



1

## Introdução

O projeto final vai implementar um TAD Grafo com métodos de consulta, inserção, remoção e pesquisa, que vai ser usado numa aplicação prática, usando dados de ligações (seguido-seguidor) do [GitHub](#).



Os grafos são uma das estruturas de dados mais comuns e a modelação de problemas baseada em grafos é um componente chave em muitos dos sistemas que usamos todos os dias, tais como o encaminhamento no Google Maps ou a recomendação de amigos nas redes sociais.

As redes de partilha e colaboração são importantes nos mais variados domínios (pesquisa científica, construção de novos produtos ou serviços, convívio, informação, saúde, governação, etc.) e a rede do GitHub é apenas um exemplo. A Análise de Redes Sociais é um ramo da sociologia que explora estruturas sociais através do uso de ferramentas analíticas, tais como grafos. Nesta tarefa, irá também implementar algoritmos básicos para a análise de redes sociais aplicando-os num grafo com informações extraídas da rede social do GitHub.

2

Em geral, quando se estuda uma destas redes, pretende-se identificar vértices influentes ou importantes. A identificação destes vértices usa medidas de centralidade com base nas diferentes noções de importância dos vértices ou arestas.

Algumas destas medidas são: a centralidade de grau (degree **centrality**), a de proximidade (**closeness**) e a de intermediação (**betweenness**). Nestes casos, a importância de um vértice deve-se ao seu número de ligações e à assunção de que importante será um nó por onde muitos dos caminhos no grafo passem.



#### Degree Centrality:

A **centralidade de grau** um vértice é calculada como a quantidade de vizinhos (nós adjacentes), ou seja o grau de um nó. Se o grafo for dirigido, é importante saber a centralidade de entrada e a de saída, uma vez que a primeira mede a popularidade do nó e a segunda o grau de influência.

#### Closeness:

A centralidade de **proximidade** de um vértice é calculada como o inverso da soma das distâncias do vértice a todos os restantes. Quanto mais próximo estiver um nó de todos os restantes, mais importante é na rede.

#### Betweenness:

A centralidade de **intermediação** de um vértice é calculada como a quantidade de vezes que o vértice intervém nos caminhos mais curtos entre os restantes. Um nó por onde passam muitos dos caminhos da rede tem, nessa rede, um papel central.

3

## Descrição do trabalho: estudo e implementação – fase 1

### • Fase 1:

1. **Implementação do TAD Grafo** numa classe em Python, de acordo com a implementação prática nos exercícios do **módulo 7**: ou 3.a) (usando dicionários de vértices e de adjacências) ou 3.b) (um dicionário de vértices e uma estrutura de lista ligada para as adjacências implementada pelo grupo). Devem usar a interface e as classes Vertex e Edge, como indicado na aula.
  
2. **Preparação de um método de carregamento de dados de um ficheiro csv** qualquer, que obedeça ao seguinte formato:
  1. por linha existem 3 valores de dados - o 1.<sup>º</sup> e o 2.<sup>º</sup> indicam nomes de vértices e o 3.<sup>º</sup> um peso.
  2. A 1.<sup>ª</sup> linha do ficheiro indica o nome das colunas.

**Nota importante:** caso não exista 3<sup>a</sup> coluna, devem assumir que o grafo não é pesado e lidar com essa situação sem erro.
  
3. **Proceda ao Carregamento de dados do ficheiro Github.csv (no e-Learning).**

4

## Descrição do trabalho: estudo e implementação – fase 1

4. **Estudo** de algoritmos para resolver a determinação do(s) **caminho(s) mais curtos** num grafo (com e sem pesos nas arestas)
  5. **Implementação** de métodos para determinar caminhos mais curtos num grafo: (a) sem usar os pesos nas arestas); (b) usando os pesos nas arestas
  6. **Implementação**, pelo menos, das **medidas de centralidade**: centralidade de grau (**degree centrality**) e centralidade de proximidade (**closeness**).
  7. Indique:
    - a) O Top 10 dos vértices mais interligados na sua rede.
    - b) O Top 10 dos vértices que apresentam maior proximidade aos restantes.
    - c) Outras análises que julgue interessantes.
- Nota: para leitura de dados pode ser usado o módulo CSV do Python.
  - **Atenção: não podem** recorrer a módulos de computação em redes, tais como Networkx ou outras, para substituir a sua implementação. Podem, no entanto, usar algum desses módulos para os seus testes, para comparação e validação dos seus resultados.

5

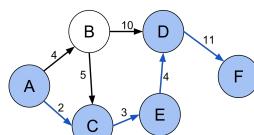
## Para a procura do caminho mais curto

Na segunda fase do projeto será necessário implementar um algoritmo que permita determinar qual o caminho mais curto entre quaisquer dois vértices de origem e destino.

Este é um problema com aplicações práticas óbvias, não apenas em análise de redes sociais, mas também para navegação em aplicações de transportes, na gestão de tráfego de dados em redes de computadores, ou até, em trocas financeiras (como, por exemplo, de moeda).

Para uma ilustração de alguns métodos de determinação de caminhos mais curtos podem usar, por exemplo,

- Cap. 14, Data Structures and Algorithms in Python, Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser, especialmente as secções 14.3.3 e 14.6;
- Materiais indicados no e-Learning.



6

## Para informação sobre medidas de centralidade

Na segunda fase do projeto terão ainda que implementar métodos de cálculo de medidas de centralidade.

Mais uma vez, estas medidas têm utilidade em outras aplicações para além da análise de redes sociais, como em redes de transportes, fragilidades de redes de conectividade, entre outras áreas de aplicação.

Para uma ilustração podem consultar, por exemplo,

- <https://core.ac.uk/download/pdf/303042278.pdf>
- <https://arxiv.org/pdf/1901.07901.pdf>
- [https://ocw.mit.edu/courses/civil-and-environmental-engineering/1-022-introduction-to-network-models-fall-2018/lecture-notes/MIT1\\_022F18\\_lec4.pdf](https://ocw.mit.edu/courses/civil-and-environmental-engineering/1-022-introduction-to-network-models-fall-2018/lecture-notes/MIT1_022F18_lec4.pdf)

7

## Relatório:

- Execução:

Desenvolvimento da implementação, teste, escrita e entrega de relatório.

- Relatório:

- Capa onde deve constar: o número do grupo e a sua composição, o nome da unidade curricular e um título para o trabalho.
- Utilização de secções onde devem documentar, SINTETICAMENTE:
  - \* Uma introdução (contextualização de possíveis utilizações da implementação construída e outros pontos relevantes para a compreensão do trabalho).
  - \* Secções separadas com **os algoritmos** que utilizaram nos pontos 5) e 6).
  - \* Apresentação dos resultados de testes, incluindo resultados não pedidos mas que julgem ser interessantes ou possam ajudar a documentar o trabalho produzido.
- A bibliografia e webgrafia consultadas para a elaboração do trabalho.

8