

Estruturas de Dados e Algoritmos

Trabalho de Grupo Final:
Utilização de Grafos para estudo de
comunidades na rede social *Facebook*

16 de Maio de 2022

Equipa de docentes de EDA

LCD, 2021-2022

Versões com correções
* 18/5/22, corrigida a data
de entrega que é 1/Junho



Objetivos

O trabalho final tem como objetivos principais:

- 1) Implementar um TAD Grafo, com uma função para obter uma árvore de cobertura mínima (usando o algoritmo de Kruskal).
- 2) Aplicar o TAD Grafo para estudar as comunidades que se podem obter de um conjunto de dados sobre *profiles* e *circles (friends lists)* no *Facebook* (<https://www.facebook.com/>); para tal será preciso:
 - a) desenvolver um algoritmo de agrupamento (*clustering*) hierárquico;
 - b) criar os (sub)grafos para cada *cluster*;
 - c) usar o NetworkX (<https://networkx.org/>) para o estudo de sub-comunidades em cada comunidade (representada por um grafo).



<https://fortune.com/2015/11/09/barack-obama-facebook/>

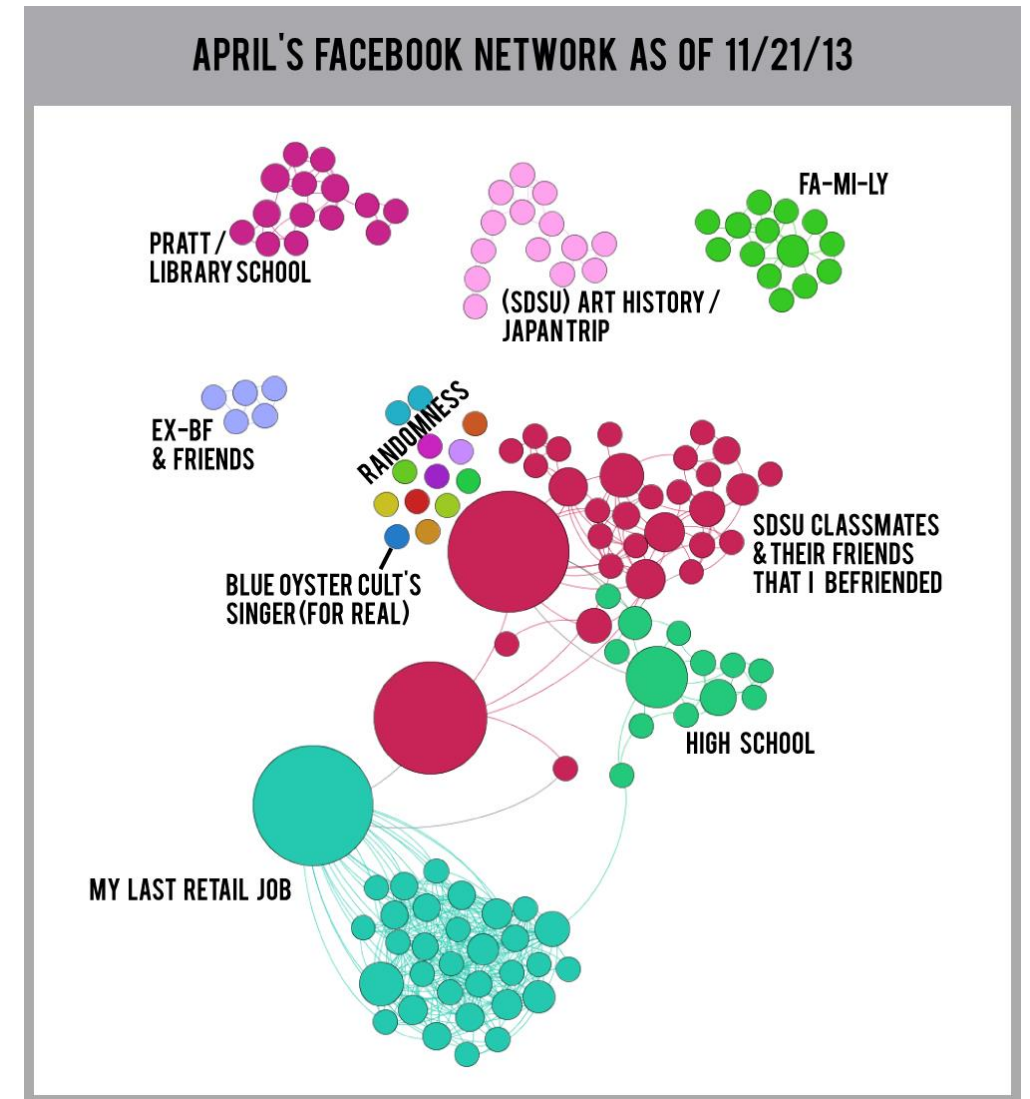


<https://fmsasg.com/socialnetworkanalysis/facebook/>

Introdução

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 *Facebook* 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 *Facebook*, para o estudo de comunidades e sub-comunidades no *Facebook*.



Descrição do trabalho – fase 1

- a) **Implementar o TAD Grafo numa classe em Python**, de acordo com a implementação feita em aula nos exercícios do módulo 10, exercícios 1, 2 e 3. a) (usando a representação de mapas de adjacências). Devem usar a interface e as classes Vertex e Edge, como indicado na aula.

- b) **Estudar e implementar o algoritmo de Kruskal**, para obter uma árvore de cobertura mínima (https://algorithms.discrete.ma.tum.de/graph-algorithms/mst-kruskal/index_en.html). Esta função deverá ser um método da classe que implementa o TAD Grafo.

Descrição do trabalho – fase 2

Aplicar o TAD Grafo para estudar as comunidades que se podem obter de um conjunto de dados sobre *profiles* e *circles* (*friends lists*) no *Facebook* (<https://www.facebook.com/>). Para tal será preciso:

- a) **Estudar e implementar em Python** um algoritmo de agrupamento (*clustering*) hierárquico.
- b) **Criar os (sub)grafos para cada *cluster***, obtidos pela aplicação do algoritmo desenvolvido em a) aos dados sobre *profiles* e *circles* em *Facebook* fornecidos no ficheiro *facebook_netwrok.csv*. Isto é:
 - i. Carregar os dados do ficheiro *facebook_netwrok.csv*. Para leitura de dados pode ser usado o módulo CSV do Python.
 - ii. Implementar em Python a aplicação do algoritmo de *clustering* hierárquico, desenvolvido em a), para obter o grafo para cada *cluster*.

Descrição do trabalho – fase 2

- c) **Usar o NetworkX** (<https://networkx.org/>, <https://pypi.org/project/networkx/>) para o **estudo das sub-comunidades em cada comunidade** (representada por cada grafo obtido na alínea b). Fazer o estudo do seguinte modo:
- i. Obter uma **visualização dos grafos** obtidos na alínea b) usando NetworkX.
 - ii. **Aplicar e estudar duas funções para computar e medir a estrutura das comunidades** aos grafos obtidos em b). Escolher estas duas funções de entre as disponíveis seção 3.14. Communities do documento *NetworkX Reference Release 2.8* (<https://networkx.org/documentation/stable/reference/algorithms/community.html>).

Atenção

- **Não podem** recorrer a módulos de computação em redes, tais como NetworkX ou outras, para substituir as vossas implementações. Podem, no entanto, usar algum desses módulos para os vossos testes, para comparação e validação dos vossos resultados.
- **A partilha de código** em trabalhos de grupos diferentes **será seriamente penalizada**. Serão usados os meios habituais de verificação de plágio. O código entregue deve ser integralmente da autoria dos membros do grupo.

Relatório

- Capa onde deve constar: o número do grupo e a sua composição, o nome da unidade curricular, um título para o trabalho e a data. Os alunos devem trabalhar agrupados do mesmo modo que para o trabalho de grupo intermédio. Se houver alguma razão para não seguir esta regra a autorização deve ser pedida por mail enviado a: maria.albuquerque@iscte-iul.pt.
- Utilização de seções onde devem documentar, SINTETICAMENTE:
 - * Uma introdução (contextualização de possíveis utilizações da implementação construída e outros pontos relevantes para a compreensão do trabalho).
 - * Secções separadas para as diferentes fases e sub-fases **do trabalho produzido**.
 - * Apresentação de justificações e resultados, incluindo resultados não pedidos mas que julguem ser interessantes ou possam ajudar a documentar o trabalho produzido.
- A bibliografia e webgrafia consultadas para a elaboração do trabalho (ver slide seguinte).

Referências consultadas

- Todas as referências e recursos consultados devem ser referenciados no relatório e, caso faça sentido, na apresentação. Podem encontrar mais informação em:

<https://libguides.library.usyd.edu.au/citation>

(se escolherem um estilo escolham IEEE).

- A seguinte tabela pode também ser útil:

Common elements for inclusion											
	Author	Year of publication	Title of article/chapter	Title of Publication	Issue information (vol/part: if known)	Place of publication	Publisher	Edition	Page number(s)	URL/dio	Date accessed/downloaded
Book	✓	✓		✓		✓	✓				
Chapter of book	✓	✓	✓	✓		✓	✓		✓		
Ebook	✓	✓		✓						URL if required	✓
Journal article (print & electronic)	✓	✓	✓	✓	✓				✓	doi if required	
Webpage	✓	✓		✓						✓	✓
Newspaper article (print & electronic)	✓	✓	✓	✓	✓				✓		

Entrega do trabalho

- Os alunos devem trabalhar agrupados do mesmo modo que para o trabalho de grupo intermédio. Se houver algum problema relativo a esta regra deve ser discutido por mail enviado a: maria.albuquerque@iscte-iul.pt.
- Entrega até 4ª feira, 1 de Junho de 2022 às 23h59 de um ficheiro zip com o relatório e ficheiros .py para as implementações.
- Entrega no *Blackboard* no link de cada grupo acessível a partir da seção Grupos de trabalho.
- Apresentação e discussão oral individual na 4ª feira, 8 de Junho de 2022. O horário para cada grupo será disponibilizado no *Blackboard* mais perto da data.
- Falha na entrega de relatório ou recusa de apresentação ou discussão significa classificação de F neste trabalho.