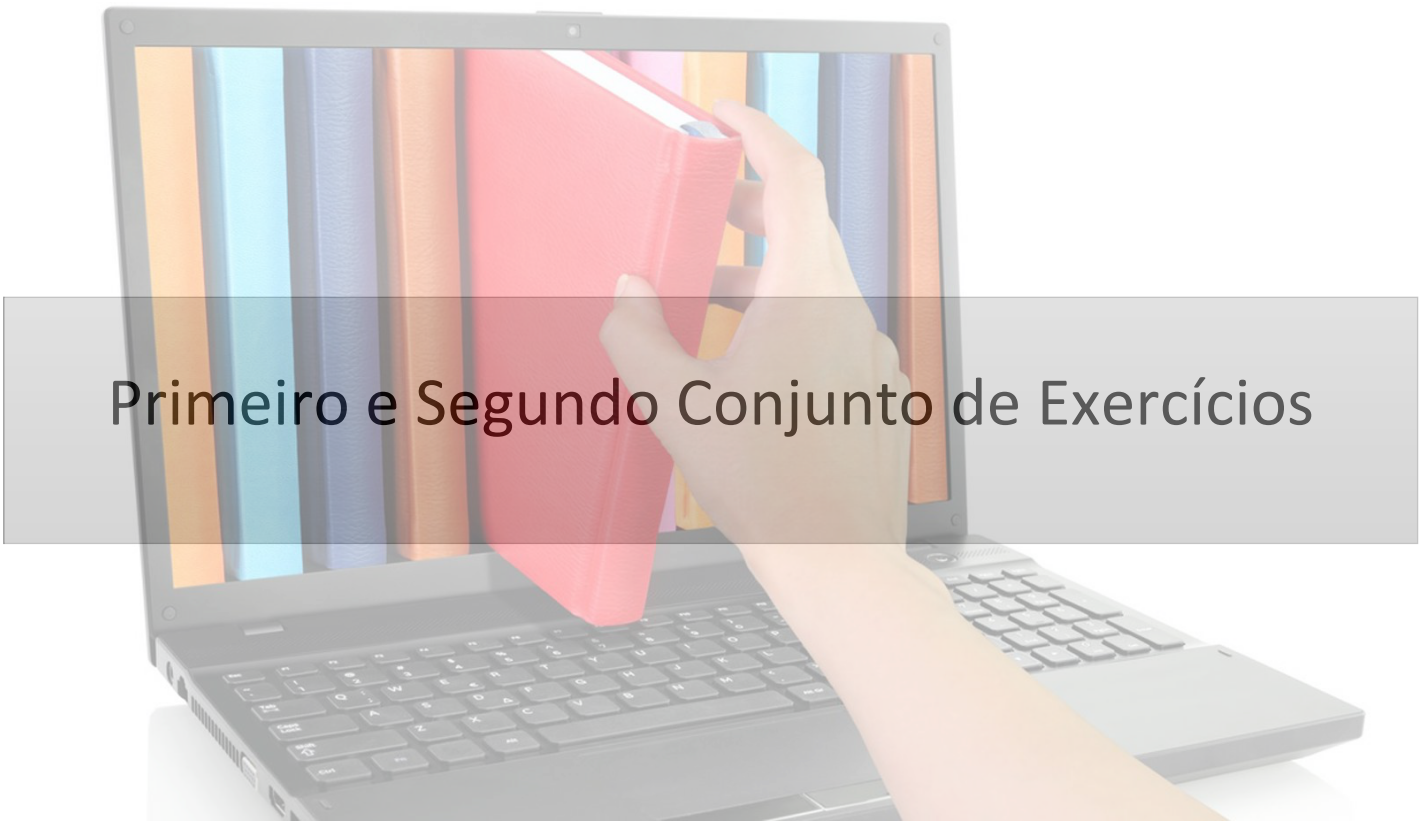
Universidade do Minho

Mestrado Integrado em Engenharia Biomédica – Ramo de Informática Médica

Unidade Curricular: Bibliotecas Digitais



# Primeiro e Segundo Conjunto de Exercícios

**Discente**: Ana Sousa (74753), Ana Machado (75088), Ana Ramos (74727)

**Docente:** Joaquim Melo Macedo

Braga

Maio 2018

# Índice

# Primeiro Conjunto

## Exercício 1 - Definições

**Dê a definição dos seguintes termos:**

- **Índice invertido**

O índice invertido, no mundo da computação é visto como uma estrutura de mapeamento de dados e das suas ocorrências. Esta estratégia de indexação é muito utilizada, pela sua rapidez na pesquisa de dados na base de dados. Funciona como que uma matriz em que apenas se apresentem as posições em que um determinado termo se encontra.

Assim sendo, pode afirmar-se que vão existir 2 estruturas, um dicionário onde se alocam todos os termos/palavras existentes e posting-list's onde estarão as ocorrências de cada termo. A posting list pode ser continua, se as palavras estiverem em disco. Caso isto não aconteça (palavras estiverem em memória), esta estrutura é complexa e de tamanho variável.

A construção de um índice deste género é também das mais simples. Tendo-se um documento, extraem-se as palavras que nele existem. Em seguida, dividem-se e indexam-se as palavras para que depois se possam registar as duas ocorrências. Por último, constrói-se uma tabela com uma coluna com todos os termos que existem e outra coluna com os DocID em que estes aparecem. Esta tabela deve estar ordenada alfabeticamente. No final, a primeira coluna corresponde ao dicionário e a segunda coluna corresponde às posting list's

- **Lista de ocorrência (Posting List)**

Tal como já mencionado, aquando a construção de um índice invertido elaboram-se duas estruturas, uma delas é a posting list. Esta é definida como a estrutura onde se vão registar, para cada termo, os documentos em que este ocorre. Pode também ser definido como um tuplo (termo, docID) em que o primeiro parâmetro identifica a palavra em questão e o segundo os ID dos documentos onde ocorre. Ao conjunto de todas as posting list's dá-se o nome de posting.

- **Relevância**

Num modelo clássico de pesquisa, existem vários passos que vão definir a qualidade do resultado, desde a definição da tarefa pretendida até à definição da query que irá executar a pesquisa tudo é bastante importante. Assim sendo a relevância é uma das caraterística que se quer encontrar aquando a receção do resultado da pesquisa. Em ciências da computação e mais especificamente em bibliotecas digitais a relevância é definida como a caraterística que vai qualificar se o que se recebeu é ou não importante, se é o que se pretendia ou se pelo menos ajudou de algum modo a tirar conclusões sobre o assunto em questão.

- **Precisão**

Após a descoberta dos documentos tidos como relevantes na pesquisa, existem duas medidas, simples, mas muito importantes para a avaliação da qualidade da informação que se obteve. Uma delas é a precisão.

Em engenharia, a precisão é vista como o grau de variação de um determinado resultado de uma medição. Em computação não se pode afirmar que se tem uma medida, mas sim uma pesquisa. Assim sendo, a precisão que aqui se fala vai "medir", perante os ficheiros de uma determinada coleção a relevância que estes têm para a necessidade de informação do utilizados. Matematicamente expressa-se da seguinte forma:

$$Precisão = \frac{TP}{TP + FP}$$

Em que TP são os verdadeiros positivos e FP os falsos positivos.

- **Cobertura (Recall)**

*Recall* é a segunda medida que anteriormente se mencionou. Esta define-se como sendo a fração de instâncias relevantes. Muitas vezes esta medida é denominada de sensibilidade, medindo a probabilidade que um documento relevante tem de ser captado pela query de pesquisa.

$$Recall = \frac{TP}{TP + FN}$$

- **Necessidade de Informação**

Um sistema de recuperação de informação recupera informação que se possa relevar útil para um utilizador, para isso é necessário realizar a construção de *query's*. O mais complicado nestes sistemas é realmente identificar qual a necessidade de informação que o utilizador tem. Muitas das vezes a necessidade da informação é um hábito criado pelo ser humana aquando as suas vivencias, para suportar determinados acontecimentos práticos.

- **Operadores de proximidade**

A definição de operadores nem sempre foi consensual. Estes são termos que podem ser usados para informar os sistemas de pesquisa sobre a relação entre as palavras que foram digeridas. Estes operadores são divididos em duas categorias: Operadores lógicos e Operadores de proximidade.

Operadores de proximidade permitem que o utilizador determine a distância máxima que duas palavras possuem entre si. Estes são bastante úteis uma vez que tornam a pesquisa muito mais objetiva. Como exemplo de operadores de proximidade tem-se o "NEAR/x", o "SAME", entre outros.

# Exercício 2 – Indexação de Documentos

**Considere a seguinte coleção composta pelos seguintes documentos (1 por linha)**

- **Doc1: out of the clear blue sky**
- **Doc2: the blue car next to the entrance**
- **Doc3: sky news: information retrieval is nice**

**Proponha uma stop-list e o índice da coleção para esta stoplist**

**Apresente o índice posicional para esta coleção**

**Que modificações haveria na lista de termos se fossem normalizados usando um lematizador ou um stemmer?**

A Stop-list proposta é: {out, of,the,next,to,is,} Se esta for utilizada, cada documento terá a sua própria lista

Doc1: clear blue sky

Doc2: blue car entrance

Doc3: sky news information retrieval nice

Os termos a serem utilizados e respetivas indexações estão representados na tabela abaixo:

**Blue**: 2; <1:2>; <2:1>

**Car**: 1; <2:2>

**Clear:** 1; <1:1>

**Entrance:** 1; <2:3>

**Information:** 1; <3:3>

**News:** 1; <3:2>

**Nice:** 1; <3:5>

**Retrieval:** 1; <3:4>

**Sky:** 2; <1:3; 3:1>

Deve notar-se que os termos foram ordenados por ordem alfabética.

# Exercício 3 – Índice Posicional

**Considere o seguinte índice**

- **love: <d1,12> <d2,23-32-43> <d3,53>**
- **hell: <d1,25> <d2,34-40> <d5,38>**

**DI é o documento I e o resto são posições**

**Há algum documento com a frase "love is hell"?**

No que diz respeito à recuperação de informação, os índices posicionais são os mais comumente empregados. Assim sendo, estes tem para cada termo presente no vocabulário, as ocorrências que este tem, ordenadas pelas posições.

Analisando o excerto "love is hell" pode dizer-se que a palavra "love" terá sempre a posição mais baixa, e a palavra "hell" duas posições acima da anterior. Assim sendo, no documento 1 esta frase não aparece, pois, as palavras "love" e "hell" estão distanciadas de 13 posições. Já no documento 2 pode observar-se que a palavra "love" se encontra na posição 32 e, duas posições à frente, encontra-se a palavra "hell" (posição 34). Para o documento 3, a posição da palavra "love" é maior do que a da palavra "hell" e por isso a frase não está presente.

Em suma, pode dizer-se que o único documento que poderá conter a frase "love is hell" é o documento 2.

# Exercício 4 – Índice Posicional

**Considere o seguinte índice**

- **universidade:<d1,12><d2,23-32-43><d3,53><d5,36-42-48>**
- **minho: <d1,25> <d2,34-40> <d5,38-51>**

**DI é o documento I e o resto são posições**

**O operador infixo NEAR/x refere à proximidade x entre dois termos**

a) **Dê as soluções à interrogação "universidade NEAR/2 minho"**

b) **Dê os pares (x,docids) para cada x tal que a interrogação universidade NEAR/x minho tem pelo menos uma solução**

c) **Proponha um algoritmo para encontrar documentos que unifiquem com este operador.**

a)

O operados "Near/x" é utilizado para encontrar registos onde os termos unidos pelo operados estejam a um determinado número de palavras de cada um, sendo que o x é o número máximo de palavras que separam os termos. Este operador é muito útil e tem prioridade de tratamento perante todos os outros.

Assim sendo, o pretendido é encontrar-se no índice posicional fornecido, documentos para os quais as palavras "Universidade" e "Minho" estão separadas por, no máximo 2 posições.

No primeiro documento as 2 palavras estão separadas por 13 posições, logo não cumpre a interrogação.

No segundo documento a palavra "Universidade" encontra-se na posição 32 e a palavra "Minho" na posição 34, logo cumpre a interrogação.

No documento 3, a palavra "Minho" não existe, logo a interrogação torna-se impossível,

Por último, no quinto documento a palavra "Universidade" consta na posição 36 e a "Minho" na posição 38, logo confirma-se a interrogação.

b) Assim sendo os pares (x,docids) encontradas são (2,d2) e (2,d5).

c)

```
Search Neart(p1, p2, k)
answer ← < >
while p1 ≠ NIL and p2 ≠ NIL
do if docID(p1) = docID(p2)
      then l ← < >
           pp1 ← positions(p1)
           pp2 ← positions(p2)
           while pp1 ≠ NIL
           do while pp2 ≠ NIL
```

```
                    do if |pos(pp1) − pos(pp2)| ≤ k
                            then ADD(l, pos(pp2))
                            else if pos(pp2) > pos(pp1)
                                    then break
                                 pp2 ← next(pp2)
                    while l ≠ < > and |l[0] − pos(pp1)| > k
                    do Delete(l[0])
                    for each ps ∈ l
                    do ADD(answer, ádocID(p1), pos(pp1), ps ñ)
                    pp1 ← next(pp1)
                    p1 ← next(p1)
                p2 ← next(p2)
            else if docID(p1) < docID(p2)
                then p1 ← next(p1)
                else p2 ← next(p2)
return answer
```

# Exercício 5 – Modelo Booleano

**Dadas as seguintes características dos índices, com a estrutura term: #(postings):**

- **sky 189 000**
- **blue 230 000**
- **field 32 000**
- **red 453 000**
- **high 345 000**
- **low 21 000**

**Proponha uma ordem para processar a seguinte interrogação, justificando a resposta:**

**(sky OR field) AND (blue OR red) AND (high OR low)**

189000+32000 AND 230000 + 453000 AND 345000 +21000 =

221000 AND 683000 AND 366000

Os modelos de recuperação de informação estão divididos em grupos tais como: modelos clássicos (onde se encontra o booleano), modelos estruturados (onde se encontram os nós proximais), entre outros. Para o modelo que está em questão, apresenta-se normalmente um conjunto de termos que representam o "corpus" em questão. A consulta a estes é feita por termos conectados, tais como AND, OR, NOT e derivantes destes (NOT AND, por exemplo).

Neste cado admitindo que o resultado da soma dos *posting sizes* dos termos são as seguintes:

1. sky OR field - 221000
2. blue OR red. 683000
3. high OR low – 366000

e tendo em contra que o algoritmo trata os mais pequenos primeiro, a ordem de processamento seria 1;3;2

# Exercício 6 – RI tolerante

**Considere o seguinte documento: "O universo contém muitas universidades diferentes"**

**a) Quantas entradas contém um índice de trigramas para este documento?**

**b) Qual é a interrogação booleana neste índice para a interrogação uni*?**

**c) Como processaria uma interrogação com uni*e*? Apresente detalhes do processamento.**

a) O documento acima apresentado tem 6 termos, no entanto para construir trigramas, o termo deve ter pelo menos 3 letras. Assim sendo, ficamos apenas com: "universo", "contém", "muitas", "universidades" e "diferentes". Assim sendo, o índice de trigramas terá 27 entradas.

| | | |
|---|---|---|
| **ade** | universidades | |
| **con** | contém | |
| **dad** | universidades | |
| **des** | universidades | |
| **dif** | diferentes | |
| **ent** | diferentes | |
| **ere** | diferentes | |
| **ers** | universidades | universo |
| **fer** | diferentes | |
| **ida** | universidades | |
| **ife** | diferentes | |
| **ita** | muitas | |
| **ive** | universidades | universo |
| **mui** | muitas | |
| **niv** | universidades | universo |
| **nte** | contem | diferentes |
| **ont** | contem | |
| **ren** | diferentes | |
| **rsi** | universidades | |
| **rso** | universo | |
| **sid** | universidades | |
| **tas** | muitas | |
| **tem** | contem | |
| **tes** | diferentes | |
| **uit** | muitas | |

| | | |
|---|---|---|
| **uni** | universidades | universo |
| **ver** | universidades | universo |

b) A interrogação booleana neste índice que responde à interrogação uni* é:
$un and uni

c) Para a interrogação uni*e , o processamento seria o seguinte:
Uni*e -> uni*e$-> ni*e$u->i*e$un-> e$uni*

# Exercício 7 – RI tolerante

**a) Calcule a distância de edição entre a palavra filósofo e filantropia**

Em computação, distancia de edição é uma das formas de quantificar como duas palavras se relacionam. É um procedimento relativamente simples uma vez para calcular esta distância basta calcular o número mínimo de operações necessárias para transformar uma palavra na outra. Existem 3 operações que podem ser feitas e contadas para a execução deste algoritmo. São elas inserção, remoção e ainda a realocação.

Transformar filosofo em filantropia

Deste modo, e analisando as palavras "filósofo" e "filantropia", para converter a primeira na segunda, é necessário ter em conta os seguintes aspetos:

- as primeiras 3 letras são iguais e por isso não necessitam de nenhuma das operações.
- as quarta e quinta letras são substituídas por um "a" e um "n", respetivamente.
- As sexta e sétima letras são substituídas por um "t" e um "r".
- Admitindo-se o "o" que se segue na segunda palavra é o último da primeira palavra, as restantes 3 "pia" foram inseridas uma vez que a primeira palavra é mais pequena do que a segunda.

Assim sendo, a distância de edição 7

**b) Calcule o coeficiente de Jacardi entre as 2 palavras usando bigramas**

| Filósofo | Filantropia |
|---|---|
| fi | fi |
| il | il |
| lo | la |
| os | an |
| so | nt |
| of | tr |
| | ro |
| | op |
| | pi |
| | la |

| Indice de bigramas |
|---|
| **an** |
| **fi** |
| **fo** |
| **ia** |
| **il** |
| **la** |

| lo |
| --- |
| nt |
| of |
| op |
| os |
| pi |
| ro |
| so |
| tr |

$$\frac{2}{13}$$

# Segundo Conjunto

## Exercício 1

**Considere a seguinte coleção:**

**D1: jornal notícias Lisboa**

**D2: Jornal notícias Porto**

**D3: Semanário Lisboa**

**a) Assumindo que o fator de normalização é a máxima frequência de termo, calcule os pesos tf - idf para esta coleção**

O valor tf - idf tem como "tradução" a frequência do termo inverso da frequência nos documentos. Esta é uma medida estatística que tem como objetivo principal de indicar a importância de uma palavra de um determinado documento em relação a uma coleção de documentos.

Considerando, tal como diz o enunciado, que o fator de normalização é a máxima frequência do termo, o cálculo é feito através da seguinte equação:

$$ntf_t = a + (1-a)\frac{tf_{t,d}}{tf_{max}(d)}, em\ que\ a = 0{,}5$$

Aplicando esta equação aos 3 documentos, obtém-se os seguintes valores:

- Documento1

$$ntf_{jornal} = ntf_{notícias} = ntf_{Lisboa} = 0{,}5 + (1-0{,}5)\frac{1}{1} = 1$$

$$ntf_{Porto} = ntf_{Seminário} = 0{,}5 + (1-0{,}5)\frac{0}{1} = 0{,}5$$

- Documento2

$$ntf_{jornal} = ntf_{notícias} = ntf_{-porto} = 0{,}5 + (1-0{,}5)\frac{1}{1} = 1$$

$$ntf_{Lisboa} = ntf_{Seminário} = 0{,}5 + (1-0{,}5)\frac{0}{1} = 0{,}5$$

- Documento3

$$ntf_{jornal} = ntf_{notícias} = ntf_{-porto} = 0{,}5 + (1-0{,}5)\frac{0}{1} = 0{,}5$$

$$ntf_{Lisboa} = ntf_{Seminário} = 0{,}5 + (1-0{,}5)\frac{1}{1} = 1$$

Para calcular o inverso da frequência dos termos (idf), utilizou-se a seguinte equação:

$$ntf_t = \log\frac{N}{df_t}$$

Por simples aplicação e uma vez que os termos "jornal", "noticias" e "lisboa" aparecem todos 2 vezes, o seu idf é $\log\frac{3}{2}$ (0,176). Pela mesma ordem de ideias e assumindo que os termos "Porto" e "Seminário" aparecem os dois uma única vez, o sei idf é $\log\frac{3}{1}$ (0,477).

Para terminar o exercício e para calcular o valor pretendido (tf-idf) aplicou-se a seguinte equação:

$$tf - idf = ntf_t \text{ x } idf_t$$

| | | |
|---|---|---|
| **Jornal** | Documento 1 | 0,176 |
| | Documento 2 | 0,176 |
| **Porto** | Documento 2 | 0,477 |
| **Lisboa** | Documento 1 | 0,176 |
| | Documento 3 | 0,176 |
| **Seminário** | Documento 3 | 0,477 |
| **Noticias** | Documento 1 | 0,176 |
| | Documento 2 | 0,176 |

**b) Calcule o peso tf-idf usando a medida OKAPI**

O peso tf-idf, utilizando a medida OKAPI é calculado através da seguinte equação:

$$tf - idf = \log\frac{N}{df_t} \times \frac{(k_1 + 1)tf_t}{k_1\left((1-b) + b\frac{dl}{avdl}\right) + tf_t}$$

Tal como mencionado o comprimento do documento é um dos parâmetros da equação, $dl$. Deve notar-se que o comprimento do documento é definido como o somatório das frequências de todos os termos presentes no mesmo. Considera-se ainda importante definir que $avdl$ é a média do comprimento de todos os documentos considerados como coleção.

Por análise rápida percebe-se que os comprimentos dos documentos 1,2 e 3 são, respetivamente, 3, 3,2. Assim sendo o $avdl$ é (3+3+2) /2 = 2,67.

Considerando k1=1.2 e b= 0.75 e aplicando o algoritmo de frequência do termo obtém-se os seguintes valores de tf-idf.

- Documento1

$$tf - idf_{jornal} = tf - idf_{noticias} = tf - idf_{Lisboa} = idf * \frac{(1{,}2 + 1) * 1}{1{,}2 * \left((1 - 0{,}75) + 0{,}75 \times \frac{3}{2{,}67}\right) + 1} = 0{,}167$$

$$tf - idf_{Porto} = tf - idf_{Seminário} = 0$$

- Documento2

$$tf - idf_{jornal} = tf - idf_{noticias} = idf * \frac{(1{,}2+1)\times 1}{1{,}2 * \left((1-0{,}75)+0{,}75\times\frac{3}{2{,}67}\right)+1} = 0{,}167$$

$$tf - idf_{Porto} = idf * \frac{(1{,}2+1)\times 1}{1{,}2 * \left((1-0{,}75)+0{,}75\times\frac{3}{2{,}67}\right)+1} = 0{,}454$$

$$tf - idf_{Lisboa} = tf - idf_{Seminário} = 0$$

- Documento3

$$tf - idf_{Lisboa} = idf * \frac{(1{,}2+1)\times 1}{1{,}2 * \left((1-0{,}75)+0{,}75\times\frac{2}{2{,}67}\right)+1} = 0{,}196$$

$$tf - idf_{Seminário} = idf * \frac{(1{,}2+1)\times 1}{1{,}2 * \left((1-0{,}75)+0{,}75\times\frac{2}{2{,}67}\right)+1} = 0{,}531$$

$$tf - idf_{jornal} = tf - idf_{noticias} = tf - idf_{Porto} = 0$$

**c) Dada a interrogação noticias notícias semanário, ordene os documentos da coleção de acordo com o seu rank.**

$$Rank(d) = \sum_{n=1}^{m} tf_n \times idf_n$$

Rank (Documento 1) = 0,176

Rank (Documento 2) = 0,176

Rank (Documento 3) = 0,477

Obteve-se a seguinte ordem 3-1-2.

**d) Como organizaria um índice para o Modelo de Espaço Vetorial (são possíveis várias soluções)?**

O modelo de espaço vetorial propõe um ambiente sobre o qual é possível obter os documentos que respondem a uma expressão de pesquisa. Isto é realizado associando a cada termo do documento assim como para termo da expressão um peso. O resultado é um conjunto de documentos organizado pela ordem de similaridade em relação a uma expressão de pesquisa.

Um documento é representado por um vetor onde cada elemento representa o peso, do respetivo termo de indexação para o documento. Cada vetor descreve a posição do documento num espaço multidimensional, onde cada termo representa uma dimensão.

# Exercício 2

**Considere a seguinte coleção de textos financeiros**

**D1: A economia portuguesa bastante debilitada economia rating**

**D2: A instabilidade política eleições economia rating Portugal Política rating**

**D3: Dívida publica política economia Portugal produto interno bruto interno**

**D4: divida publica privada empresas rating**

**a) Represente os documentos como vetores usando tf-idf**

| | D1 | D2 | D3 | D4 | idf |
|---|---|---|---|---|---|
| **A** | 1 | 1 | 0 | 0 | 0,30103 |
| **Bastante** | 1 | 0 | 0 | 0 | 0,60206 |
| **Bruto** | 0 | 0 | 1 | 0 | 0,60206 |
| **Debilitada** | 1 | 0 | 0 | 0 | 0,60206 |
| **Divida** | 0 | 0 | 1 | 1 | 0,30103 |
| **Economia** | 2 | 1 | 1 | 0 | 0,124939 |
| **Eleições** | 0 | 1 | 0 | 0 | 0,60206 |
| **empresas** | 0 | 0 | 0 | 1 | 0,60206 |
| **instabilidade** | 0 | 1 | 0 | 0 | 0,60206 |
| **Interno** | 0 | 0 | 2 | 0 | 0,60206 |
| **Política** | 0 | 2 | 1 | 0 | 0,30103 |
| **Portugal** | 0 | 1 | 1 | 0 | 0,30103 |
| **Portuguesa** | 1 | 0 | 0 | 0 | 0,60206 |
| **privada** | 0 | 0 | 0 | 1 | 0,60206 |
| **Produto** | 0 | 0 | 1 | 0 | 0,60206 |
| **Publica** | 0 | 0 | 1 | 1 | 0,30103 |
| **Rating** | 1 | 2 | 0 | 1 | 0,124939 |

| | D1 | D2 | D3 | D4 |
|---|---|---|---|---|
| **A** | 0,30103 | 0,30103 | 0 | 0 |
| **Bastante** | 0,60206 | 0 | 0 | 0 |
| **Bruto** | 0 | 0 | 0,60206 | 0 |
| **Debilitada** | 0,60206 | 0 | 0 | 0 |
| **Divida** | 0 | 0 | 0,30103 | 0,30103 |
| **Economia** | 0,249877 | 0,124939 | 0,124939 | 0 |
| **Eleições** | 0 | 0,60206 | 0 | 0 |
| **empresas** | 0 | 0 | 0 | 0,60206 |
| **instabilidade** | 0 | 0,60206 | 0 | 0 |
| **Interno** | 0 | 0 | 1,20412 | 0 |
| **Política** | 0 | 0,60206 | 0,30103 | 0 |
| **Portugal** | 0 | 0,30103 | 0,30103 | 0 |
| **Portuguesa** | 0,60206 | 0 | 0 | 0 |
| **privada** | 0 | 0 | 0 | 0,60206 |
| **Produto** | 0 | 0 | 0,60206 | 0 |
| **Publica** | 0 | 0 | 0,30103 | 0,30103 |
| **Rating** | 0,124939 | 0,249877 | 0 | 0,124939 |

A tabela 2 representa os documentos como vetores usando tf-idf.

**b) Crie um vetor para a interrogação rating divida**

| | tf | df | idf | weight |
|---|---|---|---|---|
| **Rating** | 1 | 3 | 0,125 | 0,125 |
| **Divida** | 1 | 2 | 0,301 | 0,301 |

**c) Calcule a pontuação dos documentos para a interrogação usando a fórmula de similaridade do cosseno**

$$similaridade = \sum produto$$

**Documento 1**

| | query | | | | Documento 1 | | Produto |
|---|---|---|---|---|---|---|---|
| | tf | df | idf | weight | weight | N'lized | |
| **A** | 0 | 2 | 0,301 | 0 | 0,30103 | 0,268 | 0 |
| **Bastante** | 0 | 1 | 0,602 | 0 | 0,60206 | 0,537 | 0 |
| **Bruto** | 0 | 1 | 0,602 | 0 | 0 | 0 | 0 |
| **Debilitada** | 0 | 1 | 0,602 | 0 | 0,60206 | 0,537 | 0 |
| **Divida** | 1 | 2 | 0,301 | 0,301 | 0 | 0 | 0 |
| **Economia** | 0 | 4 | 0 | 0 | 0,249877 | 0,223 | 0 |
| **Eleições** | 0 | 1 | 0,602 | 0 | 0 | 0 | 0 |
| **empresas** | 0 | 1 | 0,602 | 0 | 0 | 0 | 0 |
| **instabilidade** | 0 | 1 | 0,602 | 0 | 0 | 0 | 0 |
| **Interno** | 0 | 2 | 0,602 | 0 | 0 | 0 | 0 |
| **Política** | 0 | 3 | 0,125 | 0 | 0 | 0 | 0 |
| **Portugal** | 0 | 2 | 0,301 | 0 | 0 | 0 | 0 |
| **Portuguesa** | 0 | 1 | 0,602 | 0 | 0,60206 | 0,537 | 0 |
| **privada** | 0 | 1 | 0,602 | 0 | 0 | 0 | 0 |
| **Produto** | 0 | 1 | 0,602 | 0 | 0 | 0 | 0 |
| **Publica** | 0 | 2 | 0,301 | 0 | 0 | 0 | 0 |
| **Rating** | 1 | 3 | 0,125 | 0,125 | 0,124939 | 0,111 | 0,014 |

Similaridade = 0,014

**Documento 2**

| | query | | | | Documento 2 | | Produto |
|---|---|---|---|---|---|---|---|
| | tf | df | idf | weight | weight | N'lized | |
| A | 0 | 2 | 0,301 | 0 | 0,30103 | 0,733 | 0 |
| Bastante | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Bruto | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Debilitada | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Divida | 1 | 2 | 0,301 | 0,301 | 0 | 0,000 | 0 |
| Economia | 0 | 4 | 0 | 0 | 0,124939 | 0,304 | 0 |
| Eleições | 0 | 1 | 0,602 | 0 | 0,60206 | 1,466 | 0 |
| empresas | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| instabilidade | 0 | 1 | 0,602 | 0 | 0,60206 | 1,466 | 0 |
| Interno | 0 | 2 | 0,602 | 0 | 0 | 0,000 | 0 |
| Política | 0 | 3 | 0,125 | 0 | 0,60206 | 1,466 | 0 |
| Portugal | 0 | 2 | 0,301 | 0 | 0,30103 | 0,733 | 0 |
| Portuguesa | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| privada | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Produto | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Publica | 0 | 2 | 0,301 | 0 | 0 | 0,000 | 0 |
| Rating | 1 | 3 | 0,125 | 0,125 | 0,249877 | 0,608 | 0,067 |

Similaridade = 0,067

**Documento3**

| | query | | | | Documento 3 | | Produto |
|---|---|---|---|---|---|---|---|
| | tf | df | idf | weight | Weight | N'lized | |
| A | 0 | 2 | 0,301 | 0 | 0 | 0,000 | 0 |
| Bastante | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Bruto | 0 | 1 | 0,602 | 0 | 0,60206 | 4,819 | 0 |
| Debilitada | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Divida | 1 | 2 | 0,301 | 0,301 | 0,30103 | 2,409 | 0,725 |
| Economia | 0 | 4 | 0 | 0 | 0,124939 | 1,000 | 0 |
| Eleições | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| empresas | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| instabilidade | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Interno | 0 | 2 | 0,602 | 0 | 1,20412 | 9,638 | 0 |
| Política | 0 | 3 | 0,125 | 0 | 0,30103 | 2,409 | 0 |
| Portugal | 0 | 2 | 0,301 | 0 | 0,30103 | 2,409 | 0 |
| Portuguesa | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| privada | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Produto | 0 | 1 | 0,602 | 0 | 0,60206 | 4,819 | 0 |
| Publica | 0 | 2 | 0,301 | 0 | 0,30103 | 2,409 | 0 |
| Rating | 1 | 3 | 0,125 | 0,125 | 0 | 0 | 0 |

Similaridade = 0,725

**Documento 4**

| | query | | | | Documento 4 | | Produto |
|---|---|---|---|---|---|---|---|
| | tf | df | idf | weight | weight | N'lized | |
| A | 0 | 2 | 0,301 | 0 | 0 | 0,000 | 0 |
| Bastante | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Bruto | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Debilitada | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Divida | 1 | 2 | 0,301 | 0,301 | 0,30103 | 2,409 | 0,725 |
| Economia | 0 | 4 | 0 | 0 | 0 | 0,000 | 0 |
| Eleições | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| empresas | 0 | 1 | 0,602 | 0 | 0,60206 | 4,819 | 0 |
| instabilidade | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Interno | 0 | 2 | 0,602 | 0 | 0 | 0,000 | 0 |
| Política | 0 | 3 | 0,125 | 0 | 0 | 0,000 | 0 |
| Portugal | 0 | 2 | 0,301 | 0 | 0 | 0,000 | 0 |
| Portuguesa | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| privada | 0 | 1 | 0,602 | 0 | 0,60206 | 4,819 | 0 |
| Produto | 0 | 1 | 0,602 | 0 | 0 | 0,000 | 0 |
| Publica | 0 | 2 | 0,301 | 0 | 0,30103 | 2,409 | 0 |
| Rating | 1 | 3 | 0,125 | 0,125 | 0,124939 | 1,000 | 0,125 |

Similaridade = 0,725+0,125=0,85

**d) Construa uma matriz de coocorrência dos termos nos documentos**

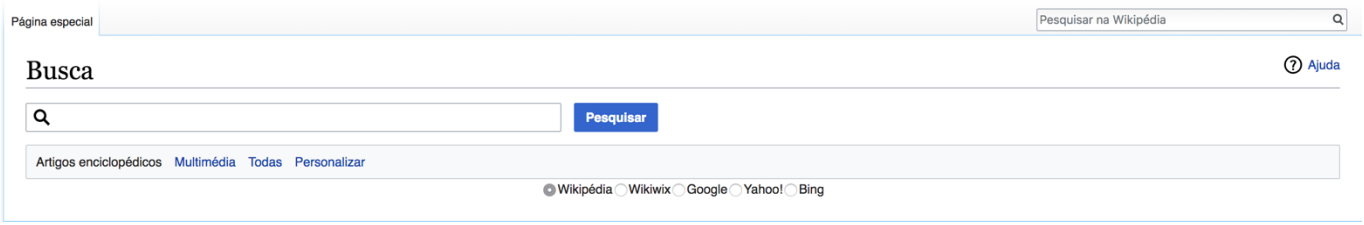
| | a | bastante | bruto | debilitada | dívida | economia | eleições | empresas | instabilidade | interno | politica | portugal | portuguesa | privada | produto | publica | rating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bastante | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bruto | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Debilitada | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Divida | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| Economia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| Eleições | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| empresas | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| instabilidade | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Interno | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Política | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Portugal | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Portuguesa | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| privada | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Produto | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Publica | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| Rating | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

# Exercício 3

**Faça cinco interrogações no domínio pt.wikipedia.org nos dois principais motores de busca (Google, Bing) e também usando o motor de busca interno da Wikipédia. Diga qual deles dá a melhor resposta para cada interrogação?**

| Interrogação | Google | Bing | Wikipédia |
|---|---|---|---|
| **Universidade do Minho** | 127000000 | 488000 | 1540 |
| **Engenharia Biomédica** | 426000 | 219000 | 876 |
| **Bibliotecas Digitais** | 446000 | 292000 | 410 |
| **Information Retrieval** | 242000000 | 8980000 | 2465 |
| **Indice Invertido** | 5590000 | 24000 | 311 |



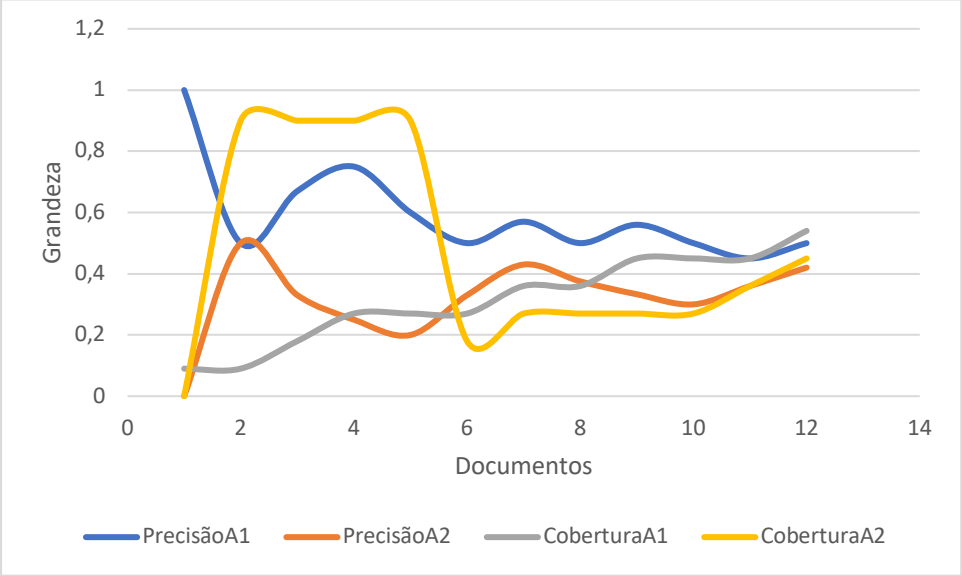
# Exercício 4

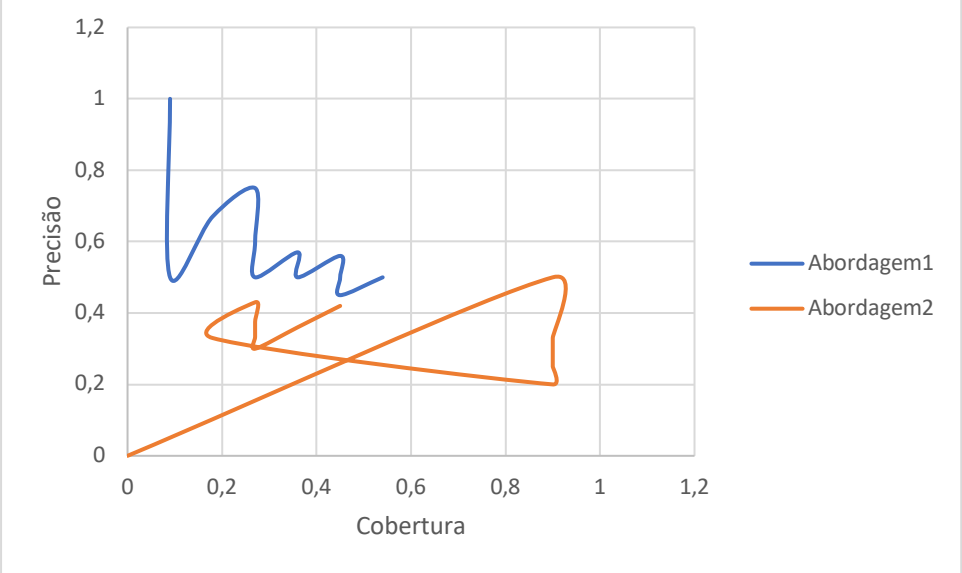
**Calcule a precisão e cobertura para os motores de busca A1 e A2 e faça os respetivos gráficos de precisão e cobertura baseado nos 11 níveis de cobertura discutidos na aula. A coluna R contém a lista dos documentos relevantes. A coluna A1 e A2 as respostas dos respetivos motores de busca.**

| Coleção | R | A1 | A2 |
|---|---|---|---|
| D1 | D2 | D2 | D27 |
| D2 | D6 | D12 | D6 |
| D3 | D9 | D9 | D4 |
| D4 | D15 | D22 | D12 |
| D5 | D22 | D25 | D1 |
| D6 | D23 | D17 | D31 |
| D7 | D26 | D23 | D26 |
| D8 | D28 | D4 | D7 |
| D9 | D29 | D2 | D11 |
| D10 | D31 | D30 | D19 |
| D11 | D34 | D8 | D34 |
| D12 | | D6 | D26 |
| D13 | | | |
| D14 | | | |
| D15 | | | |
| D16 | | | |
| D17 | | | |
| D18 | | | |
| D19 | | | |
| D20 | | | |
| D21 | | | |
| D22 | | | |
| D23 | | | |
| D24 | | | |
| D25 | | | |
| D26 | | | |
| D27 | | | |
| D28 | | | |
| D29 | | | |
| D30 | | | |
| D31 | | | |
| D32 | | | |
| D33 | | | |
| D34 | | | |
| D35 | | | |

| Documentos Devolvidos | Abordagem1 | Abordagem2 |
|---|---|---|
| 1 | **R** | **N** |
| 2 | **N** | **R** |
| 3 | **R** | **N** |
| 4 | **R** | **N** |
| 5 | **N** | **N** |
| 6 | **N** | **R** |
| 7 | **R** | **R** |
| 8 | **N** | **N** |
| 9 | **R** | **N** |
| 10 | **N** | **N** |
| 11 | **N** | **R** |
| 12 | **R** | **R** |

| Abordagem1 | Precisão | Cobertura | Abordagem2 | Precisão | Cobertura |
|---|---|---|---|---|---|
| 1 | 1 | 0,09 | **1** | 0 | 0 |
| 2 | 0,5 | 0,09 | **2** | 0,5 | 0,9 |
| 3 | 0,67 | 0,18 | **3** | 0,33 | 0,9 |
| 4 | 0,75 | 0,27 | **4** | 0,25 | 0,9 |
| 5 | 0,6 | 0,27 | **5** | 0,2 | 0,9 |
| 6 | 0,5 | 0,27 | **6** | 0,33 | 0,18 |
| 7 | 0,57 | 0,36 | **7** | 0,43 | 0,27 |
| 8 | 0,5 | 0,36 | **8** | 0,375 | 0,27 |
| 9 | 0,56 | 0,45 | **9** | 0,333 | 0,27 |
| 10 | 0,5 | 0,45 | **10** | 0,3 | 0,27 |
| 11 | 0,45 | 0,45 | **11** | 0,36 | 0,36 |
| 12 | 0,5 | 0,54 | **12** | 0,42 | 0,45 |


1,2
1
0,8
0,6
0,4
0,2
0
Grandeza
0
2
4
6
8
10
12
14
Documentos
PrecisãoA1
PrecisãoA2
CoberturaA1
CoberturaA2


1,2
1
0,8
0,6
0,4
0,2
0
Precisão
0
0,2
0,4
0,6
0,8
1
1,2
Cobertura
Abordagem1
Abordagem2

# Exercício 5

**Suponha um sistema de RI contem apenas 1000 documentos. Uma interrogação tem os seguintes 27 documentos relevantes: {d1, d5, d7, d10, d88, d151, d200, d211, d250, d300, d399, d401, d405, d450, d473, d500, d501, d530, d545, d590, d600, d735, d700, d720, d800, d888, d900}.**

**São usadas duas diferentes abordagens para obter documentos seriados para esta interrogação. Cada sistema devolve apenas os 10 dez primeiros documentos de topo. As abordagens 1 e 2 devolve documentos um de cada vez na seguinte ordem:**

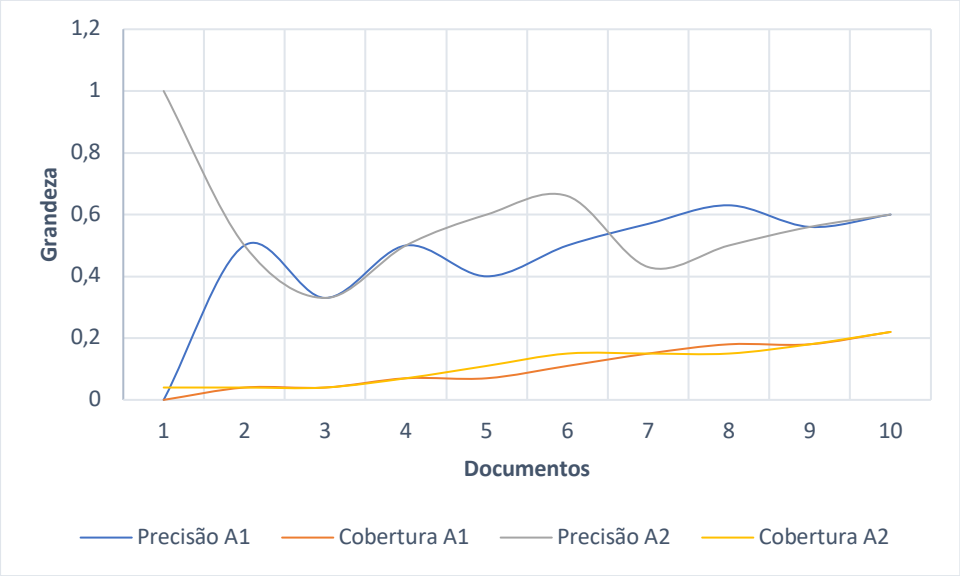
**Abordagem 1: d122, d211, d150, d88, d37, d1, d501, d800, d201, d5.**

**Abordagem 2: d10, d700, d6, d250, d88, d600, d59, d422, d500, d7.**

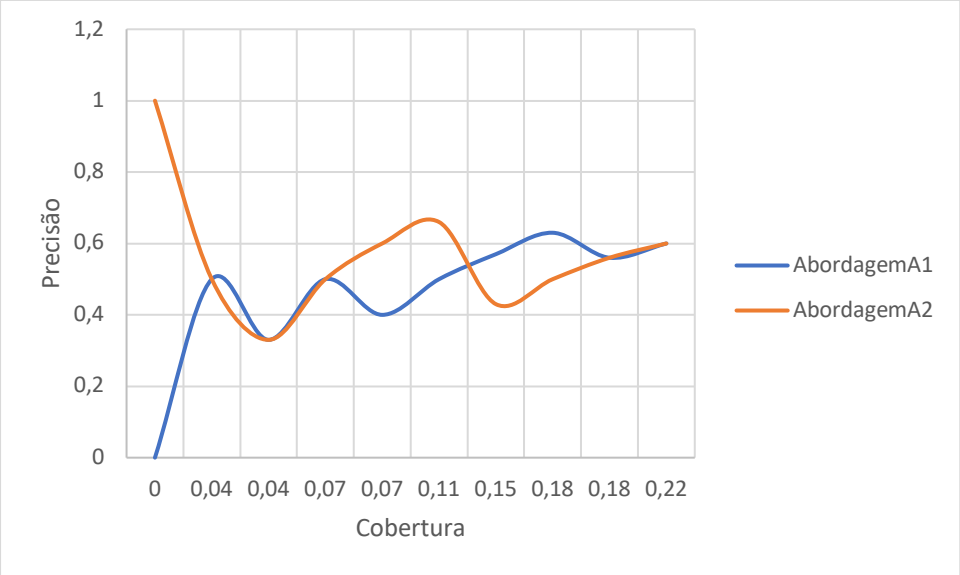
a) **Apresente um gráfico de precisão e cobertura para cada abordagem como uma função do número de documentos devolvidos**

| Documentos Devolvidos | Abordagem1 | Abordagem2 |
|---|---|---|
| 1 | N | R |
| 2 | R | N |
| 3 | N | N |
| 4 | R | R |
| 5 | N | R |
| 6 | R | R |
| 7 | R | N |
| 8 | R | N |
| 9 | N | R |
| 10 | R | R |

| A1 | Precisão | Cobertura | A2 | Precisão | Cobertura |
|---|---|---|---|---|---|
| 1 | 0 | 0 | **1** | 1 | 0,04 |
| 2 | 0,5 | 0,04 | **2** | 0,5 | 0,04 |
| 3 | 0,33 | 0,04 | **3** | 0,33 | 0,04 |
| 4 | 0,5 | 0,07 | **4** | 0,5 | 0,07 |
| 5 | 0,4 | 0,07 | **5** | 0,6 | 0,11 |
| 6 | 0,5 | 0,11 | **6** | 0,66 | 0,15 |
| 7 | 0,57 | 0,15 | **7** | 0,43 | 0,15 |
| 8 | 0,63 | 0,18 | **8** | 0,5 | 0,15 |
| 9 | 0,56 | 0,18 | **9** | 0,56 | 0,18 |
| 10 | 0,6 | 0,22 | **10** | 0,6 | 0,22 |




**b) Apresente o gráfico de precisão contra cobertura para as abordagens 1 e 2 usando os resultados da alínea anterior**




**c) Qual a melhor abordagem? Justifique a sua resposta.**

Considera-se que a melhor Abordagem é a segunda, pois tem melhor precisão e cobertura ao longo de toda a coleção.

# Exercício 6

**Considere uma interrogação que tem 5 documentos relevantes numa coleção: um é perfeito (P), dois são excelentes (E) e dois são bons (B). Os restantes são não relevantes (N). Suponha que os documentos devolvidos em resposta à interrogação são seriados da seguinte forma da esquerda para a direita:**

**P N B E N E B N N …**

**Observe que os documentos P, E e B são relevantes.**

**a) Qual é a precisão aos 10? Como mudaria se o sistema de RI fosse ideal?**

A precisão às K é definida como a precisão na posição k, neste caso 10. Ela é bastante mais intuitiva se se pensar na definição básica de precisão. Supondo que o sistema recupera os 10 documentos, os relevantes encontram-se nas posições 1,3,6,7,9 e o que nos queremos saber é a precisão que o sistema vai ter na posição 5. Então o que se vai fazer é calcular a razão entre o número de documentos relevantes até a posição 5 incluindo esta e o número de documentos até esta posição. Ou seja, seria 2/5.

No presente caso, uma vez que o K é 10, e o total de documentos relevantes até esta posição é 5. A precisão é ½.

**b) Qual a R-precision desta resposta? O que é a R-Precision?**

R precision é definida por $\frac{r}{R}$, ou seja, a razão entre os documentos recuperados relevantes, até ao momento da classificação, que é igual ao número de documentos da coleção e o número de documentos relevantes. Por exemplo, se existir uma coleção com 100 documentos, dos quais 30 são relevantes, recuperam-se os primeiros 30 documentos. Destes 30, apenas 10 são relevantes. Então a R-Precision é 1/3.

Na coleção apresentada têm-se 9 documentos, dos quais apenas 5 são relevantes. Recuperando os primeiros cinco obtém-se PNBEN, destes, apenas 3 são relevantes. Então a R-Precision é 5/3.

**c) Qual vantagem da R-Precision relativamente à precisão às 10?**

Tendo em atenção as definições dadas anteriormente, pode afirmar-se que a R-precision é uma medida que permite uma avaliação mais geral, ou seja, ela permite avaliar todo o sistema enquanto que a K-Precision só avalia até os documentos que se tem em consideração. Ainda assim, ambas são utilizadas dependendo da avaliação que se está a fazer.

**d) Qual é MAP desta lista de resultados?**

A *Mean Average Precision* é a média entre os valores de precisão das posições onde se encontram os documentos relevantes. É a mais utilizadas nas pesquisas assim como na redação dos artigos científicos, uma vez que consegue agregar todos os documentos relevantes e por isso avaliar a integridade do sistema.

Assim sendo e uma vez que temos 9 documentos e os relevantes encontram-se nas posições 1,3,4,6,7 a MAP resultante é :

$$\frac{1Precision + 3Precision + 4Precision + 6Precision + 7Precision}{5} = 0{,}7595$$

**e) Calcule o ganho cumulativo (CG). Assuma B=1, E=10, P=100. Qual é o objetivo das medidas com desconto DCG e nDCG?**

$$CG = r_1 + r_2 + \ldots r_n$$

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i}$$

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{log(1+i)}$$

# Exercício 7

**A relevância dos documentos devolvidos foi avaliada por 2 juízes, da seguinte forma (+ significa relevante e – significa não relevante)**

**Juiz A: 1+, 2+, 3-, 4- Juiz B: 1+, 2-, 3-, 4-**

**Dê a medida Kappa de concordância entre os dois juízes.**

| Posição | Juiz1 | Juiz3 |
|---|---|---|
| **1** | 1+ | 1+ |
| **2** | 2+ | 2- |
| **3** | 3- | 3- |
| **4** | 4- | 4- |

$$P(A) = 3/4 = 0,75$$

$$P(\text{nonrelevant}) = (2+3) / (8) = 0,625$$

$$P(\text{relevant}) = (2+1) / (8) = 0,375$$

$$P(E) = 0.625*0.375 + 0.375*0.625 = 0.469$$

$$\text{Kappa} = (P(A)-P(E)) / (1-P(E)) = (0,75-0,469) / (1-0,469) = 0,281/0,53 = 0,529$$

# Exercício 8

**Dada a seguinte rede social:**

**Nós: P1- João, P2- António, P3-Manuel, P4- Sebastião, P5- Catarina, P6- Bernardo, P7- Jorge, P8- Micaela, P9- Joana, P10- Domingos**

**Ramos**

**P1 ↔P2, P1 ↔P4, P1 ↔P5**

**P2 ↔P3, P2 ↔P4**

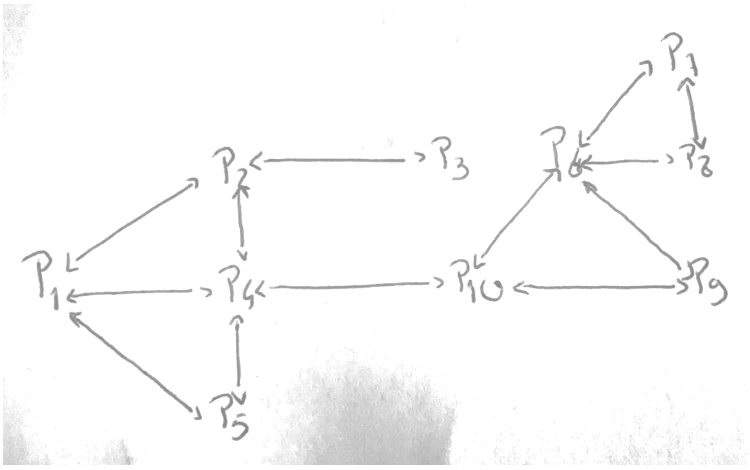
**P4 ↔P5, P4 ↔P10**

**P6 ↔P7, P6 ↔P8, P6 ↔P9, P6 ↔P10**

**P7 ↔P8**

**P9 ↔P10**

a) **Apresente a rede social como um grafo e como uma matriz**



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |

b) **Calcule as seguintes métricas: grau de cada vértice, coeficiente de aglomeração, percurso mais curo entre cada dois pares, centralidade. ( Betweenness)**

- **Grau de cada vértice**

Deg(P1) = 3 Deg(P2) = 3

Deg(P3) = 1 Deg(P4) = 4

Deg(P5) = 2 Deg(P6) = 4

Deg(P7) = 2 Deg(P8) = 2

Deg(P9) = 2 Deg(P10) = 3

- **O Coeficiente de Aglomeração é dado por:**

$$CC(i) = \frac{2 \times N_i}{K_1(K_i - 1)}$$

Sendo que, i corresponde ao nó, Ki é o número de vértices lidados ao vértice i, Ni o número de ligações realizadas em vértices vizinhos. Os coeficientes de aglomeração calculados para os vértices apresentam-se nas próximas linhas.

$$CC(P1) = \frac{2 \times 2}{3(3-1)} = 0{,}667 \qquad CC(P2) = \frac{2 \times 1}{3(3-1)} = 0{,}333$$

$$CC(P3) = \frac{2 \times 0}{1(1-1)} = 0 \qquad CC(P4) = \frac{2 \times 2}{4(4-1)} = 0{,}333$$

$$CC(P5) = \frac{2 \times 1}{2(2-1)} = 1 \qquad CC(P6) = \frac{2 \times 2}{4(4-1)} = 0{,}333$$

$$CC(P7) = \frac{2 \times 1}{2(2-1)} = 1 \qquad CC(P8) = \frac{2 \times 1}{2(2-1)} = 1$$

$$CC(P9) = \frac{2 \times 1}{2(2-1)} = 1 \qquad CC(P10) = \frac{2 \times 1}{3(3-1)} = 0{,}333$$

- **Percurso mais curto entre cada 2 pares**

- **Centralidade (Beetweness)**

**Para O Vértice P1:**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| (P2,P3) | 1 | 0 | 0 |
| (P2,P4) | 1 | 0 | 0 |
| (P2,P5) | 2 | 1 | 0,5 |
| (P2,P6) | 1 | 0 | 0 |
| (P2,P7) | 1 | 0 | 0 |
| (P2,P8) | 1 | 0 | 0 |
| (P2,P9) | 1 | 0 | 0 |
| (P2,P10) | 1 | 0 | 0 |
| (P3,P4) | 1 | 0 | 0 |
| (P3,P5) | 2 | 1 | 0,5 |
| (P3,P6) | 1 | 0 | 0 |
| (P3,P7) | 1 | 0 | 0 |
| (P3,P8) | 1 | 0 | 0 |
| (P3,P9) | 1 | 0 | 0 |
| (P3,P10) | 1 | 0 | 0 |
| (P4,P5) | 1 | 0 | 0 |
| (P4,P6) | 1 | 0 | 0 |
| (P4,P7) | 1 | 0 | 0 |
| (P4,P8) | 1 | 0 | 0 |
| (P4,P9) | 1 | 0 | 0 |
| (P4,P10) | 1 | 0 | 0 |
| (P5,P6) | 1 | 0 | 0 |
| (P5,P7) | 1 | 0 | 0 |
| (P5,P8) | 1 | 0 | 0 |
| (P5,P9) | 1 | 0 | 0 |
| (P5,P10) | 1 | 0 | 0 |
| (P6,P7) | 1 | 0 | 0 |
| (P6,P8) | 1 | 0 | 0 |
| (P6,P9) | 1 | 0 | 0 |
| (P6,P10) | 1 | 0 | 0 |
| (P7,P8) | 1 | 0 | 0 |
| (P7,P9) | 1 | 0 | 0 |
| (P7,P10) | 1 | 0 | 0 |
| (P8,P9) | 1 | 0 | 0 |
| (P8,P10) | 1 | 0 | 0 |
| (P9,P10) | 1 | 0 | 0 |

Centralidade P1 = 0,5+0,5 = 1

**Para O Vértice P2:**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| (P1,P3) | 1 | 1 | 1 |
| (P1,P4) | 1 | 0 | 0 |
| (P1,P5) | 1 | 0 | 0 |
| (P1,P6) | 1 | 0 | 0 |
| (P1,P7) | 1 | 0 | 0 |
| (P1,P8) | 1 | 0 | 0 |
| (P1,P9) | 1 | 0 | 0 |
| (P1,P10) | 1 | 0 | 0 |
| (P3,P4) | 1 | 1 | 1 |
| (P3,P5) | 2 | 2 | 1 |
| (P3,P6) | 1 | 1 | 1 |
| (P3,P7) | 1 | 1 | 1 |
| (P3,P8) | 1 | 1 | 1 |
| (P3,P9) | 1 | 1 | 1 |
| (P3,P10) | 1 | 1 | 1 |
| (P4,P5) | 1 | 0 | 0 |
| (P4,P6) | 1 | 0 | 0 |
| (P4,P7) | 1 | 0 | 0 |
| (P4,P8) | 1 | 0 | 0 |
| (P4,P9) | 1 | 0 | 0 |
| (P4,P10) | 1 | 0 | 0 |
| (P5,P6) | 1 | 0 | 0 |
| (P5,P7) | 1 | 0 | 0 |
| (P5,P8) | 1 | 0 | 0 |
| (P5,P9) | 1 | 0 | 0 |
| (P5,P10) | 1 | 0 | 0 |
| (P6,P7) | 1 | 0 | 0 |
| (P6,P8) | 1 | 0 | 0 |
| (P6,P9) | 1 | 0 | 0 |
| (P6,P10) | 1 | 0 | 0 |
| (P7,P8) | 1 | 0 | 0 |
| (P7,P9) | 1 | 0 | 0 |
| (P7,P10) | 1 | 0 | 0 |
| (P8,P9) | 1 | 0 | 0 |
| (P8,P10) | 1 | 0 | 0 |
| (P9,P10) | 1 | 0 | 0 |

Centralidade P2 = 8

**Para O Vértice P3:**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| (P1,P2) | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |
| **(P1,P6)** | 1 | 0 | 0 |
| **(P1,P7)** | 1 | 0 | 0 |
| **(P1,P8)** | 1 | 0 | 0 |
| **(P1,P9)** | 1 | 0 | 0 |
| **(P1,P10)** | 1 | 0 | 0 |
| **(P2,P4)** | 1 | 0 | 0 |
| **(P2,P5)** | 1 | 0 | 0 |
| **(P2,P6)** | 1 | 0 | 0 |
| **(P2,P7)** | 1 | 1 | 0 |
| **(P2,P8)** | 1 | 1 | 0 |
| **(P2,P9)** | 1 | 1 | 0 |
| **(P2,P10)** | 1 | 1 | 0 |
| **(P4,P5)** | 1 | 0 | 0 |
| **(P4,P6)** | 1 | 0 | 0 |
| **(P4,P7)** | 1 | 0 | 0 |
| **(P4,P8)** | 1 | 0 | 0 |
| **(P4,P9)** | 1 | 0 | 0 |
| **(P4,P10)** | 1 | 0 | 0 |
| **(P5,P6)** | 1 | 0 | 0 |
| **(P5,P7)** | 1 | 0 | 0 |
| **(P5,P8)** | 1 | 0 | 0 |
| **(P5,P9)** | 1 | 0 | 0 |
| **(P5,P10)** | 1 | 0 | 0 |
| **(P6,P7)** | 1 | 0 | 0 |
| **(P6,P8)** | 1 | 0 | 0 |
| **(P6,P9)** | 1 | 0 | 0 |
| **(P6,P10)** | 1 | 0 | 0 |
| **(P7,P8)** | 1 | 0 | 0 |
| **(P7,P9)** | 1 | 0 | 0 |
| **(P7,P10)** | 1 | 0 | 0 |
| **(P8,P9)** | 1 | 0 | 0 |
| **(P8,P10)** | 1 | 0 | 0 |
| **(P9,P10)** | 1 | 0 | 0 |

Centralidade P3 = 0

**Para O Vértice P4:**

| **(u,w)** | $\sigma_{uw}$ | $\sigma_{uw}$**(V)** | $\sigma_{uw}$**(V)**/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P1,P6)** | 1 | 1 | 1 |
| **(P1,P7)** | 1 | 1 | 1 |
| **(P1,P8)** | 1 | 1 | 1 |
| **(P1,P9)** | 1 | 1 | 1 |
| **(P1,P10)** | 1 | 1 | 1 |
| **(P2,P3)** | 1 | 1 | 1 |
| **(P2,P5)** | 2 | 1 | 0,5 |
| **(P2,P6)** | 1 | 1 | 1 |
| **(P2,P7)** | 1 | 1 | 1 |
| **(P2,P8)** | 1 | 1 | 1 |
| **(P2,P9)** | 1 | 1 | 1 |
| **(P2,P10)** | 1 | 1 | 1 |
| **(P3,P5)** | 2 | 1 | 0,5 |
| **(P3,P6)** | 1 | 1 | 1 |
| **(P3,P7)** | 1 | 1 | 1 |
| **(P3,P8)** | 1 | 1 | 1 |
| **(P3,P9)** | 1 | 1 | 1 |
| **(P3,P10)** | 1 | 1 | 1 |
| **(P5,P6)** | 1 | 1 | 1 |
| **(P5,P7)** | 1 | 1 | 1 |
| **(P5,P8)** | 1 | 1 | 1 |
| **(P5,P9)** | 1 | 1 | 1 |
| **(P5,P10)** | 1 | 1 | 1 |
| **(P6,P7)** | 1 | 0 | 0 |
| **(P6,P8)** | 1 | 0 | 0 |
| **(P6,P9)** | 1 | 0 | 0 |
| **(P6,P10)** | 1 | 0 | 0 |
| **(P7,P8)** | 1 | 0 | 0 |
| **(P7,P9)** | 1 | 0 | 0 |
| **(P7,P10)** | 1 | 0 | 0 |
| **(P8,P9)** | 1 | 0 | 0 |
| **(P8,P10)** | 1 | 0 | 0 |
| **(P9,P10)** | 1 | 0 | 0 |

Centralidade P4 = 22

**Para O Vértice P5**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P6)** | 1 | 0 | 0 |
| **(P1,P7)** | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P1,P8)** | 1 | 0 | 0 |
| **(P1,P9)** | 1 | 0 | 0 |
| **(P1,P10)** | 1 | 0 | 0 |
| **(P2,P3)** | 1 | 0 | 0 |
| **(P2,P4)** | 1 | 0 | 0 |
| **(P2,P6)** | | | |
| **(P2,P7)** | 1 | 0 | 0 |
| **(P2,P8)** | 1 | 0 | 0 |
| **(P2,P9)** | 1 | 0 | 0 |
| **(P2,P10)** | 1 | 0 | 0 |
| **(P3,P4)** | 1 | 0 | 0 |
| **(P3,P6)** | 1 | 0 | 0 |
| **(P3,P7)** | 1 | 0 | 0 |
| **(P3,P8)** | 1 | 0 | 0 |
| **(P3,P9)** | 1 | 0 | 0 |
| **(P3,P10)** | 1 | 0 | 0 |
| **(P4,P6)** | 1 | 0 | 0 |
| **(P4,P7)** | 1 | 0 | 0 |
| **(P4,P8)** | 1 | 0 | 0 |
| **(P4,P9)** | 1 | 0 | 0 |
| **(P4,P10)** | 1 | 0 | 0 |
| **(P6,P7)** | 1 | 0 | 0 |
| **(P6,P8)** | 1 | 0 | 0 |
| **(P6,P9)** | 1 | 0 | 0 |
| **(P6,P10)** | 1 | 0 | 0 |
| **(P7,P8)** | 1 | 0 | 0 |
| **(P7,P9)** | 1 | 0 | 0 |
| **(P7,P10)** | 1 | 0 | 0 |
| **(P8,P9)** | 1 | 0 | 0 |
| **(P8,P10)** | 1 | 0 | 0 |
| **(P9,P10)** | 1 | 0 | 0 |

Centralidade P5 = 0

**Para O Vértice P6**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |
| **(P1,P7)** | 1 | 1 | 1 |
| **(P1,P8)** | 1 | 1 | 1 |
| **(P1,P9)** | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P1,P10)** | 1 | 0 | 0 |
| **(P2,P3)** | 1 | 0 | 0 |
| **(P2,P4)** | 1 | 0 | 0 |
| **(P2,P5)** | 2 | 0 | 0 |
| **(P2,P7)** | 1 | 1 | 1 |
| **(P2,P8)** | 1 | 1 | 1 |
| **(P2,P9)** | 1 | 0 | 0 |
| **(P2,P10)** | 1 | 0 | 0 |
| **(P3,P4)** | 1 | 0 | 0 |
| **(P3,P5)** | 2 | 0 | 0 |
| **(P3,P7)** | 1 | 1 | 1 |
| **(P3,P8)** | 1 | 1 | 1 |
| **(P3,P9)** | 1 | 0 | 0 |
| **(P3,P10)** | 1 | 0 | 0 |
| **(P4,P5)** | 1 | 0 | 0 |
| **(P4,P7)** | 1 | 1 | 1 |
| **(P4,P8)** | 1 | 1 | 1 |
| **(P4,P9)** | 1 | 0 | 0 |
| **(P4,P10)** | 1 | 0 | 0 |
| **(P5,P7)** | 1 | 0 | 0 |
| **(P5,P8)** | 1 | 1 | 1 |
| **(P5,P9)** | 1 | 1 | 1 |
| **(P5,P10)** | 1 | 0 | 0 |
| **(P7,P8)** | 1 | 0 | 0 |
| **(P7,P9)** | 1 | 1 | 1 |
| **(P7,P10)** | 1 | 1 | 1 |
| **(P8,P9)** | 1 | 1 | 1 |
| **(P8,P10)** | 1 | 1 | 1 |
| **(P9,P10)** | 1 | 0 | 0 |

Centralidade P6 = 14

**Para O Vértice P7**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |
| **(P1,P6)** | 1 | 0 | 0 |
| **(P1,P8)** | 1 | 0 | 0 |
| **(P1,P9)** | 1 | 0 | 0 |
| **(P1,P10)** | 1 | 0 | 0 |
| **(P2,P3)** | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P2,P4)** | 2 | 0 | 0 |
| **(P2,P5)** | 1 | 0 | 0 |
| **(P2,P6)** | 1 | 0 | 0 |
| **(P2,P8)** | 1 | 0 | 0 |
| **(P2,P9)** | 1 | 0 | 0 |
| **(P2,P10)** | 1 | 0 | 0 |
| **(P3,P4)** | 1 | 0 | 0 |
| **(P3,P5)** | 2 | 0 | 0 |
| **(P3,P6)** | 1 | 0 | 0 |
| **(P3,P8)** | 1 | 0 | 0 |
| **(P3,P9)** | 1 | 0 | 0 |
| **(P3,P10)** | 1 | 0 | 0 |
| **(P4,P5)** | 1 | 0 | 0 |
| **(P4,P6)** | 1 | 0 | 0 |
| **(P4,P8)** | 1 | 0 | 0 |
| **(P4,P9)** | 1 | 0 | 0 |
| **(P4,P10)** | 1 | 0 | 0 |
| **(P5,P6)** | 1 | 0 | 0 |
| **(P5,P8)** | 1 | 0 | 0 |
| **(P5,P9)** | 1 | 0 | 0 |
| **(P5,P10)** | 1 | 0 | 0 |
| **(P6,P8)** | 1 | 0 | 0 |
| **(P6,P9)** | 1 | 0 | 0 |
| **(P6,P10)** | 1 | 0 | 0 |
| **(P8,P9)** | 1 | 0 | 0 |
| **(P8,P10)** | 1 | 0 | 0 |
| **(P9,P10)** | 1 | 0 | 0 |

Centralidade P7 = 0

**Para O Vértice P8**

| **(u,w)** | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |
| **(P1,P6)** | 1 | 0 | 0 |
| **(P1,P7)** | 1 | 0 | 0 |
| **(P1,P9)** | 1 | 0 | 0 |
| **(P1,P10)** | 1 | 0 | 0 |
| **(P2,P3)** | 1 | 0 | 0 |
| **(P2,P4)** | 1 | 0 | 0 |
| **(P2,P5)** | 2 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P2,P6)** | 1 | 0 | 0 |
| **(P2,P7)** | 1 | 0 | 0 |
| **(P2,P9)** | 1 | 0 | 0 |
| **(P2,P10)** | 1 | 0 | 0 |
| **(P3,P4)** | 1 | 0 | 0 |
| **(P3,P5)** | 2 | 0 | 0 |
| **(P3,P6)** | 1 | 0 | 0 |
| **(P3,P7)** | 1 | 0 | 0 |
| **(P3,P9)** | 1 | 0 | 0 |
| **(P3,P10)** | 1 | 0 | 0 |
| **(P4,P5)** | 1 | 0 | 0 |
| **(P4,P6)** | 1 | 0 | 0 |
| **(P4,P7)** | 1 | 0 | 0 |
| **(P4,P9)** | 1 | 0 | 0 |
| **(P4,P10)** | 1 | 0 | 0 |
| **(P5,P6)** | 1 | 0 | 0 |
| **(P5,P7)** | 1 | 0 | 0 |
| **(P5,P9)** | 1 | 0 | 0 |
| **(P5,P10)** | 1 | 0 | 0 |
| **(P6,P7)** | 1 | 0 | 0 |
| **(P6,P9)** | 1 | 0 | 0 |
| **(P6,P10)** | 1 | 0 | 0 |
| **(P7,P9)** | 1 | 0 | 0 |
| **(P7,P10)** | 1 | 0 | 0 |
| **(P9,P10)** | 1 | 0 | 0 |

Centralidade P8 = 0

**Para O Vértice P9**

| (u,w) | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |
| **(P1,P6)** | 1 | 0 | 0 |
| **(P1,P7)** | 1 | 0 | 0 |
| **(P1,P8)** | 1 | 0 | 0 |
| **(P1,P10)** | 1 | 0 | 0 |
| **(P2,P3)** | 1 | 0 | 0 |
| **(P2,P4)** | 1 | 0 | 0 |
| **(P2,P5)** | 2 | 0 | 0 |
| **(P2,P6)** | 1 | 0 | 0 |
| **(P2,P7)** | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P2,P8)** | 1 | 0 | 0 |
| **(P2,P10)** | 1 | 0 | 0 |
| **(P3,P4)** | 1 | 0 | 0 |
| **(P3,P5)** | 2 | 0 | 0 |
| **(P3,P6)** | 1 | 0 | 0 |
| **(P3,P7)** | 1 | 0 | 0 |
| **(P3,P8)** | 1 | 0 | 0 |
| **(P3,P10)** | 1 | 0 | 0 |
| **(P4,P5)** | 1 | 0 | 0 |
| **(P4,P6)** | 1 | 0 | 0 |
| **(P4,P7)** | 1 | 0 | 0 |
| **(P4,P8)** | 1 | 0 | 0 |
| **(P4,P10)** | 1 | 0 | 0 |
| **(P5,P6)** | 1 | 0 | 0 |
| **(P5,P7)** | 1 | 0 | 0 |
| **(P5,P8)** | 1 | 0 | 0 |
| **(P5,P10)** | 1 | 0 | 0 |
| **(P6,P7)** | 1 | 0 | 0 |
| **(P6,P8)** | 1 | 0 | 0 |
| **(P6,P10)** | 1 | 0 | 0 |
| **(P7,P8)** | 1 | 0 | 0 |
| **(P7,P10)** | 1 | 0 | 0 |

Centralidade P9 = 0

**Para o vértice P10**

| **(u,w)** | $\sigma_{uw}$ | $\sigma_{uw}$(V) | $\sigma_{uw}$(V)/$\sigma_{uw}$ |
|---|---|---|---|
| **(P1,P2)** | 1 | 0 | 0 |
| **(P1,P3)** | 1 | 0 | 0 |
| **(P1,P4)** | 1 | 0 | 0 |
| **(P1,P5)** | 1 | 0 | 0 |
| **(P1,P6)** | 1 | 1 | 1 |
| **(P1,P7)** | 1 | 1 | 1 |
| **(P1,P8)** | 1 | 1 | 1 |
| **(P1,P9)** | 1 | 1 | 1 |
| **(P2,P3)** | 1 | 0 | 0 |
| **(P2,P4)** | 1 | 0 | 0 |
| **(P2,P5)** | 2 | 0 | 0 |
| **(P2,P6)** | 1 | 1 | 1 |
| **(P2,P7)** | 1 | 1 | 1 |
| **(P2,P8)** | 1 | 1 | 1 |
| **(P2,P9)** | 1 | 1 | 1 |
| **(P3,P4)** | 1 | 0 | 0 |

| | | | |
|---|---|---|---|
| **(P3,P5)** | 2 | 0 | 0 |
| **(P3,P6)** | 1 | 1 | 1 |
| **(P3,P7)** | 1 | 1 | 1 |
| **(P3,P8)** | 1 | 1 | 1 |
| **(P3,P9)** | 1 | 1 | 1 |
| **(P4,P5)** | 1 | 0 | 0 |
| **(P4,P6)** | 1 | 1 | 1 |
| **(P4,P7)** | 1 | 1 | 1 |
| **(P4,P8)** | 1 | 1 | 1 |
| **(P4,P9)** | 1 | 1 | 1 |
| **(P5,P6)** | 1 | 1 | 1 |
| **(P5,P7)** | 1 | 1 | 1 |
| **(P5,P8)** | 1 | 1 | 1 |
| **(P5,P9)** | 1 | 1 | 1 |
| **(P6,P7)** | 1 | 0 | 0 |
| **(P6,P8)** | 1 | 0 | 0 |
| **(P6,P9)** | 1 | 0 | 0 |
| **(P7,P8)** | 1 | 0 | 0 |
| **(P7,P9)** | 1 | 0 | 0 |
| **(P8,P9)** | 1 | 0 | 0 |

Centralidade P10 = 20

# Exercício 9

**Para o seguinte grafo direto modificado**

**Nós: P1- João, P2- António, P3-Manuel, P4- Sebastião, P5- Catarina, P6- Bernardo, P7- Jorge, P8- Micaela, P9- Joana, P10- Domingos**

**Ramos**

**P1→P4**

**P2→P1, P2→P4**

**P3→P2**

**P4→P10**

**P5→ P4, P5→P1**

**P6→P7, P6→P8, P6→P9**

**P7→P8**

**P10→P6, P10→P9**

**Usando uma folha Excel**

a) **Calcule o PageRank das páginas ao fim de 10 iterações, usando 1 como valor de PageRank inicial**

| MATRIZ A | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 |
| | P1 | 0 | 0 | 0 | 1,00 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P2 | 0,50 | 0 | 0 | 0,50 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P3 | 0 | 1,00 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1,00 |
| | P5 | 0,50 | 0 | 0 | 0,50 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P6 | 0 | 0 | 0 | 0 | 0 | 0 | 0,33 | 0,33 | 0,33 | 0 |
| | P7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | P8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P10 | 0 | 0 | 0 | 0 | 0 | 0,50 | 0 | 0 | 0,50 | 0 |

| MATRIZ A-Transposta | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 |
| | P1 | 0 | 0,5 | 0 | 0 | 0,50 | 0 | 0 | 0 | 0 | 0 |
| | P2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P4 | 1 | 0,5 | 0 | 0 | 0,50 | 0 | 0 | 0 | 0 | 0 |
| | P5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,50 |
| | P7 | 0 | 0 | 0 | 0 | 0 | 0,33 | 0 | 0 | 0 | 0 |
| | P8 | 0 | 0 | 0 | 0 | 0 | 0,33 | 1 | 0 | 0 | 0 |
| | P9 | 0 | 0 | 0 | 0 | 0 | 0,33 | 0 | 0 | 0 | 0,33 |
| | P10 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

| MATRIZ1 | |
|---|---|
| P1 | 0,85 |
| P2 | 0,85 |
| P3 | 0 |
| P4 | 1,7 |
| P5 | 0 |
| P6 | 0,425 |
| P7 | 0,28 |
| P8 | 0,283333333 |
| P9 | 0,708333333 |
| P10 | 0,85 |

| MATRIZ4 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0 |
| P5 | 0 |
| P6 | 0,6375 |
| P7 | 0,28333333 |
| P8 | 0,425 |
| P9 | 0,92083333 |
| P10 | 0,425 |

| MATRIZ7 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0 |
| P5 | 0 |
| P6 | 0,10625 |
| P7 | 0 |
| P8 | 0 |
| P9 | 0,10625 |
| P10 | 0 |

| MATRIZ10 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0 |
| P5 | 0 |
| P6 | 0 |
| P7 | 0 |
| P8 | 0 |
| P9 | 0 |
| P10 | 0 |

| MATRIZ2 | |
|---|---|
| P1 | 0,425 |
| P2 | 0 |
| P3 | 0 |
| P4 | 1,275 |
| P5 | 0 |
| P6 | 0,425 |
| P7 | 0,141666667 |
| P8 | 0,425 |
| P9 | 0,566666667 |
| P10 | 1,7 |

| MATRIZ5 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0,2125 |
| P5 | 0 |
| P6 | 0 |
| P7 | 0 |
| P8 | 0 |
| P9 | 0 |
| P10 | 0 |

| MATRIZ8 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0 |
| P5 | 0 |
| P6 | 0 |
| P7 | 0,03541667 |
| P8 | 0,03541667 |
| P9 | 0,03541667 |
| P10 | 0 |

| MATRIZ3 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0,425 |
| P5 | 0 |
| P6 | 0,85 |
| P7 | 0,141666667 |
| P8 | 0,283333333 |
| P9 | 0,991666667 |
| P10 | 1,275 |

| MATRIZ6 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0 |
| P5 | 0 |
| P6 | 0 |
| P7 | 0 |
| P8 | 0 |
| P9 | 0 |
| P10 | 0,2125 |

| MATRIZ9 | |
|---|---|
| P1 | 0 |
| P2 | 0 |
| P3 | 0 |
| P4 | 0 |
| P5 | 0 |
| P6 | 0 |
| P7 | 0 |
| P8 | 0,03541667 |
| P9 | 0 |
| P10 | 0 |

## b) Repita alínea anterior usando 0,15 como valor de PageRank inicial

MATRIZ A

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 |
|---|---|---|---|---|---|---|---|---|---|---|
| P1 | 0 | 0 | 0 | 0,15 | 0 | 0 | 0 | 0 | 0 | 0 |
| P2 | 0,075 | 0 | 0 | 0,075 | 0 | 0 | 0 | 0 | 0 | 0 |
| P3 | 0 | 0,15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,15 |
| P5 | 0,075 | 0 | 0 | 0,075 | 0 | 0 | 0 | 0 | 0 | 0 |
| P6 | 0 | 0 | 0 | 0 | 0 | 0 | 0,05 | 0,05 | 0,05 | 0 |
| P7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,15 | 0 | 0 |
| P8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P10 | 0 | 0 | 0 | 0 | 0 | 0,08 | 0 | 0 | 0,08 | 0 |

MATRIZ A-Transposta

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 |
|---|---|---|---|---|---|---|---|---|---|---|
| P1 | 0 | 0,075 | 0 | 0 | 0,075 | 0 | 0 | 0 | 0 | 0 |
| P2 | 0 | 0 | 0,15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P4 | 0,15 | 0,075 | 0 | 0 | 0,075 | 0 | 0 | 0 | 0 | 0 |
| P5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,075 |
| P7 | 0 | 0 | 0 | 0 | 0 | 0,05 | 0 | 0 | 0 | 0 |
| P8 | 0 | 0 | 0 | 0 | 0 | 0,05 | 0,15 | 0 | 0 | 0 |
| P9 | 0 | 0 | 0 | 0 | 0 | 0,05 | 0 | 0 | 0 | 0,05 |
| P10 | 0 | 0 | 0 | 0,15 | 0 | 0 | 0 | 0 | 0 | 0 |

| MATRIZ1 | | MATRIZ4 | | MATRIZ7 | | MATRIZ10 | |
|---|---|---|---|---|---|---|---|
| P1 | 0,13 | P1 | 0 | P1 | 0 | P1 | 0 |
| P2 | 0,13 | P2 | 0 | P2 | 0 | P2 | 0 |
| P3 | 0 | P3 | 0 | P3 | 0 | P3 | 0 |
| P4 | 0,255 | P4 | 0 | P4 | 0 | P4 | 0 |
| P5 | 0 | P5 | 0 | P5 | 0 | P5 | 0 |
| P6 | 0,06375 | P6 | 0,00032273 | P6 | 1,8154E-07 | P6 | 0 |
| P7 | 0,04 | P7 | 0,00014344 | P7 | 0 | P7 | 0 |
| P8 | 0,0425 | P8 | 0,00021516 | P8 | 0 | P8 | 0 |
| P9 | 0,10625 | P9 | 0,00046617 | P9 | 1,8154E-07 | P9 | 0 |
| P10 | 0,1275 | P10 | 0,00021516 | P10 | 0 | P10 | 0 |

| MATRIZ2 | | MATRIZ5 | | MATRIZ8 | |
|---|---|---|---|---|---|
| P1 | 0,0095625 | P1 | 0 | P1 | 0 |
| P2 | 0 | P2 | 0 | P2 | 0 |
| P3 | 0 | P3 | 0 | P3 | 0 |
| P4 | 0,0286875 | P4 | 1,6137E-05 | P4 | 0 |
| P5 | 0 | P5 | 0 | P5 | 0 |
| P6 | 0,0095625 | P6 | 0 | P6 | 0 |
| P7 | 0,0031875 | P7 | 0 | P7 | 9,0769E-09 |
| P8 | 0,0095625 | P8 | 0 | P8 | 9,0769E-09 |
| P9 | 0,01275 | P9 | 0 | P9 | 9,0769E-09 |
| P10 | 0,03825 | P10 | 0 | P10 | 0 |

| MATRIZ3 | | MATRIZ6 | | MATRIZ9 | |
|---|---|---|---|---|---|
| P1 | 0 | P1 | 0 | P1 | 0 |
| P2 | 0 | P2 | 0 | P2 | 0 |
| P3 | 0 | P3 | 0 | P3 | 0 |
| P4 | 0,00143438 | P4 | 0 | P4 | 0 |
| P5 | 0 | P5 | 0 | P5 | 0 |
| P6 | 0,00286875 | P6 | 0 | P6 | 0 |
| P7 | 0,00047813 | P7 | 0 | P7 | 0 |
| P8 | 0,00095625 | P8 | 0 | P8 | 1,3615E-09 |
| P9 | 0,00334688 | P9 | 0 | P9 | 0 |
| P10 | 0,00430313 | P10 | 2,4205E-06 | P10 | 0 |

**c) Compare o comportamento da convergência para os dois valores iniciais anteriores**

Ambas as abordagens convergem até ao “ponto objetivo” em que todos os vértices atingem o 0. No entanto, uma vez que o segundo valor é bastante mais pequeno do que o primeiro, a convergência é mais rápida e atinge valores que se poderiam considerar finais, que são aproximadamente 0, logo na iteração 5.

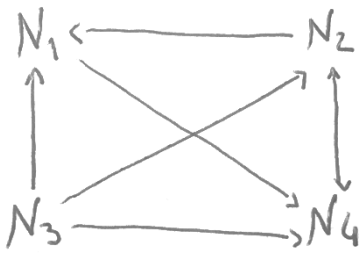
# Exercício 10

**Calcule o PakeRank e o HUBs e Authorities em apenas três iterações para páginas do grafo direto. definido pela seguinte matriz de adjacências**

## HUB's and Authorities

| | $N1$ | $N2$ | $N3$ | $N4$ |
|---|---|---|---|---|
| $N1$ | **0** | **0** | **0** | **1** |
| $N2$ | **1** | **0** | **0** | **1** |
| $N3$ | **1** | **1** | **0** | **1** |
| $N4$ | **0** | **1** | **0** | **0** |

Assim sendo obtém-se a seguinte o gráfico e a tabela que a seguir se apresentam:




| Nodos | Hub | Authority |
|---|---|---|
| N1 | 1 | 1 |
| N2 | 2 | 2 |
| N3 | 3 | 0 |
| N4 | 1 | 3 |

A partir deste ponto devem tirar-se conclusões sobre a organização do Hub assim como do Authority.

**Hub:** N3,N2,N1,N4{TIE}

**Authority:** N4,N1,N2{TIE},N3

Neste ponto, deve calcular-se a matriz transposta da matriz inicial, para que assim se possam calcular os novos valores tanto para o *Hub* como para a *Authority.* A matriz obtida é :

$$\begin{matrix} \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{matrix}$$

Os novos valores para a Authority foram calculados assumindo que o vetor inicial Hub é 1. Assim sendo e assumindo que o Hub e a Authority são representados por u e v , respetivamente, obtêm-se os seguintes valores:

$v = A^T \times u$, sendo que u=1

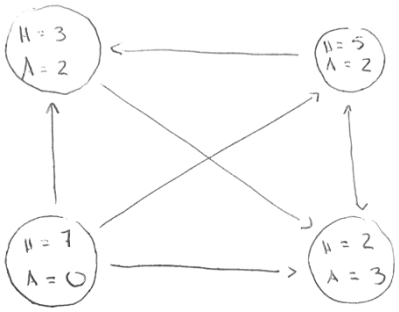
$$\begin{matrix} \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{matrix} \times \begin{matrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{matrix} = \begin{matrix} \mathbf{2} \\ \mathbf{2} \\ \mathbf{0} \\ \mathbf{3} \end{matrix}$$

Estes valores vão corresponder aos valores de Authority para a primeira iteração.

$$u = A \times v$$

$$\begin{matrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \end{matrix} \times \begin{matrix} \mathbf{2} \\ \mathbf{2} \\ \mathbf{0} \\ \mathbf{3} \end{matrix} = \begin{matrix} \mathbf{3} \\ \mathbf{5} \\ \mathbf{7} \\ \mathbf{2} \end{matrix}$$

**Primeira Iteração – K=1**




| Nodos | Hub | Authority |
|---|---|---|
| N1 | 3 | 2 |
| N2 | 5 | 2 |
| N3 | 7 | 0 |
| N4 | 2 | 3 |

**Hub:** N3,N2,N1,N4

**Authority:** N4,N1,N2{TIE},N3

- Novos valores para Authority

$$\mathrm{v1} = 2^2 + 2^2 + 0^2 + 3^2 = 17$$

$$\mathrm{N1} \rightarrow \frac{2}{\sqrt{17}} = 0{,}485$$

$$\mathrm{N2} \rightarrow \frac{2}{\sqrt{17}} = 0{,}485$$

$$\mathrm{N3} \rightarrow \frac{0}{\sqrt{17}} = 0$$

$$\mathrm{N4} \rightarrow \frac{3}{\sqrt{17}} = 0{,}727$$

- Novos valores para Hub

$$\mathrm{u1} = 3^2 + 5^2 + 7^2 + 2^2 = 87$$

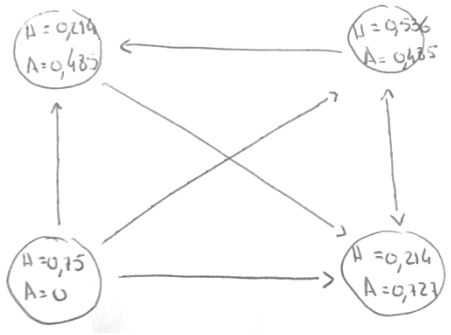
$$\mathrm{N1} \rightarrow \frac{3}{\sqrt{87}} = 0{,}214$$

$$\mathrm{N2} \rightarrow \frac{5}{\sqrt{87}} = 0{,}536$$

$$\mathrm{N3} \rightarrow \frac{7}{\sqrt{87}} = 0{,}750$$

$$\mathrm{N4} \rightarrow \frac{2}{\sqrt{87}} = 0{,}214$$

**Segunda Iteração – K=2**




| Nodos | Hub | Authority |
|---|---|---|
| N1 | 0,214 | 0,485 |
| N2 | 0,536 | 0,485 |
| N3 | 0,750 | 0 |
| N4 | 0,214 | 0,727 |

**Hub:** N3,N2,N1,N4{TIE}

**Authority:** N4,N1,N2{TIE},N3

- Novos valores para Authority

$$u1 = 0{,}485^2 + 0{,}485^2 + 0^2 + 0{,}727^2 = 1$$

$$N1 \rightarrow \frac{0{,}485}{\sqrt{1}} = 0{,}214$$

$$N2 \rightarrow \frac{0{,}485}{\sqrt{1}} = 0{,}536$$

$$N3 \rightarrow \frac{0}{\sqrt{1}} = 0{,}750$$

$$N4 \rightarrow \frac{0{,}727}{\sqrt{1}} = 0{,}214$$

- Novos valores para Hub

$$u1 = 0{,}214^2 + 0{,}536^2 + 0{,}75^2 + 0{,}214^2 = 0{,}841$$

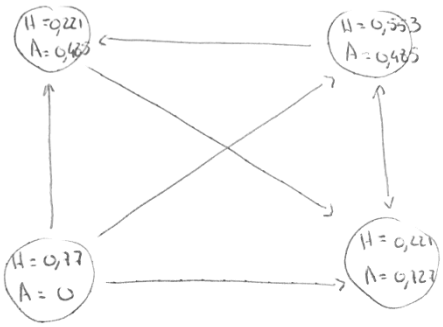
$$N1 \rightarrow \frac{0{,}214}{\sqrt{0{,}841}} = 0{,}221$$

$$N2 \rightarrow \frac{0{,}536}{\sqrt{0{,}841}} = 0{,}553$$

$$N3 \rightarrow \frac{0{,}75}{\sqrt{0{,}841}} = 0{,}77$$

$$N4 \rightarrow \frac{0{,}214}{\sqrt{0{,}841}} = 0{,}221$$

**Terceira Iteração – K=3**

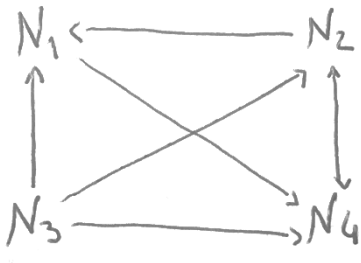


| Nodos | Hub | Authority |
|---|---|---|
| N1 | 0,221 | 0,485 |
| N2 | 0,553 | 0,485 |
| N3 | 0,77 | 0 |
| N4 | 0,221 | 0,727 |

**Hub:** N3,N2,N1,N4{TIE}

**Authority:** N4,N1,N2{TIE},N3

# PakeRank




Assumindo que o PageRank inicial é 1 e o Damping Factor é 0,85.

A Matriz Inicial M é: $\begin{matrix} 0 & 0 & 0 & 1 \\ 1/2 & 0 & 0 & 1/2 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1 & 0 & 0 \end{matrix}$

Fazendo a Transposta desta Matriz e multiplicação Pelo fator de teleporte obtém-se a primeira iteração:

**Primeira Iteração**

$$\begin{matrix} 0 & 1/2 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1/2 & 1/3 & 0 \end{matrix} \times \begin{matrix} 0,85 \\ 0,85 \\ 0,85 \\ 0,85 \end{matrix} = \begin{matrix} 0,71 \\ 1,13 \\ 0 \\ 1,56 \end{matrix}$$

**Segunda Iteração**

$$\begin{matrix} 0 & 1/2 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1/2 & 1/3 & 0 \end{matrix} \times \begin{matrix} 0,71 \\ 1,13 \\ 0 \\ 1,56 \end{matrix} = \begin{matrix} 0,565 \\ 1,56 \\ 0 \\ 1,275 \end{matrix}$$

**Segunda Iteração**

$$\begin{matrix} 0 & 1/2 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1/2 & 1/3 & 0 \end{matrix} \times \begin{matrix} 0,565 \\ 1,56 \\ 0 \\ 1,275 \end{matrix} = \begin{matrix} 0,78 \\ 1,275 \\ 0 \\ 1,345 \end{matrix}$$