been done more thoroughly. The resulting genres where usable, although there were still too many genres, arguably. This is mainly because of subgenres. There might be a way to categorize these subgenres under their respective main genres: most of the time these subgenres are recognizable by their modifier preceding the genre name itself. For example: *Atmospheric* Black Metal, *Technical* Death Metal and *Brutal* Death Metal. That way bands can be categorized as pertaining to a main genre, even though they play a subgenre. Again, this may remove some semantic meaning, but it will certainly simplify matters.

When it comes to the actual scraping of the data, not much can be said. There was a brief discussion of using a pre-made scraper to grab the data, but that ran into problems with saving the data. At the end of the day, we needed some utilities to query for the data when running analyses against it. The chosen approach, a custom built scraper, probably took longer that it should have, but turned out to be a boon as it became pretty straightforward to incorporate analysis functionality into the scraper itself. We didn't have to generate a dataset, lug it over to another program or machine and then run the analysis, we could do it right in the scraper program. Because everything was done in the scraper program, including analysis, this means any algorithm used was implemented too. This doesn't mean that everything was built from scratch, obviously, where code could be copied from other sources or pseudo-implementations where available these where used. However, this ran a fine line. Luckily, this study didn't need very complicated algorithms. Anyone with a modicum of programming ability can implement a hierarchical clustering method or a Floyd-Warshall algorithm. But since this was an exploratory study it could very well have turned out differently. It's suspected that these explorative studies will have this problem. One does jump from one path to the next in an explorative study, which can seriously disorganize the tool-cabinet so to speak. The art lies in navigating between the two extremes, one extreme being the absolute patchwork of existing programs and protocols, the other being the giant time-sink when implementing everything yourself. For this study it seems to have worked out well enough.

5.3 Results

In this study we have taken an exploratory approach and attempted to see if a genre hierarchy can be observed in the social network structure. Unfortunately, the approach with genre-distribution and dendrograms hasn't provided the hoped-for results, instead providing a null-result. At this stage, it is unclear whether or not social network structure is influenced by the genre hierarchy. One of the main problems was that of visualization. In order to reconcile the genre distribution with the visualization of the clustering itself, the histogram approach proved to be too unclear. Different ways of visualizing hierarchical data can be tried to achieve the goal this study has laid out such as using a radial dendrogram. As stated before in the result section, the method used to compare the two hierarchies was that of visual inspection. It was quickly concluded that the two hierarchies don't share many similarities. This is down to, among other reasons, the nature of the two hierarchies, one contains distributions while the other is plain hierarchy of tags. The distribution matter is probably the hardest to *fix*, if such a word is appropriate, because it's a logical by product of the structure of the network. Genres aren't distributed evenly throughout the dataset, and can't be balanced without losing the network's cohesion. The genre hierarchy itself might be weighted, but then the question is to what criteria the weighting would be performed.

The Word2Vec and supplementary t-SNE plot did provide some inside into the organization of genre. It seems that genres are pretty heavily clustered around their stylistic counterparts. Such as all *punk* genres clustered together, all extreme forms of *death metal* clustered together and all *pagan* genre types clustered together. This is fairly surprising, since the analysis was done with a genre co-occurrence list. This does point to the idea that bands will choose a certain abstract subject, such as politics, horror or philosophy, and then pick a style of metal to explore said subject, not the other way around. It seems more logical that a band plays a certain way because of the preference of the musicians, regardless of subject matter. The clusters found in the t-SNE plot don't necessarily support this notion. One argument for this is the inclusion of *RAC* in the *punk* and *trash* cluster and the inclusion of *Funeral Doom* and *Depressive Black* in the *ambient* cluster. Apart from this it is difficult to find any hierarchy in this plot, as is connectedness between clusters, if any.

Also, while 50% is a decent sample of the full dataset, it may leave out important bands that may have a large impact on the shape of the network. Therefore, the full dataset should be acquired and applied.

5.4 Future work

One thing that must be addressed is the role of time. A genre hierarchy such as the one used for this study does exist in time; the popularity of genres rise and fall when time progresses and new genres are formed. Therefore, taking the evolution of the social network through time into account when comparing it to the genre hierarchy may yield interesting results. Several challenges are included in this, such as the hierarchical clustering and visualization of the evolving network.

As discussed before, there are different ways of classifying the bands by their genres. These should be investigated to see if they yield a better balance between resolution and semantic meaning. The approach used in this study did yield a list of genres which was fairly useful, but it may be useful to take subgenres into account and group subgenres with their respective main genres. This could hypothetically give an even more discrete list of genres, with added flexibility as it allows one to doubly classify a band: first to its 'real' genre (which could be a subgenre) and secondly to any main genre the subgenre may come from.

Other types of future work focuses more on the data contained in the dataset which wasn't used for this explorative study. There is a fair corpus of lyrics contained within the dataset which can be used for semantic analysis. It might be interesting to look at the lyrical content of songs based on location, as there seems to be a strong linkage between regional mythology and the metal bands from said region. This links back to *Unheilige Allianzen* **[48]** which lays out the case that, among others, ethno-paganism is very popular theme in Heavy Metal. Also, as stated in the introduction, one of the initial interests that lie at the basis of this study is the presence of all manner of extremism and radicalism within Heavy Metal. One could envision a study that uses the corpus of lyrics within the dataset along with the lyrical themes of the bands to gain some insight into the more extreme elements of Heavy Metal, how wide-spread it might be and how these sub-communities are connected to the wider Heavy Metal community. It might seem logical that these communities are very isolated, but perhaps that isn't the case. Also, as was briefly touched on in the results section, it seems strong local sub scenes can develop and become isolated from the wider metal community, this can prove to be a measure of identifying the more extreme sub scenes.

6 References

[1] P.M. Gleiser, L. Danon, *Community Structure in Jazz*, Advanced Complex Systems, vol 6 issue 4 pages 565-573, 2003, link

[2]M Girvan, M.E.J. Newman, *Community structure in social and biological networks*, PNAS 99 issue 12 pages 7821-7826, 2002, link

[3] A. L. Barabasi, E Bonabeu, *Scale-free networks*, Scientific American, 2003, link

[4] H. Jeong, B Tombor, R. Albert, Z. N. Oltvai, A. L. Barabasi, *The large-scale organization of metabolic networks*, Nature, 2000, link

[5] P Erdös, A Rényi, *On the evolution of random graphs*, Institute of Mathematics & Hungarian Academy of Sciences, 1960, link

[6] D. Watts, S Strogatz, *Collective dynamics of small-world networks*, Nature, 1998, link

[7] Encyclopedia Metallum, https://www.metal-archives.com

[8] M. E. J. Newman, *Detecting community structure in networks,* The European Physical Journal B, vol 38 issue 2 pages 321-330, 2004, link

[9] L. C. Freeman, *A Set of Measures of Centrality Based on Betweenness,* Sociometry col 40 issue 1 pages 35-41, 1977, link

[10] V. D. Blondel, J. L. Guillaume, R. Lambiotte, E. Lefebvre, *Fast unfolding of communities in large networks*, Journal of Statistical Mechanics: Theory and Experiment, 2008, link

[11] B. H. Good, Y-A. de Montjoye, A Clauset, *The performance of modularity maximization in practical contexts,* Physical Review, 2010, https://arxiv.org/pdf/0910.0165.pdf

[12] J. Wallch, H. M. Berger, P. D. Greene, *Metal Rules the Globe: Heavy Metal Music Around the World*, pages 180 – 183, 2011, Duke University Press

[13] S. Stack, J. Gundlach, *The Effect of Country Music on Suicide*, Social Forces Vol 71 issue 1 pages 211-218, 1992, link

[14] Eric Lacourse, Michel Claes, Martine Villeneuve, *Heavy Metal Music and Adolescent Suicidal Risk*, Journal of Youth and Adolescence, vol 30 issue 3 pages 321-332, link

[15] B. Valyou, B. Dickinson, W. Hu, *Determining Leaders and Communities on Networks using neighborhood similarity*, Social Networking vol 3 issue 1, 2014, link

[16] T. Makkonen, *North from here: The collaboration networks of Finnish metal music genre superstars*, Creative Industries Journal vol 10, pages 104-118, 2017, link

[17] Robert L Gross, *Heavy Metal Music: A New Subculture in American Society*, Journal of Popular Culture vol 24 issue 1, 1990, link

[18] S. Stack, J. Gundlach, J. L. Reeves, *The Heavy Metal Subculture and Suicide*, Suicide and Life Threatening Behavior vol 24 issue 1 pages 15-23, 1994, link

[19] K. R. Scheel, J. R. Westefeld, *Heavy Metal Music and Adolescent Suicidality, An Empirical Investigation*, Adolescence vol 34 issue 134 pages 253-273, 1999, link

[20] J. Arnett, *Heavy metal music and reckless behavior among adolescents*, 1991, Journal of Youth and Adolescence, link

[21] T. Makkonen, *Tales from the Thousand Lakes: placing the creative network of metal music in Finland*, Environment and Planning A: Economy and Space, 2014, link

[22] V. Venkatesh, D. Perri, J. S. Podoshen, K. Urbaniak, *From Pride To Prejudice To Shame, Multiple Faces of the Black Metal scene within and without Online Environments*, Educational Psychological and Behavioral Considerations in Niche Online Communities, 2014, link

[23] B. Hickam, *From Shamans to Cyborgs, the use of masks and shadow in branding heavy metal*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[24] C. Hoad, *He can be whatever you want him to be: Identity and intimacy in the masked performance of Ghost*, Popular Music vol 37 issue 2 pages 175-192, 2018, link

[25] T. Helyar-Cardwell, *Battle Jackets, Craft and Folk Culture: Research through Creative Practice*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[26] A. Berg, T. Gulden, V. Kiort af Ornas, N. Pavel, V. Sjovoll, *The Metal T-shirt: Transmedia Storytelling in Products*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[27] R. Platz Cortsen, *By The Cake Of The Dark Lord! Metal Cultures in Three Nordic Comics*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[28] N. Scott, T. O'Boyle, *Doom Values: Deceleration Promoting a Philosophy of Progression Through Opposites*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[29] T. Hjelm, *Controversial Metal and Controversial Religion, Theorising The Connections*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[30] C. N. Davis, *When Marduk Read Romans: Biblical Text as a Hermeneutic in Heavy Metal, Heavy Metal as Biblical Exegesis*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[31] I. Abraham, *Innovation and Standardization in Christian Metalcore: The Cultural Influences of Church and Market*, Modern Heavy Metal: Markets, Practices and Cultures conference, 2015, link

[32] Peter Doherty, *The Beginner's Guide to Winning the Nobel Prize*, Columbia University Press, 2008, link

[33] Encyclopaedia Metallum statistics, https://www.metal-archives.com/stats

[34] I. Husbyn Aarsand, *Imagining Cascadia: Bioregionalism as Environmental Culture in the Pacific Northwest*, 2013, link

[35] G. Horetua, Bullets, *Drugs, and Rock and Roll: Colombian Punk Rock, Heavy Metal culture in a time of Revolt and Terrorism*, Proquest Dissertations Publishing, 2013, link

[36] Genre hierarchy, Bound By Metal, retrieved from https://www.boundbymetal.com/en/common/metal-genres-graph

[37] A. Pradhan, G. Mahinthakumar. *Finding all-pairs shortest path for a large-scale transportation network using parallel Floyd-Warshall and parallel Dijkstra algorithms*, Journal of Computing in Civil Engineering vol 27 issue 3 pages 263-273, 2012, link

[38] R. W. Floyd, *Algorithm 97: Shortest path*, Communications of the ACM vol 5, 1962, link

[39] D. Kulkarni, N. Sharma, P. Shinde, V. Varma, *Parallelization of Shortest Path Finder on GPU: Floyd-Warshall*, International Journal of Computer Applications vol 110, 2015, link

[40] G. J. Katz, J. T. Kider, *All-Pairs Shortest-Paths for Large Graphs on the GPU*, Proceedings of the 23rd ACM/SIGGRAPH/EUROGRAPHICS symposium, 2008, link

[41] J. Ward, *Hierarchical Grouping to Optimize an Objective Function*, Journal of the American Statistical Association vol 58, issue 301, 1963, link

[42] B. Bollobas, *The Evolution of Random Graphs – the Giant Component,* Random Graphs, 2001, link

[43] M. Bastian, S, Heyamnn, J. Mathieu, *Gephi, an open source software for exploring and manipulating networks,* Icwsm 8 pages 361-362, 2009, link

[44] K Zuse. *Der Plankalkül*, Gesellschaft für Mathematik und Datenverarbeitung vol 63, 1972.

[45] V. I. Levenshtein, *Binary codes capable of deletions, insertions and reversals*, Doklady Akademii vol 10, issue 8, 1965, link

[46] L. R. Dice, *Measures of the amount of ecologic association between species*, Ecology vol 26, issue 3, pages 297-302, 1945, link

[47] C. Iseman, *Medellin Ultra Metal*, Invisible Oranges, 2015, Retrieved from http://www.invisibleoranges.com/medellin-ultra-metal/

[48] C. Dornbush, H.P. Killguss, *Unheilige Allianzen*, Unrast-Verlag, 2005

[49] M. Gardell, *Gods of the Blood: the pagan revival and white seperatism*, Duke University Press, 2003

[50] R. E. Gonzalez, *Woodland Cathedral: Cascadian Black Metal and implicit religion,* University of Texas, 2015, link

[51] B. Hedge Olsen, *I am the black wizards: Multiplicity, Mysticism and Identity in Black Metal Music and Culture*, 2008, Bowling Green University, link

[52] M. Phillipov, *Extreme music for extreme people? Norwegian black metal and transcendent violence,* Popular Music History vol 6 issue 1 pages 1501-163, 2011, link

[53] *Satan Rides the Media,* directed by T. Gruden, 1998, link

[54] *Until The Light Takes Us*, directed by A. Aites and A. Ewell, Variance Films, 2009

[55] *Once upon a time in Norway,* directed by P. Aasdal and M Ledang, Grezenlos Productions, 2007, link

[56] D. Weinstein, *Heavy Metal: the music and it's culture*, Chapter 2, Da Capo Press, 2000

[57] M. E. Ballard, S. Coates, *The Immediate Effects of Homicidal, Suicidal, and Nonviolent Heavy Metal and Rap Songs on the Moods of College Students*, Youth & Society Vol 2 issue 2 pages 148-168, December 1ste 1995, link

[58] S. Stack, *Heavy Metal, Religiosity and Suicide Acceptability*, Suicide and Life Threatening Behavior vol. 28, issue 4, pages 388-394, 1998, link

[59] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, *Distributed Representations of Words and Phrases and their Compositionality,* NIPS vol 26, 2013, link

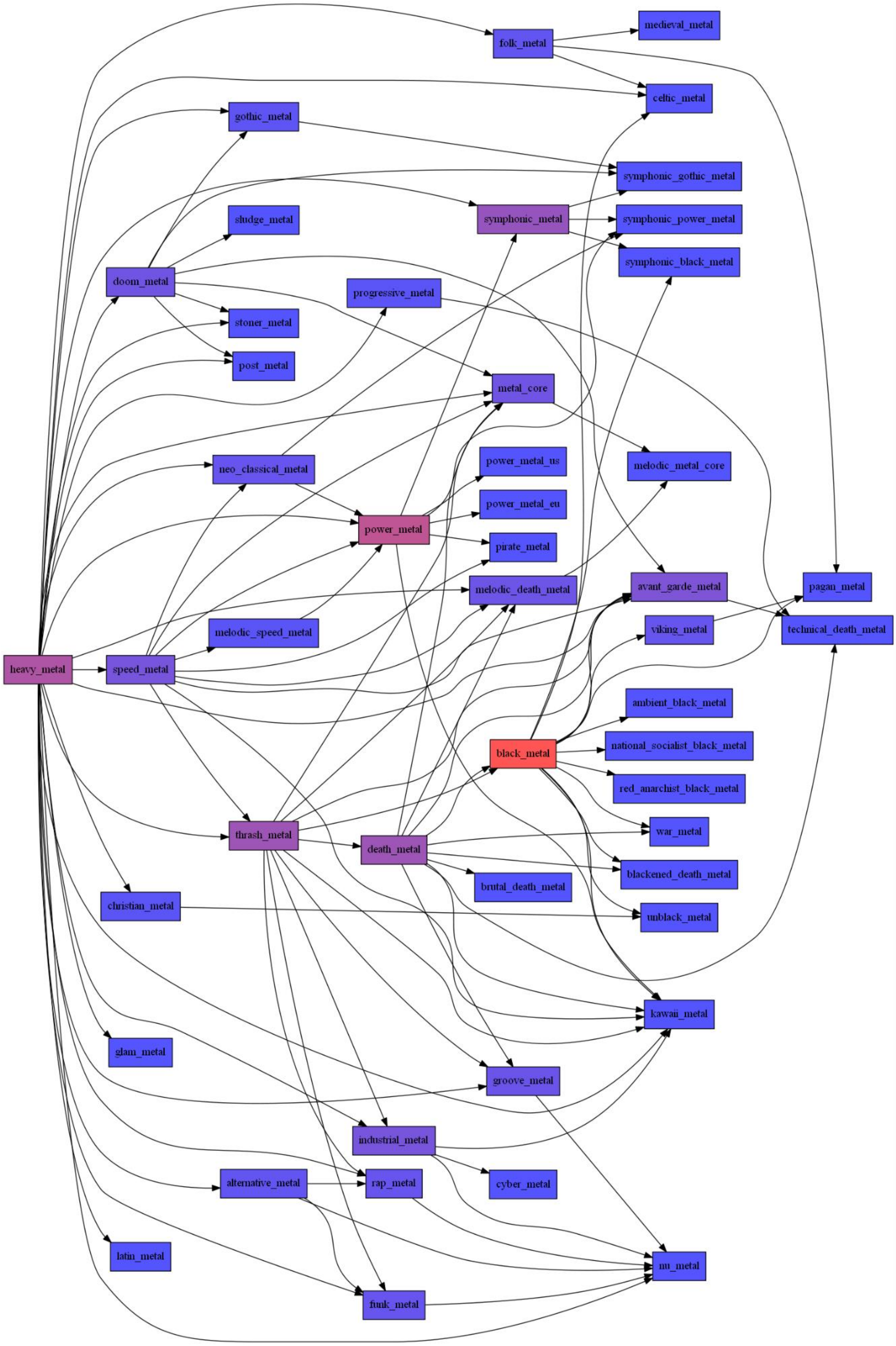
[60] L. van der Maaten, G. Hinton, *Visualizing data using t-SNE*, Journal of Machine Learning Research, 2008, link

[61] R. Rehurek, P. Sojka, *Software framework for topic modelling with large corpora,* LREC, 2010, link

[62] Pedregosa et al, *Scikit-learn: Machine Learning in Python,* JMLR vol 12, pages 2825-2830, 2011, link

[63] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman, *Indexing by Latent Semantic Analysis,* Journal of the American society for information science 41, pages 391-407, 1990, link

7 Appendix



folk_metal
medieval_metal
celtic_metal
gothic_metal
symphonic_gothic_metal
symphonic_metal
symphonic_power_metal
sludge_metal
symphonic_black_metal
doom_metal
progressive_metal
stoner_metal
post_metal
metal_core
power_metal_us
melodic_metal_core
neo_classical_metal
power_metal_eu
power_metal
pirate_metal
melodic_death_metal
avant_garde_metal
pagan_metal
melodic_speed_metal
viking_metal
technical_death_metal
heavy_metal
speed_metal
ambient_black_metal
national_socialist_black_metal
black_metal
red_anarchist_black_metal
war_metal
thrash_metal
death_metal
blackened_death_metal
brutal_death_metal
christian_metal
unblack_metal
kawaii_metal
glam_metal
groove_metal
industrial_metal
alternative_metal
rap_metal
cyber_metal
latin_metal
nu_metal
funk_metal