

Exercício Prático 03

Prazo para a Entrega: Definido na Programação de Aulas e na Tarefa do Classroom para esta prática

Formato de Entrega: Através do formulário anexado a Tarefa do Classroom para esta prática

Reposição: Prazo definido na programação de aulas

Usando o código fornecido em <https://repl.it/@pdlm/EPG03-Template> e no repositório GraphTheory-JGraphT (ver folder src/main/java/epgs e src/test/java/graphs), analise o problema abaixo respondendo as questões com base nos dados fornecidos através da execução do referido código.

Rede de Troca de E-mails

Neste exercício, estudaremos os dados de troca de email em uma instituição de pesquisa europeia obtidos neste link: (<https://snap.stanford.edu/data/email-Eu-core-temporal.html>)

Como a base original é muito extensa, estaremos utilizando os dados disponíveis neste link:

https://drive.google.com/file/d/1eS5PS-cNgZYouDJINWY_y9mBHKST6sTo/view?usp=sharing

Para a execução das questões abaixo, use os seguintes arquivos:

- Grafo no formato graphml (Para carregar na yEd Live ou yEd desktop):
<https://drive.google.com/file/d/1eLksa9J20SB0kEwIGcmLy21dpoBokxD/view?usp=sharing>
- Grafo no formato gml (Para carregar no código JGraphT ou yEd desktop):
<https://drive.google.com/file/d/1ELbBbaq9pv4C7NMTsg9VDkYI-nd8Asn8/view?usp=sharing> (este arquivo já está disponível no repl.it e repositório JGraphT-GraphTheory).

Questões

1. Carregue o grafo na ferramenta yEd e encontre uma visualização gráfica adequada para o mesmo. Justifique escolha com base no domínio do problema. Exporte o resultado no formato de imagem (.jpg ou .png)

Segundo a documentação do site que fornece a ferramenta yEd, a representação Organic geralmente expõe a simetria inerente em grafos de modo que estruturas de clusters (com distribuição de vértices bem equilibrada e poucos cruzamentos em arestas) são geradas.

Além disso, embora a representação Organic seja indicada para grafos não-direcionados, sabe-se que as métricas de Clustering são calculadas sobre o grafo base. Assim, objetivando-se destacar os “Natural Clusters” contidos no grafo em análise, optou-se pela visualização gráfica do tipo Organic.

Fonte:

https://yed.yworks.com/support/manual/layout_smartorganic.html

2. Que usuários podem ser considerados bons disseminadores de informação nesta rede (*hubs*)? Indique os 5 com maior *score*. Justifique observando o domínio do problema e indicando a(s) métrica(s) considerada(s).

Para determinar quais usuários apresentam maiores níveis de disseminação de informação, deve-se quantificar a influência que os vértices (representativos dos usuários) tem sobre o fluxo de informação no grafo correspondente. Sabendo-se que a métrica de centralidade *Betweenness Centrality* fornece a extensão da presença de um vértice nos menores caminhos entre pares de outros vértices, pode-se estabelecer mediante ao uso de tal parâmetro a medida do controle que cada vértice tem na disseminação de informação. Por conseguinte, determina-se os bons disseminadores de informação da rede.

Assim, mediante a execução do código fornecido, obtém-se os cinco *hubs* com maior *score* em ordem decrescente, de acordo com a *Betweenness Centrality* de cada vértice, como a seguir:

1. Cleary: 0.13;
2. Burns: 0.09;
3. Bicknell: 0.07;
4. Bain: 0.05;
5. Bice: 0.04.

Fontes:

<https://www.sci.unich.it/~francesco/teaching/network/betweenness.htm>

<https://neo4j.com/docs/graph-algorithms/current/labs-algorithms/betweenness-centrality/>

<https://www.hindawi.com/journals/ijcom/2014/241723/>

3. Que usuários são mais influentes? Indique os 5 mais influentes. Justifique observando o domínio do problema e indicando a(s) métrica(s) considerada(s).

Para determinar quais usuários são mais influentes deve-se avaliar as conexões de cada usuário com os outros usuários. De modo que, para dois usuários com o mesmo número de conexões, um usuário será mais influente que outro se as conexões de um dos usuários forem mais influentes que as conexões do outro.

A métrica de centralidade *EigenVector Centrality* permite mensurar a influência de um vértice em um grafo com base em pontuação. O princípio desta mensuração estabelece que conexões a vértices com alta pontuação contribuem mais para a pontuação de um vértice de que conexões a vértices com baixa pontuação. Assim, pode-se utilizar tal parâmetro para determinação dos usuários mais influentes na rede.

Como a medida de centralidade Alpha-centrality é um tipo específico de *EigenVector Centrality* disponível na JgraphT, pode-se utilizar este parâmetro para o estabelecimento dos cinco usuários mais influentes dispostos abaixo:

1. Arrault: 2.26;
2. Bourne: 2.18;
3. Berntssen: 1.70;
4. Chartman: 1.66;
5. Adenet: 1.54

Fontes:

[http://en.wikipedia.org/wiki/Centrality#:~:text=Eigenvector%20centrality%20\(also%20called%20eigencentrality,connections%20to%20low-scoring%20nodes](http://en.wikipedia.org/wiki/Centrality#:~:text=Eigenvector%20centrality%20(also%20called%20eigencentrality,connections%20to%20low-scoring%20nodes)

<https://www.strategic-planet.com/2019/07/understanding-the-concepts-of-eigenvector-centrality-and-pagerank/>

<https://www.sciencedirect.com/topics/computer-science/eigenvector-centrality>

4. O perfil de envio de e-mails dos usuários é semelhante? Justifique observando o domínio do problema e indicando a(s) métrica(s) considerada(s).

Já que os vértices do grafo representam usuários da rede, verificar se o envio de e-mails dos usuários da rede é semelhante equivale a verificar se os vértices interconectados no grafo apresentam graus semelhantes.

A métrica de conexão denominada Assortatividade, é utilizada para quantificar a semelhança no padrão de conexão de cada par de vértices a partir de um coeficiente assortativo baseado em alguma medida de similaridade (geralmente os graus dos vértices). O coeficiente assortativo r assume valores entre -1 e 1, de modo que quando $r = 1$ o grafo é perfeitamente assortativo, quando $r = 0$ não é assortativo, e quando $r = -1$ o grafo é completamente não assortativo.

Nota-se que a classe MeasureUtil, utilizada no programa fornecido, possui um método denominado assortativityCoefficient baseado nos graus dos vértices do grafo, logo o

valor de assortatividade do referido programa pode ser usado para avaliar o perfil de e-mails dos usuários. Como o valor resultante é 0.22, o grafo apresenta baixa assortatividade, e consequentemente, o perfil de envio de e-mails dos usuários é pouco semelhante.

Fontes: <https://en.wikipedia.org/wiki/Assortativity>

<https://methods.sagepub.com/base/download/DatasetStudentGuide/assortativity-in-ukfaculty-2008-python>,

5. É possível identificar comunidades ou clusters diferentes? Justifique observando o domínio do problema e indicando a(s) métrica(s) considerada(s).

É notório que Coeficiente de Clustering é uma métrica de segmentação que quantifica a tendência dos vértices de um grafo de se organizarem em clusters. Adicionalmente, sabe-se que existem duas versões deste parâmetro: global, e local. A versão global estabelece a tendência geral de formação de clusters em um grafo, enquanto que a versão local estabelece a tendência de integração entre vértices individuais.

Ao analisar o programa fornecido, nota-se que a implementação do Coeficiente de Clustering é correspondente a versão global. Logo, tal coeficiente evidencia que embora existam comunidades que são altamente conectadas internamente (como pode ser constatado devido aos “Natural Clusters” estabelecidos pela ferramenta yEd), a tendência destas comunidades formarem *clusters* entre si é baixa (como pode também pode ser constatado com a ferramenta yEd devido à baixa interação entre os clusters). Assim, pode-se visualizar 6 diferentes clusters, porém com baixo acoplamento entre si.

https://en.wikipedia.org/wiki/Clustering_coefficient