

# Análise de Redes Avançada

Projeto Final para Avaliação

Ano 2024/2025

## 1 Introdução

Este projeto final tem como objetivo colocar em prática o conhecimento de análise de redes que foi ensinado em aula, nomeadamente ao nível da manipulação, exploração, visualização e tratamento de dados por elas representados. Seu objetivo principal é avaliar a capacidade de pensar sobre as entidades e as relações que as redes representam e retirar informações/conclusões que se podem obter ao explorar os dados usando conceitos próprios da Ciência das Redes.

Não é objetivo deste projeto avaliar a vossa capacidade de programação de linguagens como o Python ou do uso de software como o Gephi. No entanto, aconselhamos o uso de ferramentas de análise de redes, como o Gephi, o Visone ou o Cytoscape que foram demonstrados em aula, e scripts em Python ou R.

A secção a seguir servirá como indicador e diretriz. Os passos mencionados não são obrigatórios nem exclusivos. Eles servirão, no entanto, como orientação para a forma como vocês podem fazer um projeto completo de Análise de Redes, num contexto mais geral de Análise de Dados.

## 2 Proposta Indicativa de Projeto

### 2.1 Definição do Problema e Análise do que já foi estudado sobre o problema

É verdade que definir o problema, o que queremos alcançar e/ou que questões pretendemos responder não é fácil. Nesta fase é essencial pensarmos no contexto que vamos analisar, se estamos à vontade com ele, que informação já temos, e como podemos obter mais informação através das ferramentas e do conhecimento de que dispomos.

Um passo importante é o levantamento daquilo que já se sabe sobre o problema. Se as soluções são triviais, ou o problema já foi muito estudado, o nosso contributo pode não ser muito grande e o problema pode até já estar resolvido. É importante ler e aprender com o trabalho que já foi desenvolvido em torno do problema.

### 2.2 Aquisição de Dados

Embora isso possa parecer um passo simples para aqueles que têm experiência anterior com dados (ou que trabalhar com eles diariamente), pode ser um desafio significativo para quem quer investigar e resolver um problema específico pelo que envolve de recolha de informação. Neste projeto os dados a utilizar já estão disponíveis em *datasets* normalmente bem formatados. No entanto, dependendo do problema colocado e da forma como o vão resolver, podem ter de recolher informação adicional em *datasets* adicionais e de ajustar formatos ou conciliar os dados entre *datasets*.

### 2.3 Tratamento e Análise Exploratória dos Dados

Este processo consiste em limpar, manipular e transformar dados considerados brutos, em dados que podem ser lidos/processados por, por exemplo, *scripts* em Python. É um processo necessário quando

queremos 'filtrar' nossos dados para servir de base à resolução do problema, eliminando possíveis dados incorretos e/ou nulos, formatos incorretos e/ou inválidos, selecionando apenas colunas interessantes e descartando aqueles que não são interessantes, detetando possíveis *outliers*, entre outros. É o resultado deste processo que nos permite obter visualizações limpas dos dados, bons modelos e também o contrário: pode ser a razão pela qual os algoritmos não têm os resultados esperados. É na sequência desta fase que podemos tirar as primeiras conclusões sobre nossos dados. Nesta fase são calculadas métricas e estatísticas sobre os dados que permitem uma visão geral e a detecção de particularidades no conjunto dos dados. Uma análise bem feita dos dados não se deve basear unicamente sobre resultados quantitativos, mas também na análise das possíveis relações qualitativas existentes nos mesmos através do exame criterioso daquilo que os dados representam

## 2.4 Análise e Modelação dos dados

É na sequência desta fase que podemos tirar conclusões sobre nossos dados. Nesta fase são utilizados algoritmos e técnicas que são aplicadas sobre os dados através de ferramentas de software, ou através de scripts programados para o efeito. É nesta fase que são calculadas métricas e estatísticas sobre os dados que vão suportar a relevância estatística dos resultados. Uma análise bem feita dos dados não se deve basear unicamente sobre resultados quantitativos, mas também na análise das possíveis relações qualitativas existentes nos mesmos através do exame criterioso daquilo que os dados representam.

## 2.5 Recolha de resultados, discussão e conclusões

Quando se tratam dados em grande quantidade a recolha de resultados pode constituir uma tarefa morosa. Desaconselhamos de todo o uso de *datasets* muito grandes. A recolha de resultados nestes casos pode constitui um passo importante. É também nesta fase que são agregados resultados para visualização gráfica.

O processo culmina na extração de conclusões sobre os resultados da nossa análise. Se tivermos muitas informações sobre os dados, agora podemos passar para a etapa que nos permite apresentar a solução final para o nosso problema, que pode dar resposta às questões nele colocadas.

Nesta fase também se colocam mais hipóteses sobre outras alternativas de trabalho, discute-se a validade das conclusões em função das circunstâncias em que elas foram tomadas e desenha-se um trabalho futuro.

## 2.6 Reporte e Visualização

Este procedimento vem completar o nosso processo de análise com o reporte escrito do mesmo e a elaboração de gráficos e outras visualizações. Esta fase é muito importante porque a qualidade do reporte tem um grande impacto naquilo que o vosso trabalho pode representar. Devemos procurar ser claros, objetivos e ler o nosso relatório através do olhar dos outros procurando ter em conta o seu sentido subjetivo e a sua facilidade de apreensão.

### 3 Enunciado

Utilizando um dos *datasets* de redes mencionado na página Moodle da cadeira, ou um *dataset* por vós recolhido, analise um problema que é representado por uma rede de entidades/nós/atores e relações. Use as métricas, conceitos e algoritmos que foram estudadas na cadeira, ferramentas de software como o Gephi ou o Cytoscape, estatísticas sobre os dados do *dataset*, ou informação adicional sobre o problema, sempre tendo em conta as relações existentes na rede, para extrair conclusões.

Por exemplo:

Utilizando a rede de coautoria de artigos científicos de Paul Erdős

<http://vlado.fmf.uni-lj.si/pub/networks/data/Erdos/Erdos02.net>

Procurar analisar o universo de publicações à sua volta, relativamente por exemplo às áreas preferenciais de investigação, se essas áreas dependem do número de Erdos ([https://pt.wikipedia.org/wiki/N%C3%BAmero\\_de\\_Erd%C5%91s](https://pt.wikipedia.org/wiki/N%C3%BAmero_de_Erd%C5%91s)), e à reciprocidade entre investigadores.

Para isso iríamos utilizar medidas de centralidade, medidas de transitividade e recolher, na medida do possível quais são as áreas de atuação de cada um dos investigadores. Para isso podíamos sublinhar os investigadores unicamente à distância de 1 de Erdős mais prolíferos e recolher essa informação na Internet.

Seu projeto final deve ser concluído em um grupo de duas a três pessoas e deve concluir com a elaboração de um relatório que deverá ter no máximo 20 páginas A4, e deverá ter o formato PDF. Eventualmente o relatório poderá ser entregue em ficheiro zipado se as imagens do mesmo forem entregues em ficheiros separados, as mesmas deverão, no entanto, ser bem referenciadas.

O relatório deverá incluir pelo menos as seguintes seções:

1. Introdução (na qual serão abordadas as seções 2.1 e 2.2 referidas atrás)
2. Métodos (na qual serão abordadas as seções 2.3 e 2.4 referidas atrás)
3. Discussão e Conclusões (na qual será abordada a seção 2.5 e 2.6 referidas atrás)

Não serão avaliados ficheiros ou relatórios sem a indicação clara dos nomes e números dos elementos do grupo.

### 4 Entrega

O projeto deverá ser entregue até ao dia 22 de dezembro de 2024, às 23:59. O projeto poderá ser entregue com atraso com o número de valores de penalização igual ao número de dias de atraso.

A submissão do projeto é feita via Moodle, na pasta Projeto Final. Os alunos podem enviar quantas vezes quiserem, mas apenas a última submissão é considerada para avaliação.

### 5 Discussão

Seu projeto será apresentado e discutido oralmente, via PowerPoint ou em outro qualquer formato que beneficie a apresentação. A duração da apresentação é no máximo 15 minutos e todos os membros do grupo devem intervir.

## 6 Critérios de avaliação

Os critérios de avaliação incluirão:

1. O carácter inovador da análise em face do trabalho que outros já desenvolveram
2. A relevância e realismo da análise
3. O rigor e a correção da análise na utilização dos conceitos falados na aula
4. A boa apresentação e reporte

Serão avaliados os seguintes itens:

1. Introdução e compreensão do problema
2. Descrição e compreensão dos dados
3. Preparação dos dados e análise exploratória
4. Modelação e avaliação
5. Conclusões
6. Apresentação e Discussão Oral