ANÁLISE DE REDES
Conceitos Básicos

Licenciatura em
Ciência de Dados

1

Conceitos Básicos

- **Rede**
- Uma **rede** ou **grafo** (*network* ou *graph*) G :
  - é constituída por dois conjuntos – V e A;
  - V é o conjunto de **nodos** (*nodes*), ou conjunto de **vértices** (*vertices*, *vertex* (*sing.*));
  - A é o conjunto de **ligações** (*links*), é constituído por pares ordenados de nodos;
  - A também pode ser designado por conjunto de **arcos** (*arcs*) ou conjunto de **arestas** (*edges*).

2

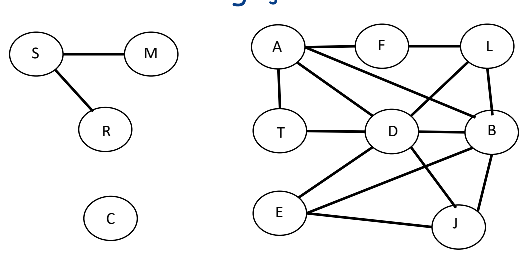
Conceitos Básicos

- **Rede**
- N representa o número de nodos;
- L representa o número de ligações.

3

Conceitos Básicos

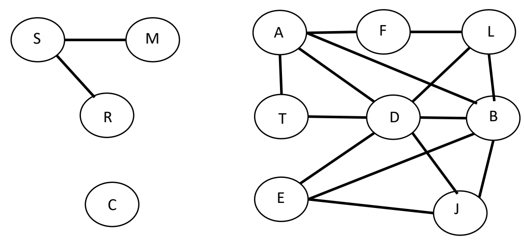
- **Exemplo de uma Rede**
- Nesta rede representam-se um conjunto de 12 pessoas e as suas ligações no *facebook*.



4

Conceitos Básicos

* **Exemplo de uma Rede (continuação)**



* $V = \{A, B, C, D, E, F, J, L, M, R, S, T\}$; $N = 12$;
* $(S,M) \in A$ (ou $(M,S) \in A$); $L = 16$.

5

Conceitos Básicos

* **Rede**
* Um **<u>lacete</u>** (*loop*) é uma ligação que une o mesmo nodo.
* A ligação (1,1) é um lacete.

6

Conceitos Básicos

* **Redes Orientadas e Redes não Orientadas**
* Se as ligações não têm orientação, a rede é **<u>não orientada</u>** (*undirected network*).
* Se as ligações têm orientação, a rede é **<u>orientada</u>** (*directed network*).
* As redes orientadas também são designadas por **<u>digrafos</u>** (*digraphs*).

7

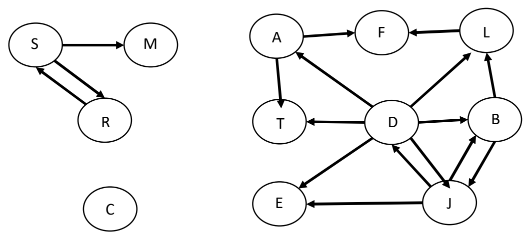
Conceitos Básicos

* **Redes Orientadas e Redes não Orientadas**
* A rede do exemplo anterior é não orientada.
* As ligações de uma rede não orientada também são designadas por **<u>arestas</u>** (*edges*).
* As ligações de uma rede orientada também são designadas por **<u>arcos</u>** (*arcs*).

8

Conceitos Básicos

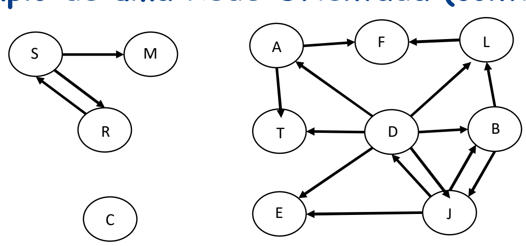
* **Exemplo de uma Rede Orientada**
* Nesta rede representam-se as mensagens enviadas por um conjunto de 12 pessoas.



9

Conceitos Básicos

* **Exemplo de uma Rede Orientada (continuação)**

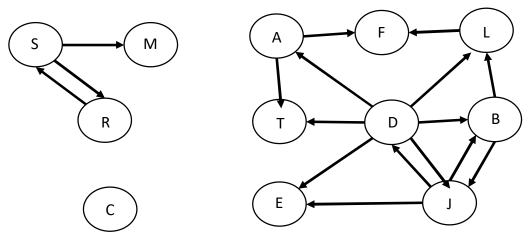


* N = 12, L = 17.

10

Conceitos Básicos

- **Exemplo de uma Rede Orientada (continuação)**



- A ligação (D,A) representa o envio de uma ou mais mensagens de D para A.

11

Conceitos Básicos

- **Exemplos de redes orientadas:**
  - a *Web*,
  - a *Wikipedia*,
  - o *Twitter*,
  - a rede de *emails*,
  - as redes biológicas,
  - as redes de transportes.

12

Conceitos Básicos

* **Redes com pesos associados às ligações**
* Diversas redes não têm pesos associados às ligações (*unweighted networks*).
* Outras redes têm atributos associados às ligações (*weighted networks*).
* Estes pesos podem representar a interacção entre os diferentes nodos, que pode não ser igual para todos os pares de nodos.
* Torna-se então relevante representar a intensidade ou o valor do fluxo entre os nodos através de pesos associados às ligações.

13

Conceitos Básicos

* **Redes com pesos associados às ligações**
* Os pesos podem representar:
  * a antiguidade, em número de anos, da relação de amizade no Facebook;
  * o número de mensagens enviadas ou o volume de *email*,
  * o número de *retweets*,
  * o número de passageiros, numa rede de transporte.

14

Conceitos Básicos

- **Nodos Adjacentes**
- Se (i,j) é uma ligação da rede G então nodos i e j são **nodos adjacentes** (*adjacent nodes*), *i* é adjacente de *j* e *j* é adjacente de *i*.
- Em inglês também se pode utilizar o termo *neighbor* para nodo adjacente.

15

Conceitos Básicos

- **Densidade e Esparsidade de Redes**
- **Rede Completa** (*complete network*)
  - Uma rede é **completa** (*complete network*) se existe uma ligação entre cada nodo e cada um dos restantes.
  - Representa-se o número máximo de ligações por L_{max}.

16

Conceitos Básicos

- **Densidade e Esparsidade de Redes**
- **Rede Completa** (*complete network*)
- Para redes não orientadas

  $L_{max} = N(N-1)/2$;
- Para redes orientadas

  $L_{max} = N(N-1)$.

17

Conceitos Básicos

- **Densidade e Esparsidade de Redes**
- **Densidade**
- A **densidade** (*density*), d, é uma medida relativa que relaciona o número de ligações existentes numa rede com o número máximo possível de ligações.

$$d = \frac{L}{L_{max}}$$

18

Conceitos Básicos

- **Densidade e Esparsidade de Redes**
- **Densidade**
- Para redes não orientadas

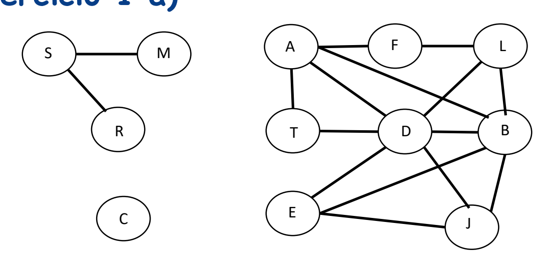
  $\mathrm{d} = 2L/N(N-1)$;
- Para redes orientadas

  $\mathrm{d} = L/N(N-1)$.

19

Conceitos Básicos

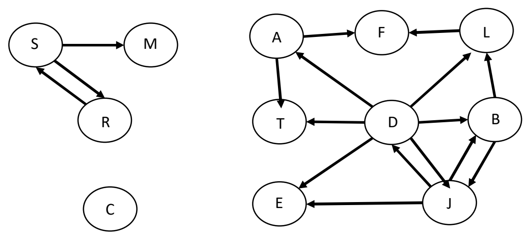
- **Densidade e Esparsidade de Redes**
- **Densidade**
- **Exercício 1 a)**



20

Conceitos Básicos

- **Densidade e Esparsidade de Redes**
- **Densidade**
- **Exercício 2 a)**



Conceitos Básicos

- **Densidade e Esparsidade de Redes**
- **Redes Esparsas e Redes Densas**
- Uma rede é **esparsa** (*sparse*) se o número de ligações for bastante inferior a L_{max}

  (se $L \ll L_{max}$).
- Caso contrário, a rede é **densa** (*dense*).
- Em geral, as redes reais são esparsas.

Conceitos Básicos

* **Densidade e Esparsidade de Redes**
* **Redes Esparsas e Redes Densas**
* *As redes anteriores são esparsas, porque*
  16 « 66 e 17 « 132.

23

Conceitos Básicos

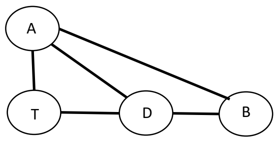
* **Subredes**
* Uma **<u>subrede</u>** (*subnetwork*) consiste numa rede que se obtém considerando apenas um subconjunto de nodos e todas as ligações da rede original que unem os elementos do subconjunto.
* Este conceito aplica-se tanto a redes não orientadas como a redes orientadas.

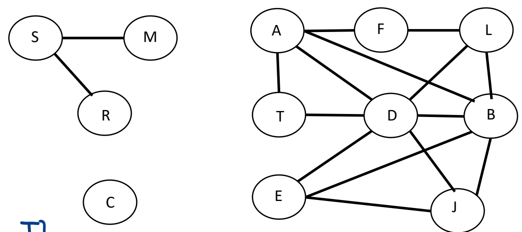
24

Conceitos Básicos

* **Exemplo**

é uma subrede de





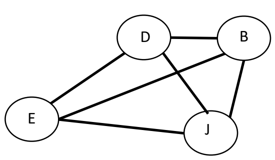
gerada por V_1= {A, B, D, T}.

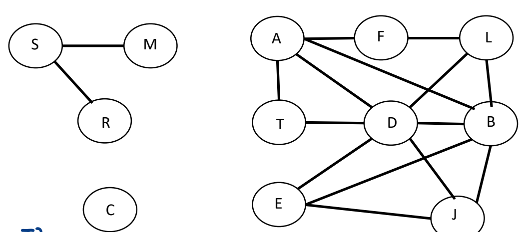
25

Conceitos Básicos

* **Exemplo**

é uma subrede de





gerada por V_2= {B, D, E, J}.

26

Conceitos Básicos

* **Subredes**
* Uma **clique** (clique) é uma subrede completa.
* A subrede do exemplo anterior é uma clique.

27

Conceitos Básicos

* **Subredes**
* Um caso particular de subredes é a **rede ego** (*ego network*) de um nodo.
* Esta subrede é gerada pelo subconjunto de nodos cujos elementos são o nodo escolhido e os seus adjacentes.
* Estas redes são frequentemente estudadas em ciências sociais.

28

Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede não orientada**
- O **grau** (*degree*) de um nodo i de uma rede não orientada é o número de ligações incidentes nesse nodo e representa-se por k_i.
- Na determinação do grau, um lacete é contabilizado como duas ligações.

29

Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede não orientada**
- A média dos graus dos nodos de uma rede é designada por **grau médio** (*average degree*) e representa-se por $< k >$.

$$< k >= \frac{\sum_i k_i}{N}$$

é equivalente a

$$< k >= \frac{2\,L}{N}$$

30

Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede não orientada**
* Atendendo à definição de densidade, o grau médio também pode ser expresso como:

$$< k >= d\ (N-1).$$

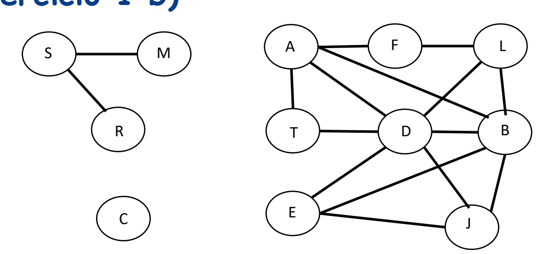
* Equivalentemente

$$d = \frac{< k >}{N-1}.$$

31

Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede não orientada**
* **Exercício 1 b)**



32

Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede orientada**
* Numa rede orientada, a orientação das ligações é relevante.
* Assim, definem-se o grau incidente e o grau divergente.

33

Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede orientada**
* O **grau incidente** (*in-degree*) de um nodo i de uma rede orientada é o número de ligações incidentes nesse nodo (que se dirigem ao nodo) e representa-se por k_i^{in}.
* O **grau divergente** (*out-degree*) de um nodo i de uma rede orientada é o número de ligações divergentes desse nodo (que partem do nodo) e representa-se por k_i^{out}.

34

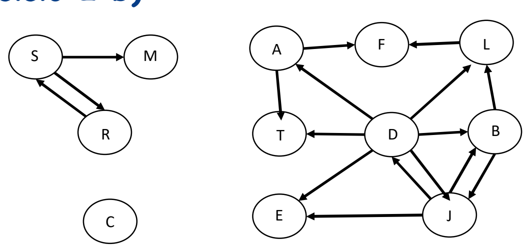
Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede orientada**
- O **<u>grau médio incidente</u>** (*average in-degree*) [**<u>grau médio divergente</u>** (*average out-degree*)] pode ser calculado dividindo a soma dos graus incidentes [graus divergentes] pelo número de nodos e representa-se por $< k^{in} >$ [$< k^{out} >$].

35

Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede orientada**
- **Exercício 2 b)**



36

Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede não orientada com pesos associados às ligações**
* Seja G uma rede não orientada com pesos associados às ligações, representados por w_{ij}: $(i, j) \in A$.
* O **grau ponderado** ou **força** de um nodo (*weighted degree* ou *strength*) é dado por:

$$s_i = \sum_j w_{ij}.$$

37

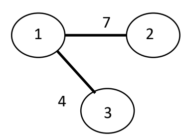
Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede não orientada com pesos associados às ligações**
* O **grau médio ponderado** ou **força média** (*average weighted degree* ou *average strength*) pode ser calculado dividindo a soma dos graus ponderados pelo número de nodos e representa-se por $< s >$.

38

Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede não orientada com pesos associados às ligações**
- **Exercício 3 a)**



Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede orientada com pesos associados às ligações**
- Seja G uma rede orientada com pesos associados às ligações, representados por w_{ij} $em\ que\ (i,j) \in A$.
- *O* **grau incidente ponderado** ou **força incidente** de um nodo (*weighted in-degree* ou *in-strength*) é dado por:

$$s_j^{in} = \sum_i w_{ij}.$$

Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede orientada com pesos associados às ligações**
- *O* **<u>grau divergente ponderado</u>** ou **<u>força divergente</u>** de um nodo (*weighted out-degree* ou *out-strength*) é dado por:

$$s_i^{out} = \sum_j w_{ij}.$$

41

Conceitos Básicos

- **Grau**
- **Grau de um nodo de uma rede orientada com pesos associados às ligações**
- *O* **<u>grau médio incidente ponderado</u>** ou **<u>força média incidente</u>** (*average weighted in-degree ou average in-strength*) pode ser calculado dividindo a soma dos graus incidentes ponderados pelo número de nodos e representa-se por $< s^{in} >$.

42

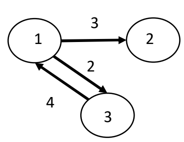
Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede orientada com pesos associados às ligações**
* O **grau médio divergente ponderado** ou **força média divergente** (*average weighted out-degree ou average out-strength*) pode ser calculado dividindo a soma graus divergentes ponderados pelo número de nodos e representa-se por $< s^{out} >$.

43

Conceitos Básicos

* **Grau**
* **Grau de um nodo de uma rede orientada com pesos associados às ligações**
* **Exercício 4 a)**



44

Conceitos Básicos

* **Utilização dos Conceitos de Grau para caracterizar a *Web***
* A *Web* pode ser representada numa rede.
* Os documentos são representados nos nodos e as ligações da rede representam as hiperligações.
* Esta rede é designada por ***grafo da Web*** (*Web graph*).
* Diversos estudos sobre os graus das páginas da Web têm sido realizados utilizando *crawlers*.

45

Conceitos Básicos

* **Utilização dos Conceitos de Grau para caracterizar a *Web***
* Estes estudos indicam que o grau médio incidente (número de hiperligações para uma página) se situa entre 10 e 30.
* Contudo, o desvio padrão é ainda superior.
* Assim, a média não é uma boa medida para caracterizar esta rede.

46

Conceitos Básicos

* **Utilização dos Conceitos de Grau para caracterizar a *Web***
* Passados poucos anos após o surgimento desta rede, verificou-se que a distribuição do grau incidente era assimétrica.
* Com o aumento da dimensão da *Web*, a ordem de grandeza de muitos nodos da rede cresceu, mas a assimetria permaneceu.

47

Conceitos Básicos

* **Utilização dos Conceitos de Grau para caracterizar a *Web***
* A análise da distribuição do grau divergente é mais difícil.
* É fácil encontrar páginas com muitas ligações dirigidas para outras páginas, contudo a distribuição do grau divergente não mostra a mesma ordem de grandeza que a distribuição de grau incidente.

48

Conceitos Básicos

* **Utilização dos Conceitos de Grau para caracterizar a *Web***
* Isto pode dever-se também ao facto dos *crawlers* frequentemente truncarem as páginas muito longas, não se obtendo informação fiável sobre o grau divergente.
* Por outro lado, uma página com demasiadas ligações orientadas para outras páginas é típica de um comportamento de *spam*.

49

Conceitos Básicos

* **Utilização dos Conceitos de Grau para caracterizar a *Web***
* Quando uma página ou documento tem ligações incidentes que provêm de páginas com grau incidente nulo, então está-se provavelmente perante *spam*.
* Pelo contrário, se uma página ou documento tem ligações incidentes que provêm de páginas com grau incidente elevado, então é expectável que a página ou documento tenha prestígio ou qualidade elevada.

50

Conceitos Básicos

* **Redes Multicamada**
* As ligações de cada uma das redes já consideradas representam apenas um tipo de relação entre os nodos.
* Pode ser vantajoso representar numa mesma rede diversos tipos de relação entre os nodos.
* Por exemplo, para o transporte entre diversas cidades, algumas ligações podem representar o transporte aéreo, enquanto que outras representam o transporte ferroviário.

51

Conceitos Básicos

* **Redes Multicamada**
* Neste caso, em cada camada (*layer*) da rede as ligações representam um destes tipos de transporte e obtém-se uma **rede multicamada** (*multilayer network*).
* Outro exemplo pode ser obtido considerando um conjunto de pessoas. Pode ser necessário representar, em simultâneo, diversos tipos de relação (amizade, laços familiares, relações laborais,...).

52

Conceitos Básicos

* **Redes Temporais**
* Outras redes, ainda não mencionadas, são as **redes temporais** (*temporal networks*).
* Nestas redes as ligações são dinâmicas, no sentido em que as interações entre os nodos ocorrem e mudam ao longo do tempo.

53

Conceitos Básicos

* **Redes Temporais**
* Podem considerar-se alguns intervalos de tempo. A rede associada a cada intervalo pode ser considerada uma camada (*layer*) de uma rede multicamada.
* A agregação de todos os intervalos de tempo permite obter uma rede estática.

54

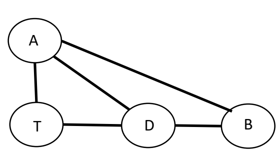
Conceitos Básicos

- **Estruturas de dados utilizadas para representar redes**
- **Matriz de Adjacência**
- Na **<u>matriz de adjacência</u>** (*adjancency matrix*), usualmente representada por A, o elemento a_{ij} toma o valor 1 se existe a ligação (i,j) e o valor 0, no caso contrário.

55

Conceitos Básicos

- **Estruturas de dados utilizadas para representar redes**
- **Exemplo (Matriz de Adjacência)**



$$\begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{vmatrix}$$

56

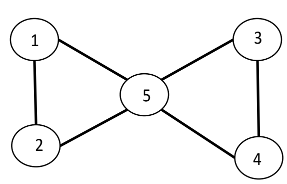
Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Matriz de Adjacência**
* Para redes não orientadas contém informação em duplicado.
* Para redes esparsas contém muitos zeros.
* É uma estrutura bastante conhecida, mas não é eficiente.

57

Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Matriz de Adjacência**
* **Exercício 5 a)**



58

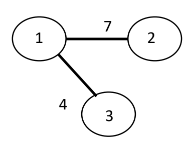
Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Matriz de Adjacência para redes com pesos associados às ligações**
* Os elementos iguais a 1 são substituídos pelo peso da ligação associada.

59

Conceitos Básicos

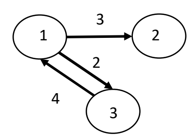
* **Estruturas de dados utilizadas para representar redes**
* **Matriz de Adjacência para redes com pesos associados às ligações**
* **Exercício 3 b)**



60

Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Matriz de Adjacência para redes com pesos associados às ligações**
* **Exercício 4 b)**



61

Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Lista de Ligações**
* *A* **lista de ligações** (*edge list* ou *arc list*) consiste em dois vectores.
* Num dos vectores indica-se um dos extremos de cada ligação. No outro vector, na componente homóloga, indica-se o outro extremo da ligação.

62

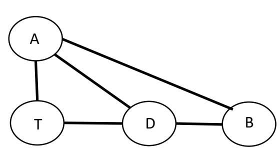
Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Lista de Ligações**
* A lista de ligações também pode ser representada numa matriz com duas colunas.

63

Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Exemplo (Lista de Ligações)**



A	A	A	B	D
B	D	T	D	T

A ligação (A,B) está representada apenas uma vez.

64

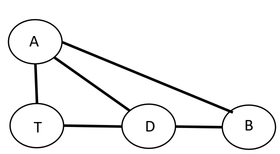
Conceitos Básicos

- **Estruturas de dados utilizadas para representar redes**
- **Lista de Adjacência**
- *A* **lista de adjacência** (*adjacency list*) consiste em dois vectores.

65

Conceitos Básicos

- **Estruturas de dados utilizadas para representar redes**
- **Lista de Adjacência**
- **Exemplo**



1	4	6	9	11

B	D	T	A	D	A	B	T	A	D

66

Conceitos Básicos

- **Estruturas de dados utilizadas para representar redes**
- **Lista de Adjacência**
- **Exemplo**

1	4	6	9	11

B	D	T	A	D	A	B	T	A	D

- No segundo vector, nas componentes de 1 a (4-1) estão os nodos adjacentes de A, de 4 a (6-1) estão os adjacentes de B,...

67

Conceitos Básicos

- **Estruturas de dados utilizadas para representar redes**
- **Lista de Adjacência**
- **Exemplo**

1	4	6	9	11

B	D	T	A	D	A	B	T	A	D

- Nesta lista, a ligação (A,B) está representada duas vezes. B é nodo adjacente de A (primeira componente do segundo vetor) e A é nodo adjacente de B (quarta componente do segundo vetor)

68

Conceitos Básicos

* **Estruturas de dados utilizadas para representar redes**
* **Listas para redes com pesos associados às ligações**
* Acrescenta-se um terceiro vector com os pesos das ligações.

69

Conceitos Básicos

* **Representação Gráfica de Redes**
* Pode ser útil para identificar e visualizar algumas propriedades e características das redes.
* Por exemplo, quais são os nodos com grau bastante mais elevado do que os restantes.

70

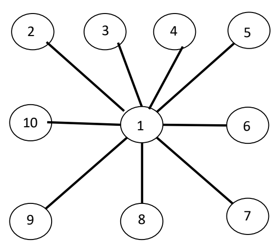
Conceitos Básicos

* **Casos Particulares de Redes**
* **Rede em Estrela** (*star network*)
  * Um e um só nodo está ligado a todos os restantes;
  * Não existem outras ligações, além das anteriores.

71

Conceitos Básicos

* **Casos Particulares de Redes**
* **Rede em Estrela** (*star network*)
* **Exemplo**



72

Conceitos Básicos

* **Casos Particulares de Redes**
* **Redes Bipartidas** (*bipartite networks*)
  * o conjunto dos nodos está dividido em dois subconjuntos, um com N_1 nodos e o outro com N_2 nodos;
  * os nodos adjacentes pertecem a conjuntos diferentes, o que significa que não existem ligações entre os nodos de um mesmo subconjunto.

73

Conceitos Básicos

* **Casos Particulares de Redes**
* **Redes Bipartidas** (*bipartite networks*)
* **Exemplo**
* Um dos subconjuntos de nodos representa as doenças humanas;
* O outro subconjuntos é constituído pelos genes cujas mutações originam e infuenciam as doenças;
* As ligações unem doenças e genes.

74

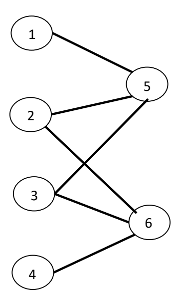
Conceitos Básicos

* **Casos Particulares de Redes**
* **Redes Bipartidas** (*bipartite networks*)
* **Exemplo**
* Um dos subconjuntos de nodos representa atores;
* O outro subconjuntos é constituído por filmes;
* Cada ligação une um ator a um filme se o ator representou um papel nesse filme.

75

Conceitos Básicos

* **Casos Particulares de Redes**
* **Redes Bipartidas** (*bipartite networks*)
* **Exemplo**



76

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- A partir de uma rede bipartida, podem obter-se outras duas redes em que o conjunto dos nodos é apenas um dos subconjuntos da rede bipartida.
- Cada umas destas novas redes é designada por **projeção** (*projection*).

77

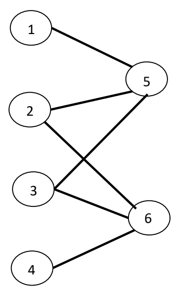
Conceitos Básicos

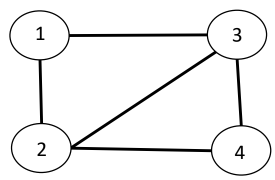
- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- As ligações de uma projeção são os pares ordenados de nodos que partilham pelo menos um nodo adjacente na rede bipartida.

78

Conceitos Básicos

* **Casos Particulares de Redes**
* **<u>Redes Bipartidas</u>** (*bipartite networks*)
* **Exemplo**





79

Conceitos Básicos

* **Casos Particulares de Redes**
* **<u>Redes Bipartidas</u>** (*bipartite networks*)
* Por exemplo, se numa rede bipartida se tem um conjunto de doenças humanas e um conjunto de genes, as ligações unem as doenças e os genes que as originam.
* Então uma projeção é obtida para as doenças. Nesta rede, uma ligação entre doenças indica que ambas estão associadas a mutações em, pelo menos, um gene comum.

80

Conceitos Básicos

- **Casos Particulares de Redes**
- **<u>Redes Bipartidas</u>** (*bipartite networks*)
- A outra projeção refere-se aos genes. Uma ligação entre dois genes indica que uma ou mais doenças são provocadas por mutações nos dois genes.

81

Conceitos Básicos

- **Casos Particulares de Redes**
- **<u>Redes Bipartidas</u>** (*bipartite networks*)
- **Redes Co-ocorrentes**
- Se a projeção resultante for uma rede com pesos associados às ligações então é designada por **<u>rede co-ocorrente</u>** ou **<u>rede de co-ocorrência</u>** (*co-occurent network*).
- Na projeção da rede bipartida das doenças humanas, pode ter-se o número de genes em comum causadores do par de doenças ou o número de doenças causadas por mutações nos dois genes.

82

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- As redes bipartidas também podem representar consumo de bens ou serviços. Neste caso, as projecções podem ser úteis para identificar padrões de consumo.
- **Exercício 8**

83

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- Se N_1 e N_2 representam, respectivamente, o número de nodos do primeiro e do segundo subconjuntos então

  $L_{max} = N_1 x\ N_2$.
- A densidade de uma rede bipartida é dada por

  $\mathrm{d} = L/(N_1 x N_2)$.

84

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- As redes tripartidas são uma generalização das redes bipartidas.
- Nestas redes o conjunto de nodos é constituído por três subconjuntos de nodos, por exemplo, V_1, V_2 e V_3.
- Existem dois subconjuntos de ligações – ligações entre V_1 e V_2; e entre V_2 e V_3.

85

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- Estas redes tripartidas podem ser utilizadas representar os nutrientes presentes em receitas de culinária.
- Um dos subconjuntos representa as receitas, outro representa os ingredientes e o terceiro subconjunto refere-se aos nutrientes.

86

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- Em alguns contextos, pode obter-se uma rede bipartida, a partir de uma rede tripartida.
- Também se podem obter projeções de redes tripartidas num só conjunto de nodos.

87

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- Por exemplo, a partir da redes em que representam as receitas, os ingredientes e os nutrientes pode obter-se uma rede em que se representam apenas os ingredientes.
- Entre os ingredientes que partilham, pelo menos, um nutriente existe uma ligação.

88

Conceitos Básicos

- **Casos Particulares de Redes**
- **Redes Bipartidas** (*bipartite networks*)
- Os nodos podem ser representados com dimensões diferentes. A dimensão representará a prevalência do ingrediente nas receitas.
- A espessura das ligações pode ser proporcional ao número de nutrientes partilhados.

89

ANÁLISE DE REDES
Small Worlds

Licenciatura em
Ciência de Dados

Small Worlds

- Vamos estudar três características das redes:
  - semelhança entre nodos adjacentes;
  - existência de caminhos curtos entre os nodos;
  - existência de triângulos formados por nodos adjacentes.

2

Small Worlds

- As redes sociais são algumas das redes em que as características mencionadas estão presentes.
- Numa rede social:
  - as pessoas são representadas nos nodos,
  - as ligações representam as relações (amizade, laços familiares, conhecimento, trabalho) entre as pesooas.

3

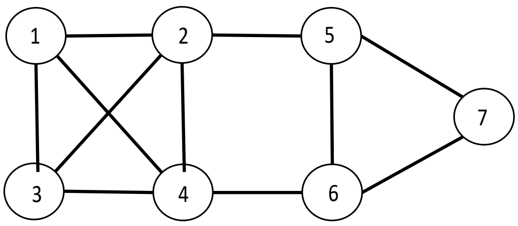
Small Worlds

- **Associação** (*Assortativity*)
- Em muitas redes existem ligações entre nodos que representam entidades com algumas características comuns.
- Esta **associação** é designada em inglês por *assortativity*.

4

Small Worlds

- **Exemplo 1**
- Considere-se a rede com 7 nodos e 11 ligações:

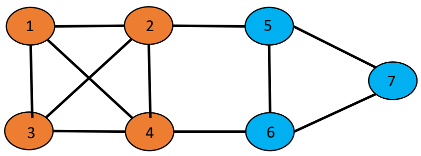


Small Worlds

- **Exemplo 1**
- Verifica-se que existe uma ligação entre cada par de nodos do subconjunto {1, 2, 3, 4}.
- Apenas 2 e 4 estão ligados a nodos que não pertecem ao subconjunto.
- Para o subconjunto {5, 6, 7} ocorre algo semelhante.

Small Worlds

* **Exemplo 1**
* Pode concluir-se que a maioria das ligações desta rede unem elementos do mesmo subconjunto.



7

Small Worlds

* As redes sociais são exemplos de redes em que as ligações representam relações que traduzem alguma semelhança entre os nodos.
* Em geral, os amigos partilham características, opiniões comuns e tendem a tornar-se mais parecidos ao longo do tempo.
* Este fenómeno pode dever-se ao facto de as semelhanças aproximarem as pessoas ou as semelhanças ocorrerem por influência das pessoas com quem convivem.

8

Small Worlds

- Por outro lado, em algumas destas redes, verifica-se uma divisão dos nodos em subconjuntos, havendo uma polarização.
- Este fenómeno **pode ter consequências negativas** (manipulação de opinião, desinformação, ...), que têm sido estudadas.

9

Small Worlds

- **Associação de grau** (*degree assortativity*)
- A associação (*assortativity*) não é exclusiva das redes sociais.
- Em algumas redes, a associação está presente no grau dos nodos.
- Este tipo de associação é designado por **associação de grau** (*degree assortativity*) ou **correlação de grau** (*degree correlation*).

10

Small Worlds

- **Associação de grau** (*degree assortativity*)
- Ocorre quando os nodos de maior grau estão ligados principalmente a nodos de grau elevado e os de menor grau estão unidos principalmente a nodos de grau reduzido.

11

Small Worlds

- **Rede associativa** (*assortative network*)
- Uma rede com a propriedade anterior é designada por **rede associativa** (*assortative network*).
- Nestas redes, a parte mais densa é contituída pelos nodos de maior grau, enquanto que os de menor grau constituem a estrutura periférica.

12

Small Worlds

- **Rede associativa** (*assortative network*)
- No caso dos nodos de maior grau estarem ligados aos de menor grau, a rede é **não associativa** (*disassortative*).
- Uma rede que não é associativa nem não associativa é uma rede **neutra** (*neutral network*).

13

Small Worlds

- **Rede associativa** (*assortative network*)
- A *Web*, a rede de emails e as redes biológicas são redes não associativas.
- A rede de electricidade é neutra.

14

Small Worlds

* **Rede associativa** (*assortative network*)
* A associação baseada no grau pode ser medida de duas formas:
  * a primeira consiste em determinar o coeficiente de correlação de Pearson para os graus de nodos adjacentes;
  * a segunda forma é baseada no grau médio dos nodos adjacentes de cada nodo.

15

Small Worlds

* **Rede associativa** (*assortative network*)
* O coeficiente de correlação de Pearson toma valores entre -1 e 1.
* Se o valor for nulo, significa que não existe correlação.
* Se tomar o valor 1, a correlação é positiva perfeita. Para o valor -1, a correlação é negativa perfeita.

16

Small Worlds

* **Rede associativa** (*assortative network*)
* Uma correlação entre duas variáveis é positiva se valores elevados de uma das variáveis estão associados a valores elevados da outra variável.
* É negativa se valores elevados de uma das variáveis estão associados a valores reduzidos da outra variável.

17

Small Worlds

* **Rede associativa** (*assortative network*)
* Se o coeficiente toma um valor significativo positivo então a rede é associativa.
* Se o valor do coeficiente for significativo e negativo então a rede é não associativa.
* **Ser não associativa ≠ não ser associativa**

18

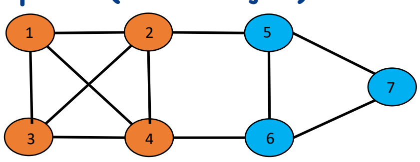
Small Worlds

- **Rede associativa** (*assortative network*)
- O valor do coeficiente pode não ser significativo ou difícil de interpretar, sendo necessária outra abordagem.
- **Ser não associativa ≠ não ser associativa**

19

Small Worlds

- **Rede associativa** (*assortative network*)
- **Exemplo 1 (continuação)**

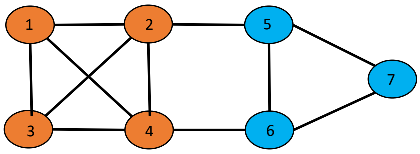


- O valor do coeficiente de correlação de Pearson é igual a 0,043.
- Conclusão?

20

Small Worlds

- **<u>Rede associativa</u>** (*assortative network*)
- **Exemplo 1 (continuação)**

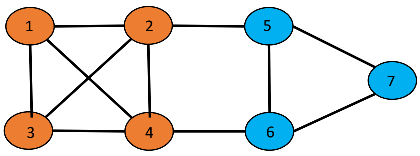


- Trata-se de um valor positivo mas pouco signficativo, o que não permite considerar que a rede é associativa.

21

Small Worlds

- **<u>Rede associativa</u>** (*assortative network*)
- **Exemplo 1 (continuação)**



- Este resultado não surpreende, pois dois dos sete nodos estão ligados não só aos nodos com maior grau, mas também ao nodo com menor grau. A rede é neutra.

22

Small Worlds

- **Rede associativa** (*assortative network*)
- **Exemplo 1 (continuação)**

Nodo i	Nodo j	Grau de i	Grau de j
1	2	3	4
1	3	3	3
1	4	3	4
2	1	4	3
…	…	…	…

- Correlação entre o grau de i e o grau de j.

23

Small Worlds

- **Rede associativa** (*assortative network*)
- Para medir a associação de grau com base no grau médio dos nodos adjacentes considere-se

$$k_{nn}(i) = \frac{1}{k_i} \sum_j a_{ij}\, k_j$$

em que $a_{ij} = 1$ se i e j são adjacentes e 0, no caso contrário.

24

Small Worlds

* **Rede associativa** (*assortative network*)
* Seja também a função

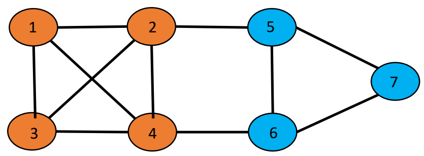
$$\langle k_{nn}\ (k)\rangle$$

que consiste na média dos valores $k_{nn}(i)$ calculada para os nodos i com grau igual a k.
* Se esta função é crescente então a rede é associativa.
* Se a função é decrescente então a rede é não associativa.

25

Small Worlds

* **Rede associativa** (*assortative network*)
* **Exemplo 1 (continuação)**



$$k_{nn}(1) = \frac{1}{3}(4+3+4) = \frac{11}{3};\ k_{nn}(2) = \frac{1}{4}(3+3+4+3) = \frac{13}{4}$$

26

Small Worlds

- **Rede associativa** (*assortative network*)
- **Exemplo 1 (continuação)**

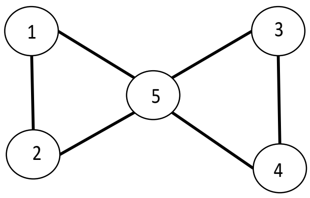
Nodos	k_{nn}	Nodos	k_{nn}	Grau	$<k_{nn}>$
1	11/3	5	3	2	3
2	13/4	6	3	3	10/3
3	11/3	7	3	4	13/4
4	13/4				

- **Conclusão?**

27

Small Worlds

- **Rede associativa** (*assortative network*)
- **Exercício 1 a) – c)**



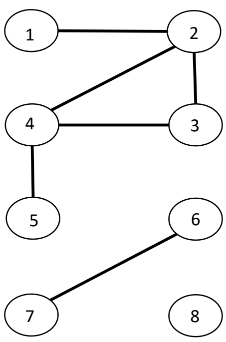
28

Small Worlds

- **Caminhos e Distâncias**
- Numa rede existem L ligações.
- Estas ligações permitem unir diversos pares de nodos através de uma sequência de ligações.
- Um **caminho** (*path*) é uma sequência de ligações que une um par de nodos. Cada par de ligações consecutivas tem, pelo menos, um extremo em comum.
- No caso das redes orientadas, esta sequência tem que ter em conta a orientação das ligações.

Small Worlds

- **Caminhos e Distâncias**

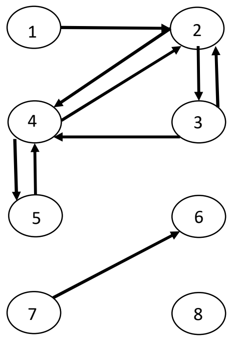


{(1,2), (2,3), (3,4), (4,5)} e {(1,2), (2,4), (4,5)} são dois caminhos entre 1 e 5.

Não existem caminhos entre 1 e 6.

Small Worlds

* **Caminhos e Distâncias**



{(1,2), (2,3), (3,4), (4,5)} e {(1,2), (2,4), (4,5)} são dois caminhos entre 1 e 5.

Não existem caminhos entre 5 e 1.

31

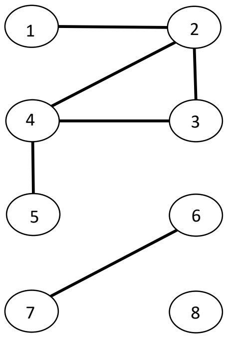
Small Worlds

* **Caminhos e Distâncias**
* Para redes sem pesos associados às ligações, o **comprimento de um caminho** (*path length*) é o número de ligações consideradas no caminho.
* Para redes com pesos associados às ligações, o **comprimento de um caminho** é a soma dos pesos das ligações que constituem o caminho.

32

Small Worlds

* **Caminhos e Distâncias**



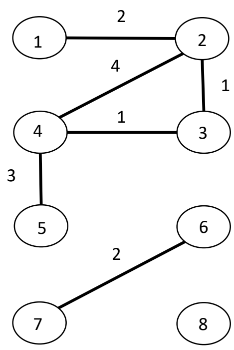
{(1,2), (2,3), (3,4), (4,5)} tem comprimento igual a 4.

{(1,2), (2,4), (4,5)} tem comprimento igual a 3.

33

Small Worlds

* **Caminhos e Distâncias**



{(1,2), (2,3), (3,4), (4,5)} tem comprimento igual a 7.

{(1,2), (2,4), (4,5)} tem comprimento igual a 9.

34

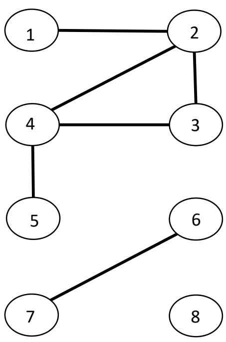
Small Worlds

- **Caminhos e Distâncias**
- O **caminho mais curto** (*shortest path*) entre um par de nodos é o caminho entre esse par de nodos com menor comprimento.
- Pode existir mais do que um caminho mais curto entre um par de nodos.
- Um **caminho simples** (*simple path*) é um caminho em que não se repetem ligações.
- Um **ciclo** (*cycle*) é um caminho que parte de um nodo e regressa a esse nodo.

35

Small Worlds

- **Caminhos e Distâncias**



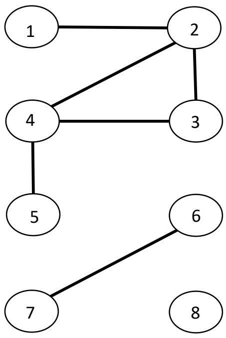
{(1,2), (2,3), (3,4), (4,5)} e {(1,2), (2,4), (4,5)} são caminhos simples entre 1 e 5.

{(1,2), (2,4), (4,5)} é o caminho mais curto entre 1 e 5.

36

Small Worlds

* **Caminhos e Distâncias**



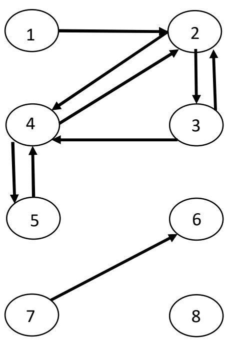
A sequência {(2,3), (3,4), (4,2)} constitui um ciclo.

Existem também caminhos entre 5 e 1.

37

Small Worlds

* **Caminhos e Distâncias**



{(1,2), (2,3), (3,4), (4,5)} e {(1,2), (2,4), (4,5)} são dois caminhos entre 1 e 5.

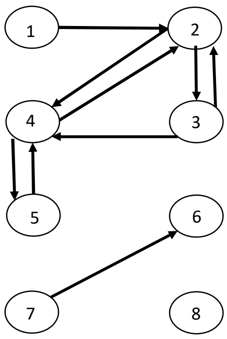
{(1,2), (2,3), (3,4), (4,5)} tem comprimento igual a 4.

{(1,2), (2,4), (4,5)} tem comprimento igual a 3.

38

Small Worlds

* **Caminhos e Distâncias**



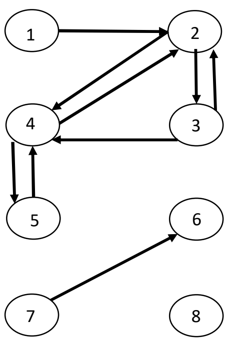
{(1,2), (2,3), (3,4), (4,5)} e {(1,2), (2,4), (4,5)} são caminhos simples entre 1 e 5.

{(1,2), (2,4), (4,5)} é o caminho mais curto entre 1 e 5.

39

Small Worlds

* **Caminhos e Distâncias**

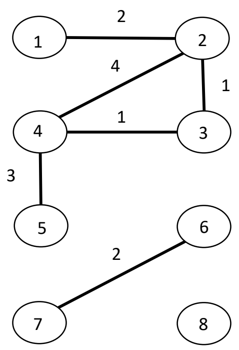


A sequência {(2,3), (3,4), (4,2)} constitui um ciclo.

40

Small Worlds

* **Caminhos e Distâncias**



{(1,2), (2,3), (3,4), (4,5)} é um caminho com comprimento igual a 7.

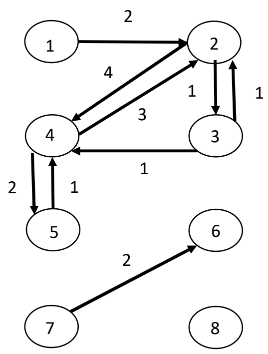
{(1,2), (2,4), (4,5)} é um caminho com comprimento igual a 9.

{(1,2), (2,3), (3,4), (4,5)} é o caminho mais curto entre 1 e 5.

41

Small Worlds

* **Caminhos e Distâncias**



{(1,2), (2,3), (3,4), (4,5)} é um caminho com comprimento igual a 6.

{(1,2), (2,4), (4,5)} é um caminho com comprimento igual a 8.

{(1,2), (2,3), (3,4), (4,5)} é o caminho mais curto entre 1 e 5.

42

Small Worlds

- **Caminhos e Distâncias**
- Vamos agora considerar o problema de determinação do número de caminhos, existentes numa rede, entre dois nodos i e j com k ligações.
- Este problema pode ser resolvido recorrendo à multiplicação de matrizes de adjacência.

43

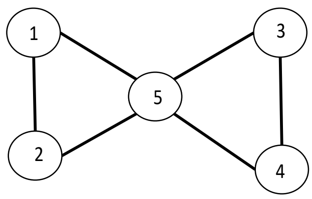
Small Worlds

- **Caminhos e Distâncias**
- Seja A a matriz de adjacência de uma rede sem pesos associados às ligações.
- Seja A^k a matriz que se obtém multiplicando k vezes a matriz A . O elemento da linha i e coluna j da matriz A^k indica o número de caminhos entre i e j com k ligações.

44

Small Worlds

- **Caminhos e Distâncias**
- **Exercício 1 d)**



Small Worlds

- **Caminhos e Distâncias**
- Os comprimentos dos caminhos entre os nodos podem ser utilizados para caracterizar a rede.
- Estes comprimentos podem medir a distância entre os nodos.
- Podem utilizar-se medidas de distância agregada, como a **média dos comprimentos dos caminhos mais curtos** e o **diâmetro da rede**.

Small Worlds

- **Caminhos e Distâncias**
- Para uma rede não orientada, **a média dos comprimentos dos caminhos mais curtos** (*average path length* ou *average path length*) é definida como:

$$\langle l \rangle = \frac{2\sum_{i,j} l_{ij}}{N(N-1)}$$

em que l_{ij} é o comprimento do caminho mais curto entre i e j, e N representa o número de nodos da rede.

47

Small Worlds

- **Caminhos e Distâncias**
- No somatório, que figura no numerador do quociente anterior, para cada par de nodos (i,j), considera-se l_{ij} ou l_{ji} (**apenas um** destes valores).

48

Small Worlds

- **Caminhos e Distâncias**
- Para uma rede orientada, **a média dos comprimentos dos caminhos mais curtos** é definida como:

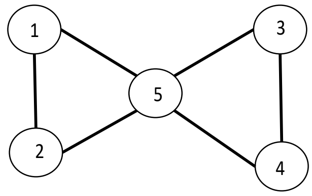
$$\langle l \rangle = \frac{\sum_{i,j} l_{ij}}{N(N-1)}$$

- Neste caso, no somatório, que figura no numerador, consideram-se **ambos** os valores l_{ij} e l_{ji}.

49

Small Worlds

- **Caminhos e Distâncias**
- **Exercício 1 e)**



50

Small Worlds

* **Caminhos e Distâncias**
* No caso de não existir um caminho entre dois nodos i e j, numa rede não orientada, pode considerar-se $l_{ij} = \infty. (1/l_{ij}) = 0$, e

$$\langle l \rangle = \left(\frac{2\left(\sum_{i,j} 1/l_{ij}\right)}{N(N-1)} \right)^{-1}.$$

* Para calcular esta média, começa-se por determinar a média dos inversos dos comprimentos. A média dos comprimentos será o inverso da média anterior.

51

Small Worlds

* **Caminhos e Distâncias**
* Para as redes orientadas, a adaptação é semelhante.

52

Small Worlds

- **Caminhos e Distâncias**
- Outras alternativas, para medir a distância, para redes em que não existem caminhos entre todos os pares de nodos serão sugeridas posteriormente.

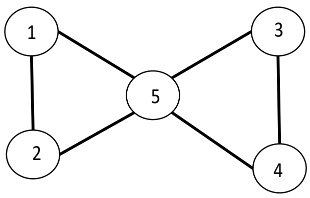
- O **diâmetro** (*diameter*) de uma rede é definido como:

$$l_{max} = \max_{i,j} l_{ij}.$$

53

Small Worlds

- **Caminhos e Distâncias**
- **Exercício 1 f)**



54

Small Worlds

- **Conectividade e Componentes**
- O número máximo de ligações de uma rede depende do número de nodos.
- Muitas redes são esparsas, têm muito menos ligações do que o número máximo possível.
- Se o número de ligações for muito reduzido, poderão não existir caminhos entre alguns pares de nodos, não se verificando a conectividade.

55

Small Worlds

- **Conectividade e Componentes**
- Uma rede não orientada diz-se **conexa** (*connected*) se existe, pelo menos, um caminho entre qualquer par de nodos.
- Caso contrário, a rede diz-se **desconexa** (*disconnected*).
- Uma rede não orientada desconexa é composta por **componentes conexas**.

56

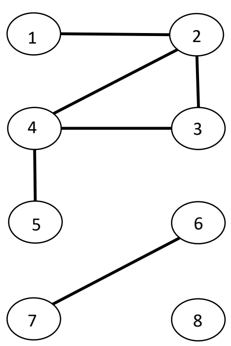
Small Worlds

- **Conectividade e Componentes**
- Uma **componente conexa** ou **componente** (*connected component* ou *component*) é uma subrede conexa.
- Se a maior das componentes conexas de uma rede envolve uma parte substancial da rede então é designada por **componente gigante** (*giant component*).

57

Small Worlds

- **Conectividade e Componentes**



A rede é desconexa.

É composta por três componentes conexas:
- a subrede gerada por {1, 2, 3, 4, 5};
- a subrede gerada por {6, 7};
- a subrede gerada por {8}.

58

Small Worlds

* **Conectividade e Componentes**
* Uma rede orientada é **fortemente conexa** (*strongly connected*) se existe, pelo menos, um caminho entre qualquer par de nodos (i, j), de i para j e de j para i.
* Uma rede orientada é **fracamente conexa** (*weakly connected*) se, quando se ignora a orientação das ligações, existe, pelo menos, um caminho entre qualquer par de nodos (i, j).

59

Small Worlds

* **Conectividade e Componentes**
* Uma subrede de uma rede orientada é uma **componente fortemente conexa** (*strongly connected component*) se existe, pelo menos, um caminho entre qualquer par de nodos da subrede, em ambos os sentidos.

60

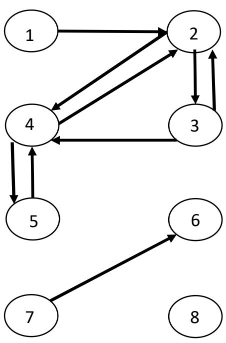
Small Worlds

- **Conectividade e Componentes**
- Uma subrede de uma rede orientada é uma **componente fracamente conexa** (*weakly connected component*) se, quando se ignora a orientação das ligações, existe, pelo menos, um caminho entre qualquer par de nodos da subrede.

61

Small Worlds

- **Conectividade e Componentes**

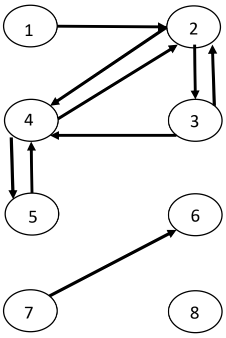


A rede não é fortemente conexa nem fracamente conexa, porque não existem caminhos para alguns pares de nodos (por exemplo, entre 1 e 6).

62

Small Worlds

* **Conectividade e Componentes**

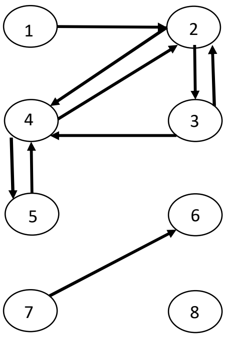


As subredes geradas por {1, 2, 3, 4, 5} e por {6, 7} são componentes fracamente conexas.

63

Small Worlds

* **Conectividade e Componentes**



A subrede gerada por {2, 3, 4, 5} é uma componente fortemente conexa (que designaremos por S), existe, pelo menos, um caminho entre qualquer par de nodos deste subconjunto, qualquer que seja a orientação considerada.

64

Small Worlds

- **Conectividade e Componentes**
- Sejam:
  - V_1 um subconjunto de nodos de uma rede orientada;
  - S uma componente fortemente conexa desta rede.
- Se:
  - existe um caminho entre cada elemento de V_1 e os elementos de S;
  - não existe nenhum caminho entre os elementos de S e cada um dos elementos de V_1

  Então
  - V_1 é o **conjunto incidente** (*in-component*) em S.

65

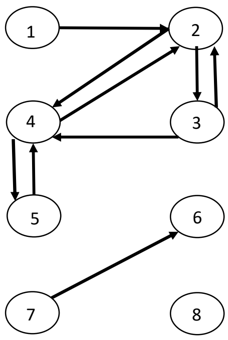
Small Worlds

- **Conectividade e Componentes**
- Se V_1 é o **conjunto incidente** (*in-component*) em S então os elementos de S podem ser alcançados a partir de V_1, mas não podem alcançar os elementos de V_1
- Analogamente, o conjunto de nodos que podem ser alcançados (usando caminhos) a partir de S, mas não podem alcançar os elementos de S é designado por **conjunto divergente** (*out-component*) de S.

66

Small Worlds

* **Conectividade e Componentes**



O conjunto {1} é o conjunto incidente em *S* (componente fortemente conexa {2, 3, 4, 5}), existe um caminho entre 1 e qualquer elemento de S e não existem caminho entre os nodos de S e 1.

67

Small Worlds

* **Alternativas para medir a distância numa rede em redes desconexas**
* Em redes desconexas, a média dos comprimentos dos caminhos mais curtos pode ser calculada considerando apenas os nodos da componente gigante.
* Outra alternativa consiste em calcular a média das distância tomando apenas os pares de nodos para os quais existem caminhos.

68

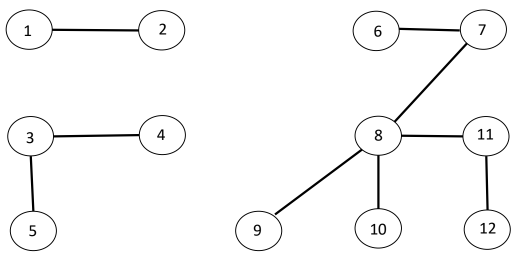
Small Worlds

* **Alternativas para medir a distância numa rede em redes desconexas**
* Para o diâmetro, pode determinar-se o diâmetro de cada componente conexa e atribuir ao diâmetro da rede o máximo dos diâmetros.

69

Small Worlds

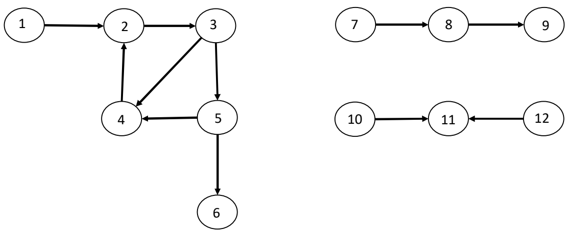
* **Conectividade e Componentes**
* **Exercício 2**



70

Small Worlds

- **Conectividade e Componentes**
- **Exercício 3**



71

Small Worlds

- ***Web*, Caminhos e Componentes**
- A maior das componentes do grafo da *Web* inclui mais de 90% das páginas, sendo por isso uma componente gigante.
- Além desta componente, existem muitas componentes de pequena dimensão.
- Também se observam muitas componentes fracamente conexas cuja dimensão apresenta uma distribuição enviesada (ou assimétrica).

72

Small Worlds

- ***Web*, Caminhos e Componentes**
- Na componente gigante é possível identificar uma componente gigante fortemente conexa, designada usualmente por *core*.
- Existem também um conjunto incidente neste *core* (*in-component*) e um conjunto divergente deste *core* (*out-component*).

73

Small Worlds

- ***Web*, Caminhos e Componentes**
- Frequentemente nodos que representam entidades com semelhanças ou características comuns estão unidos por ligações.
- Esta propriedade é designada por **homofilia**.
- Os nodos de redes de informação como as páginas da *Web*, os artigos da *Wikipedia* e artigos de publicações científicas contêm informação sobre um determinado tema ou assunto.

74

Small Worlds

* ***Web, Caminhos e Componentes***
* Com base no conteúdo é possível determinar o tema ou tópico abordado.
* Por outro lado, páginas ou artigos sobre tópicos relacionados podem estar ligadas entre si ou ter um caminho com poucas ligações entre elas.
* Quando isto acontece diz-se que a rede tem *topical locality*.

75

Small Worlds

* ***Web, Caminhos e Componentes***
* Esta propriedade surge quando os autores de novas páginas, artigos ou publicações os unem a outros com informação relevante sobre a temática ou tópico abordado.
* Para quantificar a *topical locality*, pode medir-se a semelhança, baseada no tema ou tópico abordado, entre uma determinada página (ou um determinado artigo), que designaremos por origem (*source*), e as páginas (ou artigos) a uma dada distância.

76

Small Worlds

* ***Web*, Caminhos e Componentes**
* É expectável que a fração de páginas ou artigos, a uma distância não superior a duas ligações da origem, que partilham o mesmo tema ou tópico da origem seja superior a páginas escolhidas ao acaso.
* Páginas ou documentos mais afastados da origem terão menor possibilidade de partilhar o tema ou tópico. Este fenómeno é designado por *topic drift*.

77

Small Worlds

* ***Web*, Caminhos e Componentes**
* Além da partilha de temas ou tópicos em páginas ou documentos próximos, também se verifica outra propriedade.
* A qualidade das páginas ou documentos próximas de uma origem ou semente (*seed*) de boa qualidade não será significativamente inferior à da origem ou semente.
* Esta é a razão pela qual os *crawlers* (programas que recolhem informação) utilizam métodos de pesquisa em largura (*breath-first search*).

78

Small Worlds

- **Árvores**
- Uma **árvore** (*tree*) é uma rede não orientada, conexa e sem ciclos.
- Uma árvore também pode ser definida como uma rede não orientada conexa, com N nodos e $N-1$ ligações.
- Se se acrescentar alguma ligação à árvore, a rede passará a ter um ciclo.
- Se se remover alguma ligação da árvore, a rede deixará de ser conexa.

79

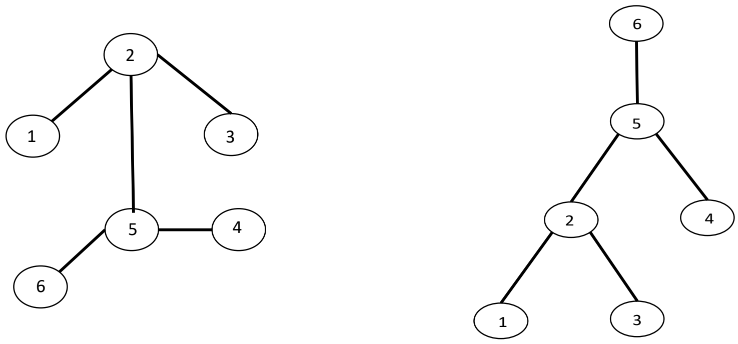
Small Worlds

- **Árvores**
- Apesar de uma árvore ser uma rede não orientada, pode induzir-se uma hierarquia nesta rede.
- Esta hierarquização será feita escolhendo um dos nodos para **raiz** (*root*) da árvore.

80

Small Worlds

* **Árvores**



81

Small Worlds

* ***Árvores***
* No exemplo anterior a mesma rede está representada graficamente de duas formas. Esta rede é uma árvore (é uma rede conexa e sem ciclos).
* Na representação à esquerda o nodo 2 é a raiz da árvore. Este nodo é **pai** (*parent*) dos nodos 1 e 3.
* Na representação à direita o nodo 6 é a raiz da árvore. Este nodo é **pai** (*parent*) do nodo 5.

82

Small Worlds

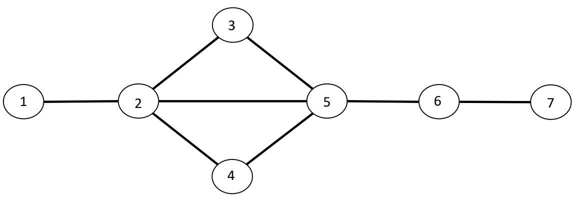
- **Árvores**
- Alguns algoritmos utilizam a estrutura hierárquica obtida numa árvore.

- Se uma rede orientada não tem ciclos e existe um caminho entre a raiz e cada um dos restantes nodos, então essa rede é uma **<u>arborescência</u>** (*arborescence* ou *directed tree*).

83

Small Worlds

- **Árvores**
- **Exercício 4**



84

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- Existem diversas algoritmos e *software* para determinar os camnhos mais curtos numa rede.
- A rede pode ter pesos associados às ligações ou não.

85

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- Considere-se uma rede sem pesos associados às ligações.
- O algoritmo que vamos estudar permite determinar os caminhos mais curtos entre um nodo, que designaremos por **origem** (*source*), e todos os restantes da rede.
- Este algoritmo considera uma representação hierárquica da rede, com diversos **níveis** ou **camadas** (*layers*).

86

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- No nível 0, encontra-se apenas a origem.
- No nível 1, colocam-se os nodos adjacentes da origem.
- Cada novo nível é preenchido com os nodos adjacentes dos elementos do nível anterior, caso ainda não tenham sido representados.
- Nesta representação, as ligações entre os nodos são as ligações da rede original.

87

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- A pesquisa começa na origem, no nível 0.
- Percorrem-se as ligações entre o nível 0 e o nível 1.
- A distância entre a origem e cada um dos nodos do nível 1 é igual a 1.

88

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- De seguida, escolhe-se um nodo do nível 1 e percorrem-se as ligações entre este nodo e os seus adjacentes do nível 2.
- A distância entre a origem e cada um destes últimos nodos é igual a 2.
- Caso existam, no nível 1, nodos ainda não considerados para pesquisa, escolhe-se um destes nodos.

89

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- Todos os nodos do nível 1 terão que ser escolhidos para efectuar pesquisa.
- Passa-se depois ao nível seguinte e procede-se de forma semelhante.
- Repete-se a pesquisa até se analisarem todos os níveis da componente ou da rede.

90

Small Worlds

* **Determinação dos Caminhos mais Curtos**
* Só se efectua a pesquisa a partir de um determinado nível após a inspecção completa do nível anterior.
* Este tipo de pesquisa é designado, em Inglês, por *breadth-first search*.
* Começaremos por representar todas as ligações a tracejado. À medida que as ligações são percorridas passam a ser representadas a cheio.

91

Small Worlds

* **Determinação dos Caminhos mais Curtos**
* **Descrição do algoritmo *breadth-first search***
  * Nesta descrição utiliza-se um conjunto S de nodos. Considerem-se os seguintes passos:
  * 1 – Constrói-se a árvore de acordo com a descrição anterior e com as ligações a tracejado.
  * 2 – Seja s o nodo origem. Introduz-se s em S, $l(s, s) = 0$.

92

Small Worlds

* **Determinação dos Caminhos mais Curtos**
* **Descrição do algoritmo *breadth-first search***
  * 3 – Se $S = \emptyset$ então o algoritmo termina.

    Caso contrário, retira-se de S um dos nodos com menor índice de nível na árvore. Seja i esse nodo e k o seu nível na árvore.

93

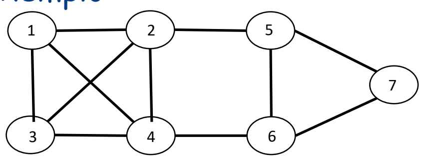
Small Worlds

* **Determinação dos Caminhos mais Curtos**
* **Descrição do algoritmo *breadth-first search***
  * 4 – Incluem-se em S os nodos adjacentes de i que estão no nível $k + 1$. Para cada um destes nodos, que se representam por j, $l(s,j) = l(s,i) + 1$ e desenham-se as ligações (i,j) a cheio.
  * 5 – Volta-se ao Passo 3.

94

Small Worlds

* **Determinação dos Caminhos mais Curtos**
* Exemplo



* Quais são os caminhos mais curtos entre 1 e cada um dos restantes nodos?

95

Small Worlds

* **Determinação dos Caminhos mais Curtos**
* Pode existir mais do que um caminho mais curto. O algoritmo determina apenas um um dos caminhos.
* Se a rede tiver mais do que uma componente conexa, não existem todos os caminhos e o algoritmo termina sem encontrar alguns caminhos.

96

Small Worlds

- **Determinação dos Caminhos mais Curtos**
- Para redes com pesos associados às ligações existem outros algoritmos de determinação dos caminhos mais curtos.

97

Small Worlds

- **Distância Social**
- Algumas redes apresentam caminhos longos, por exemplo, as redes de estradas e as redes de abastecimento.
- No entanto, verifica-se que as redes sociais apresentam, em geral, caminhos mais curtos com poucas ligações. Como exemplos, podem mencionar-se as redes de colaboração científica e as redes de actores.

98

Small Worlds

* **Distância Social**
* Nestas redes as pessoas são representadas nos nodos.
* Uma ligação entre dois nodos representa uma colaboração científica (coautoria de um artigo científico) ou a participação no mesmo filme ou série.

Small Worlds

* **Distância Social**
* No caso da colaboração científica, entre matemáticos, pode determinar-se a distância entre dois matemáticos ou entre um matemático e Paul Erdös em https://mathscinet.ams.org/mathscinet/collaborationDistance.html.
* (Paul Erdös foi um matemático Húngaro que publicou bastantes artigos científicos.)

Small Worlds

- **Distância Social**
- No caso das redes de actores, há um jogo - *Six Degrees of Kevin Bacon*, disponível em *oracleofbacon.org* - que indica a distância de Kevin Bacon a outro ator ou entre dois atores, apresentando um caminho.
- Para obter estas distâncias, consideram-se apenas filmes.

101

Small Worlds

- **Distância Social**
- Verifica-se que é difícil encontrar caminhos longos nestas redes e noutras redes reais.
- Por este motivo, estas redes são classificadas como **mundos pequenos** (*small worlds*).

102

Small Worlds

* **Seis Graus de Separação**
* O nome do jogo *Six Degrees of Kevin Bacon* é baseado no conceito de seis graus de separação (*six degrees of separation*).
* Este conceito tem origem no facto de num mundo pequeno quaisquer dois elementos estão unidos por uma sequência de ligações com poucos elementos.
* Numa rede social, em geral, as distâncias são curtas e a média é ainda menor.

103

Small Worlds

* **Seis Graus de Separação**
* O número seis foi proposto pelo autor Húngaro Frigyes Karinthy em "Cadeias", em 1929.
* Contudo, já no início do século XX, Marconi tinha uma ideia semelhante sobre distância social.
* Só na década de 60 do século XX se obteve a primeira evidência empírica de mundos pequenos.

104

Small Worlds

- **Seis Graus de Separação**
- O psicólogo Stanley Milgram propôs uma experiência para medir a distância social entre estranhos.
- Solicitou a 160 habitantes do Nebranka e do Kansas que enviassem uma carta a um conhecido, com instruções de que carta deveria chegar a uma determinada pessoa do Massachusetts.
- Cada pessoa que recebia a carta deveria enviá-la a um conhecido seu.

105

Small Worlds

- **Seis Graus de Separação**
- Apenas 42 cartas chegaram ao destino.
- Os comprimentos dos caminhos da rede associada às cartas que chegaram ao destino variavam entre 3 e 12 e a média foi pouco superior a 6, o que poderá ter originado a expressão “seis graus de separação”.

106

Small Worlds

- **Seis Graus de Separação**
- Esta experiência foi repetida com emails em 2003. Consideraram-se 18 destinatários em 13 países. Das mais de 24 000 cadeias iniciadas, apenas 384 ficaram completas.
- As cadeias completas apresentavam uma distância média de 4.
- Tendo em conta muitas cadeias interrompidas, os autores do estudo estimaram que a mediana da distância estaria entre 5 e 7.

Small Worlds

- **Seis Graus de Separação**
- Mais recentemente, em 2011, um estudo realizado na Universidade de Milão, sobre o *Facebook*, permitiu concluir que as distâncias entre utilizadores ativos (várias centenas de milhão) apresentavam uma média de 4,74.

Small Worlds

- **Seis Graus de Separação**
- Nos exemplos anteriores referiu-se a existência de distâncias pequenas.
- No entanto, o que pode ser considerado "pequeno"?
- Como se define "pequeno"?
- Esta definição deve ter em conta o número de nodos da rede.

109

Small Worlds

- **Seis Graus de Separação**
- Pode dizer-se que a distância média é pequena se cresce muito lentamente com o número de nodos da rede.
- Usualmente compara-se a distância média com o valor do logaritmo (de base 10) do número de nodos, que é uma função que cresce muito lentamente.
- Considera-se então que a distância média é pequena se for próxima do valor do logaritmo de N

$$\langle l \rangle \sim \log N$$

110

Small Worlds

- **Seis Graus de Separação**
- Os caminhos curtos não são exclusivos das redes sociais (colaboração científica, redes de actores, Facebook, ...).
- A rede de tráfego aéreo, a *Wikipedia* e a *Web* também apresentam caminhos com distâncias pequenas.

111

Small Worlds

- ***Coeficientes de Clustering***
- Considere-se uma rede social. Se a Alexandra e a Beatriz são amigas do Pedro então existe a possibilidade de a Alexandra e de a Beatriz serem amigas.
- Pode mesmo dizer-se que há fortes possibilidades de um amigo de um dos meus amigos ser também meu amigo.
- Este fenómeno observa-se no *Linkedin*, *Facebook* e noutras redes (amizade, pessoas que conhecemos, colegas,...).

112

Small Worlds

- ***Coeficientes de Clustering***
- Numa rede social, podem observar-se diversos **triângulos** (*triangles*) – conjuntos de três nodos em que existe uma ligação entre cada par de nodos.
- A conectividade entre os nodos adjacentes de outros nodos é uma característica relevante da estrutura local da rede, porque indica como se agrupam os nodos.

113

Small Worlds

- ***Coeficientes de Clustering***
- O coeficiente de *clustering* de um nodo é a fracção de pares de nodos adjacentes desse nodo que estão ligados entre si.
- Equivalentemente, é igual ao rácio entre o número de triângulos que incluem o nodo e o número máximo de triângulos envolvendo o nodo que poderiam existir.

114

Small Worlds

* ***Coeficientes de Clustering***
* O coeficiente de *clustering* do nodo i $-C(i)-$ é dado por

$$\frac{\tau(i)}{\tau_{max}} = \frac{2\tau(i)}{k_i(k_i - 1)}$$

* $\tau(i)$ representa o número de triângulos envolvendo o nodo i
* τ_{max} representa o número máximo de triângulos envolvendo o nodo i que poderiam existir na rede.

115

Small Worlds

* ***Coeficientes de Clustering***
* Este número máximo é dado pelo número de subconjuntos com dois elementos do conjunto de nodos adjacentes de i.
* O coeficiente anterior está definido apenas para nodos com grau superior a 1.

116

Small Worlds

- ***Coeficientes de Clustering***
- O coeficiente de *clustering* médio é a média dos coeficientes de clustering dos nodos e é dado por

$$\langle C \rangle = \frac{\sum_{i:k_i>1} C(i)}{N_{k>1}}$$

- No cálculo desta média podem considerar-se apenas os nodos com grau superior a 1 ou todos os nodos da rede.

Small Worlds

- ***Coeficientes de Clustering***
- Para definir o coeficiente de *clustering* da rede, considerem-se as subredes conexas de dimensão três.
- Estas subredes são designadas por ternos conexos (*triplet*).

Small Worlds

- ***Coeficientes de Clustering***
- Nestas subredes podem existir duas ou três ligações, sendo designadas por ternos conexos abertos (*open triplets*) ou ternos conexos fechados (*closed triplets*).
- No caso de existirem três ligações, é possível identificar um triângulo.
- A cada triângulo estão associados três ternos conexos, cada terno conexo está centrado num nodo.

119

Small Worlds

- ***Coeficientes de Clustering***
- O coeficiente de *clustering* da rede é dado pelo rácio

$$C = \frac{número\ de\ ternos\ conexos\ fechados}{número\ total\ de\ ternos\ conexos}$$

- que é equivalente a

$$C = \frac{3\ x\ número\ de\ triângulos}{número\ total\ de\ ternos\ conexos}$$

120

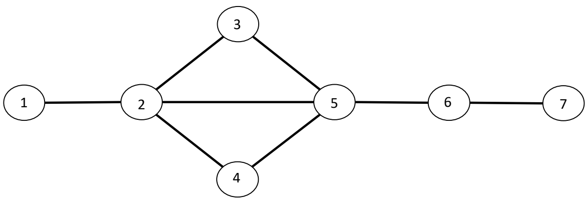
Small Worlds

- ***Coeficientes de Clustering***
- Ou ainda, equivalente a

$$C = \frac{traço(A^3)}{\sum_i k_i(k_i - 1)}$$

Small Worlds

- ***Coeficientes de Clustering***
- **Ex. 6 g)**



Small Worlds

* ***Coeficientes de Clustering***
* Esta definição de coeficiente de clustering apenas se aplica a redes não orientadas.
* No caso de redes orientadas, será necessário definir em que consiste um triângulo, se é baseado numa relação de transitividade ou não. A escolha depende da rede em análise.

123

Small Worlds

* ***Coeficientes de Clustering***
* Um valor elevado, próximo de 1, indica a existência de muitos triângulos.
* As redes sociais apresentam coeficientes de *clustering* significativos.
* Isto deve-se ao facto de as pessoas conhecerem pessoas a partir dos seus contactos, obtendo-se assim triângulos.
* Este mecanismo é designado por **fecho triádico** (*triadic closure*).

124

ANÁLISE DE REDES

Hubs e Heterogeneidade de Pesos

Licenciatura em
Ciência de Dados

Hubs

- Em diversas redes, alguns nodos têm bastante mais ligações do que os restantes:
  - aeroportos com muito mais ligações do que a maioria;
  - pessoas com bastante mais visibilidade do que a maioria.
- Nestes casos, estamos perante **heterogeneidade** da rede.

Hubs

- A importância de um nodo ou ligação de uma rede pode ser estimada determinando a sua centralidade.
- O grau de um nodo é uma das medidas de centralidade.
- Os ***hubs*** são nodos com grau elevado, bastante superior ao da maior parte dos restantes nodos da rede.
- A existência de *hubs* induz propriedades particulares nas redes.

3

3

Hubs

- Medidas de Centralidade
  - Grau
  - Proximidade e Centralidade de Proximidade
  - Intermediação e Centralidade de Intermediação.

4

4

Hubs

* **Grau**
* O grau de um nodo de uma rede não orientada é o número de ligações incidentes nesse nodo.
* O grau médio indica como, em média, os nodos de uma rede estão ligados entre si.
* Por se tratar de um valor médio, pode não ser representativo da distribuição de grau.

Hubs

* **Proximidade e Centralidade de Proximidade**
* A **proximidade** (*closeness*) é uma medida de centralidade que mede o quão "próximo" está um nodo dos restantes.
* A proximidade de um nodo é baseada na soma das distâncias entre o nodo e cada um dos restantes.
* Se o total for reduzido então o nodo tem uma centralidade elevada, está próximo dos restantes.

Hubs

- **Proximidade e Centralidade de Proximidade**
- A **centralidade de proximidade** (*closeness centrality*) é definida como o inverso da soma das distâncias entre o nodo e todos os restantes:

$$g_i = \frac{1}{\sum_{j \neq i} l_{ij}}$$

em que l_{ij} representa a distância de i para j, o comprimento do caminho mais curto entre i e j.

7

7

Hubs

- **Proximidade e Centralidade de Proximidade**
- Esta medida pode depender da dimensão da rede (número de nodos da rede).
- Torna-se difícil comparar redes de diferentes dimensões.
- Como alternativa, pode multiplicar-se esta medida por $(N-1)$:

$$\tilde{g}_i = \frac{N-1}{\sum_{j \neq i} l_{ij}} \quad \text{ou} \quad \tilde{g}_i = \frac{1}{\sum_{j \neq i} l_{ij}/(N-1)}$$

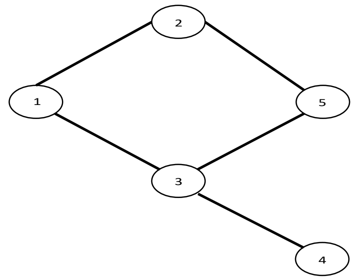
- O denominador representa a distância média entre o nodo i e os restantes nodos da rede.

8

8

Hubs

* **Proximidade e Centralidade de Proximidade**
* **Ex. 1**



a) e **b)**.

Hubs

* **Intermediação e Centralidade de Intermediação**
* Muitos fenómenos são baseados em processos de difusão.
* Esta difusão pode ocorrer não diretamente, mas através de caminhos.
* Nodos com grau pouco elevado podem ser relevantes em processos de difusão, se passam muitos caminhos mais curtos por estes nodos.
* Para avaliar a intermediação dos nodos, considera-se uma medida baseada no número de caminhos mais curtos.

Hubs

- **Intermediação e Centralidade de Intermediação**
- Sejam σ_{hj} o número de caminhos mais curtos entre h e j, $\sigma_{hj}(i)$ o número destes caminhos que passam pelo nodo i. A **centralidade de intermediação** (*betweenness centrality*) de i é dada por:

$$b_i = \sum_{(h,j):h,j\neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}$$

11

11

Hubs

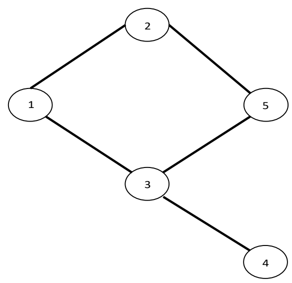
- **Intermediação e Centralidade de Intermediação**
- Esta medida é adequada quando o número de caminhos mais curtos dá uma boa aproximação da frequência de utilização do nodo, o que é comum, por exemplo, em redes de transportes.

12

12

Hubs

- Intermediação e Centralidade de Intermediação
- Ex. 1 c)

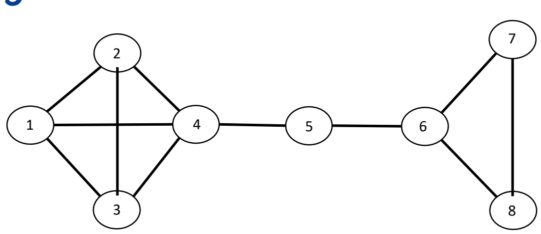


13

13

Hubs

- Medidas de Centralidade
- Ex. 5



a), b), c) e d).

14

14

Hubs

- **Intermediação e Centralidade de Intermediação**
- A centralidade de intermediação normalizada de um nodo *i* obtém-se dividindo b_i por
$$\frac{(N-1)(N-2)}{2}$$
- O número máximo de caminhos, entre nodos distintos de i, que passam pelo nodo i, é igual a
$$\frac{(N-1)(N-2)}{2}$$

15

15

Hubs

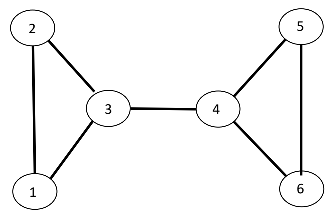
- **Intermediação e Centralidade de Intermediação**
- A **centralidade de intermediação de uma ligação** é a fracção de caminhos mais curtos que incluem esta ligação, de entre todos os caminhos mais curtos existentes na rede.

16

16

Hubs

* **Intermediação e Centralidade de Intermediação**
* **Exemplo**



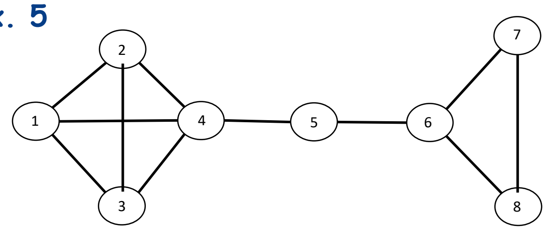
* *A centralidade de intermediação de (1,3) é 4/15.*
* *A centralidade de intermediação de (3,4) é 9/15 = 0,6.*

17

17

Hubs

* **Intermediação e Centralidade de Intermediação**
  * **Ex. 5**



e) Determine a medida de centralidade de intermediação para a ligação (4,5).

18

18

Hubs

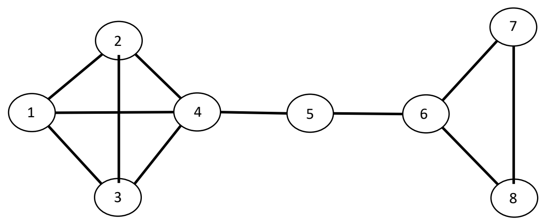
* **Intermediação e Centralidade de Intermediação**
* Ligações com uma centralidade de intermediação elevada frequentemente unem partes da rede que constituem comunidades.

Hubs

* **Distribuição de Centralidade**
* Para redes de dimensão elevada, será necessário uma abordagem estatística para estudar a distribuição de grau.
* Podem considerar-se tabelas de frequências e representações gráficas, por exemplo histogramas.

Hubs

* Distribuição de Centralidade
  Exemplo



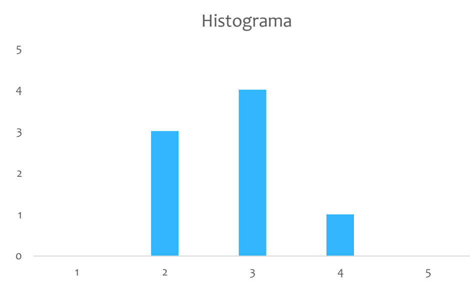
Hubs

* Distribuição de Centralidade
  Exemplo (continuação)

Grau	Frequências Absolutas	Frequências Relativas
1	0	0
2	3	0,375 (3/8)
3	4	0,5
4	1	0,125 (1/8)
		1

Hubs

* **Distribuição de Centralidade**
  Exemplo (continuação)



Hubs

* **Distribuição de Centralidade**
* Quando estão presentes ordens de grandeza muito diversas, podem considerar-se **escalas logarítmicas**.

Hubs

- **Distribuição de Centralidade**
- **Heterogeneidade**
- A amplitude de variação na distribuição poderá dar uma ideia sobre a heterogeneidade.
- Uma representação gráfica da distribuição de grau com uma cauda longa ou pesada (*heavy-tailed*) revela heterogeneidade nos valores dos graus.
- Neste caso, muitos nodos têm poucos adjacentes enquanto que poucos nodos têm muitos adjacentes.

Hubs

- **Distribuição de Centralidade**
- **Heterogeneidade**
- Nodos com grau bastante superior ao da maior parte dos restantes nodos da rede são designados por ***hubs***.
- Para avaliar a heterogeneidade de uma rede, pode considerar-se o parâmetro

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}.$$

Hubs

- **Distribuição de Centralidade**
- **Heterogeneidade**
- Se o valor de κ está próximo de 1 então não há heterogeneidade, a distribuição de grau estará concentrada à volta de um determinado valor.
- Quando o valor de κ é bastante superior a 1, estamos perante *hubs*.

27

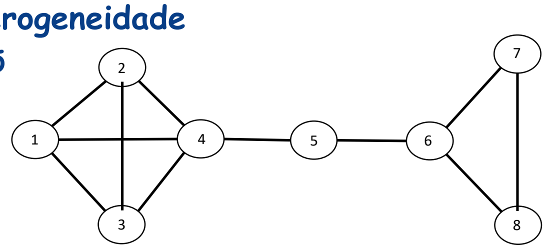
27

Hubs

- **Distribuição de Centralidade**

  Heterogeneidade

  Ex. 5



f) Determine o parâmetro de heterogeneidade.

28

28

Hubs

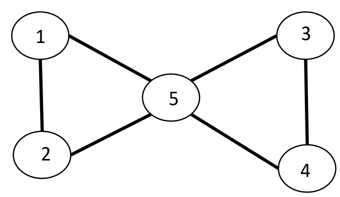
* **Distribuição de Centralidade**
* **Heterogeneidade**
* Para redes orientadas, o estudo da heterogeneidade será feito considerando o grau incidente e o grau divergente.

Hubs

* **Paradoxo da Amizade**
* Numa rede de dimensão N, quando se escolhe aleatoriamente um nodo, a probabilidade de o nodo escolhido ser o de maior grau é igual a $1/N$.
* No entanto, se se escolher aleatoriamente um nodo adjacente de um nodo escolhido ao acaso, a probabilidade de obter o nodo com maior grau aumenta.
* Designemos a escolha aleatória de nodos por primeiro método e o segundo procedimento de escolha de um nodo por segundo método.

Hubs

* **Paradoxo da Amizade**
* **Exemplo**



* 5 é o nodo com maior grau (1/5 vs. 2/5).

31

31

Hubs

* **Paradoxo da Amizade**
* Quando se considera o segundo método, em vez de se escolher um nodo, procuram-se ligações.
* Assim, quanto maior o grau de um nodo, maior será a probabilidade de ser o escolhido.

32

32

Hubs

* **Paradoxo da Amizade**
* Se se generalizar o segundo método, passando diversas vezes de um nodo para os seus adjacentes, a probabilidade de se obter um *hub* aumenta.
* Em diversos contextos, a determinação de *hubs* pode ser necessária (para tentar interromper cadeias de transmissão, difundir mais rapidamente informação).

33

33

Hubs

* **Paradoxo da Amizade**
* A escolha baseada nas ligações, em vez dos nodos, tem outra implicação.
* O grau médio é inferior à média dos graus dos nodos adjacentes (vizinhos).
* Pode ser enunciado como "Os nossos amigos têm, em média, mais amigos do que nós." (**Paradoxo da Amizade)**

34

34

Hubs

* **Paradoxo da Amizade**
* Quando se calcula o grau médio dos nodos, o grau de cada nodo é considerado apenas uma vez.
* Quando se calcula a média dos graus médios dos nodos adjacentes, cada nodo é considerado tantas vezes quantas o seu grau.

Hubs

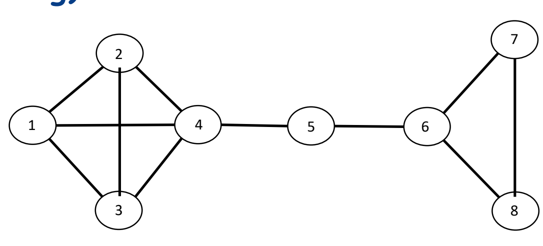
* **Paradoxo da Amizade**
* Pode considerar-se que o cálculo das duas médias é baseado em amostragem.
* No primeiro caso, utiliza-se amostragem uniforme.
* No segundo caso, a cada nodo corresponde uma proporção baseada no seu grau.

Hubs

- **Paradoxo da Amizade**
- Quando os graus dos nodos diferem pouco, as duas médias são aproximadas.
- Pelo contrário, quando existem nodos com grau bastante superior ao da maioria (existem *hubs*) a diferença entre as duas médias acentua-se.

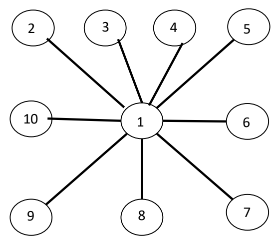
Hubs

- **Paradoxo da Amizade**
- **Ex. 5 g)**



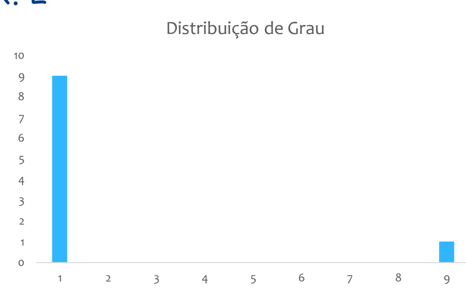
Hubs

- Resolver **Ex. 3 a) – c)**



Hubs

- Ex. 2



Hubs

- ***Ultra-Small Worlds***
- Muitas redes reais são *small worlds* (qualquer par de nodos está ligado por um número reduzido de ligações).
- Em redes em que existem *hubs*, muitos caminhos mais curtos passam provavelmente por um ou mais *hubs* (porque estes nodos têm muitas ligações).
- A inclusão de *hubs* nos caminhos permite reduzir as distâncias.

41

41

Hubs

- ***Ultra-Small Worlds***
- Frequentemente, em redes com maior diferença entre o grau máximo e o grau mínimo (redes com *hubs*), as distâncias entre nodos são muito reduzidas.
- Esta propriedade é designada por *ultra-small worlds*.

42

42

Hubs

- ***Ultra-Small Worlds***
- Quando se comparam redes com o mesmo número de nodos e de ligações, com e sem *hubs*, espera-se que a distância média seja menor em redes com *hubs* do que em redes sem *hubs*.
- *O Twitter* é um exemplo de *ultra-small world*.

Hubs

- ***Web e Hubs***
- Como já foi referido, a *Web* pode ser representada numa rede orientada.
- A distribuição do grau incidente apresenta uma cauda longa ou pesada.
- Além disso, o parâmetro de heterogeneidade toma um valor elevado.
- Estas duas características indicam a presença de *hubs* de dimensão considerável.

Hubs

* ***Web* e *Hubs***
* A *Web* apresenta a estrutura de mundos ultra pequenos (*ultra-small worlds*), devido à presença de *hubs*.
* Por exemplo, considerando dados de 2012, constatou-se que na maior das componentes fortemente conexa, que incluía 1800 milhões de páginas, a distância média era inferior a 13.

Hubs

* ***Web* e *Hubs***
* A *Wikipedia* e os *blogs* também apresentam uma distribuição com cauda longa do grau incidente e da dimensão das componentes.
* A estrutura de mundos ultra pequenos (*ultra-small worlds*) também está presente nestas redes de informação.

Hubs

- **Redes *Scale-Free***
- Uma rede *scale-free* é uma rede cuja distribuição de grau segue uma lei de potência (*power law*).
- Trata-se de uma rede cuja distribuição de grau é bem representada ou bem aproximada por

$$k^{-\gamma}$$

($p_k \sim k^{-\gamma}$) em que k representa os valores do grau.

Hubs

- **Redes *Scale-Free***
- É expectável que uma rede com esta propriedade tenha *hubs*.
- Por outro lado, pode provar-se que quanto maior for a dimensão da rede maior será o grau do nodo com grau máximo.

Hubs

* **Robustez**
* Um sistema é robusto se a falha de alguma das suas componentes não afecta o seu funcionamento.
* Genericamente, a robustez depende de quais as componentes que falham e dos danos resultantes.

Hubs

* **Robustez**
* Como se pode definir a robustez de uma rede?
* A robustez de uma rede pode ser avaliada tendo em conta as consequências da remoção de alguns nodos e das ligações incidentes nestes nodos.

Hubs

- **Robustez**
- A remoção destes nodos pode não afectar a conectividade e o sistema resultante associado mantém-se em funcionamento.
- Ou pode deixar a rede desconexa e, consequentemente, o funcionamento do sistema será severamente comprometido.

Hubs

- **Robustez**
- O teste *standard* à robustez consiste em determinar como a conectividade é afectada à medida que o número de nodos removidos aumenta.
- Para estimar o impacto da remoção de nodos, determina-se o rácio entre o número de nodos da componente gigante e o número original de nodos da rede.

Hubs

* **Robustez**
* Se a rede se mantém conexa, o valor deste rácio diminui lentamente devido à redução do número de nodos na rede.
* Se, pelo contrário, a rede deixa de ser conexa, o valor do rácio reduz significativamente.

53

53

Hubs

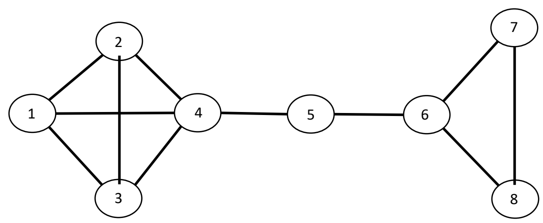
* **Robustez**
* A remoção pode dever-se a falhas (*failures*) ou ataques (*attacks*).
* No caso da remoção ser aleatória, devido a uma falha, espera-se que o número de *hubs* se mantenha significativo. Assim, a dimensão da componente gigante diminuirá lentamente.
* No caso de ataques, os *hubs* serão os alvos e a rede tornar-se-á rapidamente desconexa.

54

54

Hubs

- **Robustez**
- **Ex. 5 h), i)**



Hubs

- **Robustez**
- As redes reais são, em geral, robustas a falhas e vulneráveis a ataques, porque possuem *hubs*.

Hubs

- **Decomposição da Rede (*Core Decomposition*)**
- Em redes de grande dimensão pode ser útil considerar apenas a parte mais densa (*core*).
- O grau dos nodos pode ser utilizado para dividir a rede em partes, designadas por **<u>conchas</u>** (*shells*), e baseadas na estrutura e localização periféricas da rede.

57

57

Hubs

- **Decomposição da Rede (*Core Decomposition*)**
- Para obter a parte mais densa da rede, baseada no grau dos nodos, começa-se por remover os nodos com menor grau.
- O algoritmo de decomposição *k-core decomposition* é um algoritmo iterativo que começa por iniciar o valor de *k* a 0 e consiste nos seguintes passos:

58

58

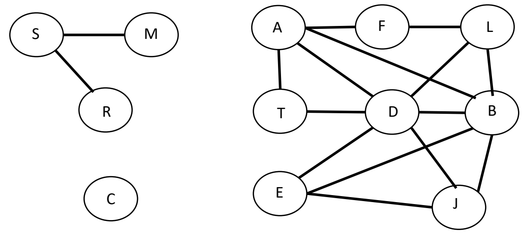
Hubs

* **Decomposição da Rede (*Core Decomposition*)**
  * 1 – Remover recursivamente todos os nodos com grau igual a *k*, de forma que todos os restantes nodos (se existirem) tenham grau superior a *k*;
  * 2 – Os nodos removidos constituem a *k*-concha (*k-shell*) e os restantes nodos constituem o *(k+1)-core*, porque têm grau maior ou igual a *k*+1.
  * 3 – Se não existem mais nodos, o algoritmo termina; caso contrário, incrementa-se *k* e volta-se ao passo 1.

Esta decomposição é bastante útil para remover os nodos periféricos.

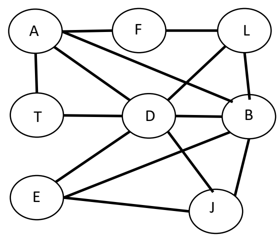
Hubs

* **Decomposição da Rede (*Core Decomposition*)**
* **Exemplo**
* O 2-*core* da seguinte rede



Hubs

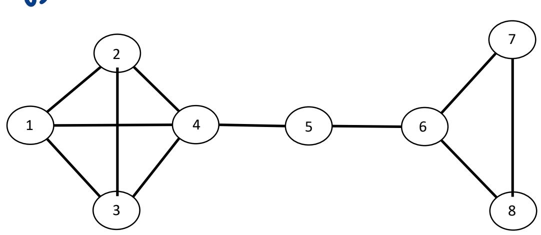
- **Decomposição da Rede (*Core Decomposition*)**
- **Exemplo (Continuação)**
- é



61

Hubs

- **Decomposição da Rede (*Core Decomposition*)**
- **Ex. 5 j)**



62

Heterogeneidade de Pesos

- A visualização de redes densas não é fácil.
- O seu estudo também pode ser difícil.
- Muitas das ligações de redes densas com pesos associados às ligações não são relevantes e poderão ser removidas.

Heterogeneidade de Pesos

- Pode pensar-se em generalizar a decomposição de *core* estudada (que se baseia no grau dos nodos).
- Assim, as ligações com pesos inferiores a um valor escolhido seriam removidas.
- Este procedimento pode ser adequado em alguma situações.

Heterogeneidade de Pesos

- Noutras, em que se verifica uma grande heterogeneidade, pode não ser indicado.
- Por exemplo, se a força de alguns nodos for bastante inferior ao da maior parte dos nodos, deve considerar-se outra escala para as ligações incidentes nestes nodos.

Heterogeneidade de Pesos

- Assim, pode considerar-se, em vez de um valor absoluto, um valor relativo:
  - manter apenas uma percentagem das ligações incidentes num nodo (as com maior peso)
  - ou manter as ligações de maior peso que totalizem uma percentagem da força do nodo.

Heterogeneidade de Pesos

- Estas escolhas não garantem que se mantenham as ligações mais relevantes.
- Outra alternativa consiste em determinar as ligações que têm associadas uma fracção desproporcional da força de cada nodo. Estas serão as ligações a manter.

Heterogeneidade de Pesos

- **"Espinha Dorsal" da Rede (*Network Backbone*)**
- Sejam i um nodo da rede, k_i e s_i o seu grau e a sua força, respectivamente.
- Seja ainda α o nível de significância escolhido.
- Cada ligação será avaliada considerando que os pesos das ligações incidentes em i se distribuem aleatoriamente pelas k_i ligações e que a força de i é s_i.

Heterogeneidade de Pesos

- **“Espinha Dorsal” da Rede (*Network Backbone*)**
- Assumindo esta hipótese, a probabilidade da ligação (i, j) ter peso igual ou superior a w_{ij} é dada por

$$p_{ij} = \left(1 - \frac{w_{ij}}{s_i}\right)^{k_i - 1}$$

- Se $p_{ij} < \alpha$ então mantém-se a ligação. Caso contrário, a ligação é removida.

69

ANÁLISE DE REDES
Redes Aleatórias

Licenciatura em
Ciência de Dados

Redes Aleatórias

- Muitas redes reais partilham um conjunto de propriedades:
  - os caminhos mais curtos entre dois nodos têm poucas ligações;
  - têm muitos triângulos, tendo por isso coeficientes de *clustering* elevados;
  - apresentam heterogeneidade quando se consideram os graus dos nodos e os pesos associados às ligações.

Redes Aleatórias

- Paul Erdös e Alfréd Rényi iniciaram o estudo da Teoria de Grafos Aleatórios, para estudar como surgem estas propriedades.
- Este estudo é baseado nas **redes aleatórias**, também conhecidas como **redes de Erdös-Rényi**.

Redes Aleatórias

- Uma rede de Erdös-Rényi é obtida escolhendo o número de nodos (N) e o número de ligações (L).
- Para escolher as ligações, geram-se aleatoriamente L pares de nodos.

Redes Aleatórias

- Outras redes aleatórias, propostas por E. N. Gilbert, podem ser obtidas escolhendo o número de nodos e uma probabilidade de inclusão de cada ligação (p).

Redes Aleatórias

- Neste caso, as redes serão obtidas aplicando o seguinte procedimento:
  - 1 – seleccione-se um par de nodos (i, j);
  - 2 – gere-se um número aleatório r, entre 0 e 1. Se $r < p$ então inclui-se a ligação (i, j);
  - 3 – repetem-se os passos 1 e 2 para todos os pares de nodos.

Redes Aleatórias

- A aplicação deste procedimento, diversas vezes, para valores fixos de N e p pode gerar redes com diferentes números de ligações.
- Contudo, para valores de N suficientemente grandes, espera-se que o número de ligações seja aproximado.

Redes Aleatórias

- Suponha agora que se está a gerar uma rede aleatória.
- Inicialmente a rede é constituída apenas por nodos.
- À medida que se adicionam ligações, pares de nodos são ligados.
- Após a introdução das primeiras ligações, a rede é composta por subredes de pequena dimensão.

Redes Aleatórias

- Quando se forma a componente gigante?
- Erdös e Rényi descobriram que a componente gigante forma-se quando $\langle k \rangle = 1$.
- A configuração da rede muda repentinamente quando $\langle k \rangle$ passa de um valor inferior a 1 para igual a 1.
- Por outro lado, a dimensão desta componente gigante cresce rapidamente com o aumento do grau médio (para valores superiores a 1).

- Ver NetLogo Giant Component

9

9

Redes Aleatórias

- **Densidade das Redes Aleatórias**
- Considere-se o método de geração de uma rede aleatória baseado na escolha do número de nodos e da probabilidade de inclusão de cada ligação (Modelo de Gilbert).
- Este processo de geração de uma rede aleatória é semelhante à experiência de lançar repetidas vezes uma moeda.

10

10

Redes Aleatórias

- **Densidade das Redes Aleatórias**
- Suponha-se que p representa a probabilidade de obter uma cara.
- O número esperado de caras obtidas será dado por pn, em que n representa o número de lançamentos.

11

11

Redes Aleatórias

- **Densidade das Redes Aleatórias**
- No caso da rede aleatória, o número de lançamentos corresponde a

$$\binom{N}{2} = \frac{N(N-1)}{2}$$

que representa o número máximo de ligações de uma rede não orientada com N nodos.

12

12

Redes Aleatórias

* **Densidade das Redes Aleatórias**
* Então:

$$\langle L \rangle = \frac{pN(N-1)}{2}.$$

* Atendendo à expressão para o grau médio, tem-se:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1).$$

13

13

Redes Aleatórias

* **Densidade das Redes Aleatórias**
* Além disso, a densidade será dada por:

$$d = \frac{\langle L \rangle}{L_{max}} = \frac{pN(N-1)/2}{N(N-1)/2} = p.$$

* Como as redes reais são esparsas, uma rede aleatória adequada para o estudo de redes reais terá associada uma probabilidade pequena.

14

14

Redes Aleatórias

- **Distribuição de Grau das Redes Aleatórias**
- A distribuição de grau de uma rede aleatória será determinada pelo cálculo das probabilidades de um qualquer nodo da rede ter k adjacentes.
- Seja i um nodo qualquer de uma rede com N nodos. Qualquer um dos restantes $N-1$ nodos pode ser adjacente de i.

15

15

Redes Aleatórias

- **Distribuição de Grau das Redes Aleatórias**
- Para que cada um dos restantes nodos seja adjacente de i, é necessário que exista uma ligação entre i e o nodo.
- No processo de geração da rede aleatória, a decisão de incluir uma ligação é independente da decisão de incluir ou não incluir cada uma das restantes.

16

16

Redes Aleatórias

* **Distribuição de Grau das Redes Aleatórias**
* Se a probabilidade de dois nodos serem adjacentes for p então a probabilidade de um nodo ter grau igual a k é dada por:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

17

17

Redes Aleatórias

* **Distribuição de Grau das Redes Aleatórias**
* Trata-se da distribuição Binomial, de parâmetros $(N-1)$ e p, que tem valor esperado dado por $(N-1)p$.
* Assim, para valores elevados de N e com Np constante (e não muito pequeno) pode considerar-se que $Np \approx \langle k \rangle$.

18

18

Redes Aleatórias

* **Distribuição de Grau das Redes Aleatórias**
* Na distribuição Binomial os valores que acumulam maior probabilidade estão concentrados à volta do valor esperado.
* Além disso, o gráfico que representa as probabilidades não apresenta uma cauda longa.
* Podemos então concluir que nas redes aleatórias **<u>não se verifica a presença de *hubs*</u>**.

Redes Aleatórias

* **Caminhos mais curtos**
* Já se viu que os graus dos nodos de uma rede aleatória não diferem muito.
* Para estudar a existência de caminhos mais curtos com poucas ligações, considere-se uma rede conexa em que todos os nodos têm grau k.

Redes Aleatórias

- **Caminhos mais curtos**
- Considere-se ainda um nodo qualquer da rede.
- Este nodo está:
  - a uma distância igual a 1 de k nodos;
  - a uma distância igual a 2 de $k(k-1)$ nodos;
  - a uma distância igual a 3 de $k(k-1)^2$ nodos.

21

21

Redes Aleatórias

- **Caminhos mais curtos**
- Assim, o nodo está a uma distância igual a l de $k(k-1)^{l-1}$ nodos.
- O valor real de nodos pode ser inferior a $k(k-1)^{l-1}$ porque alguns dos nodos a uma distância não superior a l podem estar repetidos.

22

22

Redes Aleatórias

* **Caminhos mais curtos**
* Se o valor de k não for muito pequeno então pode considerar-se $k \approx k-1$ e a estimativa para o número de nodos a uma distância igual a l será aproximada por k^l.
* Seja l_{max} a maior das distâncias – o diâmetro da rede –, então esta distância permite alcançar todos os nodos da rede e

$$k^{l_{max}} = N.$$

Redes Aleatórias

* **Caminhos mais curtos**
* Tem-se então $l_{max} = log_k N = \frac{\log N}{\log k}$.
* Dado que os graus dos nodos de uma rede aleatória não diferem muito, a expressão anterior constitui uma boa aproximação quando se tomam valores de k em torno de $\langle k \rangle$.

Redes Aleatórias

* **Caminhos mais curtos**
* Como a função logarítmica cresce lentamente, o valor de l_{max} cresce lentamente quando N aumenta.
* Conclui-se então que nas redes aleatórias existem caminhos mais curtos com poucas ligações.

Redes Aleatórias

* **Caminhos mais curtos**
* Considere-se uma rede aleatória para representar a rede mundial de contactos sociais.
* Com base no número de Dunbar, considere-se ainda que o grau médio é 150.
* Tem-se que $150^5 = 75\,937{,}5$ milhões, que é nove vezes superior à população mundial.
* Este resultado é compatível com a experiência de Stanley Milgram.

Redes Aleatórias

- ***Coeficiente de Clustering***
- O coeficiente de *clustering* de um nodo mede a fracção de nodos adjacentes unidos por uma ligação.
- Uma ligação entre dois nodos adjacentes forma um triângulo.

Redes Aleatórias

- ***Coeficiente de Clustering***
- Numa rede aleatória, gerada considerando o número de nodos e uma probabilidade p de inclusão de cada ligação, a probabilidade de dois nodos adjacentes estarem unidos é igual a p.
- O coeficiente de *clustering* de cada nodo pode não ser exatamente igual a p.
- Contudo, espera-se que a média dos coeficientes de *clustering* dos nodos seja bem aproximada por p.

Redes Aleatórias

* ***Coeficiente de Clustering***
* Se a rede aleatória for esparsa então o valor de p será pequeno e a rede terá poucos triângulos.
* Se se aumentar o valor da probabilidade então o número de triângulos aumenta, assim como a densidade da rede (o que não será realista para simular redes sociais reais).

Redes Aleatórias

* **Mundos Pequenos (*Small Worlds*)**
* Com vista a obter redes aleatórias com coeficientes de *clustering* mais elevados, Duncan J. Watts e Steven H. Strogatz desenvolveram o Modelo de Mundos Pequenos (*small-world model*), também conhecido como modelo Watts-Strogatz.

Redes Aleatórias

- **Mundos Pequenos (*Small Worlds*)**
- Começaram por gerar uma rede em que cada nodo está ligado aos 4 nodos mais próximos. Esta rede apresenta um coeficiente de *clustering* de 0,5.
- Contudo, a distância média não é pequena, devido às distâncias entre os nodos mais afastados entre si.

Redes Aleatórias

- **Mundos Pequenos (*Small Worlds*)**
- Para reduzir a distância média, algumas ligações da rede serão substituídas. Esta substituição consiste em manter um dos nodos (um dos extremos da ligação) e modificar o outro nodo.
- A probabilidade de substituir cada ligação será representada por p.

Redes Aleatórias

* **Mundos Pequenos (*Small Worlds*)**
* A substituição das ligações irá reduzir o número de triângulos, mas reduzirá algumas distâncias e a distância média.
* Quais os valores de p que permitem reduzir a distância média sem diminuir drasticamente o coeficiente de *clustering*?

Redes Aleatórias

* **Mundos Pequenos (*Small Worlds*)**
* A probabilidade deve ser pequena para manter muitos triângulos.
* Por exemplo, alguns testes sugerem que valores entre 0,01 e 0,1 permitem reduzir a distância média e manter um número significativo de triângulos.
* Ver NetLogo Small Worlds

Redes Aleatórias

- ***Mundos Pequenos (Small Worlds)***
- Outra possibilidade para a substituição de cada ligação consiste em escolher aleatoriamente os dois nodos unidos pela ligação, não mantendo nenhum dos nodos iniciais.
- Em vez de substituir ligações, podem ser adicionadas ligações escolhidas ao acaso.

Redes Aleatórias

- ***Mundos Pequenos (Small Worlds)***
- Também podem ser consideradas outras configurações iniciais. Por exemplo, uma grelha em que os nodos internos têm grau igual a 6 e os nodos na fronteira têm grau igual a 2, 3 ou 4.
- O modelo Watts-Strogatz não gera redes com *hubs*.

Redes Aleatórias

* **Modelo de Configuração (*Configuration Model*)**
* Dada uma sequência, será possível gerar uma rede em que os graus dos nodos são os elementos da sequência?
* Para que exista uma rede nestas condições, a soma dos elementos da sequência terá que ser um número par.

Redes Aleatórias

* **Modelo de Configuração (*Configuration Model*)**
* Uma solução simples é dada pelo Modelo de Configuração (*Configuration Model*).
* Suponha que se tem um conjunto de nodos e uma sequência que representa os graus dos nodos.

Redes Aleatórias

* **Modelo de Configuração (*Configuration Model*)**
* Em cada nodo, desenha-se um número de linhas incidentes (*stubs*) igual ao grau do nodo.
* Estas linhas não representam ligações porque incidem apenas num nodo.

Redes Aleatórias

* **Modelo de Configuração (*Configuration Model*)**
* O modelo consiste em:
  * 1 – escolher aleatoriamente um par de linhas;
  * 2 – ligar as duas linhas escolhidas, obtendo-se uma ligação.
* Este procedimento é repetido até todas as linhas ficarem unidas.

Redes Aleatórias

- **Modelo de Configuração (*Configuration Model*)**
- O modelo pode originar soluções que apresentem algumas desvantagens: mais do que uma ligação entre o mesmo par de nodos e existência de lacetes (*loops*).

41

41

Redes Aleatórias

- **Modelo de Configuração (*Configuration Model*)**
- Este modelo pode ser utilizado para verificar se uma propriedade de uma rede é consequência da distribuição de grau.
- Se a propriedade da rede em estudo também estiver presente em todas as redes geradas pelo modelo então a propriedade resulta da distribuição de grau.

42

42

Redes Aleatórias

- **Modelo de Configuração (*Configuration Model*)**
- Para assegurar a validade desta conclusão, o modelo terá que gerar todas as redes com a distribuição de grau pretendida.
- Por exemplo, pretende-se averiguar se o coeficiente de *clustering* da rede resulta da distribuição de grau.

Redes Aleatórias

- **Modelo de Configuração (*Configuration Model*)**
- Utilizando o Modelo de Configuração (*Configuration Model*) obtêm-se todas as redes com a distribuição de grau pretendida.
- Determina-se o coeficiente de *clustering* de cada uma das redes obtidas e verifica-se se é igual ao da rede inicial.

Redes Aleatórias

- **Modelo de Configuração (*Configuration Model*)**
- Os Modelos Aleatórios Exponenciais (*Exponential Random Models*) permitem obter redes com outras propriedades ou com conjuntos de propriedades (por exemplo, redes com determinado coeficiente médio de clustering e determinada densidade).

45

45

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- Todas as redes já consideradas são redes estáticas (*static networks*), porque o número de nodos não sofre alterações durante o processo de geração, apenas se acrescentam ou substituem ligações.

46

46

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- Usualmente, as redes reais são dinâmicas (*dynamic networks*), no sentido em que nodos e ligações podem ser acrescentados ou removidos.
- *A Web*, a *Wikipedia*, o *Facebook* são alguns exemplos de redes dinâmicas. Apesar de algumas remoções, a dimensão destas redes tem crescido.

47

47

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- Os modelos dinâmicos tipicamente incorporam alguma forma de crescimento da rede.
- Começa-se com uma configuração inicial, frequentemente considera-se uma clique de pequena dimensão (uma clique é uma rede ou subrede completa).
- Depois acrescentam-se os nodos um a um.

48

48

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- Cada novo nodo é unido aos nodos já existentes de acordo com uma regra, que caracteriza o modelo.
- As redes aleatórias e modelos já estudados caracterizam-se pela falta de *hubs*, porque a probabilidade de uma ligação ser escolhida não difere da probabilidade associada a qualquer outra ligação.

49

49

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- O próximo modelo procura gerar redes com *hubs*.
- Para tal, os novos nodos serão ligados aos nodos existentes, privilegiando as ligações a alguns dos nodos.
- Este mecanismo é designado por **<u>ligação preferencial</u>** (*preferential attachment*).

50

50

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- O mecanismo privilegia as ligações a nodos com maior grau.
- Este mecanismo baseia-se no princípio: quanto mais se tem, mais se recebe.
- O método de ligação preferencial (*preferential attachment)* mais conhecido para redes foi proposto por Barabási e Albert, sendo designado por modelo de Barabási-Albert ou modelo BA.

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- Começa-se com uma rede completa com m_0 nodos. Cada iteração consiste em dois passos:
  - 1 – Adiciona-se um novo nodo i à rede com $m \leq m_0$ ligações. O parâmetro m representa o grau médio da rede inicial.
  - 2 – A probabilidade de unir o novo nodo a um nodo j já existente é dada por

$$\frac{grau\ de\ j}{soma\ dos\ graus\ dos\ nodos}.$$

Redes Aleatórias

* **Ligação Preferencial (*Preferential Attachment*)**
* O número de iterações realizadas é escolhido de forma a obter o número de nodos pretendidos.
* A aplicação deste método começa com graus iguais.
* A introdução de novos nodos e de novas ligações leva ao aumento dos graus de alguns nodos.
* Desde as primeiras iterações, o grau dos primeiros nodos será maior do que o dos nodos introduzidos posteriormente.

Redes Aleatórias

* **Ligação Preferencial (*Preferential Attachment*)**
* Os nodos com maior grau têm maior probabilidade de serem escolhidos, logo o seu grau tende a aumentar bastante mais do que o dos restantes nodos.
* Assim, surgem a heterogeneidade e os *hubs*.

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- No modelo Ligação Preferencial (*preferential attachment)*, a escolha dos nodos adjacentes de cada novo nodo, com base no grau dos nodos já existentes, é relevante para a obtenção de *hubs*.
- No caso de a escolha dos nodos adjacentes ser aleatória, sem ter em conta o grau, observar-se-ia um crescimento da rede sem heterogeneidade.

55

55

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- O método Ligação Preferencial (*preferential attachment)* foi utilizado para modelar a população das cidades, a concentração de riqueza individual, a dimensão de empresas, a produção científica e outros fenómenos.
- Ver NetLogo Preferential Attachment.

56

56

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- O modelo Ligação Preferencial apresenta um conjunto de desvantagens:
  - o padrão da distribuição de grau não se altera qualquer que seja a escolha dos parâmetros;
  - os *hubs* pertencem ao conjunto dos primeiros nodos; nenhum dos nodos acrescentados a partir de determinada iteração será um *hub*;
  - não forma muitos triângulos;

Redes Aleatórias

- **Ligação Preferencial (*Preferential Attachment*)**
- O modelo Ligação Preferencial apresenta um conjunto de desvantagens:
  - só se adicionam nodos e ligações e não há remoções de nodos nem de ligações;
  - forma uma rede conexa.

Redes Aleatórias

- **Outros Modelos**
- Uma extensão do modelo anterior utiliza uma potência do grau, deixando a preferência de ser linear. Assim, a probabilidade anterior pode ser substituída por

$$\frac{k_j^\alpha}{\sum_l k_l^\alpha}$$

Redes Aleatórias

- **Outros Modelos**
- Se o valor de α for inferior a 1 então as probabilidades não crescem na mesma proporção que os graus e os *hubs* tendem a desaparecer.
- Se, pelo contrário, o valor de α for superior a 1, então os nodos com maior grau acumulam mais ligações muito mais rapidamente. Assim, m nodos serão adjacentes da maioria dos nodos.

Redes Aleatórias

- **Outros Modelos**
- Pode concluir-se que o modelo Ligação Preferencial está dependente da preferência linear (o expoente dos graus é igual a 1) para gerar diversos *hubs*. Esta é outra desvantagem do modelo.

Redes Aleatórias

- **Outros Modelos**
- ***Attractiveness Model***
- Outra variante do modelo Ligação Preferencial modifica a probabilidade adicionando uma constante A aos graus dos nodos. A probabilidade será dada por

$$\frac{A + k_j}{\sum_l (A + k_l)}$$

Redes Aleatórias

- **Outros Modelos**
- ***Attractiveness Model***
- O parâmetro A é positivo e é designado por parâmetro de atratividade.
- A ideia consiste em escolher nodos com base, não apenas no grau, mas também na sua atratividade (por exemplo, citações de obras).

63

63

Redes Aleatórias

- **Outros Modelos**
- ***Attractiveness Model***
- Ao contrário do modelo Ligação Preferencial, este modelo pode ser aplicado a configurações iniciais com nodos com grau nulo, redes orientadas e permite obter diversas distribuições de grau.
- São vantagens face ao modelo Ligação Preferencial.

64

64

Redes Aleatórias

- **Outros Modelos**
- ***Fitness Model***
- O "Fitness Model", proposto por Bianconi e Barabási, considera uma modificação no cálculo das probabilidades.
- Considere-se uma função que toma valores para cada um dos nodos.

Redes Aleatórias

- **Outros Modelos**
- ***Fitness Model***
- O grau de cada nodo é multiplicado pelo valor da função para o nodo. O valor da função traduz o apelo do nodo.
- A probabilidade será dada por:

$$\frac{n_j k_j}{\sum_l (n_l k_l)}$$

Redes Aleatórias

* **Outros Modelos**
* ***Fitness Model***
* Quanto maior for o valor, maior será o apelo. A função é designada por função de *fitness*.
* Este modelo permite obter diversos *hubs*, se os valores da função forem limitados.
* Os valores da função podem permitir a competição entre os nodos iniciais e os novos nodos.

67

67

Redes Aleatórias

* **Outros Modelos**
* ***Fitness Model***
* Como exemplos de aplicação pode referir-se a *Web*, a *Wikipedia* e publicações científicas. As páginas e os artigos com mais ligações, ou mais citados, não são necessariamente os mais antigos.

68

68

Redes Aleatórias

* **Outros Modelos**
* ***Fitness Model***
* Os valores da função não sofrem alterações ao longo do tempo, o que pode constituir uma desvantagem do método.

Redes Aleatórias

* **Outros Modelos**
* **Modelo Passeio Aleatório (*Random Walk Model*)**
* As redes geradas pelo modelo BA apresentam coeficientes de *clustering* reduzidos, porque a probabilidade de um nodo receber uma ligação é proporcional ao seu grau e não tem em conta a existência de nodos adjacentes.

Redes Aleatórias

- **Outros Modelos**
- ***Modelo Passeio Aleatório (Random Walk Model)***
- Para aumentar o número de triângulos, é necessário um mecanismo que favoreça a introdução de ligações entre nodos adjacentes.

Redes Aleatórias

- **Outros Modelos**
- ***Modelo Passeio Aleatório (Random Walk Model)***

- A formação de triângulos resultante da adição de uma ligação é designada por **fecho triádico** (*triadic closure*) e é um dos mecanismos mais relevante para a formação de ligações numa rede social.

Redes Aleatórias

- **Outros Modelos**
- ***Modelo Passeio Aleatório (Random Walk Model)***
- Vamos considerar a implementação mais intuitiva deste mecanismo, que é designada por **Modelo Passeio Aleatório** (*random walk model*).

Redes Aleatórias

- **Outros Modelos**
- ***Modelo Passeio Aleatório (Random Walk Model)***
- A ideia consiste em unir um novo nodo não só a um já existente mas também a um ou mais nodos adjacentes deste último.

Redes Aleatórias

* **Outros Modelos**
* **Modelo Passeio Aleatório (*Random Walk Model*)**
* Considere-se uma rede qualquer de dimensão pequena. Cada iteração consiste nos passos:
  * 1 -Um novo nodo i é acrescentado à rede com $m > 1$ ligações;

75

75

Redes Aleatórias

* **Modelo Passeio Aleatório (*Random Walk Model*)**
  * 2 – A primeira ligação será (i, j) em que j é um nodo já existente, escolhido aleatoriamente;
  * 3 - Cada uma das restantes ligações une i a um dos adjacentes de j com probabilidade p ou une i a um nodo escolhido aleatoriamente com probabilidade $1 - p$.

76

76

Redes Aleatórias

- ***Modelo Passeio Aleatório (Random Walk Model)***
- O número de triângulos formados depende da probabilidade p considerada.
- Se esta probabilidade não for demasiado pequena, este modelo também vai gerar *hubs*.
- A probabilidade de um nodo receber ligações será proporcional ao seu grau, como no modelo BA.

Redes Aleatórias

- ***Modelo Passeio Aleatório (Random Walk Model)***
- Contudo, os novos nodos não escolhem os seus adjacentes com base no grau. A escolha do primeiro adjacente de cada novo nodo é aleatória.
- O fecho triádico induz a ligação preferencial.

Redes Aleatórias

* ***Modelo Passeio Aleatório (Random Walk Model)***
* Este modelo foi proposto em 2003.
* Contudo, em 1973, Mark S. Granovetter publicou o artigo "*The strength of weak ties*", em que estabeleceu uma relação estreita entre três componentes fundamentais das redes sociais: triângulos, peso das ligações e comunidades.

79

79

Redes Aleatórias

* ***Modelo Passeio Aleatório (Random Walk Model)***
* Introduziu o princípio do fecho triádico forte para explicar a formação das ligações nas redes sociais.

80

80

Redes Aleatórias

* ***Modelo Passeio Aleatório (Random Walk Model)***
* Neste artigo, o sociólogo argumentou que ligações fortes estão presentes em comunidades sociais, enquanto que as ligações fracas unem comunidades sociais.
* Estas ligações fracas permitem a circulação de "informação" na rede.

Redes Aleatórias

* ***Modelo Cópia (Copy Model)***
* O modelo Cópia consiste numa variante do modelo Passeio Aleatório, em que cada nodo é ligado a um nodo escolhido aleatoriamente ou a algum (ou a alguns) dos seus adjacentes.

Redes Aleatórias

* ***Modelo Cópia (Copy Model)***
* O modelo Cópia procura modelar cenários em que muitos triângulos podem resultar de *cópias* de contactos.

Redes Aleatórias

* ***Modelo Cópia (Copy Model)***
* Como exemplos, podem referir-se:
  * duplicação de genes (as cópias vão interagir com as mesmas proteínas);
  * conhecimento de novas obras a partir de listas bibliográficas ou lista de obras da mesma colecção;

Redes Aleatórias

* ***Modelo Cópia (Copy Model)***
* Como exemplos, podem referir-se:
  * criação de conteúdos na *Web* e cópia das hiperligações.
* A escolha entre um nodo e alguns dos seus adjacentes permite obter *hubs* mas o número de triângulos é reduzido.

Redes Aleatórias

* ***Modelo Ordenação (Rank Model)***
* No modelo Ligação Preferencial, frequentemente designado por BA, a escolha das ligações dos novos nodos é baseada no grau dos restantes nodos.
* Este modelo requer o conhecimento dos valores absolutos dos graus.

Redes Aleatórias

* ***Modelo Ordenação (Rank Model)***
* O modelo Ordenação permite considerar outras propriedades, além do grau.
* Baseia-se numa ordenação dos nodos considerando a propriedade escolhida.

Redes Aleatórias

* ***Modelo Ordenação (Rank Model)***
* O modelo pode começar com uma qualquer rede de pequena dimensão com m_0 nodos.
* Uma propriedade dos nodos (grau, idade, ...) é escolhida para ordenar os nodos.

Redes Aleatórias

- ***Modelo Ordenação (Rank Model)***
- Cada iteração consiste nos passos:
  - 1 – Todos os nodos são ordenados com base na propriedade. Atribuem-se os valores *R* = 1, 2, 3, ... aos nodos. O nodo *l* da lista ordenada recebe *R* = *l*.

89

89

Redes Aleatórias

- ***Modelo Ordenação (Rank Model)***
- Cada iteração consiste nos passos:
  - 2 – Adiciona-se um novo nodo i à rede com $m \leq m_0$ novas ligações.

90

90

Redes Aleatórias

* ***Modelo Ordenação (Rank Model)***
* Cada iteração consiste nos passos:
  * 3 – A probabilidade de unir o novo nodo a um nodo j já existente é dada por
$$\frac{R_j^{-\alpha}}{\sum_l R_l^{-\alpha}}$$
em que $\alpha > 0$ é um parâmetro.

91

91

Redes Aleatórias

* ***Modelo Ordenação (Rank Model)***
* Os nodos poderão ter que ser reordenados após cada iteração, se a propriedade escolhida depender das novas ligações adicionadas.
* Por exemplo, a escolha do grau obriga a reordenar os nodos.

92

92

Redes Aleatórias

- ***Modelo** **Ordenação** (**Rank Model**)*
- Um nodo nas primeiras posições da lista ordenada terá maior probabilidade de receber uma nova ligação do que um dos nodos nas últimas posições da lista ordenada.
- Se os nodos tiverem sido ordenados pelo seu grau, os nodos com maior grau terão probabilidades associadas mais elevadas do que os com menor grau.

93

93

Redes Aleatórias

- ***Modelo** **Ordenação** (**Rank Model**)*
- No entanto, a probabilidade depende da posição na ordenação e não será proporcional aos graus.
- A taxa de decréscimo da probabilidade depende do expoente considerado.
- Este modelo gera redes com distribuição de grau com caudas longas, qualquer que seja a propriedade escolhida e o expoente considerado.

94

94

Redes Aleatórias

- ***Modelo Ordenação (Rank Model)***
- Com escolhas diferentes do expoente obtêm-se distribuições de grau com formas diferentes, o que permite reproduzir distribuições empíricas.
- Este modelo permite criar *hubs*, apesar de não incorporar informação detalhada sobre o sistema.

Redes Aleatórias

- ***Modelo Ordenação (Rank Model)***
- Um exemplo de aplicação deste modelo é relativo à ligação de novos artigos da *Wikipedia*.
- Os autores de artigos procuram ligá-los a artigos relevantes.

Redes Aleatórias

- ***Modelo Ordenação (Rank Model)***
- A utilização de motores de busca permite obter uma lista de artigos ordenada por relevância.
- Usualmente escolhem-se os artigos da primeira página da lista ordenada.
- Esta escolha leva à criação de *hubs*.

97

97

Redes Aleatórias

- **Algumas Aplicações de Redes Aleatórias**
- As redes aleatórias podem ser utilizadas para modelar e estudar alguns sistemas complexos.
- No caso das redes sociais, alguns estudos permitiram concluir que determinados comportamentos humanos são adoptados mais facil e rapidamente em redes com *clusters* (ou grupos) do que em redes aleatórias.

98

98

Redes Aleatórias

- **Algumas Aplicações de Redes Aleatórias**
- No caso da propagação de doenças, é relevante identificar as subredes em que o contágio ocorrerá mais cedo para desenvolver estratégias de prevenção e tratamento.
- A estrutura da rede em estudo poderá ter influência nos resultados, pelo que deve representar bem a população em estudo.

99

99

Redes Aleatórias

- **Algumas Aplicações de Redes Aleatórias**
- Outras aplicações consistem no estudo:
  - da robustez de redes de abastecimento (eletricidade, gás, ...);
  - na propagação de virus em redes;
  - na identificação de pontos sensíveis e fracos numa rede.

100

100

Redes Aleatórias

* **Algumas Aplicações de Redes Aleatórias**
* No campo da neurociência, o cérebro pode ser representado por uma rede de neurónios.
* Há evidência de que as alterações provocadas pelo Alzheimer condicionam a actividade na rede de neurónios, estando mais próxima da actividade numa rede aleatória do que numa rede representativa de um cérebro saudável.

101

101

Redes Aleatórias

* **Algumas Aplicações de Redes Aleatórias**
* Quais são as vantagens em utilizar redes aleatórias para modelar e estudar sistemas complexos?

102

102

ANÁLISE DE REDES
Comunidades

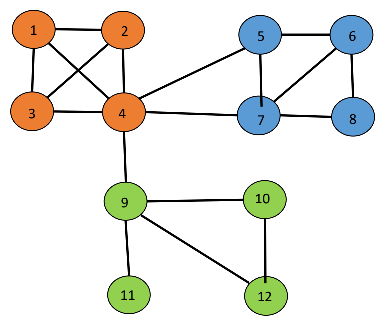
Licenciatura em
Ciência de Dados

Comunidades

* Numa rede, os nodos podem estar agrupados em subconjuntos.
* Se num subconjunto se observa que os seus elementos têm mais ligações dentro do subconjunto do que ligações para outros nodos, então o subconjunto constitui uma **comunidade**, ***cluster*** ou **módulo**.

Comunidades

* **Exemplo**



Comunidades

* As comunidades estão presentes:
  * nas redes sociais (as pessoas ou utilizadores apresentam semelhanças);
  * nas redes biológicas podem representar a interacção das proteínas (grupos de proteínas que interagem e participam na mesma função biológica);

Comunidades

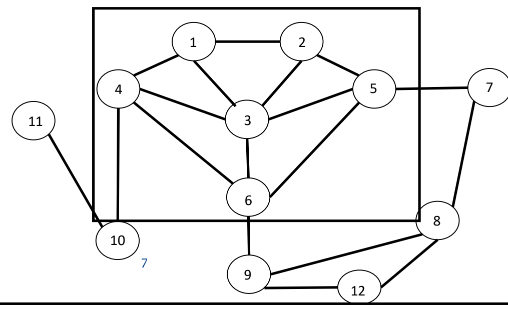
* As comunidades estão presentes:
  * na *Web* (conjuntos de documentos com muitas hiperligações entre si são usualmente sobre temas relacionados).
* As comunidades também podem ser identificadas noutras redes.
* Um exemplo bastante conhecido é *Zachary's karate club (no package igraph do R, graph("Zachary"))*.

Comunidades

* A estrutura da comunidade permite identificar quais os nodos que têm ligações apenas na comunidade – o *core* da comunidade -, e os nodos que estão na fronteira – os que estão unidos a nodos da comunidade e a outros nodos.
* Neste segundo subconjunto estão os elementos cruciais para a transmissão de "informação".

Comunidades

* **Exemplo**
* Considere que os nodos 1, 2, 3, 4, 5 e 6 pertencem à mesma comunidade.

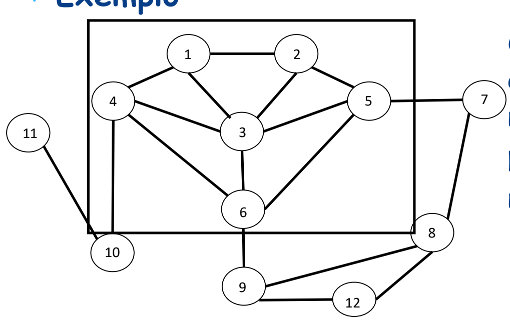


7

7

Comunidades

* **Exemplo**



O *core* da comunidade é constituído pelos nodos 1, 2 e 3.

Na fronteira estão os nodos 4, 5 e 6.

8

8

Comunidades

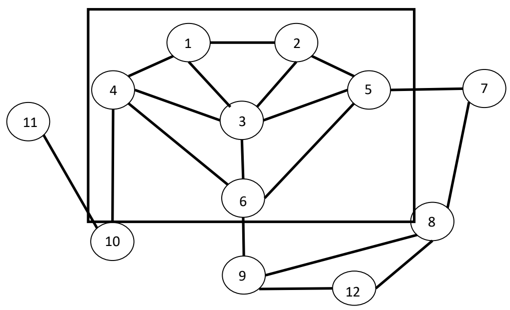
- A identificação das comunidades é relevante. Em algumas redes este processo é simples e fácil, noutras é mais difícil.
- Existem diversas definições de comunidades, umas mais restritivas do que outras.

Comunidades

- **Conceitos Básicos**
- Uma comunidade é uma subrede conexa. Além disso, existem outras condições que terão que ser verificadas.
- Seja C uma comunidade.
  - O número de nodos que pertencem a C é representado por N_C.
  - As **ligações internas** de C são as ligações entre os nodos de C. O número de ligações internas em C é representado por L_C.

Comunidades

* **Conceitos Básicos**



* $N_C = 6$.
* $L_C = 10$
* (1,2), (3,6), (2,5) são algumas das ligações internas.

11

11

Comunidades

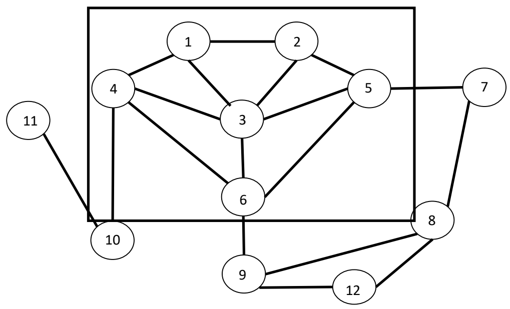
* **Conceitos Básicos**
* O **grau interno** (*internal degree*) de um nodo i relativo à comunidade C - k_i^{int} - é dado pelo número de ligações entre este nodo e os nodos de C.
* O **grau externo** (*external degree*) de um nodo i relativo à comunidade C - k_i^{ext} - é dado pelo número de ligações entre este nodo e os nodos que não pertencem a C.
* Naturalmente tem-se $k_i = k_i^{int} + k_i^{ext}$.

12

12

Comunidades

* **Conceitos Básicos**



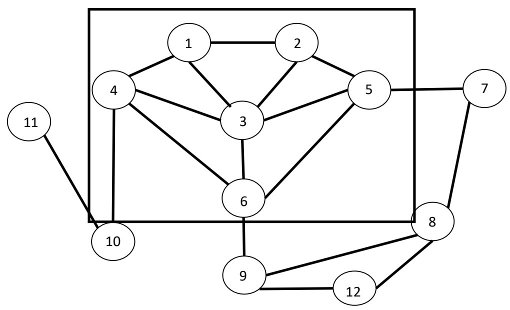
* k_1^{int}= 3, k_1^{ext}= 0;
* k_4^{int}= 3, k_4^{ext}= 1;
* k_7^{int}= 1, k_7^{ext}= 1;
* k_8^{int}= 0, k_8^{ext}= 2.

Comunidades

* **Conceitos Básicos**
* Se $k_i^{ext} = 0$ e k_i^{int}> 0 então o nodo i é um **nodo interno** de C. Este nodo pertence ao *core* da comunidade.
* Se $k_i^{ext} > 0$, k_i^{int}> 0 e i pertence a C então o nodo i pertence à **fronteira** da comunidade (*boundary node*).
* Se $k_i^{int} = 0$ então i não pertence a C e não tem ligações à comunidade.

Comunidades

* **Conceitos Básicos**



* O nodo 1 é interno.
* O nodo 4 está na fronteira.
* 7 não pertence à comunidade mas tem ligações a C.
* O nodo 8 não pertence à comunidade nem tem ligações à comunidade.

15

15

Comunidades

* **Conceitos Básicos**
* A **densidade interna** (*internal link density*) da comunidade C (relativamente às ligações) é dada por

$$d_C^{int} = \frac{L_C}{\binom{N_C}{2}} = \frac{2L_C}{N_C(N_C-1)}.$$

* O **grau da comunidade** C (*degree community*) ou **volume** (*volume*) é a soma dos graus dos nodos de C,

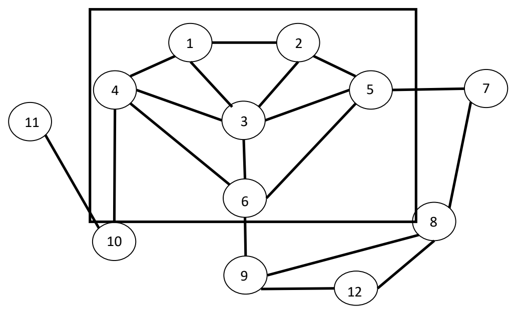
$$k_C = \sum_{i \in C} k_i.$$

16

16

Comunidades

- **Conceitos Básicos**



- A densidade interna de C é $\frac{2*10}{6*5} = \frac{2}{3}$.
- O grau de C é $k_1 + k_2 + k_3 + k_4 + k_5 + k_6 = 23$.

17

17

Comunidades

- **Conceitos Básicos**
- As definições anteriores aplicam-se a redes não orientadas e sem pesos associados às ligações.
- No caso da existência de pesos, os grau são substituídos pelo grau ponderado ou força. Assim, ao grau interno corresponde a **força interna** (*internal strength*) e ao grau externo corresponde a **força externa** (*external strength*).
- Para redes orientadas, a aplicação destes conceitos pode carecer de sentido.

18

18

Comunidades

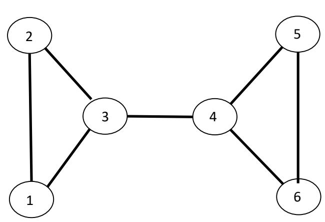
* **Definições de Comunidade**
* Usualmente considera-se que as comunidades são subredes:
  * com coesão elevada – têm muitas ligações internas;
  * apresentam uma separação elevada – estão ligadas, a outras comunidades ou subredes, por poucas ligações.

19

19

Comunidades

* **Definições de Comunidade**
* **Exemplo**



* As subredes geradas por {1,2,3} e por {4,5,6} são redes completas (apresentam uma coesão muito elevada).
* Além disso, estas duas subredes estão unidas apenas por uma ligação, estão bem separadas.

20

20

Comunidades

* **Definições de Comunidade**
* As definições de comunidades baseadas apenas na coesão tratam a comunidade como um sistema sem ter em conta o resto da rede.
* Quando se define uma comunidade como uma clique, está a considerar-se apenas a coesão.
* Contudo, em muitas situações, as comunidades são menos densas e alguns nodos têm associado um papel diferente do dos outros membros da comunidade.

21

21

Comunidades

* **Definições de Comunidade**
* Quando se consideram tanto a coesão como a separação espera-se que o número de ligações internas seja superior ao número de ligações externas.
* Podem definir-se:
  * uma **comunidade forte** (*strong community*) como uma subrede em que cada nodo tem mais adjacentes na subrede que no resto da rede (o grau interno de cada nodo da subrede é superior ao grau externo);

22

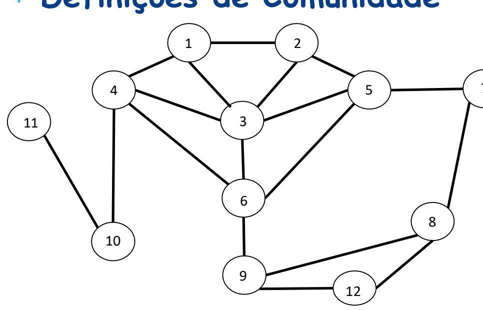
22

Comunidades

* **Definições de Comunidade**
* Podem definir-se:
  * uma **comunidade fraca** (*weak community*) como uma subrede em que a soma dos graus internos dos nodos é superior à soma dos graus externos dos nodos.
* Uma comunidade forte é também uma comunidade fraca, mas uma comunidade fraca pode não ser uma comunidade forte.

Comunidades

* **Definições de Comunidade**



* Se a subrede gerada por {1,2,3,4,5,6,9} é uma comunidade.
* É uma comunidade fraca porque $\sum_{i \in C} k_i^{int} > \sum_{i \in C} k_i^{ext}$.
* Não é forte porque $k_9^{int} < k_9^{ext}$.

Comunidades

* **Definições de Comunidade**
* As definições anteriores centram-se na comunidade e separam-na do resto da rede.
* É expectável que o resto da rede seja constituído por outras comunidades.
* Esta observação sugere outras definições de comunidades.

Comunidades

* **Definições de Comunidade**
* Naturalmente que se continua a ter em conta a coesão.
* Desta forma, se a subrede C é uma comunidade então espera-se que cada um dos nodos de C tenha ligações mais fortes com os elementos de C do que com os elementos dos restantes grupos.

Comunidades

* **Definições de Comunidade**
* Tendo em conta estas observações, propõem-se as novas definições:
  * uma **<u>comunidade forte</u>** (*strong community*) é uma subrede em que cada nodo tem mais adjacentes na subrede do que em cada uma das restantes comunidades;
  * uma **<u>comunidade fraca</u>** (*weak community*) é uma subrede em que a soma dos graus internos dos nodos é superior à soma dos graus externos dos nodos relativos a cada uma das restantes comunidades.

27

27

Comunidades

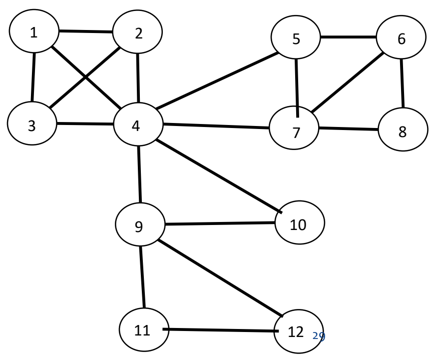
* **Definições de Comunidade**
* Estas definições são menos restritivas do que as anteriores.
* Podem identificar-se subredes que satisfazem as condições das últimas definições, mas não verificam as condições das primeiras definições.

28

28

Comunidades

* **Definições de Comunidade**



29

29

Comunidades

* **Definições de Comunidade**
* No exemplo anterior, a subrede gerada por {1,2,3,4} é uma clique.
* Esta subrede é uma comunidade forte de acordo com a segunda definição (se se considerar que cada uma das subredes {5,6,7,8} e {9,10,11,12} é uma comunidade).
* No entanto, não é uma comunidade forte de acordo com a primeira, porque o nodo 4 tem três nodos adjacentes na subrede e quatro nodos adjacentes que não pertencem à subrede.

30

30

Comunidades

- **Definições de Comunidade**
- No exemplo anterior, a subrede gerada por {5,6,7,8} é uma comunidade forte de acordo com as duas definições:

$$k_5^{int} = 2,\ k_5^{ext} = 1;\ k_6^{int} = 3,\ k_6^{ext} = 0;$$
$$k_7^{int} = 3, k_7^{ext} = 1; k_8^{int} = 2,\ k_8^{ext} = 0.$$

- O grau interno de cada nodo é superior ao grau externo, quer se considere o resto da rede ou cada um dos suconjuntos de nodos do resto da rede.

Comunidades

- **Definições de Comunidade**
- No exemplo anterior, a subrede gerada por {9,10,11,12} é uma comunidade fraca de acordo com as duas definições:

$$k_9^{int} = 3,\ k_9^{ext} = 1;\ k_{10}^{int} = 1,\ k_{10}^{ext} = 1;$$
$$k_{11}^{int} = 2, k_{11}^{ext} = 0;\ k_{12}^{int} = 2,\ k_{12}^{ext} = 0.$$

- A soma dos graus internos é superior à soma dos graus externos, quer se considere o resto da rede ou um subconjunto dos restantes nodos.
- A subrede não pode ser uma comunidade forte porque o grau interno do nodo 10 é igual ao grau externo.

Comunidades

- **Definições de Comunidade**

Comunidade Forte	Comparação baseada nos graus de cada nodo	Comparação da comunidade com o resto da rede	Definição mais restritiva
Comunidade Fraca	Comparação baseada na soma dos graus dos nodos	Comparação da comunidade com cada uma das restantes comunidades	Definição menos restritiva

33

33

Comunidades

- **Definições de Comunidade**
- As definições de comunidades já apresentadas baseiam-se no número de ligações e na comparação entre o número de ligações internas e externas de subredes.
- Contudo, o número de ligações pode depender da dimensão da subrede, o que pode dificultar ou distorcer a identificação de comunidades.

34

34

Comunidades

* **Definições de Comunidade**
* Em vez de números absolutos, seria útil considerar probabilidades.
* Alguns modelos de definição e de detecção de comunidades são baseados em probabilidades.

Comunidades

* **Definições de Comunidade**
* Por outro lado, a maior parte dos métodos para *clustering* de redes não requer uma definição precisa de comunidade.
* No entanto, a definição de critérios para comunidades pode ser útil para verificar a fiabilidade dos resultados obtidos.

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- Uma **partição** (*partition*) de uma rede é uma divisão em subredes em que cada nodo pertence a exatamente uma subrede e cada subrede tem pelo menos um nodo. Além disso, o conjunto de ligações não sofre alterações.

37

37

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- O número de possíveis partições é designado por **número de Bell** (*Bell number).*
- Este número cresce exponencialmente com o número de elementos da rede.
- Para uma rede com 10 nodos existem mais de 20 000 partições.

38

38

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- Assim, torna-se impossível obter todas as partições e escolher a melhor.
- Por este motivo, os algoritmos de *clustering* consideram apenas uma pequena fracção das partições possíveis.

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- A identificação de comunidades não sobrepostas (comunidades disjuntas) corresponde à determinação de partições com algumas propriedades.
- No caso de comunidades com possíveis sobreposições (um ou mais nodos podem pertencer a mais do que uma comunidade), a divisão da rede é designada por **cobertura** (*cover*).
- O número de coberturas é ainda superior ao número de partições.

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* As partições podem ser hierárquicas. Estas partições surgem quando a rede possui diversos níveis de organização em diferentes escalas.
* Neste caso, os grupos ou *clusters* apresentam uma estrutura em que cada comunidade pode subdividir-se noutras mais pequenas.

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* Um centro hospitalar pode ser composto por diversos hospitais. Em cada um dos hospitais há diversos grupos, comunidades profissionais ou serviços.
* Existem também ligações entre os diferentes hospitais, mas em número inferior ao de ligações em cada hospital.

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- As partições de redes reais são frequentemente heterógeneas, no sentido em que apresentam propriedades que variam muito de um grupo para outro.
- Por exemplo, muitas vezes existe uma grande diferença entre a dimensão de cada comunidade da mesma rede.

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- Na *Web* há grupos ou *clusters* (que resultam do tema abordado) com milhões de páginas e outros com milhares.
- A coesão também pode ser muito diversa, alguns grupos apresentam uma densidade interna bastante superior à de outros.

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* Usualmente as comunidades estão bem separadas de outras comunidades.
* A identificação de subredes bem separadas é o objectivo da partição de uma rede.
* A ênfase está na separação e não na coesão.

45

45

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* Por este motivo, os métodos de partição não são adequados para a detecção de comunidades.
* No entanto, algumas técnicas de partição são utilizadas, em conjunto com outros procedimentos, na detecção de comunidades.

46

46

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* O problema de partição consiste em dividir uma rede em subredes de determinada dimensão de forma a minimizar o número de ligações entre as subredes.
* Este conjunto de ligações é designado por **corte** (*cut*), porque a remoção destas ligações deixa a rede desconexa.

47

47

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* O número de elementos do corte é designado **dimensão do corte** (*cut size*).
* O problema de partição de uma rede também é conhecido como o **problema de minimização do corte**.

48

48

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* A resolução deste problema requer a indicação do número de subconjuntos e a dimensão aproximada de cada subconjunto (para evitar obter soluções triviais como um só subconjunto constituído por toda a rede ou dois subconjuntos tendo um deles apenas um dos nodos da rede).

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* Um dos algoritmos mais antigos e mais conhecidos para a bissecção de uma rede (partição em dois subconjuntos) é o algoritmo de Kerninghan-Lin.
* Considera-se uma bisecção inicial da rede. O número de elementos em cada subconjunto é igual ou difere apenas numa unidade.

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* O algoritmo consiste na troca de pares de nodos.
* Em cada troca, consideram-se dois nodos pertencentes a subconjuntos diferentes e transfere-se cada um dos nodos para o outro subconjunto. Assim, a dimensão de cada subconjunto não se altera.
* O par de nodos é escolhido de forma a reduzir o número de ligações entre os dois subconjuntos.

51

51

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* ***Algoritmo de Kerninghan-Lin***
* Começa-se com uma partição arbitrária P da rede em dois subconjuntos A e B. Por exemplo, escolhem-se aleatoriamente metade dos nodos e atribuiem-se a um dos subconjuntos. Os restantes nodos são incluídos no outro suconjnuto.

52

52

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* ***Algoritmo de Kerninghan-Lin***
* Cada iteração é constituída pelos seguintes passos:
  * 1 – Para cada par (i, j) , $i \in A$ e $j \in B$, calcula-se o impacto na dimensão do corte resultante de mudar i para B e j para A;
  * 2 – O par (i^*, j^*) com a maior redução associada é escolhido e trocam-se os nodos. O par de nodos escolhido é bloqueado e os nodos não voltarão a ser escolhidos nesta iteração;

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* ***Algoritmo de Kerninghan-Lin***
* 3 – Repetem-se os passos 1 e 2 até não se conseguir reduzir a dimensão do corte, através da troca de pares de nodos não bloqueados.
* O algoritmo termina quando não se consegue reduzir a dimensão do corte.

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- ***Algoritmo de Kerninghan-Lin***
- Pode adaptar-se este algoritmo de forma a considerar mais do que dois subconjuntos.
- As soluções determinadas por este algoritmo dependem da partição inicial.

55

55

Comunidades

- **Problemas relacionados com Comunidades**
- **Partição de Redes**
- ***Algoritmo de Kerninghan-Lin***
- Além disso, trata-se de uma heurística *greedy*, porque se escolhe a troca mais conveniente para cada uma das iterações, sem a possibilidade de alterar posteriormente as escolhas anteriores.
- Este método pode ser utilizado para melhorar soluções geradas por outros métodos.

56

56

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* A partição das redes tem a desvantagem de ter apenas como objectivo a separação, ignorando a coesão.
* Por outro lado, é necessário especificar o número de subconjuntos pretendidos em vez de escolher este número com base nos dados.

Comunidades

* **Problemas relacionados com Comunidades**
* **Partição de Redes**
* Contudo, a partição de redes pode ser aplicada a diversos problemas: processamento de imagem, dinâmica de fluidos, controlo de tráfego aéreo, computação em paralelo,...

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* As comunidades procuram agrupar nodos com caraterísticas comuns.
* Pode concluir-se que o problema de detecção de comunidades é um caso especial de um problema bastante mais geral – *data clustering* -, que procura agrupar dados com base em alguma noção de similaridade.

59

59

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* Alguns conceitos e ferramentas relativos ao *data clustering* podem ser utilizados em redes.
* Os modelos hierárquicos de *data clustering* geram um conjunto de partições, enquanto que os modelos de partição fornecem apenas uma partição.

60

60

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* Frequentemente a escolha para adaptar modelos de *data clustering* para *clustering* em redes recai sobre os modelos hierárquicos.

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* É necessária uma medida de similaridade para efectuar o agrupamento.
* Se for possível representar os nodos num plano geométrico, pode considerar-se a distância entre os nodos. Quanto mais próximos, mais similares serão os nodos.

Comunidades

- **Problemas relacionados com Comunidades**
- ***Data Clustering***
- Outra alternativa consiste em recorrer à estrutura da rede. A **equivalência estrutural** (*strutural equivalence*) baseia-se na similaridade entre os nodos adjacentes. Pode definir-se:

$$S_{ij}^{SE} = \frac{número\ de\ nodos\ adjacentes\ comuns\ de\ i\ e\ a\ j}{número\ de\ nodos\ adjacentes\ de\ i, ou\ de\ j\ ou\ de\ ambos}$$

63

63

Comunidades

- **Problemas relacionados com Comunidades**
- ***Data Clustering***
- Quanto à similaridade entre dois conjuntos de nodos, pode ser medida através das ligações entre os dois conjuntos.
- Para tal, será necessário determinar os *scores* dos pares de nodos que pertencem a conjuntos diferentes.

64

64

Comunidades

- **Problemas relacionados com Comunidades**
- ***Data Clustering***
- Dada uma medida de similaridade S e dois conjuntos de nodos G_1 e G_2, a similaridade entre G_1 e G_2 será feita com base nos *scores* dos pares de nodos. Podem considerar-se:
  - $S_{G_1G_2} = max_{i,j}S_{ij}$ (*simple linkage*);
  - $S_{G_1G_2} = min_{i,j}S_{ij}$ (*complete linkage*);
  - $S_{G_1G_2} = \langle S_{ij} \rangle_{i,j}$ (*average linkage*).

Comunidades

- **Problemas relacionados com Comunidades**
- ***Data Clustering***
- Os métodos hierárquicos de *clustering* podem ser aglomerativos (juntam grupos de nodos) ou podem consistir na divisão de grupos.
- Os métodos hierárquicos de *clustering* aglomerativos começam com uma partição com N grupos.
- Em cada grupo há apenas um nodo. Em cada passo, cada par de grupos com maior similaridade é fundido e o número de grupos diminui em uma unidade.

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* Este passo é repetido até todos os nodos estarem no mesmo grupo.
* Estes métodos geram N partições que podem ser representadas num dendrograma.

67

67

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* O *clustering* hierárquico apresenta diversas limitações.
* Gera um número de partições igual ao número de nodos sem indicação de um critério para escolher a partição mais indicada.

68

68

Comunidades

* **Problemas relacionados com Comunidades**
* ***Data Clustering***
* Os resultados destes métodos dependem da medida de similaridade adoptada.
* Estes algoritmos são muito lentos não sendo aplicáveis para redes com uma dimensão muito elevada.

Comunidades

* **Problemas relacionados com Comunidades**
* **Determinação de cliques**
* Uma clique é uma subrede completa.
* A definição de comunidade como uma clique é muito restritiva e está baseada apenas na coesão e não na separação.
* Além disso, esta definição pode gerar uma partição da rede com um número excessivo de comunidades.

Comunidades

- **Problemas relacionados com Comunidades**
- **Determinação de cliques**
- A identificação da clique com o número máximo de nodos (*maximal clique*) é um problema com aplicações em redes sociais, bioinformática, química computacional,...
- Este problema pode ser resolvido com recurso a heurísticas ou a Programação Linear Inteira.

71

71

Comunidades

- **Detecção de Comunidades**
- Existem muitos métodos de detecção de comunidades. Vamos estudar apenas alguns.
- Estudaremos também um método de avaliação.

72

72

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* Uma **ponte** (*bridge*) é uma ligação cuja remoção divide uma rede conexa em duas subredes.
* No caso de ser possível identificar todas as pontes de uma rede, a remoção destas pontes iria dividir a rede em subredes e a identificação de comunidades seria imediata.

73

73

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* Nos métodos de remoção de pontes é necessária uma medida que permita a identificação de pontes.
* No algoritmo de Girvan e Newmann, esta medida é a intermediação de ligações (baseada no número de caminhos mais curtos que passam por uma ligação).

74

74

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* É expectável que as ligações com maior intermediação sejam as que unem comunidades, em vez das ligações internas às comunidades.

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* Em cada iteração deste algoritmo, identifica-se a ligação com maior intermediação e remove-se esta ligação.
* Recalcula-se depois a intermediação das ligações, com vista a escolher a próxima ligação a ser retirada.

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* Durante a aplicação do algoritmo o número de componentes conexas vai aumentando.
* O algoritmo termina após a remoção de todas as ligações.

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* ***Algoritmo de Girvan e Newmann***
* Começa-se por determinar a intermediação de cada ligação. Cada iteração consiste nos passos:
  * 1 - Remove-se a ligação com maior intermediação. No caso de empate, escolhe-se uma das ligações com intermediação mais elevada;

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* ***Algoritmo de Girvan e Newmann***
  * 2 - Recalcula-se a intermediação para cada uma das restantes ligações.
* O algoritmo termina quando todas as ligações tiverem sido removidas.

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* Em cada iteração é necessário recalcular as medidas de intermediação, o que pode tornar o algoritmo muito lento.
* Se a rede tiver uma estrutura de comunidades fortes, rapidamente a rede fica dividida em componentes conexas.

Comunidades

- **Detecção de Comunidades**
- **Remoção de Pontes**
- Este algoritmo gera *N* partições.
- A primeira inclui todos os elementos da rede e em cada iteração cada *cluster* é dividido em dois.
- Trata-se, por isso, de um método hierárquico de *clustering* de divisão.

81

81

Comunidades

- **Detecção de Comunidades**
- **Remoção de Pontes**
- Por se tratar de um método lento, torna-se impraticável a sua aplicação para redes de dimensão elevada.
- Para tornar o método mais rápido, o cálculo da intermediação pode ser substituído por valores baseados numa amostra de ligações.

82

82

Comunidades

* **Detecção de Comunidades**
* **Remoção de Pontes**
* Outra desvantagem do método é a obtenção muitas partições sem indicação sobre a sua qualidade.
* Torna-se necessário considerar uma medida para avaliar a qualidade das partições.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Quando se pretende identificar as comunidades numa rede ou obter uma ou mais partições de uma rede, há interesse em avaliar as soluções obtidas.

Comunidades

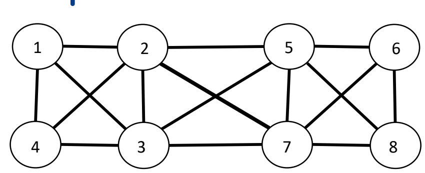
- **Detecção de Comunidades**
- **Optimização da Modularidade**
- As subredes da partição devem assemelhar-se a comunidades.
- Com vista a obter subredes coesas, pode determinar-se a densidade interna e verificar se toma valores elevados.
- No entanto, a escolha com base na densidade pode não produzir bons resultados.

85

85

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- **Exemplo**



86

86

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* **Exemplo**
* Se se considerarem duas comunidades, uma gerada por {1,2,3,4} e outra por {5,6,7,8} então a densidade interna de cada comunidade é igual a um (que é o valor máximo para a densidade).

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* **Exemplo**
* Contudo, estas duas comunidades estariam separadas por quatro ligações e cada uma teria seis ligações internas. Estariam separadas por um número excessivo de ligações.
* Considerando apenas uma comunidade, a densidade será igual a 0,57.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Outra ideia para avaliar as partições consiste em comparar com redes em que não existem comunidades.
* Nas redes aleatórias, redes em que as ligações são escolhidas aleatoriamente, não é expectável encontrar comunidades.
* Assim, pode comparar-se o número de ligações internas de cada subrede da partição com o número esperado de ligações existentes numa rede aleatória.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* A modularidade é uma medida baseada nesta ideia.
* Contudo, não se trata de uma medida que produz valores absolutos.
* Esta medida calcula um valor que pretende medir as diferenças entre a rede em estudo e a estrutura de uma rede aleatória.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Se o número de ligações internas de cada subrede for bastante superior ao número esperado numa rede aleatória então é pouco provável que a concentração elevada de ligações resulte de um processo aleatório.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* No caso de o número de ligações internas de cada subrede ser bastante superior ao número esperado numa rede aleatória, o valor da modularidade será mais elevado.
* Caso contrário, o número de ligações internas de cada subrede for aproximado do número esperado numa rede aleatória então a modularidade tomará um valor mais baixo.

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- No cálculo da modularidade de uma partição, considera-se que C representa um elemento da partição (um subconjunto) e utiliza-se uma aproximação da probabilidade de escolher ao acaso uma ligação de uma rede aleatória que una dois elementos de C.
- Se a rede for não orientada e sem pesos associados às ligações, a probabilidade aproximada de escolher aleatoriamente um dos nodos de C unidos pela ligação é dada por $\frac{k_C}{2L}$.

93

93

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- Então a probabilidade de escolher ao acaso uma ligação de uma rede aleatória que una dois elementos de C será aproximada por $\frac{k_C}{2L} \cdot \frac{k_C}{2L}$. O valor esperado de ligações internas em C será

$$L.\frac{k_C}{2L} \cdot \frac{k_C}{2L} = \frac{k_C^2}{4L}$$

94

94

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Obtém-se então

$$Q = \frac{1}{L}\sum_{C}\left(L_C - \frac{k_C^2}{4L}\right)$$

* O valor de Q será sempre inferior a 1.
* Se existir apenas um *cluster* então $Q = 0$.
* O valor de Q pode ser negativo. Se existirem N clusters então o valor de Q será negativo.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* A modularidade também pode ser definida para redes orientadas. Basta substituir o grau pelo grau incidente e grau divergente obtendo-se

$$Q_d = \frac{1}{L}\sum_{C}\left(L_C - \frac{k_C^{in} k_C^{out}}{L}\right)$$

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- Para redes com pesos associados às ligações tem-se
  - $Q_w = \frac{1}{W}\sum_C \left(W_C - \frac{s_C^2}{4W}\right)$

  em que W, W_C e s_C representam, respectivamente, a soma dos pesos das ligações da rede, a soma dos pesos das ligações internas de C e a força dos nodos em C.

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- Para redes orientadas e com pesos associados às ligações tem-se

$$Q_{dw} = \frac{1}{W}\sum_C \left(W_C - \frac{s_C^{in} s_C^{out}}{W}\right)$$

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* A modularidade foi inicialmente introduzida para avaliar as soluções obtidas pelo algoritmo de Girvan-Newmann e escolher a melhor solução.
* No entanto, a modularidade pode ser utilizada para avaliar qualquer partição. Assim, também podem considerar-se métodos de optimização de modularidade.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Como o número de partições de uma rede é bastante elevado, torna-se impossível obter e avaliar todas as partições.
* Os métodos de optimização de modularidade, consideram apenas um pequeno subconjunto de partições.

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- Um método simples, e aglomerativo, de optimização de modularidade consiste em começar com conjuntos singulares e em cada iteração fundir dois grupos.
- A partição inicial tem um valor negativo de modularidade.

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- Em cada iteração, a fusão escolhida é a que gera um maior aumento da modularidade.
- O método termina com apenas um grupo (com todos os nodos da rede).
- O método explora uma hierarquia de partições que pode ser representada por um dendrograma.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Durante a execução do método, o valor da modularidade começa por crescer até atingir o valor mais elevado.
* Depois decresce até tomar o valor nulo (quando se obtém apenas um grupo).
* A partição escolhida será a que tem o valor mais elevado de modularidade. Este método é uma heurística *greedy*.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Por se tratar de uma heurística *greedy*, o método tende a obter soluções que não atingem o valor máximo da modularidade.
* Além disso, o método pode obter soluções em que a dimensão dos grupos varia muito.
* Estas soluções podem não ser adequadas. A fusão de grupos pode mitigar esta desvantagem.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Existem diversos métodos de optimização de modularidade.
* Um dos mais utilizados é o algoritmo de Louvain, que é aglomerativo e que agrupa as comunidades em supernodos.

105

105

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* **Algoritmo de Louvain**
* A solução inicial é composta por conjuntos singulares.
* Cada iteração do algoritmo consiste em dois passos:

106

106

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* **Algoritmo de Louvain**
  * 1 - Cada nodo é atribuído à comunidade de um dos nodos adjacentes. O nodo adjacente é escolhido de forma a obter o maior aumento da modularidade face à solução actual. Os nodos são inspeccionados diversas vezes até não ser possível incrementar o valor da modularidade;

107

107

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* **Algoritmo de Louvain**
  * 2 – Transforma-se a rede numa rede com supernodos e com pesos associados às ligações. Cada comunidade determinada no passo anterior é substituída por um supernodo. As ligações terão peso igual à soma dos pesos das ligações entre as comunidades representadas nos supernodos. Criam-se lacetes com peso igual à soma dos graus internos.

108

108

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- ***Algoritmo de Louvain***
- No passo 1, a modularidade é calculada com base na rede original.
- O algoritmo termina quando não se consegue aumentar a modularidade.
- Ver exemplo de DOI: 10.1088/1742-5468/2008/10/P10008

109

109

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- ***Algoritmo de Louvain***
- Trata-se de um algoritmo *greedy*, que produz soluções que usualmente não estão próximas da óptima.
- A solução gerada pelo método depende da ordem pela qual os nodos são visitados.
- O método reduz a dimensão da rede muito rapidamente, sendo, por isso, bastante utilizado.

110

110

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* A modularidade apresenta algumas limitações.
* O valor máximo tende a ser mais elevado em redes de maior dimensão, o que não permite a comparação entre redes de dimensões diferentes.
* Apesar de a modularidade ser uma medida que procura medir as diferenças entre a rede em estudo e a estrutura de uma rede aleatória, pode assumir valores elevados para algumas redes aleatórias.

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* A medida é baseada no valor esperado de ligações em cada comunidade, mas não tem em conta flutuações aleatórias em torno do valor esperado.

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- Além disso, o valor máximo da modularidade pode não estar associado à melhor partição.
- Isto acontece porque a modularidade pode não determinar comunidades de pequena dimensão.
- As comunidades com grau inferior a $\sqrt{2L}$ podem não ser identificadas.

113

113

Comunidades

- **Detecção de Comunidades**
- **Optimização da Modularidade**
- **Exemplo**
- Considere-se uma rede constituída por 16 cliques, cada uma com 4 nodos, e com uma ligação entre cada clique.
- A partição em que cada clique corresponde a uma comunidade tem um valor de modularidade (≈0,795) inferior à partição em que cada subconjunto abrange duas cliques (≈0,804) .

114

114

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Para contornar esta desvantagem, pode introduzir-se, na definição de modularidade, um parâmetro que tenha em conta a dimensão da comunidade.
* Esta introdução requer a escolha de um valor para este parâmetro (que deve ser adequado para a rede).

Comunidades

* **Detecção de Comunidades**
* **Optimização da Modularidade**
* Por outro lado, a modularidade deve ser optimizada para diversas escolhas do parâmetro.
* Esta modificação requer algum esforço computacional. Apesar disso, esta técnica é muito utilizada.

Comunidades

- **Detecção de Comunidades**
- **Propagação de Etiquetas**
- Trata-se de um método simples e rápido de detecção de comunidades, baseado na ideia de que nodos adjacentes pertencem à mesma comunidade.

117

117

Comunidades

- **Detecção de Comunidades**
- **Propagação de Etiquetas**
- Desta forma, espera-se que a maioria das ligações seja interna e que as comunidades sejam coesas e bem separadas.

118

118

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* Em cada passo, o algoritmo inspecciona cada nodo e afeta-o à comunidade da maioria dos adjacentes.
* Começa-se por atribuir uma etiqueta distinta a cada nodo.

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* O método consiste na repetição dos seguintes passos:
  * 1 – Inspeccionam-se todos os nodos, a ordem é aleatória. Cada nodo recebe a etiqueta da maioria dos seus adjacentes. Em caso de empate, escolhe-se aleatoriamente uma das etiquetas mais frequentes.

Comunidades

- **Detecção de Comunidades**
- **Propagação de Etiquetas**
  - 2 – Se todos os nodos têm a etiqueta da maioria dos seus adjacentes (estado de estacionaridade), o algoritmo termina. Caso contrário, volta-se ao passo 1.
- As comunidades são os grupos com etiquetas iguais, no estado estacionário.

Comunidades

- **Detecção de Comunidades**
- **Propagação de Etiquetas**
- As etiquetas propagam-se durante a execução do algoritmo. Algumas desaparecem, enquanto que outras se tornam mais frequentes.

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* Durante a execução do algoritmo ocorrem diversas mudanças, no entanto, o algoritmo converge usualmente após um número pequeno de iterações, independentemente da dimensão da rede.

123

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* Na solução gerada, cada nodo tem mais adjacentes na sua comunidade do que em qualquer uma das outras comunidades.
* Por isso, este algoritmo gera comunidades fortes no sentido da definição menos restritiva.

124

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* A solução gerada pelo algoritmo pode depender da ordem de inspecção dos nodos e dos empates, o que poderá ser considerado uma desvantagem.

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* Para aplicar este algoritmo não é necessária informação sobre o número nem a dimensão das comunidades.
* O método não tem parâmetros.
* Trata-se, por isso, de um método simples, rápido e que pode ser aplicado a redes com dimensão elevada.

Comunidades

* **Detecção de Comunidades**
* **Propagação de Etiquetas**
* Se se conhecer uma estrutura de comunidades para alguns dos nodos, esta informação pode ser utilizada para definir as etiquetas iniciais.

Comunidades

* **Detecção de Comunidades**
* **Comunidades Sobrepostas**
* Em algumas situações, alguns nodos podem pertencer a mais do que uma comunidade. Estas comunidades são usualmente designadas por **comunidades sobrepostas** (*overlapping communities*).
* Existem métodos para detectar comunidades sobrepostas (Algoritmo CFinder,...). Estes métodos não pertencem aos conteúdos desta Unidade Curricular.

Comunidades

- **Avaliação dos Métodos de Detecção de Comunidades**
- Como se pode aferir a qualidade de um método de detecção de comunidades?
- Dados dois métodos, qual é o melhor?
- Estas perguntas não têm uma resposta fácil.

129

129

Comunidades

- **Avaliação dos Métodos de Detecção de Comunidades**
- Uma forma de avaliar os métodos consiste em aplicá-los a um conjunto de instâncias (*benchmark*) de redes com estruturas de comunidades naturais e analisar os resultados obtidos.
- Existem instâncias artificiais criadas para testar métodos e instâncias reais com estruturas de comunidades.

130

130

Comunidades

- **Avaliação dos Métodos de Detecção de Comunidades**
- Uma das instâncias reais mais conhecidas é a do clube de Karaté de Zachary. Um bom método deve identificar duas comunidades nesta rede.
- Além dos testes em instâncias com comunidades, podem também ser consideradas instâncias sem comunidades para avaliar o comportamento dos métodos. Caso identifiquem comunidades, os métodos não serão fiáveis.