

## Visualização em rede da pesquisa brasileira em música, através das palavras-chave

**Renato Pereira Torres Borges<sup>1</sup>**  
UNIRIO/PPGM - DOUTORADO  
SIMPOM: *Musicologia*  
renatoptborges@gmail.com

**Resumo:** Esta comunicação objetiva descrever um método de análise bibliográfica, a fim de estudar o campo de pesquisa em música no Brasil. Esse estudo abordou a produção bibliográfica como um campo rico em dados (data-rich) para diminuir as margens para impressões sobre a área. Foi produzido um teste do método em desenvolvimento, neste recorte de pesquisa de doutorado em andamento na UniRio. Dados para o teste descrito foram coletados de 129 comunicações orais e pôsteres dos Anais do III Simpósio Brasileiro de Pós-Graduandos em Música, realizados pela UniRio em 2014. Palavras-chave e suas ocorrências foram coletadas e normalizadas. A linguagem natural utilizada pelos autores nas palavras-chave foi mantida durante todo o processo. Em seguida, as palavras-chave foram organizadas visualmente, gerando um gráfico de redes por meio do D3JS. Através desse método, duas grandes redes foram identificadas no corpus empregado. Então, as redes geradas foram comentadas diante da efetividade do método em encontrá-las. Apenas termos muito abrangentes geraram conexões entre diferentes textos. Dessa maneira, textos com termos muito específicos ficaram isolados na representação visual. Isso ocorreu devido ao escopo dos dados teste e à escolha da análise da linguagem natural. As lacunas atuais do método foram comentadas, visando apresentar possíveis soluções a serem trabalhadas no desenvolvimento da pesquisa, nos próximos anos.

**Palavras-chave:** Pesquisa em música no Brasil; Palavras-chave de música; Produção bibliográfica; Redes de pesquisa.

### Network Visualization of Music Research in Brazil, Through Keywords

**Abstract:** This paper describes a method for bibliography analysis, to study the Music research field in Brazil. This study approaches the bibliographic product as a data-rich field to reduce the margins for impressions on the area. In this section of a Doctoral research, a test was made for this method. Data for the described test was gathered among 129 papers and posters from the proceedings of the III Simpósio Brasileiro de Pós-Graduandos em Música (SIMPOM), organized by UniRio in 2014. Keywords and their occurrences were collected and normalized. The natural language used by the authors for the keywords were kept during the whole process. Then, the keywords were visually organized, generating a visualization of the network with D3JS. With this method, two relevant networks were identified in the given

---

<sup>1</sup> Orientadora: Profa. Dra. Martha Tupinambá de Ulhôa. UNIRIO.

corpus. Then, the generated networks were commented regarding the effectiveness of the method in finding them. Only broad keywords generated connections between different texts. This way, texts with very specific keywords were isolated in the visual representation. This occurred because of the scope of the test data and because of the choice of analyzing natural language. The current blanks left by the method were criticized, so to present possible solutions for the future development of the research, in the following years.

**Keywords:** Music research in Brazil; Music keywords; Bibliographic production; Research Networks.

## 1. Introdução

Ferramentas de visualização do desenvolvimento em longo prazo de pesquisas estabelecem, em geral, índices a publicações e autores. Calculados pela frequência e relevância de citações, as ferramentas estabelecem redes entre publicações e/ou autores. Esses índices são largamente utilizados por agências de fomento para qualificação de periódicos, programas de pós-graduação e pesquisadores.

Essas ferramentas são proveitosas quando se busca encontrar as publicações e autores que mais influenciaram suas respectivas áreas. Nesse sentido, são fontes com potência de futuro, que mantêm sua importância por um período prolongado. O futuro proposto por uma publicação de impacto, no entanto, dialoga e se fricciona com outros, que continuam a ser propostos. Acreditar que determinada publicação de impacto muda todas as práticas de pesquisa vigentes seria o mesmo que acreditar que o desenvolvimento da dodecafonia por Arnold Schoenberg teria feito com que toda nova música do mundo se tornasse, da noite para o dia, dodecafônica. Além disso, a ideia de que um futuro específico deve ser alcançado por todos está fundamentada numa concepção positivista de pesquisa, sugerindo um desenvolvimento linear das ferramentas e objetos de estudo da área.

## 2. Questão

Como seria possível visualizar então as dimensões das pesquisas brasileiras em Música e as conexões que estas formam entre si? Em vez de buscar os momentos de novas possibilidades no campo de pesquisa, aqui interessa saber a dinâmica do campo como um todo, incluindo métodos e conceitos há muito contestados, porém recorrentes, e temas e metodologias raramente enfocados.

Essa etapa da pesquisa de doutorado, atualmente em andamento na UniRio, objetiva possibilidades de mapeamentos mais fidedignos do desenvolvimento da área, partindo de visualização de dados em grande escala (*big data*). Por isso, busca-se aqui uma

forma de visualizar essa dinâmica a fim de melhor fundamentar percepções do campo, estreitando as margens de impressões subjetivas causadas pelo conhecimento de apenas uma parcela das pesquisas realizadas – ou seja, dar a oportunidade de se adotar padrões mais rigorosos de evidências, nas palavras de David Huron (1999, p. 25). Às quinze subáreas que Huron exemplificou como beneficiadas pelos "crescentes recursos" em informações, a presente investigação soma outra: a própria pesquisa em música. Portanto, esse trabalho trata-se, então, de um de âmbito metodológico.

Embora diversos textos sobre os “rumos da pesquisa em música no Brasil” sejam frequentemente publicados e esse objeto de estudo seja rico em dados (*data-rich*, nos termos de Huron), raros são aqueles que lidam com dados bibliográficos explicitamente coletados e analisados. Entre esses, se destacam os textos de José Nunes Fernandes (2000, 2006a, 2006b) e da pesquisa desenvolvida por Lia Tomás, apoiada pela ANPPOM, *A pesquisa acadêmica na área de música: um estado da arte* (2012). Também é possível encontrar levantamentos de dados, como é o caso do texto de Ulhôa (1996).

### 3. Dados

Como primeira possibilidade de material, foram consideradas as palavras-chave de publicações. Elas foram escolhidas como principal dado a estudar a fim de evitar o enfoque dado a publicações e a autores, ao mesmo tempo em que buscam transmitir sinteticamente o objeto e os métodos empregados no texto. Para a testagem desse método, foram escolhidas as publicações dos anais do III Simpósio Brasileiro de Pós-Graduandos em Música (SIMPOM), organizado pela UniRio em 2014. Foram encampadas aqui 129 comunicações orais e pôsteres do evento, que continham 478 palavras-chave (média de 3,7 por texto), sendo 411 únicas (3,2 por texto).

Esses dados foram escolhidos tanto pelo aspecto quantitativo quanto pelo qualitativo, embora o primeiro tenha sido preponderante na decisão. O *corpus* empregado, pelo menos cinco vezes a produção anual média de um periódico brasileiro, apresenta dados suficientes para testar esta ferramenta. Do ponto de vista qualitativo, essa escolha foi fortuita já que os dados são oriundos de um mesmo momento (evitando, por ora, a ação de uma variável temporal) e de autoria de pós-graduandos, responsáveis pela maior parte da produção de pesquisa no Brasil.

Optou-se por manter as palavras-chaves o mais próximo das originais. Os termos foram somente normalizados em ortografia (correções como “não-formal” para “não formal”) e em relação ao uso de maiúsculas (“Música Brasileira” para “Música brasileira”), a fim de

diminuir redundâncias. Isso respeita os termos com que os autores identificam e classificam seus trabalhos, sem impor classificações externas a que devem se adequar, e está baseado no princípio de que o que é considerado tema ou metodologia da área deve ser controlado autônoma e coletivamente pelo próprio campo, em seus devidos debates. Pretende-se, portanto, observar o desenvolvimento da linguagem natural dos pesquisadores e não alcançar sua substituição por uma linguagem controlada (LOPES, 2002).

Por isso, quando este método for mais tarde empregado em larga escala, essa escolha também deve resultar numa proximidade maior às mudanças conceituais da área. Isso pode ser exemplificado pela frequente revisão do termo “Educação musical” com outros, utilizados mais recentemente, como “Educação e música” ou “Ensino-aprendizagem de música”. Assim, embora transformações temporais não sejam observáveis com o *corpus* do presente estudo, este método está sendo preparado para abarcá-las, em etapas posteriores.

Utilizou-se uma visualização de *force layout* para observar as conexões produzidas entre as palavras-chaves. A visualização foi gerada com o auxílio de D3JS (BOSTOCK, 2015). Cada palavra-chave foi transformada em um nó, com a palavra-chave como rótulo (Figura 1). Quando um mesmo texto utilizou duas palavras-chave, foi produzida uma aresta entre os dois nós correspondentes. Assim, o menor agrupamento de nós  $\neg$  – um par – contém três elementos: dois nós e uma aresta (Figura 2). Para textos diferentes que utilizaram uma palavra-chave em comum, estruturas mais complexas surgiram (Figura 3). Com todos os textos visualizados, por causa da quantidade e do comprimento das palavras-chave, os rótulos ficaram ilegíveis. Esse problema foi resolvido por baixar a opacidade dos rótulos, somente restaurando a opacidade dos rótulos próximos ao cursor do mouse (Figura 4). O *plug-in* de distorção de olho de peixe (*fish-eye*) do D3JS também utilizado para facilitar a visualização. O gráfico total com as 411 palavras-chave e as conexões apresentadas pelos 129 textos foi então gerado (Figura 5).

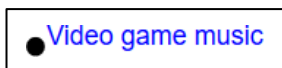


Figura 1. Exemplo de nó

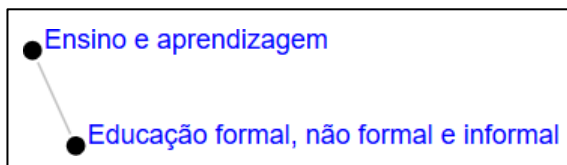


Figura 2. Nós e aresta gerada a partir do uso das palavras-chave num mesmo texto

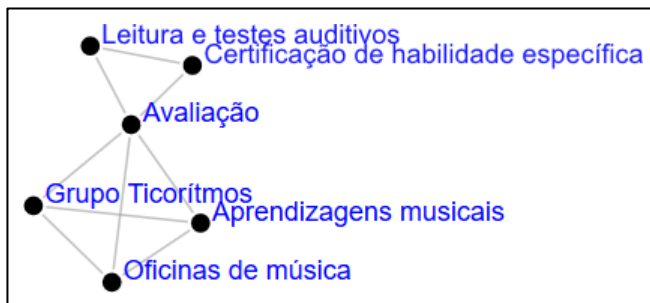


Figura 3. Rede gerada a partir de dois textos, que continham "Avaliação" como palavra-chave comum

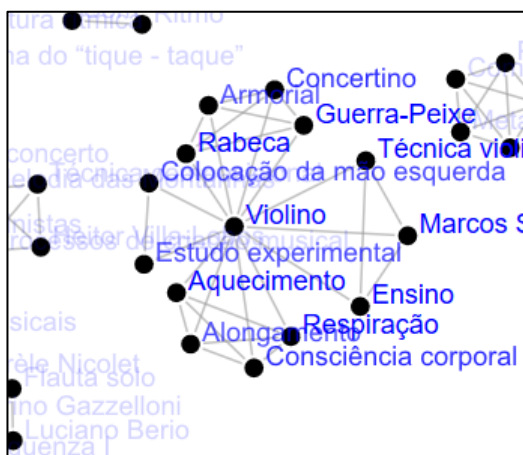


Figura 4. Visualização de estruturas complexas, com distorção de olho de peixe e alteração de cor

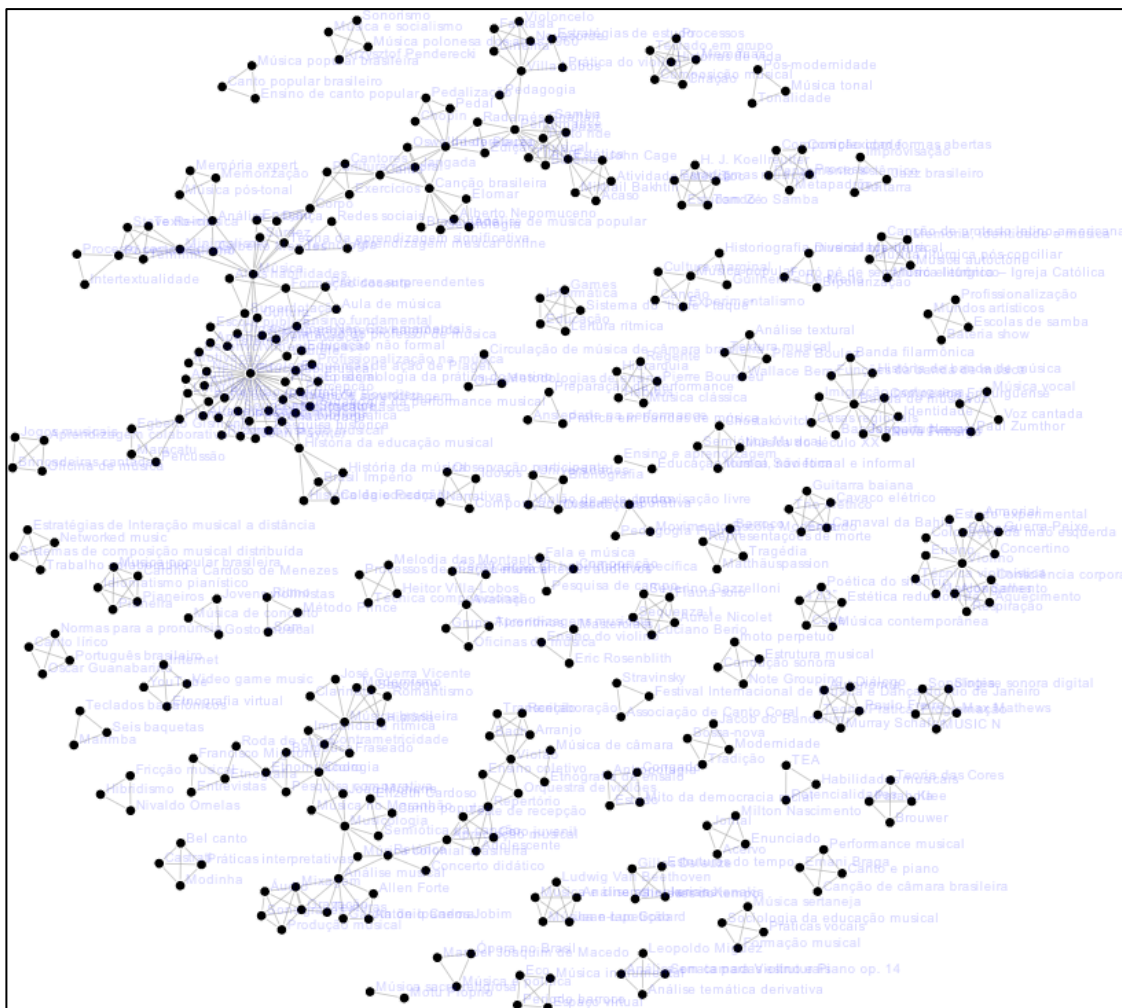


Figura 5. Redes geradas a partir dos textos do III SIMPOM

#### 4. Análise

O gráfico permitiu identificar rapidamente conexões entre temas e metodologias diversas, revelando um campo mais miscigenado do que o esperado. Duas grandes redes foram identificadas nos dados levantados. Uma delas (na parte de baixo e à esquerda, na Figura 5) apresenta uma estrutura parcialmente linear, quando lembrado que, na verdade, cada pequeno grupo de 3 a 6 nós forma uma única unidade textual. Ainda assim, observando as palavras-chave isoladamente, é possível ver o estabelecimento de redes de temas e metodologias. Outra rede apresenta estrutura semelhante mas também uma espécie de núcleo em uma de suas extremidades. Ali, um nó apresentou 41 arestas, número muito acima da média (3,56 arestas por nó).

Dos 129 textos, 45 (ou seja, 34,9%) não se articularam com outros, gerando os pequenos grupos isolados do gráfico. Nesse ponto do projeto, ainda não foi possível

aproximá-los automaticamente, por fundamentos semânticos ou aspectos afins. Vale ressaltar que o universo dos dados observados aqui é muito pequeno para concluir que tais trabalhos estejam desconectados do resto do campo de pesquisa ou mesmo que o campo, como um todo, se constitua de pesquisas desconectadas. Isso deverá ser abordado por essa pesquisa em passos posteriores.

É interessante notar que tipos de nós relacionam diferentes textos. Ao menos dentro desse pequeno recorte, fica evidente o fato de que os autores mantêm estáveis somente os termos mais abrangentes relacionados às suas pesquisas. Com o método aqui proposto, foi somente em níveis muito altos que redes surgiram. Estas se articularam com termos exemplificados na Tabela 1. Foram frequentemente utilizados nomes (54, ou 13,1%), sobretudo de compositores (33, ou 8%). Embora esse pequeno recorte só tenha apresentado uma palavra-chave de nome sendo utilizada por mais de um texto (“Villa-Lobos”), estima-se que, em estudos de larga escala, nomes também surjam como nós de muitas arestas.

Grandes áreas ou disciplinas, musicais ou relacionadas	Análise musical Estética Etnografia Etnomusicologia Interpretação Musicologia Performance
Gêneros, “pseudogêneros” e estéticas	Canção brasileira Choro Minimalismo Música brasileira Música popular Samba
Instrumentos e formações	Banda de música Canto Piano Violão Violino

**Tabela 1: Tipos de nós comuns a mais de um texto, e exemplos específicos.**

A maneira como o gráfico se desenvolveu exibe uma divisão equilibrada entre o uso e o não uso de termos muito abrangentes. De um lado, redes se organizaram através de

termos tão vastos que mal definem cada comunicação oral ou pôster. De outro, um terço dos textos não se conectou a nenhum outro, justamente pelo uso exclusivo de termos específicos. Fica evidente, então, que há uma discordância acentuada sobre o uso de termos entre os pós-graduandos inscritos naquele evento. Foi possível encontrar duas diferenciações: de nível hierárquico (mais abrangentes ou mais específicos) e de grafia. Explorar se essa inconsistência é consciente e quais motivos levam a ela escapa ao recorte desta comunicação.

Entre os aperfeiçoamentos possíveis desse método, quatro estão previstos para a sequência de seu desenvolvimento: formas de lidar com grafias diferentes, com sinonímias, com níveis hierárquicos diferentes e manutenção da singularidade dos termos. Grafias diferentes surgiram em duas categorias: uso inconsistente de plural e nomes próprios. O mesmo termo no singular e no plural ocorre frequentemente, como com “Aprendizagem musical” e “Aprendizagens musicais”. Isso também é rotineiro com nomes próprios, especialmente quando se trata de nomes muito conhecidos. No corpus analisado, foi o caso dos pares “Villa-Lobos”/“Heitor Villa-Lobos” e “Cage”/“John Cage”. Apesar da similitude, os termos desses três pares são identificados isoladamente no gráfico e, dependendo das redes em que se apresentam, podem estar em pontos opostos, exibindo uma desconexão que não se concretiza na leitura dos termos pelos pesquisadores.

O segundo problema, a sinonímia entre as palavras-chave, é de grande complexidade, pois envolve pequenas diferenças de leitura dos termos. Surgem situações como a em que o já citado “Aprendizagem musical” se mistura a outros como “Educação musical” e “Aula de música”. Nesse caso, apesar do primeiro impulso em querer unificá-los sob um único rótulo, os termos de fato podem significar coisas diferentes. A aprendizagem pode ocorrer autonomamente ou mesmo na própria prática musical, sem o intermédio de aulas. Curiosamente, não é possível também considerá-los sempre como termos exclusivos, já que muitas vezes os autores usam um desses termos como palavra-chave enquanto o corpo do texto, na verdade, lida com outro. É possível discutir partes da prática de um professor de música (plano de aula, por exemplo), o que poderia ser classificado como “Educação musical”, sem necessariamente lidar com a aprendizagem por parte dos alunos. Daí emerge a complexidade de decidir por igualar, agrupar, igualar e agrupar, ou simplesmente manter sinônimos.

Lidar com palavras-chave de níveis hierárquicos diferentes resulta, claramente, numa distorção do gráfico. Há situações de grande desequilíbrio: “Análise em camadas estruturais” não pode se relacionar de igual para igual com “Análise musical” quando é, em realidade, parte desta. Ou seja, seria necessário colocá-los em perspectiva para uma



visualização mais realista do campo: em uma segunda visualização, de termos mais abrangentes, o primeiro nó não existiria e suas arestas seriam herdadas pelo nó de “Análise musical”. Do termo mais genérico para os mais determinados, essa reclassificação seria impraticável. Cada texto precisaria ser estudado separadamente e o termo específico seria, muitas vezes, decidido sem o aval do autor. Assim, ainda correria o provável risco de distorcer o conceito por desconhecimento de causa. A direção contrária, no entanto, é possível. Estima-se que mesmo dois níveis apenas não sejam o suficiente para dar conta dos termos.

Do termo específico para o mais abrangente, essa tarefa exigirá uma classificação de grande escala: se o *corpus* aqui estudado contém 411 palavras-chave, o site *Amplificar* apresenta 4161 termos utilizados por 2052 teses, dissertações, comunicações, pôsteres e artigos, entre outros textos acadêmicos, publicados em 2013 e 2014 no Brasil na área de Música (AMPLIFICAR, 2016). Há, no entanto, uma diminuição no número de palavras-chave introduzidas por novos textos, já que a probabilidade de apresentarem termos já utilizados aumenta. Isso é constatado nas médias dos dois *corpi*: 3,18 novos termos por texto no SIMPOM 2014 e 2,02 no conteúdo indexado no *Amplificar*.

Por fim, pretende-se que seja possível visualizar (em quaisquer níveis hierárquicos que vierem a ser elaborados) versões animadas do gráfico, em que se tornem visualizáveis mudanças em eventuais surgimentos, crescimentos, estabelecimentos, diminuições e desaparecimentos de termos no campo de pesquisa, assim como de ligações entre eles. Esses gráficos podem ser gerados a partir de diferentes intervalos temporais, a fim de observar as dinâmicas do campo. Isso será debatido no desenvolver do trabalho.

### **Considerações finais**

O método descrito nesta comunicação apresentou resultados relevantes para análises iniciais do campo de pesquisa, assim como apresentou pontos passíveis de melhoria. Entre seus principais resultados, estão o estabelecimento de redes a partir de temas e metodologias, e a revelação do impacto da linguagem natural em sistemas automatizados de organização do conhecimento no campo da Música. O trabalho sobre esse método seguirá, a fim de sanar suas lacunas iniciais.

## Referências

- AMPLIFICAR. 2016. Disponível em <<http://www.amplificar.mus.br>>. Acesso em: 05 jan. 2016.
- BOSTOCK, Mike. *D3JS: Data-Driven Documents*. 2015. Disponível em <<http://d3js.org>>. Acesso em: 02 jan. 2016.
- FERNANDES, José Nunes. Pesquisa em educação musical: situação do campo nas dissertações e teses dos cursos de pós-graduação stricto sensu em educação. *Revista da ABEM*, n. 5, p. 45-57, set. 2000.
- FERNANDES, José Nunes. (Org.) *Publicações da Associação Brasileira de Educação Musical: índice de autores e assuntos: 2002-2005*. Rio de Janeiro: PPGM/UNIRIO, 2006a.
- FERNANDES, José Nunes. Pesquisa em educação musical: situação do campo nas dissertações e teses dos cursos de pós-graduação stricto sensu brasileiros. *Revista da ABEM*, n. 15, p. 11-26, set. 2006b.
- LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. *Ciência da Informação*, Jan./Abr. 2002, vol.31, no.1, p. 41-52. 2002.
- TOMÁS, Lia. *A pesquisa acadêmica na área de música: um estado da arte*. São Paulo, 2012.
- ULHÔA, Martha Tupinambá de. Dissertações de Mestrado Defendidas nos Cursos de Pós-Graduação Stricto Sensu em Música e Artes/Música até Dezembro de 1996. *Opus*, v.4, p. 80-94, 1996. v. 4. Rio de Janeiro, 1996.
- \_\_\_\_\_ (org.). *Anais do III SIMPOM*. Rio de Janeiro, 2014.