

ANÁLISE DE DADOS

**MINERAÇÃO DE DADOS
APLICADO EM REDES
SOCIAIS**

AUDIUS

CONTENTS

1 INTRODUÇÃO

PARA APLICARMOS OS CONHECIMENTOS ADQUIRIDOS EM SALA DE AULA NO QUE DIZ RESPEITO A MINERAÇÃO E ANÁLISE DE REDES SOCIAIS, FOI SOLICITADO QUE OS ALUNOS SELECCIONASSEM REDES SOCIAIS DISTINTAS E APlicassem DIVERSAS TÉCNICAS DE ANÁLISE DE REDE, LEVANTANDO DADOS, QUESTIONAMENTOS E AFINS, A RESPEITO DAQUELA REDE.

NESTE TRABALHO IREMOS ESTUDAR A AUDIUS. A AUDIUS SE TRATA DE UMA REDE DESCENTRALIZADA DE STREAMING DE MÚSICA, FAZENDO PARTE DESSA NOVA ONDA DA WEB3, A INTERNET DESCENTRALIZADA. A PLATAFORMA TEM COMO PROPOSTA GARANTIR RETORNO PARA OS PRODUTORES MUSICais E USUÁRIOS DA PLATAFORMA E ISSO ACONTECE ATRAVÉS DA TOKEN *AUDIO*. MAIS DETALHES SOBRE A DISTRIBUIÇÃO E FUNÇÃO DA TOKEN PODEM SER ENCONTRADOS NESTE PAPER

A ANÁLISE DA REDE ACONTECERÁ EM CINCO ETAPAS: EXPLORAÇÃO DOS DADOS DA REDE SOCIAL; CARACTERIZAÇÃO TOPOLOGICA DA REDE; DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS DE REDES SOCIAIS; ANÁLSIE DE TEXTO; E ANÁLISE DE COMUNIDADES E RECOMENDAÇÃO.

ACESSE A AUDIUS [AQUI](#).

2 EXPLORAÇÃO DOS DADOS DA REDE SOCIAL

A EXPLORAÇÃO SERÁ FEITA A PARTIR DA API DA AUDIUS, QUE PODE SER ACESSADA ATRAVÉS DO SITE AUDIUS API DOCUMENTATION. POR SE TRATAR DE UM PROJETO DA WEB3, É DE SE ESPERAR QUE A API SEJA ROBUSTA E FORNEÇA INFORMAÇÕES O SUFICIENTE E ISSO DE FATO OCORRE. ATRAVÉS DA API, QUE POSSUI UMA DOCUMENTAÇÃO BEM DETALHADA E DE FÁCIL ENTENDIMENTO, É POSSÍVEL CONSEGUIR INFORMAÇÕES SOBRE O USUÁRIO, SEUS SEGUIDORES, QUEM ELE SEGUE, INSCRITOS, TAGS MAIS UTILIZADAS, USUÁRIOS RELACIONADOS, FAIXAS EM ALTA, PLAYLISTS EM ALTA, DOWNLOAD DE FAIXAS, STREAMING E MAIS.

NO MOMENTO DE COLETA DE DADOS (23/10/2024), PESQUISANDO POR ARTISTAS, GÊNEROS E MÚSICAS MAIS OUVIDAS EM UMA SEMANA TIVEMOS OS SEGUINTE RESULTADOS:

2.1 ARTISTAS:

- COOP RECORDS: 11794 PLAYS;
- MORGAN PAGE: 7035 PLAYS;
- SNOOZE: 5368 PLAYS;
- BADMOUTH RECS: 4990 PLAYS;
- TRAPICAL MUSIC: 4346 PLAYS;

2.2 GÊNEROS MUSICAIS:

- ELECTRONIC: 42778 PLAYS;
- POP: 8965 PLAYS;
- DRUM & BASS: 7758 PLAYS;
- DUBSTEP: 4516 PLAYS;
- LO-FI: 4207 PLAYS;

2.3 MÚSICAS:

- MORGAN PAGE - THE LONGEST ROAD (FEAT. LISSIE) [REMIX CONTEST]: 7035 PLAYS;

- NOTION - THINK ABOUT U: 6898 PLAYS;
- MANSIONZ - THE LIFE OF A TROUBADOUR FT. SNOOZEGOD AS OLIVER: 5216 PLAYS;
- GLURCKY X DINK THE POOF - HEAD GUN: 4990 PLAYS;
- UNLEAZER - CHILDREN: 4346 PLAYS;

PELA QUANTIDADE DE "PLAYS" QUE TEMOS NOS TOP 5 DA PLATAFORMA, JÁ PERCEBEMOS QUE O VOLUME É BEM MENOR COMPARADO A OUTROS SERVIÇOS DE STREAMING, COMO SPOTIFY, SOUNDCLOUD OU YOUTUBE. TAMBÉM VEMOS UMA INFLUÊNCIA MUITO FORTE DE MÚSICAS ELETRÔNICAS, JÁ QUE DRUM & BASS, DUBSTEP E LO-FI TAMBÉM SÃO GÊNEROS DE MÚSICA ELETRÔNICA. PELA MINHA EXPERIÊNCIA, O CENÁRIO DA MÚSICA ELETRÔNICA É MUITO MAIS INDEPENDENTE, É UMA COMUNIDADE MAIS NICHADA, O QUE ACABA SE ENCAIXANDO COM O PERFIL DA REDE.

3 CARACTERIZAÇÃO TOPOLOGICA DA REDE

A CARACTERIZAÇÃO TOPOLOGICA É POSSÍVEL POIS TRANSFORMAMOS A REDE EM UM GRAFO E NESSE GRAFO CONSEGUIMOS VISUALIZAR A CONEXÃO ENTRE OS VÉRTICES. ESSA CONEXÃO PODE ACONTECER POR "N" MOTIVOS, IRÁ DEPENDER DO OBJETO DE ESTUDO DO PESQUISADOR, POIS AS CARACTERÍSTICAS QUE LIGAM OS NÓS, QUE CRIA A CONEXÃO ENTRE OS ELEMENTOS DA REDE, É DEFINIDA PELO PESQUISADOR. A NOSSA REDE BUSCARÁ ENTENDER QUAIS SÃO OS ARTISTAS MAIS RELACIONADOS E PARA MONTARMOS A REDE FOI PRECISO COLETAR DADOS ATRAVÉS DA API E PRÉ-PROCESSÁ-LOS.

3.1 PRÉ-PROCESSAMENTO:

UTILIZAMOS A API PARA COLETAR QUAIS FORAM OS ARTISTAS EM MAIOR POR TODO O TEMPO E NO ÚLTIMO MÊS. A PARTIR DOS IDS DESSES ARTISTAS, PROCURAMOS POR QUAIS ERAVAM OS SEUS SEGUIDORES E ENTÃO LIGAMOS OS USUÁRIOS OUVINTES COM OS ARTISTAS QUE SEGUIAM. O OBJETIVO É ENTENDER COMO SE DÁ A RELAÇÃO ENTRE OS ARTISTAS QUE TEM MAIS SEGUIDORES EM COMUM. PODEMOS LEVANTAR HIPÓTESES MAIS DIRETAS, POR EXEMPLO, SE O GÊNERO MUSICAL É IGUAL OU POSSUEM MUITAS PARTICIPAÇÕES. MAS PODEMOS TAMBÉM NÃO ENCONTRAR NENHUMA RELAÇÃO DIRETA E ENTENDER QUE É UM PERFIL DA REDE QUE ESTAMOS VISUALIZANDO.

DURANTE A PRÓXIMA SEÇÃO, IREMOS APRESENTAR OS RESULTADOS OBTIDOS, PRIMEIRAMENTE PARA OS DADOS COLETADOS PARA A RECOMENDAÇÃO POR TODO O TEMPO, SEGUIDO DO ÚLTIMO MÊS E A COMPARAÇÃO ENTRE ELES.

ANTES DISSO, IREMOS APRESENTAR AS REDES ALCANÇADAS:

- GRAFO PARA "EM ALTA POR TODO O TEMPO":

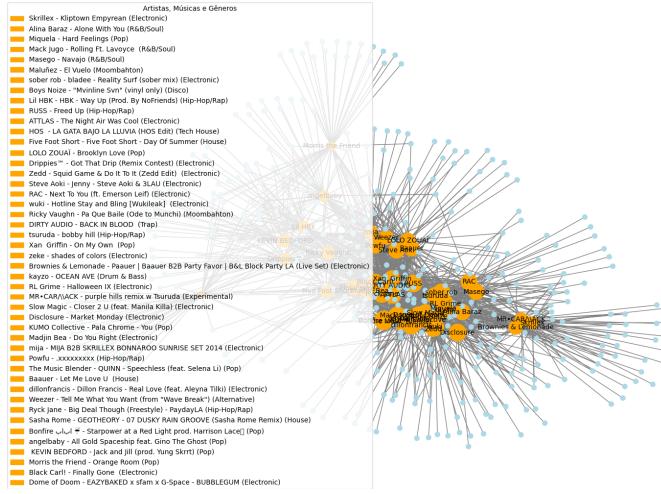


FIGURE 1: GRAFO DE "EM ALTA POR TODO O TEMPO" FEITO PELO NETWORKX UTILIZANDO CENTRALIDADE PROPORCIONAL.

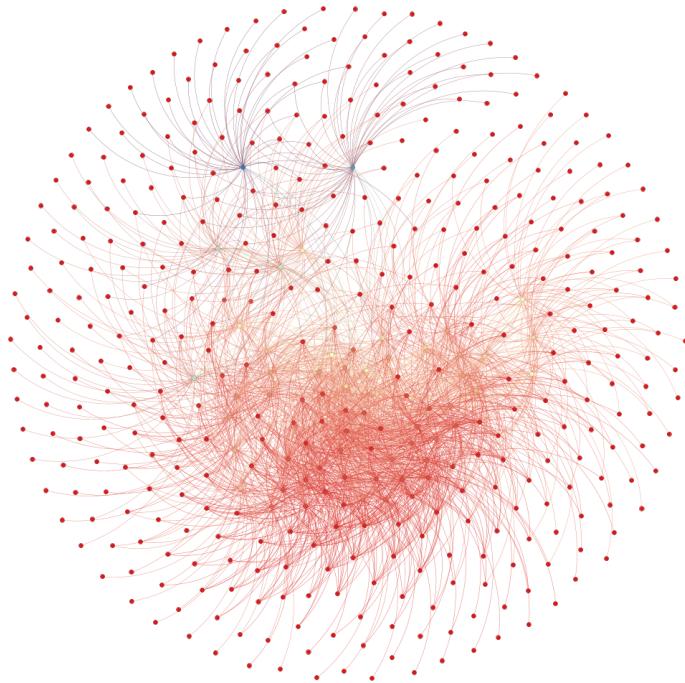


FIGURE 2: GRAFO DE "EM ALTA POR TODO O TEMPO" FEITO PELO GEPHI UTILIZANDO DISTRIBUIÇÃO DE FRUCHTERMAN REINGOLD.

APESAR DAS POSIÇÕES MAIS ALTAS SEREM PREENCHIDAS DE GÊNEROS COMO R&B/SOUL, POP, HIP-HOP, OS GÊNEROS DE MÚSICA ELETRÔNICA DOMINAM O RESTANTE DO PLACAR. SE SOMARMOS ELETRÔNICA, MOOMBAHTON, DISCO, HOUSE, TECH HOUSE, EXPERIMENTAL E DRUM & BASS, TEMOS UM TOTAL DE 29 COLOCADOS.

OS ARTISTAS QUE TIVERAM MAIS PARES EM COMUM, OU SEJA, TIVERAM MAIS SEGUIDORES EM COMUM, FORAM:

- ATTLAS - RYCK JANE: COM 41 SEGUIDORES EM COMUM. O PAR É COMPOSTO POR UM PRODUTOR DE MÚSICA ELETRÔNICA E UMA ARTISTA INDEPENDENTE QUE TOCA TROMPETE, ESCREVE RAP E PRODUZ BEATS. AO VISITAR O PERFIL DOS ARTISTAS, PERCEBEMOS QUE ATTLAS POSSUI UM REPERTÓRIO GRANDE DE MÚSICA ELETRÔNICA, NO QUE OUVIMOS, NÃO IDENTIFICAMOS MUITAS FAIXAS COM UMA COMPOSIÇÃO "MENOS ELETRÔNICA". QUERO DIZER, UMA MÚSICA COM VOCAL OU UTILIZANDO INSTRUMENTOS MUSICAIS. JÁ RYCK

JANE POSSUI UM REPÓRTO RÍTIMICO MAIS DIVERSIFICADO, ENTRETANTO, POSSUI DIVERSOS "BEATS", DIVERSAS FAIXAS APENAS DE INSTRUMENTAL, UM RITMO MAIS LO-FI E AFINS.

- SASHA ROME - MACK JUGO: COM 41 SEGUIDORES EM COMUM, NÃO CONSEGUI IDENTIFICAR RELAÇÃO ENTRE O PAR. SASHA ROME PRODUZ DISCO ENQUANTO MACK JUGO FAZ ALGO MAIS PARECIDO COM INDIE ROCK.
- RYCK JANE - BLACK CARL!: TAMBÉM COM 41 SEGUIDORES E UMA SITUAÇÃO SEMELHANTE À RELAÇÃO COM ATTLAS. BLACK CARL! TAMBÉM PRODUZ MÚSICA ELETRÔNICA.

- GRAFO PARA "EM ALTA NO ÚLTIMO MÊS":

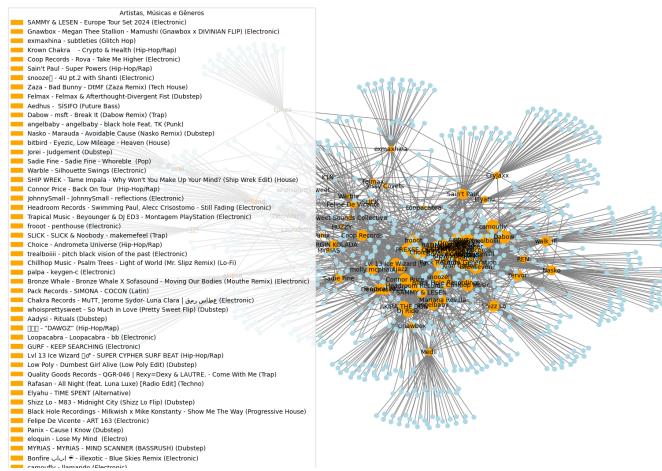


FIGURE 3: GRAFO DE "EM ALTA NO ÚLTIMO MÊS" FEITO PELO NETWORKX UTILIZANDO CENTRALIDADE PROPORCIONAL.

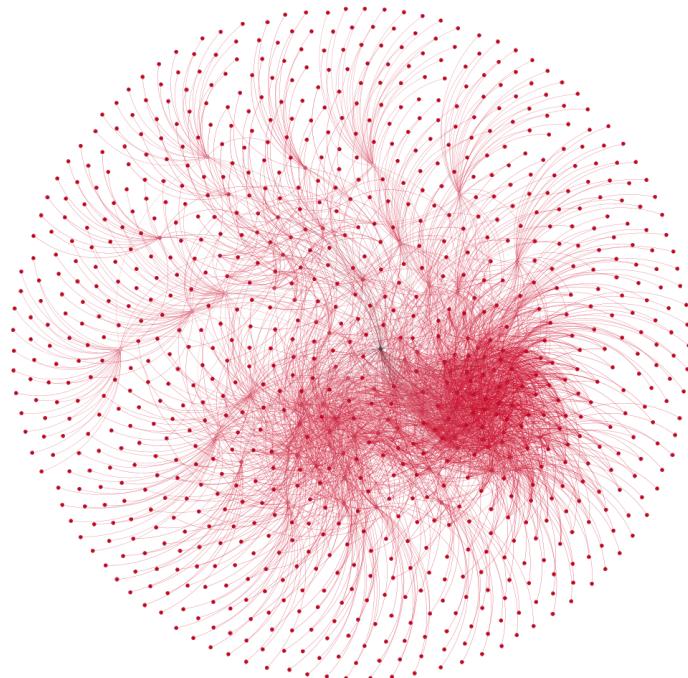


FIGURE 4: GRAFO DE "EM ALTA NO ÚLTIMO MÊS" FEITO PELO GEPHI UTILIZANDO DISTRIBUIÇÃO DE FRUCHTERMAN REINGOLD.

PARA O "EM ALTA NO ÚLTIMO MÊS" PEGAMOS UMA AMOSTRA DE 100 ARTISTAS, ENQUANTO "EM TODO O TEMPO" FOI COM 50 ARTISTAS. AQUI, NOVAMENTE, A MÚSICA ELETRÔNICA DOMINA O PLACAR COM 42 ARTISTAS SÓ DO GÊNERO ELETRÔNICA, SEM CONSIDERAR OS SUBGÊNEROS. ISSO PODE ACONTECER PORQUE TEM MAIS DE 50 ARTISTAS OU PODE INDICAR UMA TENDÊNCIA DA PRÓPRIA AUDIUS EM RELAÇÃO AOS ARTISTAS QUE PUBLICAM SEUS CONTEÚDOS POR LÁ.

OS ARTISTAS QUE TIVERAM MAIS PARES EM COMUM FORAM:

- DREAMEATER - BONFIRE バンファイア: COM 42 SEGUIDORES EM COMUM. O PAR É COMPOSTO POR UM PRODUTOR DOIS PRODUTORES DE HOUSE E AFINS. ISSO POR SI É UMA RELAÇÃO FORTE, POIS TEM O MESMO PÚBLICO.
- PALPA - LOW POLY: COM 40 SEGUIDORES EM COMUM, APESAR DE AMBOS SEREM PRODUTORES DE MÚSICA ELETRÔNICA, O ESTILO É DIFERENTE. PALPA TEM UM ESTILO MAIS EXPERI-

MENTAL ENQUANTO LOW POLY PARECE TER MAIS INFLUÊNCIA DO DUBSTEP.

- BRONZE WHALE - SHIP WREK: TAMBÉM COM 40 SEGUIDORES E DESSA VEZ AMBOS SÃO PRODUTORES DE HOUSE. A DIFERENÇA MAIS NOTÓRIA É DE QUE SHIP WREK SÓ POSSUI REMIX ENQUANTO BRONZE WHALE TEM REMIX E PRODUÇÕES AUTORAIS.

3.2 PROPRIEDADES BÁSICAS

- DISTRIBUIÇÃO DE GRAUS:
 - "EM ALTA POR TODO O TEMPO":

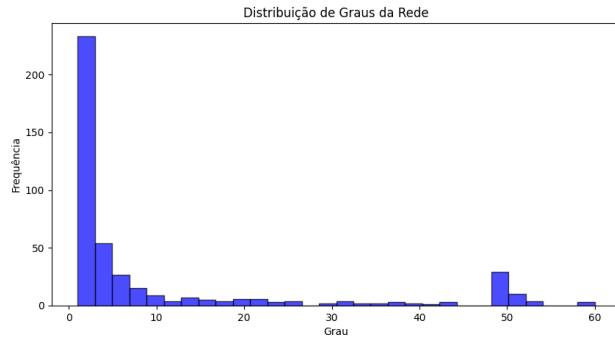


FIGURE 5: GRÁFICO DE DISTRIBUIÇÃO DE GRAUS PARA "EM ALTA POR TODO O TEMPO".

- * NÚMERO DE VÉRTICES E ARESTAS: 442 E 2292.
- * COEFICIENTE DE CLUSTERING MÉDIO: 0.1019.
- "EM ALTA NO ÚLTIMO MÊS":

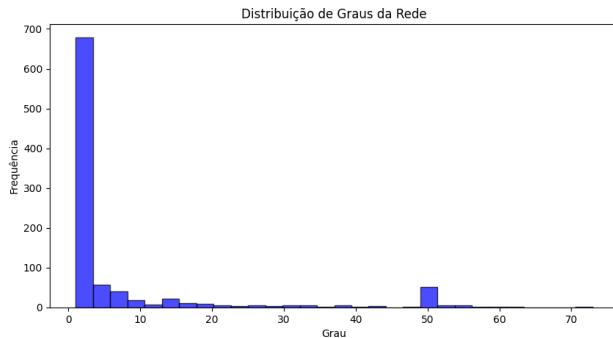


FIGURE 6: GRÁFICO DE DISTRIBUIÇÃO DE GRAUS PARA "EM ALTA NO ÚLTIMO MÊS".

* NÚMERO DE VÉRTICES E ARESTAS: 955 E 3545.

* COEFICIENTE DE CLUSTERING MÉDIO: 0.0121.

AQUI VEMOS QUE OS GRÁFICOS APRESENTAM UM COMPORTAMENTO SEMELHANTE, A QUANTIDADE DE NÓS QUE TÊM O MESMO NÚMERO DE CONEXÕES TEM COMPORTAMENTO SEMELHANTE, ENTRETANTO, PARA O PERÍODO DE "POR TODO O TEMPO" OS NÚMEROS SÃO AINDA MAiores. ISSO PODE SER EXPLICADO PELO FATO DE QUE COMO A REDE É RELATIVAMENTE NICHADA, OS ARTISTAS QUE JÁ POSSUEM INFLUÊNCIA NA REDE TEM TENDÊNCIA A PERMANECER EM ALTA, DESSA FORMA, CADA VEZ MAIS PESSOAS SEGUEM ESSES ARTISTAS QUE JÁ ESTÃO NO TOPO, LOGO, SE Torna UM SEGUIDOR EM COMUM. TENDÊNCIAS DE "EM ALTA" POR UM PERÍODO MENOR DE TEMPO DÁ UM CARÁTER DE VIRALIZAÇÃO PARA OS DADOS COLETADOS. A FAIXA ESTÁ EM ALTA POR UM PERÍODO CURTO DE TEMPO, PODE SE TRATAR DE UM "HIT WONDER", LOGO, SE TRATANDO DE UM ARTISTA DESCONHECIDO, É MENOS PROVÁVEL QUE TENHA SEGUIDORES EM COMUM COM AQUELES QUE JÁ SÃO MAIS CONHECIDOS.

ALÉM DISSO, COMO DISSEMOSS ANTERIORMENTE, O GRAFO PARA "EM ALTA NO ÚLTIMO MÊS" POSSUI O DOBRO DE NÓS. ENTRETANTO, SE PROCURARMOS A QUANTIDADE DE CONEXÕES POR NÓS,_{arestas/vrtices}, TEMOS 5.19 PARA A REDE "EM ALTA POR TODO O TEMPO" E 3.71 PARA A REDE "EM ALTA NO ÚLTIMO MÊS". ISSO PODE CONFIRMAR O NOSSO COEFICIENTE DE CLUSTERING MÉDIO, SENDO MAIOR PARA A REDE "POR TODO O TEMPO", OU SEJA, TEM MAIS CHANCE DE UM NÓ ESTAR CONECTADO A OUTRO QUANDO COMPARADO COM A OUTRA REDE.

3.3 CENTRALIDADE DOS VÉRTICES:

– GRAU MÉDIO:

* REDE "EM ALTA POR TODO O TEMPO": 10,371

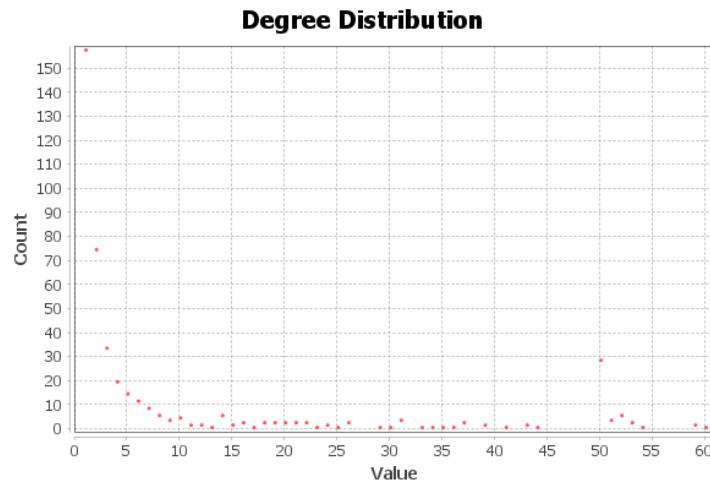


FIGURE 7: GRÁFICO DE DISTRIBUIÇÃO DE GRAUS PARA "EM ALTA POR TODO O TEMPO - GEPHI".

* REDE "EM ALTA NO ÚLTIMO MÊS": 7,424

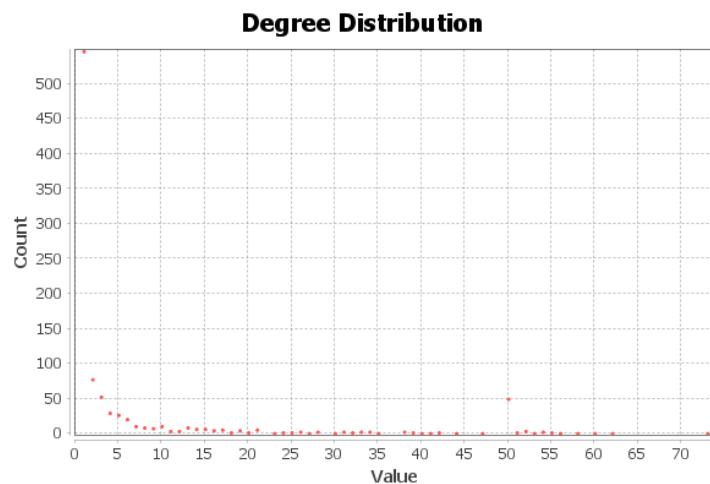


FIGURE 8: GRÁFICO DE DISTRIBUIÇÃO DE GRAUS PARA "EM ALTA NO ÚLTIMO MÊS - GEPHI".

- * EIGENVECTOR OU CENTRALIDADE DO AUTOVETOR:
 - REDE "EM ALTA POR TODO O TEMPO" EM 100 ITERAÇÕES: 0.009907

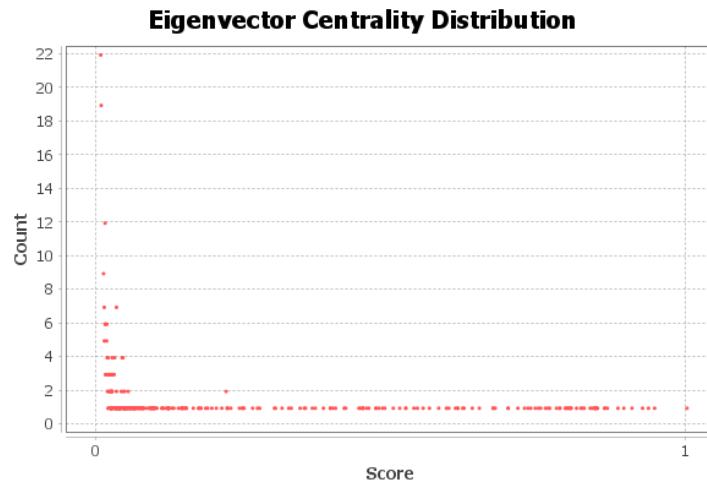


FIGURE 9: GRÁFICO DE DISTRIBUIÇÃO DE GRAUS PARA "EM ALTA POR TODO O TEMPO - GEPHI".

- REDE "EM ALTA NO ÚLTIMO MÊS" EM 100 ITERAÇÕES: 0.028205

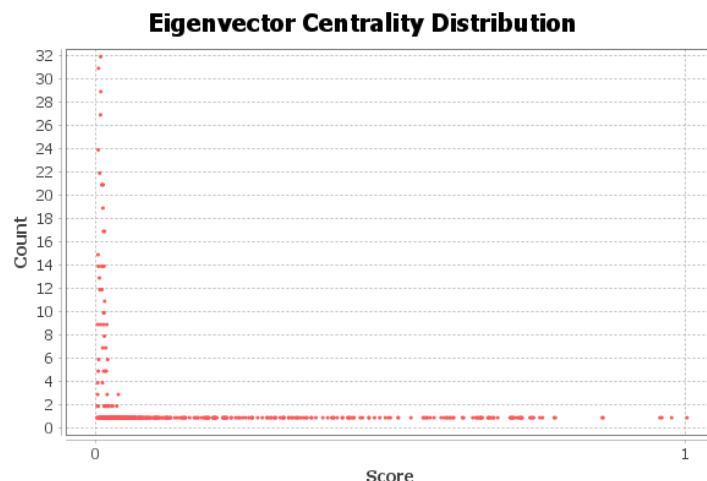


FIGURE 10: GRÁFICO DE DISTRIBUIÇÃO DE GRAUS PARA "EM ALTA NO ÚLTIMO MÊS - GEPHI".

3.4 RESULTADOS

AO REALIZARMOS A ANÁLISE DAS REDES TANTO PARA AS RECOMENDAÇÕES "POR TODO TEMPO" E "NO ÚLTIMO MÊS" IDENTIFICAMOS DIFERENÇAS ENTRE AS REDES, MESMO QUE CONTENHAM ARTISTAS EM COMUM E OUTRAS RELAÇÕES, PARA PERÍODOS MAIS CURTOS TEMOS A CHANCE DE ENCONTRAR ARTISTAS DESCONHECIDOS QUE VIRALIZARAM OU ALCANÇARAM ALGUMA NOTORIEDADE NA REDE ATRAVÉS DE ALGUM LANÇAMENTO. AGORA SE NÓS PROCURARMOS "POR TODO O TEMPO", ENCONTRAREMOS ARTISTAS COM MAIOR NOTORIEDADE, COM UMA CARREIRA MAIS CONSOLIDADA DENTRO DESSE NICHO QUE É A AUDIUS. O VOLUME DE "PLAYS" NA REDE É BEM MENOR COMPARADO AOS BILHÕES DE PLAYS DIÁRIOS DAS GRANDES COMPANHIAS DO MERCADO, E ISSO NÃO É NECESSARIAMENTE NEGATIVO, AFINAL, SE TRATA DE UMA REDE DESCENTRALIZADA, ENTÃO OS IDEIAS DAS COMPANHIAS, EM TEORIA, JÁ SÃO DIFERENTES.

4 DESCOPERTA DE CONHECIMENTO EM BASES DE DADOS DE REDES SOCIAIS E ANÁLISE DE TEXTO

4.1 SELEÇÃO

PARA O EXERCÍCIO DESTA ATIVIDADE, COLETAMOS OS DADOS REFERENTES ÀS FAIXAS DOS ARTISTAS EM ALTA "NO ÚLTIMO MÊS". ESSA DECISÃO FOI TOMADA POIS MANTÉM UM CARÁTER DE ATUALIDADE NOS OBJETOS DE ESTUDO. SE UTILIZASSEMOS AS FAIXAS "EM ALTA POR TODO O TEMPO", CORREMOS O RISCO DE CONTER TERMOS QUE TEM CONTEXTO EM UM INTERVALO ESPECÍFICO DE TEMPO. COMO A REDE SE TRATA DE UMA PLATAFORMA DE STREAMING DE MÚSICA, O CONTEÚDO TEXTUAL É PEQUENO, UTILIZAMOS O TÍTULO DAS FAIXAS, DESCRIÇÃO E TAGS. O OBJETIVO SERÁ IDENTIFICAR A RELAÇÃO ENTRE AS PALAVRAS, TAL COMO IDENTIFICAR TERMOS COMO "BIGRAMAS".

4.2 PRÉ-PROCESSAMENTO

ESTA ETAPA SE REFERE À LIMPEZA E ORGANIZAÇÃO DOS DADOS, AFIM DE REMOVER RUÍDOS E OUTROS TEXTOS QUE POSSAM INTERFERIR NA MINERAÇÃO DE DADOS. PARA ESTA ETAPA, IGNORAMOS OS RESULTADOS VAZIOS, GARANTIMOS QUE OS CONTEÚDOS SEJAM PALAVRAS. PARA FACILITAR NA IDENTIFICAÇÃO DE BIGRAMAS, SUBSTITUÍMOS PALAVRAS QUE POSSUEM HÍFEN POR UM ESPAÇO, DESTA FORMA, MANTEMOS CONTROLE DAS PALAVRAS QUE SÃO DITAS SEGUIDAS UMA DAS OUTRAS, NA TENTATIVA DE IDENTIFICAR TERMOS USADOS COM FREQUÊNCIA COMO "LIVE SET". TAMBÉM REMOVEMOS ESPAÇOS EXTRA DESNECESSÁRIOS. TODO O TEXTO EXTRAÍDO FOI CLASSIFICADO SEPARADAMENTE E EM CONJUNTO, OU SEJA, CADA SEÇÃO, TÍTULO, DESCRIÇÃO E TAGS, POSSUEM SUA CLASSIFICAÇÃO MAS TAMBÉM HÁ UM CONJUNTO QUE AGRUPA AS TRÊS SEÇÕES.

4.3 MINERAÇÃO DE DADOS

NESTA ETAPA UTILIZAMOS BERTOPIC PARA CLASSIFICAÇÃO EM RÓTULOS DAS PALAVRAS E EXTRAÍMOS OUTRAS ESTATÍSTICAS

REFERENTE ÀS PALAVRAS MAIS OCORRIDAS E EM RELAÇÃO À QUANTIDADE DE PALAVRAS CONTIDAS EM CADA SEÇÃO.

	MÉDIA	MEDIANA	DESVIO-PADRÃO	MÍNIMO	MÁXIMO
TÍTULOS	3.96	4.0	2.51	1	12
DESCRIÇÕES	19.07	11.0	27.05	0	150
TAGS	0.65	1.0	0.48	0	1

TABLE 1: ESTATÍSTICAS DOS TEXTOS ANALISADOS

4.4 TÓPICOS GERADOS PELO BERTOPIC

PARA MANTER A ORGANIZAÇÃO DO DOCUMENTO, OS RÓTULOS GERADOS PELO BERTOPIC PODEM SER ACESSADOS CLICANDO NO TÍTULO DA SEÇÃO, ACIMA.

4.4.1 TÍTULOS

O PRIMEIRO E O TERCEIRO TÓPICO PARECEM SEMELHANTES, POSSUEM TERMOS QUE CONSIGO IMAGINAR COMPONDONDO O NOME DE UMA FAIXA. JÁ O SEGUNDO TÓPICO PARECE SER COMPOSTO POR CARACTERÍSTICAS DA MÚSICA, PRINCIPALMENTE REMIX. ALGUNS REMIX COSTUMAM TER, NO TÍTULO, O NOME DO ARTISTA QUE FEZ O REMIX, JÁ QUE PODEM EXISTIR INÚMERAS VERSÕES. E BRONZE WHALE SE TRATA DE UM ARTISTA QUE APARECE "EM ALTA" E VIMOS QUE POSSUI DIVERSOS REMIX.

4.4.2 DESCRIÇÕES

NAS DESCRIÇÕES SÓ OBTIVEMOS DOIS RÓTULOS E NENHUM DOS DOIS É MUITO RELEVANTE. O PRIMEIRO PARECE SER UM CONJUNTO DE ADVÉRBIOS, ARTIGOS E PRONOMES E SÓ POSSUI A PALAVRA "FOLLOW" DE RELEVÂNCIA, TERMO COMUMENTE UTILIZADO EM REDES SOCIAIS. O SEGUNDO RÓTULO NÃO CONTEM PALAVRAS O SUFICIENTE PARA INTERPRETÁ-LO.

4.4.3 TAGS

OS RÓTULOS GERADOS PARA AS TAGS TAMBÉM NÃO É MUITO INTERESSANTE, PORÉM, O PRIMEIRO TÓPICO DESTACA A GRANDEZA

DA MÚSICA ELETRÔNICA NESSA REDE SOCIAL, CONTENDO DIVERSOS TERMOS DIRETAMENTE RELACIONADOS COM O NICHO, COMO "EDM", "ELECTRONIC", "BASS", "FESTIVAL".

4.4.4 TODOS

AQUI TAMBÉM NÃO TIVEMOS RÓTULOS MUITO INTERESSANTES. O PRIMEIRO TÓPICO SE ASSEMELHA COM O QUE FOI OBTIDO EM DESCRIÇÕES, O QUE FAZ SENTIDO POIS É A AMOSTRA DE TEXTO COM MAIOR NÚMERO AMOSTRAS, TEM MAIS PALAVRAS NAS DESCRIÇÕES QUE NAS OUTRAS PARTES. O SEGUNDO RÓTULO POSSUI ALGUMAS PALAVRAS QUE DÃO A ENTENDER QUE O ARTISTA ESTÁ TENTANDO COMUNICAR COM O PÚBLICO ATRAVÉS DO CAMPO DE DESCRIÇÃO DAS FAIXAS.

4.5 BIGRAMAS

TAMBÉM GERAMOS BIGRAMAS NA TENTATIVA DE IDENTIFICAR TERMOS UTILIZADOS NA PLATAFORMA OS BIGRAMAS QUE MAIS SE DESTACAM SÃO:

- "FEN FEN": NÃO CONSEGUI IDENTIFICAR A ORIGEM DESTE TERMO. PESQUISEI SE ERA UMA REDE SOCIAL, UMA GIRIA, O NOME DE ALGUM ARTISTA E A ÚNICA RELAÇÃO QUE ENCONTREI FOI QUE SIM, EXISTE UM ARTISTA CHAMADO FEN FEN MAS SÓ O ENCONTREI NO SPOTIFY.
- "CHAKRA RECORDS": COMO O PRÓPRIO NOME DIZ, SE TRATA DE UMA GRAVADORA.
- "EARN COOP": JÁ ACREDITO ESTAR RELACIONADO COM A CRYPTO TOKEN DA PLATAFORMA.

APARECEM OUTROS BIGRAMAS COM O NOME DE ARTISTAS, QUE PROVAVELMENTE ESTAVA ESCRITO NO TÍTULO DA MÚSICA.

4.6 NUVEM DE PALAVRAS

IREMOS DESTACAR AS PALAVRAS QUE MAIS APARECEM ATRAVÉS DE NUVEM DE PALAVRAS.

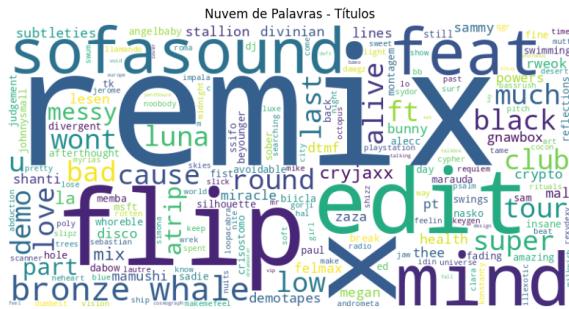


FIGURE 11: NUVEM DE PALAVRAS DOS TÍTULOS.

NOS TÍTULOS VEMOS MUITO A PRESENÇA DA PALAVRA "REMIX", "FEAT", "EDIT", TERMOS COMUNS EM TÍTULOS DE MÚSICA.



FIGURE 12: NUVEM DE PALAVRAS DAS DESCRIÇÕES.

NAS DESCRIÇÕES VEMOS A APARIÇÃO DE OUTRAS REDES, COMO FACEBOOK, TWITTER, INSTAGRAM, ATÉ MESMO SPOTIFY, SOUND-CLOUD E YOUTUBE APARECEM. PROVAVELMENTE PORQUE OS ARTISTAS ESTÃO DIVULGANDO SUAS PLATAFORMAS.

4.7 DISTRIBUIÇÃO DE SENTIMENTO

- MÉDIA DE SENTIMENTO PARA TÍTULOS: 0.01

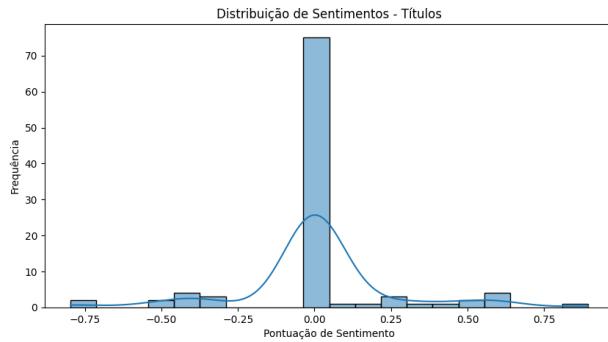


FIGURE 13: DISTRIBUIÇÃO DE SENTIMENTO DOS TÍTULOS.

– MÉDIA DE SENTIMENTO PARA DESCRIÇÕES: 0.2

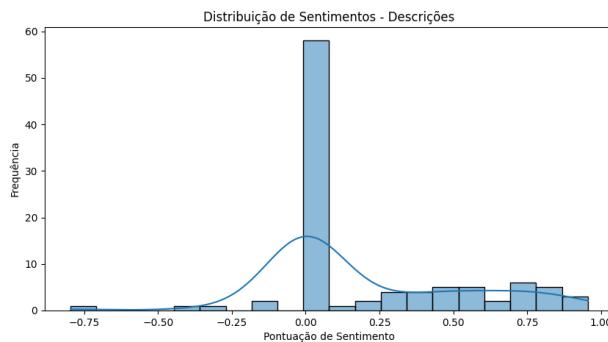


FIGURE 14: DISTRIBUIÇÃO DE SENTIMENTO DAS DESCRIÇÕES.

– MÉDIA DE SENTIMENTO PARA TAGS: 0.0

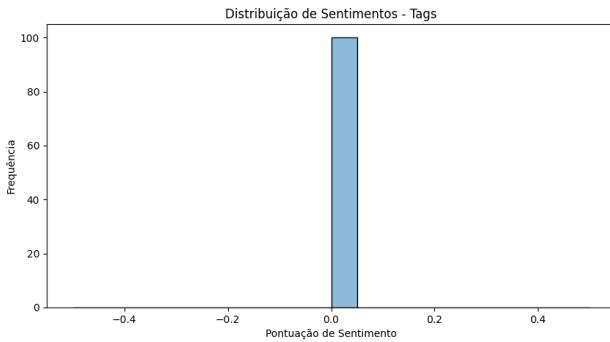


FIGURE 15: DISTRIBUIÇÃO DE SENTIMENTO DAS TAGS.

4.8 REDUÇÃO DA DIMENSIONALIDADE

AQUI PODEMOS VER COMO OS PONTOS ESTÃO SEPARADOS, ISSO SIGNIFICA QUE NÃO HÁ MUITA LIGAÇÃO ENTRE OS TERMOS, SÃO MUITO VARIADOS OU NÃO HÁ PADRÕES, PELO MENOS NO ESPAÇO REDUZIDO PARA 2D.

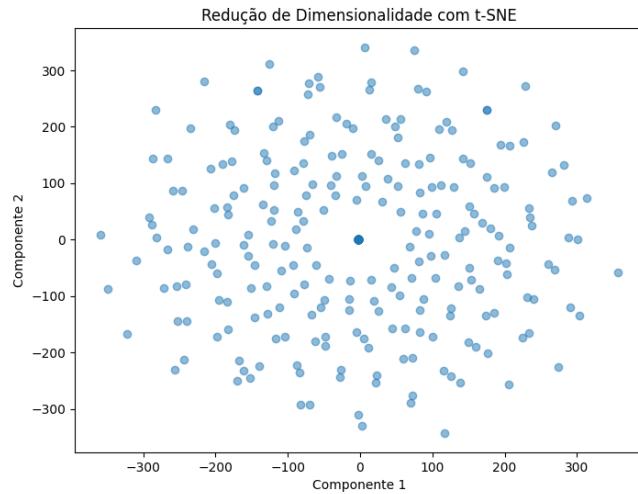


FIGURE 16: REDUÇÃO DE DIMENSIONALIDADE T-SNE.

5 DETECÇÃO DE COMUNIDADES

APESAR DE SE TRATAR DE UM STREAMING DE MÚSICA, PARTIR PARA UM SISTEMA DE RECOMENDAÇÃO COM OS DADOS QUE COLETAMOS NÃO PARECEU SER UMA ATIVIDADE FÁCIL, JÁ QUE ELES NÃO POSSUEM MINHA CONEXÃO ENTRE ELES. PORTANTO, DECIDIMOS TRABALHAR COM A DETECÇÃO DE COMUNIDADES A PARTIR DA REDE DE ARTISTAS X SEGUIDORES.

PARA IDENTIFICAÇÃO DAS COMUNIDADES, UTILIZAMOS O ALGORITMO DE LOUVAIN, PARA DESTACAR OS SUBCONJUNTOS DE NÓS COM MAIS CONEXÕES ENTRE SI. A PARTIR DESSE ALGORITMO IDENTIFICAMOS QUATRO COMUNIDADES DIFERENTES.

- NÚMERO DE COMUNIDADES DETECTADAS: 4
- TAMANHOS DAS COMUNIDADES: COUNTER(3: 172, 4: 133, 1: 71, 2: 66)

ACREDITAMOS QUE ENCONTRAMOS UM NÚMERO DE COMUNIDADES MUITO PEQUENO PARA O OBJETO DE ESTUDO. AFINAL, EXISTEM DIVERSOS GÊNEROS MUSICais. MAS COMO DESTACAMOS ANTERIORMENTE, ESSA REDE É RELATIVAMENTE NICHIADA, TENDO UMA INFLUÊNCIA MUITO FORTE DA MÚSICA ELETRÔNICA.

TAMBÉM É DIFÍCIL DESTACAR A RELAÇÃO ENTRE AS COMUNIDADES POIS DIVERSOS ARTISTAS ESTÃO COM OS DADOS COMO "GÊNERO" EM BRANCO, ENTÃO É DIFÍCIL RELACIONÁ-LOS COM A INFORMAÇÃO "VISÍVEL" QUE TEMOS. ACREDITAMOS QUE A MAIOR INFLUÊNCIA PARA A CRIAÇÃO DAS COMUNIDADES FOI DE FATO OS SEGUIDORES EM COMUM. PARA VISUALIZAR A PLANILHA COM OS DADOS DOS ARTISTAS E SUAS COMUNIDADES, CLIQUE AQUI

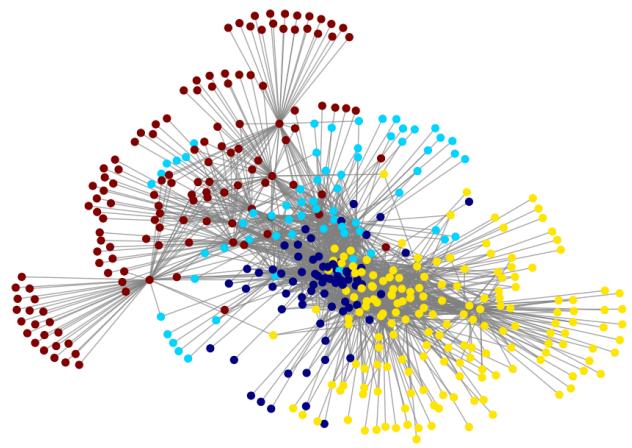


FIGURE 17: COMUNIDADES GERADAS COM O ALGORITMO DE LOUVAIN.