

Relatório 2, bolsa TT5, processo 2017/14778-4

Processo vinculado: **2017/05838-3**

Coordenadora: **Dra. Maria Cristina Ferreira de Oliveira**

Bolsista: **Dr. Renato Fabbri**

1 Informação sobre o nível e período de usufruto da Bolsa

A bolsa é TT5 (Treinamento Técnico nível 5), vigente de 01/Set/2017 até 30/Jun/2019. O período referente a este relatório é de 01/Jul/2018 até hoje, 30/Jul/2019. Eu, o bolsista, continuo em atividade, e foi submetido um pedido de renovação da bolsa por mais 2 meses. Um relatório final será submetido após estes 2 meses se aprovada a renovação.

2 Atividades realizadas e em andamento

Neste período, pude aplicar meus conhecimentos e proficiências nas atividades previstas no projeto desta bolsa TT5 e no projeto de pesquisa associado. As próximas seções detalham as atividades mais relevantes no período, algumas ainda em andamento.

2.1 Visualização interativa de redes bipartidas assistida por estratégias multinível

Desenvolvi, em conjunto com a coordenadora e o pesquisador Alan Valejo (doutorando no ICMC/USP) um método para visualização e navegação de redes bipartidas assistido por estratégias multinível [1]. Na sequência implementei o método em uma interface de visualização que tem se mostrado bastante eficiente, comportando inclusive que utilizemos mais recursos computacionais, pois não temos observados atrasos na interação com a visualização, mesmo manipulando grafos de larga escala. A interface está ilustrada na Figura 1. Apresentamos os resultados em um artigo que submetemos à conferência internacional “Graph Drawing and Network Visualization”. Como o artigo foi rejeitado, submeteremos a uma nova instância, provavelmente para a “International Conference on Information Visualization Theory and Applications”. A interface está acessível publicamente em <http://rfabbri.vicg.icmc.usp.br:3000/multilevel2/topdown>, e o vídeo expositivo em <https://youtu.be/-d9be7aTkUM>. Uma nova versão da interface foi desenvolvida utilizando layout em tempo real e arestas híbridas (ligando nós em níveis diferentes): <http://rfabbri.vicg.icmc.usp.br:3000/multilevel2/topdown2>, vídeo: <https://youtu.be/VdAVE4AHfi8>. Esta versão não será descrita aqui detalhadamente nem receberá figura devido à brevidade exigida para este relatório.

2.2 Visualização interativa de redes utilizando comunicabilidade

Iniciei recentemente uma colaboração com Ernesto Estrada, pesquisador notório em Redes Complexas, autor de livros e de diversas publicações relevantes para a área [2, 3, 4, 5]. O objetivo é desenvolver visualizações de redes complexas utilizando a medida de comunicabilidade proposta por Estrada para obter *layouts* 2D e 3D de redes e para a detecção de comunidades. O resultado da versão atual do sistema está ilustrado e descrito na Figura 2. Estrada considera que trata-se de contribuição relevante e já demos início à redação do artigo a ser submetido ao periódico “Information Visualization”. Ele manifestou intenção de apresentar a interface já na “Latin American Conference on Complex Networks” (5-9/Ago/2019): <http://rfabbri.vicg.icmc.usp.br:3000/communicability>.

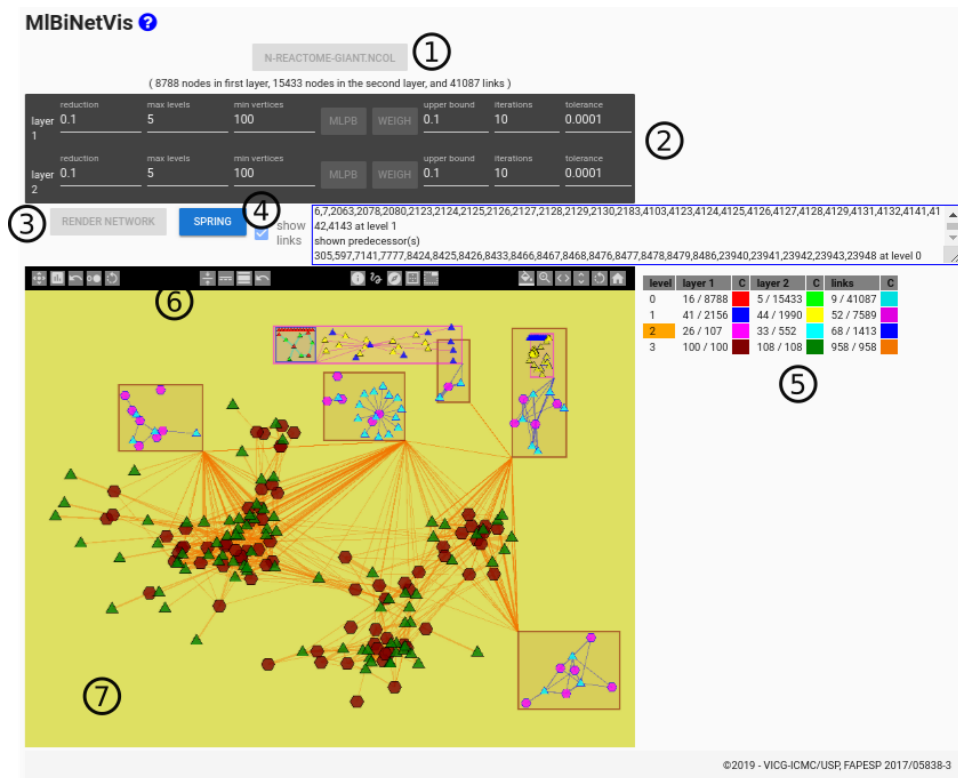


Figura 1: Interface do sistema de visualização de redes bipartidas utilizando a estratégia multinível. Esta figura é ilustrativa de uma interface de visualização de dados, com elementos para a seleção inicial dos dados e parametrização do mapeamento visual, para a especificação de transformações no mapeamento visual ao longo da interação, e para alterar as propriedades da visualização resultante. Os elementos superiores, botão no topo (1) e caixa cinza (2) permitem a parametrização dos dados de entrada do processo de mapeamento visual. Ao acionar o botão (3), a rede é renderizada, utilizando o algoritmo de *layout* de rede selecionado em (4), aonde estão reunidos outros *widgets* auxiliares para selecionar, mostrar/esconder os links, e para a interface exibir informação textual requisitada sob demanda pelo usuário ou que pode ser útil para apoiar a navegação. Em (5) tem-se uma tabela interativa com informações sobre as camadas e níveis da rede bipartida e que pode ser usada para modificar a visualização da rede apresentada no *canvas*. Em (6) está a barra de ferramentas que habilita a navegação pela rede e possibilita ajustes finos na visualização. Em (7) está o *canvas* em que a rede é desenhada na forma de diagramas nó-aresta.

2.3 Interface para visualização e análise de redes longitudinais

As redes sociais estão em constante transformação, i.e. vértices e arestas são acrescentados e removidos ao longo de sua existência. Portanto, consideramos pertinente o desenvolvimento de uma interface para análise de redes longitudinais (*longitudinal, time-evolving, dynamic networks*). Uma versão preliminar desta interface que renderiza a rede longitudinal por meio de animações e inclui algumas funcionalidades de análise está disponível em <http://rfabbri.vicg.icmc.usp.br:3000/evolution> e ilustrada na Figura 3. Submetemos um artigo com estes resultados para a conferência internacional “Graph Drawing and Network Visualization”. Como o artigo foi rejeitado, submeteremos a uma nova instância, provavelmente para a “International Conference on Information Visualization Theory and Applications”.

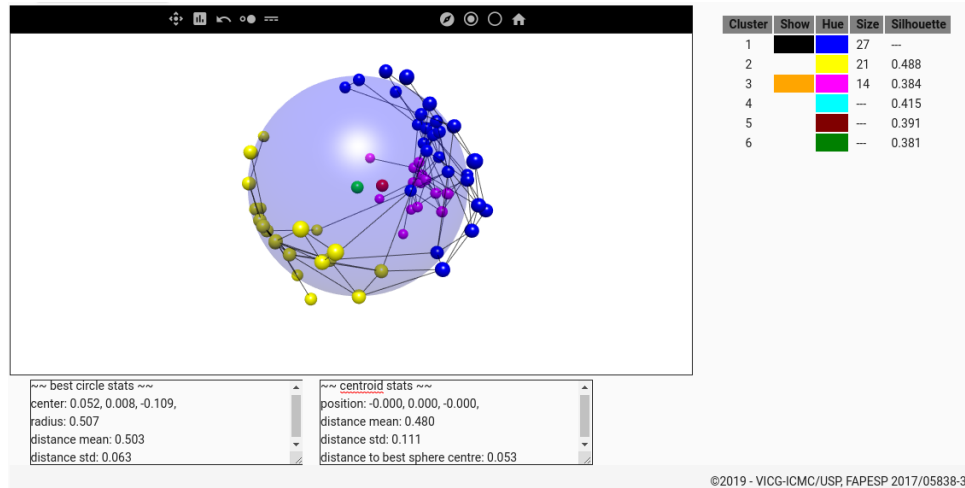


Figura 2: Layout 3D de rede obtido utilizando a medida de comunicabilidade proposta por Estrada [3, 4]. O vértice vermelho é o centroide da rede (média das posições dos vértices). O vértice verde é o centro da melhor esfera. O halo azul é a superfície da melhor esfera. Utilizando a tabela à direita, o usuário pode modificar o número de comunidades na visualização e a cor de cada comunidade. Utilizando as ferramentas na barra preta no topo do *canvas*, o usuário pode mudar o tamanho dos vértices, a transparência, a proporcionalidade do tamanho do vértice ao grau (número de vizinhos), e a transparência das arestas. Pode também mostrar ou esconder o centroide, o centro da esfera, e o halo da esfera. Pode reiniciar a visualização na configuração inicial. A interface também admite *layouts* 2D e apresenta elementos correspondentes à esfera utilizando um círculo.

2.4 Interface para visualização de redes enriquecidas com texto

Está em desenvolvimento uma interface para a análise de redes em que cada vértice possui texto associado. Esta estrutura é útil e.g. para análise de redes sociais, em que cada vértice representa um participante, e cada participante está vinculado a texto referente a suas postagens e comentários; também para redes de colaboração científica, em que cada vértice representa um pesquisador, e cada pesquisador está vinculado a texto associado a suas publicações, por exemplo. A proposta é permitir ao usuário selecionar conjuntos de vértices manualmente ou por meio do aproveitamento de métodos automatizados, e então exibir características do texto destes grupos (palavras mais incidentes, estatísticas de *synsets* da Wordnet [6], estatísticas da escrita como tamanho de palavras, percentagem de adjetivos, etc.) e também permitir a observação de diferenças entre os textos de diferentes grupos, principalmente através da distância estatística robusta, derivada do teste de Kolmogorov-Smirnov proposto anteriormente pelo bolsista [7]. Uma versão mínima desta interface já está implementada: <https://youtu.be/MH1D8S75d7E>, e pretendemos enriquecer a ferramenta com métodos já implementados em trabalhos anteriores. Pretendemos validá-la por meio de estudos de caso em redes sociais e de colaboração científica.

2.5 Visualização para auxílio à análise de paisagens sonoras

Por sugestão da coordenadora do projeto, o bolsista estudou um artigo [8] cujos autores argumentam que os métodos em uso na área de paisagens sonoras dão ênfase à descrição dos fenômenos ou no domínio do tempo ou no domínio da frequência, e por isso mostram-se incapazes de capturar informações importantes da dinâmica espectro-temporal das paisagens sonoras. Tal dinâmica espectro-temporal, segundo os autores, têm papel determinante na compreensão

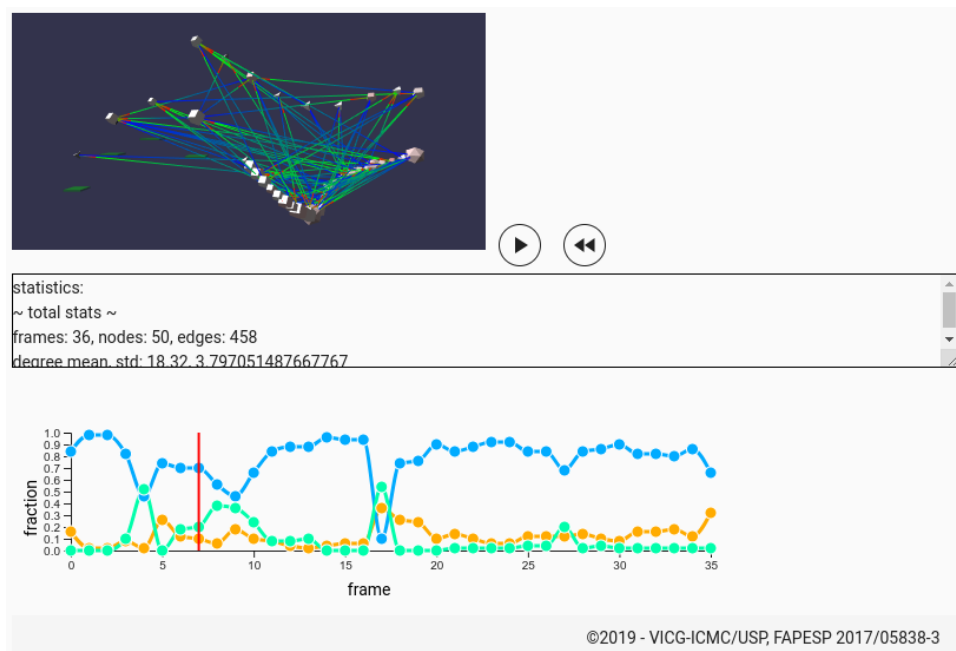


Figura 3: Na área superior da figura tem-se o canvas em que é exibida a rede, e logo abaixo um espaço para apresentação de informações textuais ao usuário, sob demanda. O gráfico na região inferior apresenta linhas do tempo com as frações de nós hubs, intermediários e periféricos. O usuário pode iniciar ou pausar a animação, e selecionar algum dos frames clicando no gráfico. Através de atalhos de teclado o usuário pode requisitar informações específicas sobre os vértices ou sobre a rede em um determinado momento.

da evolução dos ambientes ecológicos. Eles apontam a necessidade de métodos que preservem essa dinâmica espectro-temporal e sugerem um método em particular, o *Shift-Invariant Probabilistic Latent Component Analysis* em duas dimensões (SIPLC2D). Mesmo o artigo sendo relativamente recente, foi necessário atualizar os códigos disponibilizados para executar as rotinas: mudança de nome das classes, tipos numéricos, mudança na tradução de tipos de numéricos, etc. Em estudo com um arquivo de áudio com cantos de duas espécies de pássaro verificou-se que o método de fato separa as fontes sonoras, como descrito na Figura 4.

O bolsista desenvolveu uma interface que permite aos colaboradores avaliar o potencial do método no contexto dos estudos em andamento: <http://rfabbri.vicg.icmc.usp.br:3000/soundscape/min> e vídeo <https://youtu.be/bPZAGT1PzjM>. Basicamente, o usuário seleciona um trecho de áudio para análise, e a interface então apresenta as componentes detectadas. A última etapa está em desenvolvimento: o usuário poderá, por fim, aplicar a detecção das componentes a um trecho completo de áudio, de maneira a indicar todas as ocorrências encontradas. Isso é um ponto de partida para analisar a conveniência de incorporar a técnica ao ferramental de visualização sendo desenvolvido no âmbito do projeto. O bolsista estabeleceu contato com os autores do artigo [8] para explorar o método.

3 Avaliação do impacto das atividades do bolsista sobre o andamento do projeto

Desenvolvi métodos, programei interfaces, escrevi e submeti artigos. Acredito estar auxiliando no andamento do projeto, em conformidade com a bolsa TT5.

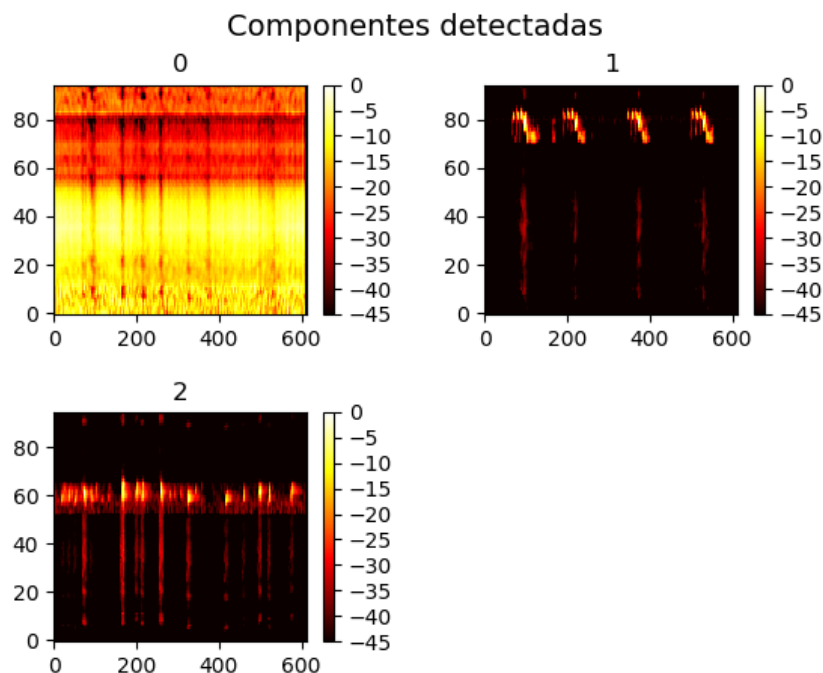


Figura 4: Espectrogramas do áudio reconstruídos utilizando apenas as componentes detectadas pelo método SIPLC2D. O espectrograma 0 corresponde ao ruído de fundo, o espectrograma 1 corresponde ao canto de uma de duas espécies de pássaros presentes no áudio original, o espectrograma 2 corresponde ao canto da outra espécie. A separação não é perfeita, mas suficiente para motivar investigações adicionais no contexto do estudo de paisagens acústicas em ecologia.

4 Mudanças ao Plano de Atividades

Foi solicitada prorrogação do projeto de pesquisa ao qual esta bolsa é vinculada até dezembro de 2019. Solicitamos a prorrogação da bolsa TT5 por dois meses, completando 24 meses. Espera-se assim viabilizar as finalizações previstas e as demais contribuições planejadas na Seção 2.

5 Apreciação do desempenho do bolsista (escrita pela coordenadora)

O bolsista tem grande conhecimento em métodos na área de redes complexas e em programação. Assim, tem contribuído significativamente no desenvolvimento de ferramentas que mostram-se fundamentais como provas-de-conceito de diversas técnicas investigadas no âmbito do projeto. Ele tem interagido mais intensamente com os pesquisadores que atuam na frente de pesquisa em métodos multinível em redes complexas, com potencial para interagir também com pesquisadores que atuam na frente de pesquisa em paisagens acústicas. Acredito também que o projeto apresentou a ele uma nova gama de conhecimentos em visualização de dados, abrindo novas perspectivas para a sua atuação futura como pesquisador. Minha avaliação é que o bolsista está contribuindo de maneira relevante para ampliar os resultados gerados no âmbito do projeto de pesquisa.

Referências

- [1] Alan Valejo, Maria Cristina Ferreira de Oliveira, PR Geraldo Filho, and Alneu de Andrade Lopes. Multilevel approach for combinatorial optimization in bipartite network. *Knowledge-Based Systems*, 151:45–61, 2018.
- [2] Ernesto Estrada. The communicability distance in graphs. *Linear Algebra and its Applications*, 436(11):4317–4328, 2012.
- [3] Ernesto Estrada. Community detection based on network communicability. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):016103, 2011.
- [4] Ernesto Estrada, Maria Fox, Desmond J Higham, and Gian-Luca Oppo. *Network science: complexity in nature and technology*. Springer Science & Business Media, 2010.
- [5] Ernesto Estrada and Philip A Knight. *A first course in network theory*. Oxford University Press, USA, 2015.
- [6] George Miller. *WordNet: An electronic lexical database*. MIT press, 1998.
- [7] Renato Fabbri and Fernando Gularte De León. A statistical distance derived from the kolmogorov-smirnov test: specification, reference measures (benchmarks) and example uses. *arXiv preprint arXiv:1711.00761*, 2017.
- [8] Alice Eldridge, Michael Casey, Paola Moscoso, and Mika Peck. A new method for ecoacoustics? toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. *PeerJ*, 4:e2108, 2016.