

Visualização Gráfica de Grafos com Comunidades Especificadas usando o Algoritmo RelaxMap

Dhruv Babani¹, Bernardo Balzan², Eduardo Cardoso³

¹Pontificia Universidade Catolica do Rio Grande do Sul(PUCRS)

d.babani001@edu.pucrs.br, b.balzan@edu.pucrs.br, Eduardo.S00@edu.pucrs.br

Resumo. *Este artigo apresenta um recurso de visualização gráfica de grafos com as suas comunidades especificadas, usando o algoritmo de RelaxMap. Sabendo disso, isso traz um melhor entendimento nos resultados que são obtidos em estudos que há grafos.*

Abstract. *This article presents a graphical visualization resource for graphs with their specified communities, using the RelaxMap algorithm. Knowing this, this brings a better understanding of the results obtained in studies that contain graphs.*

1. Introdução

Atualmente, a temática do desempenho de algoritmos que manipulam estrutura de dados, tais como grafos, atua como um intermédio para geração de resultados com a finalidade de fazer análises descritivas para fins acadêmicos. Demonstrando uma área de estudo com muito potencial e englobando áreas mais abrangentes, como a ciência de dados e as metodologias ágeis, que tem crescido de forma exponencial nos últimos anos.

Entretanto, o estudo apresenta a visualização dos grafos com as suas comunidades detectadas, utilizando tanto o algoritmo de RelaxMap quanto a biblioteca Toyplot do Python. Sabendo disso, é realizado a comparação dos dois resultados e são obtidos algumas considerações finais que vão ser abordadas no final do estudo.

2. Fundamentação Teórica

Com o intuito de realizar os estudos, entramos em contato com o Orientador Carlos Cano e o Professor César De Rose, onde eles nos ajudaram, mostrando séries de possibilidades para cumprir os objetivos da pesquisa.

Além disso, foi realizado um contato com o Gabriel Giordani, o autor do estudo "Towards a Scalable Parallel Infomap Algorithm for Community Detection"[1], com o intuito de oferecer o algoritmo de RelaxMap já implementado, para ser apto a visualizar os grafos com as suas comunidades

2.1. Seleção de Ferramentas

Através do contato realizado e as conclusões obtidas, foram escolhidas duas ferramentas para visualização de grafos:

1. NodeXL – Aplicativo
2. Infomap Online – WEB

Nesse sentido, foram testadas as respectivas alternativas com o propósito de visualização, seguindo os critérios visuais para detecção de comunidades, como por exemplo, a divergência de cores em cada comunidade ou o entendimento de cada vértice e aresta do grafo.

Sabendo disso, após todos os testes, apenas a alternativa do Infomap Online foi realmente acolhida, pelo devido fato de corresponder os seguintes critérios mencionados anteriormente. Então, com mais informações, o InfoMap é uma ferramenta que já aplica o algoritmo do Infomap, na qual faz o processo de detecção de comunidades através da entrada de um arquivo texto onde contém os dados dos vértices de origem e destino do respectivo grafo.

2.1.1. Explorando o Python

Após a escolha o Infomap, como a ferramenta primária, foi realizado mais pesquisas, com o intuito de aprimorar a visualização da estrutura de dados. Logo foi encontrado uma biblioteca da linguagem Python para manipulação de Grafos e Redes, chamada Toyplot.

Nesse sentido, percebemos que essa biblioteca está a par com os critérios definidos para visualização, além disso ela possui uma vasta implementação a séries de algoritmos de detecção de comunidades. Portanto, decidimos que vai ser feito o uso dessa biblioteca para cumprir o propósito da pesquisa.

2.2. Algoritmo RelaxMap

Após a escolha das ferramentas, seguimos para o próximo passo do estudo. Dito isso, como foi mencionado anteriormente sobre o contato com o Gabriel Giordani, utilizamos o algoritmo Relaxmap para gerar os arquivos de entrada (.clu e .tree) para a ferramenta Infomap Online, e após gerar o grafo resultante.

Com isso, foi realizado um teste de visualização entre o algoritmo de detecção da ferramenta e o do Gabriel Giordani, e percebemos que o output de grafos dos dois algoritmos apresenta um comportamento muito similar.

Sabendo disso, decidimos em optar pelo algoritmo do Gabriel Giordani e utilizar a ferramenta apenas para visualização.

3. Experimentos oficiais

Agora que temos todos os recursos para realizar os experimentos. Sabendo disso, o experimento em si, se divide em três passos:

1. Seleção dos Grafos
2. Implementação do Script em Python, usando o Toyplot
3. Representação visual dos Grafos no Infomap Online

3.1. Seleção dos Grafos

Como o objetivo do estudo é apenas fazer a visualização dos grafos, então optamos por escolha do uso de grafos de tamanho pequeno. Sabendo disso, fizemos uma pesquisa a procura de um bando de dados de grafos, e achamos o networkrepository, onde baixamos apenas dois grafos para a pesquisa. Dito isso, aqui estão os detalhes:

Graph	Nodes	Edges
bio-celegans-dir	453	4.6k
ENZYMES-g228	34	152
ia-southernwomen	50	230

3.2. Visualização - Toyplot e Infomap Online

Com os grafos selecionados, agora seguimos para a etapa final do experimento, que na qual é a visualização dos grafos pelo Toyplot e o Infomap Online. Aqui a seguir será mostrado os detalhes para a extração do resultado por eles.

3.2.1. Toyplot

Como em análise estava sendo o algoritmo RelaxMap, adaptamos o script em Python para o algoritmo. Então, em primeira parte, tivemos que implementar uma faixa de conversão, para que os resultados obtidos pelo Networkkit, pudessem ser aplicados nas estruturas necessárias para visualização.

Com isso, foi possível obter todos os nodos de vértices com as suas respectivas comunidades de forma correta. Depois disso, foi analisado que os grafos são compostos por todas as arestas (Edges), mas para que a visualização ocorresse, deve existir uma lista de tuplas, onde a primeira posição corresponde o vértice, e a segunda posição, o id do elemento na comunidade em que ele pertence.

Após disso, iniciamos a implementação para a faixa de visualização. Nesse sentido, essa secção foi simples, achamos um método (`graph()`) dentro da documentação da biblioteca, que realiza a geração de uma imagem com o grafo e as suas comunidades

3.2.2. Infomap Online

O Infomap Online é uma ferramenta WEB, então o procedimento para gerar os grafos visuais foi basicamente dividido em seguintes passos:

- Gere os arquivos clu e ftree, compilando o código do RelaxMap oferecido pelo Gabriel Giordani
- Após, escolha o arquivo ftree, onde contém todos os vértices com as suas comunidades, e coloque no parâmetro de entrada da ferramenta
- Depois disso, o grafo é originado.

3.3. Avaliação

Com o término do experimento da pesquisa, tivemos seguintes levamentos em relação à experiência durante o estudo por inteiro. Então os critérios estabelecidos foram:

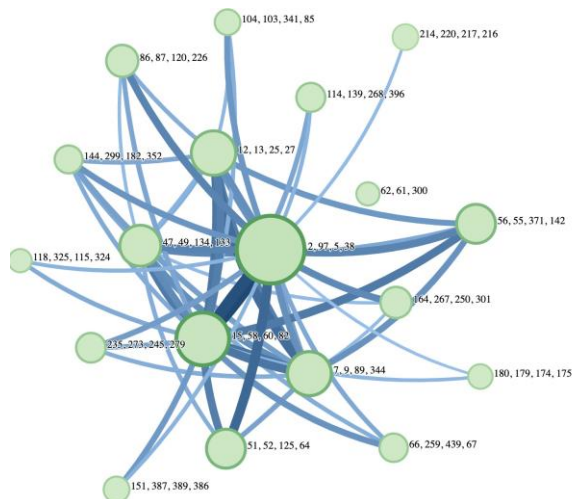
- Dificuldade na implementação do Script Python, utilizando o Toyplot
- Facilidade ao gerar os grafos pelo Infomap Online
- O entendimento dos grafos visuais são perfeitos.

4. Resultados

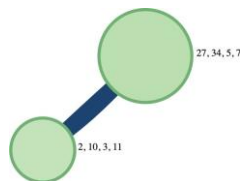
Agora com os scripts prontos e o Infomap Online inicializado, eles foram executados e os grafos foram gerados com sucesso tanto na interface gráfica (Jupyter lab) do Toyplot quanto no Infomap Online. Logo abaixo, está todos os grafos processados:

1. Grafos - Infomap Online

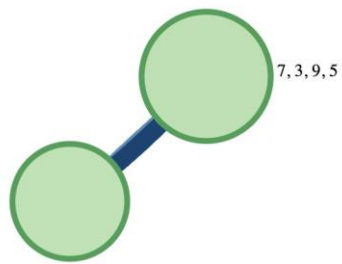
bio-celegans-dir:



ENZYMESg228:

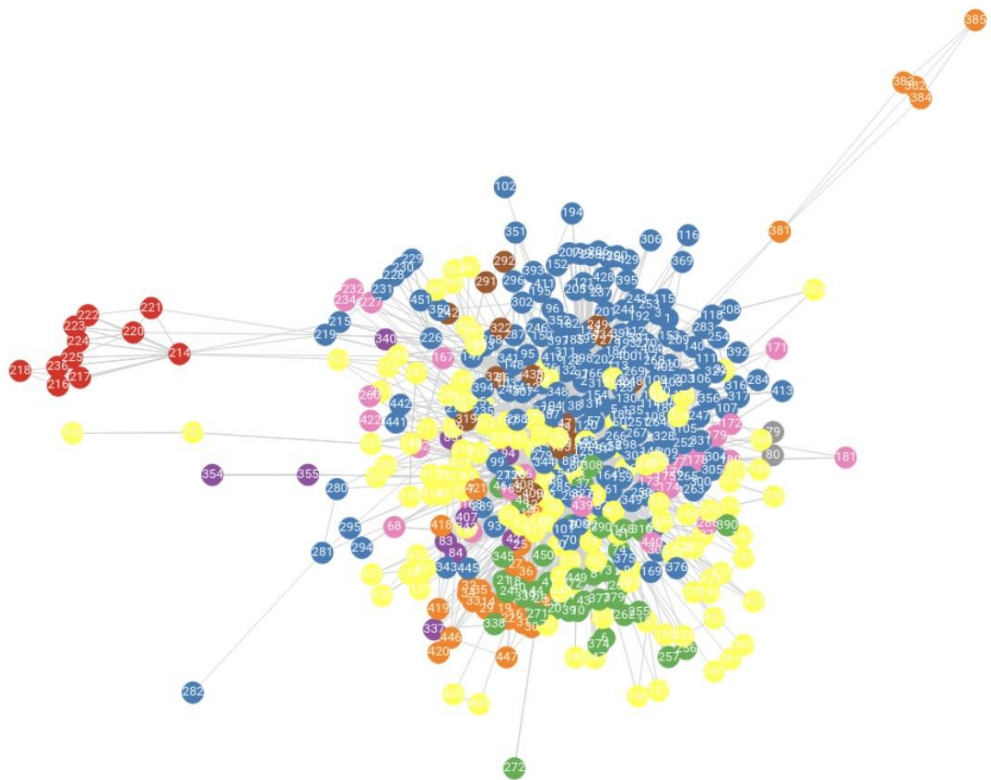


IA-southernwomen:

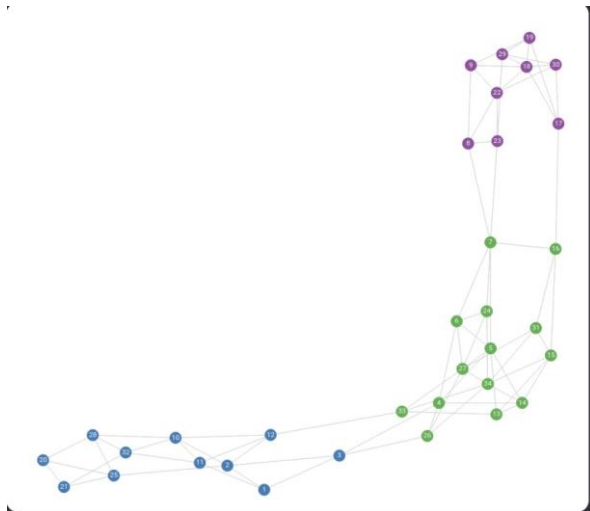


2. Grafos - Toyplot

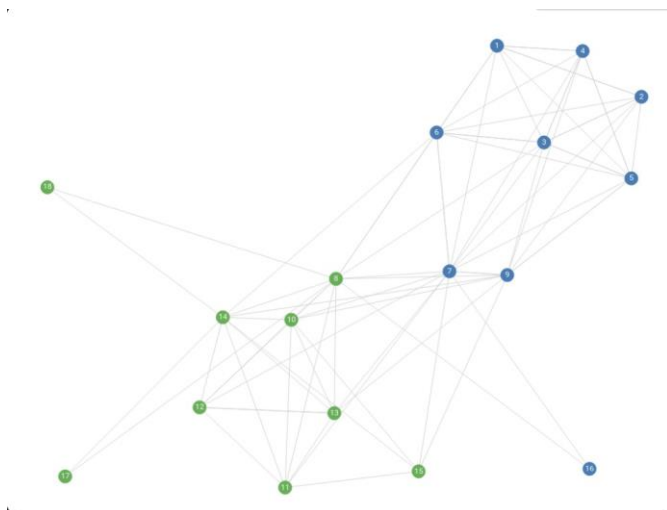
bio-celegans-dir:



ENZYMESg228:



A-southernwomen:



Como pode ser observado, os resultados gerados pelo Infomap Online armazena os vértices de cada comunidade em formato de um nodo. Já os resultados obtidos pelo Toyplot, as comunidades são especificadas através de diferentes cores.

5. Conclusões

Á nossa perspectiva, o estudo realizado neste trabalho foi muito gratificante, porque nos levou a concluir que a visualização dos grafos é realmente uma ótima alternativa, quando queremos entender um cenário dentro do universo da computação, principalmente no estudo de Escalabilidade Paralela.

Além disso, podemos concluir que entre os dois tipos de visualização gerados, os dois apresentam resultados bem satisfatórios em relação a visualização dos grafos com as suas comunidades especificadas, primeiro sendo com nodos, e o segundo através de cores. Logo, podemos afirmar que os dois resultados se comportam de forma complementar e são válidos para realizar visualização de Grafos.

References

- [1] Giordani, G. Towards a scalable parallel infomap algorithm for community detection. pages 1–10. Porto alegre, Brazil.(Artigo em preparação)