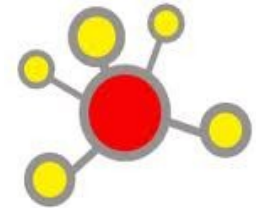




R(Igraph): Métricas para análise de Redes Complexas



Valdete Gonçalves-Almeida
valdete@dcc.ufmg.br

SUMÁRIO

Introdução

Métricas em grafos

R(Igraph)

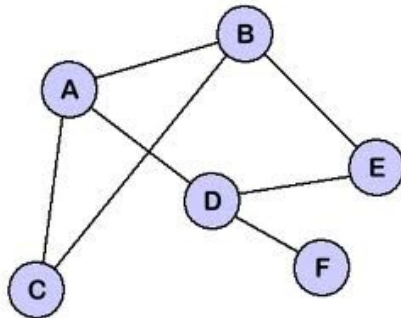
Referências

Grafos ou Redes Complexas

O que são grafos?

Um grafo é definido formalmente como $G = (V, E)$

Conjunto de vértices (entidades) conectados por arestas (relacionamento)



Grafos ou Redes Complexas

Por que estudar grafos/redes?

Importante ferramenta matemática com aplicação em diversas áreas do conhecimento

Existem centenas de problemas computacionais que usam grafos com sucesso.

Identificar a habilidade de comunicação entre duas entidades em uma rede

Criar heurísticas ótimas/sub-rotinas para realizar busca de padrões em redes reais

Grafos ou Redes Complexas

O que podemos modelar por meio de grafos?

Redes Tecnológicas

Redes Sociais

Redes de Informações

Redes Biológicas

Grafos ou Redes Complexas

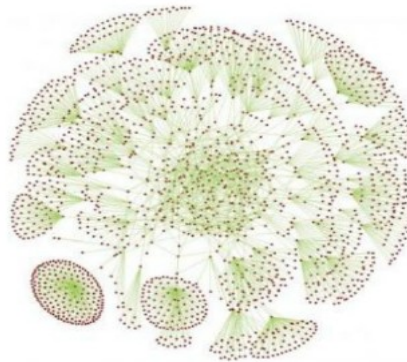
O que podemos modelar por meio de grafos?

Redes Tecnológicas

Redes Sociais

Redes de Informações

Redes Biológicas



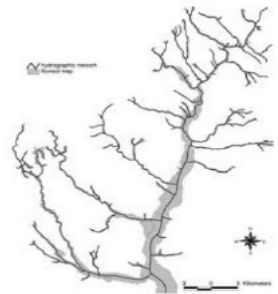
Internet

Estrutura da Internet da universidade de San Diego/California



Sistema de Metrô

Sistema de metrô de Londres



Mapa Hidrográfico

Bacia hidrográfica do rio Hérault (sul da França)

Grafos ou Redes Complexas

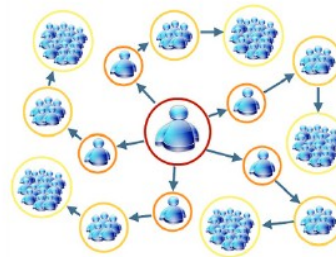
O que podemos modelar por meio de grafos?

Redes Tecnológicas

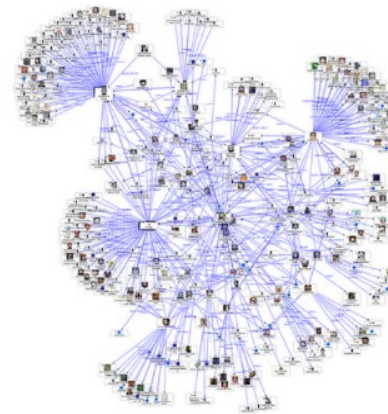
Redes Sociais

Redes de Informações

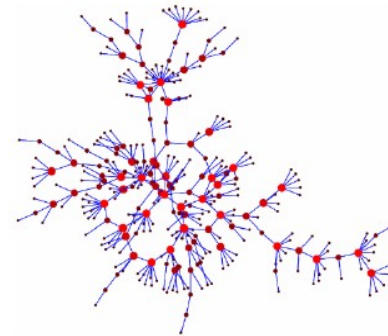
Redes Biológicas



Redes sociais (Facebook)



Crime Organizado



Contatos sexuais

Grafos ou Redes Complexas

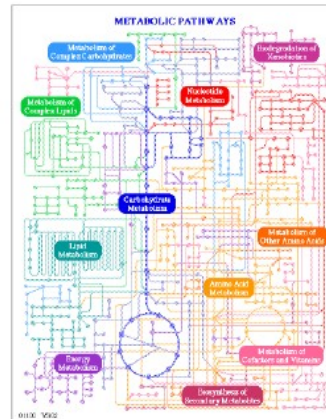
O que podemos modelar por meio de grafos?

Redes Tecnológicas

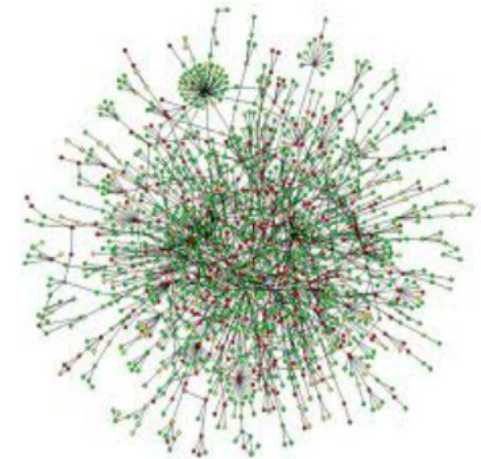
Redes Sociais

Redes de Informações

Redes Biológicas



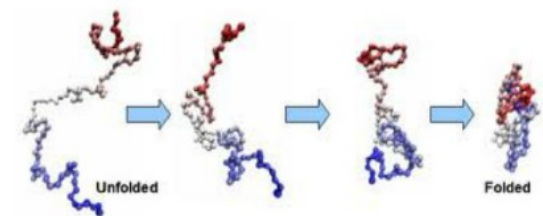
Mapa metabólico



Interações proteicas (levedura)



Etanol



Enovelamento proteico

Análises de Redes Biológicas

■ Redes de Interações de Proteínas (PPI)

■ São redes de interações de proteínas com a finalidade de desempenhar alguma função biológicas (Ex.: Replicação do DNA)

■ São conhecidas como Interatomas

■ Base de dados CORUM

■ Identificação de PPIs densamente conectadas

■ Predição e confirmação no GO (redes associadas a doenças)

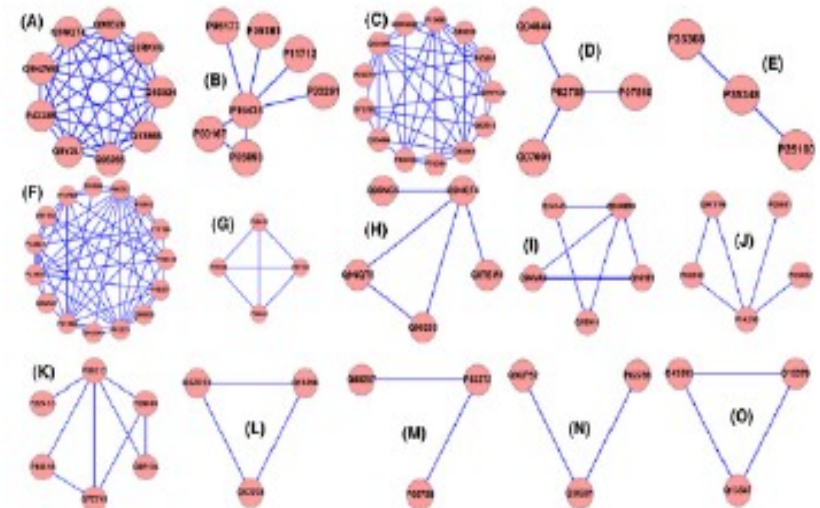


Fig. 3. Topology structure of the 10 protein complexes with the lowest P-value (p-values are performed with the CytoScape).

Análises de Redes Biológicas

■ Redes de Ecológicas

- Descreve tipos de redes ecológicas
- Redes de cooperação entre espécies (mutualismo)
- Grafos Bipartidos

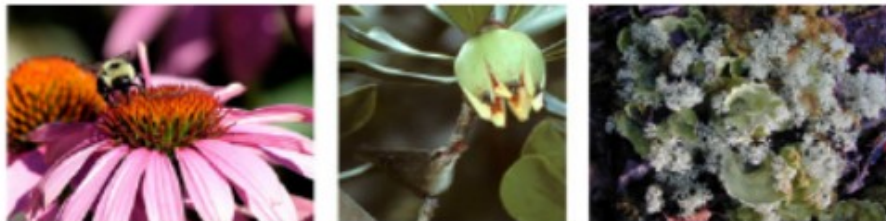


Figure 1 Examples of mutualistic relationships in nature. From left to right: plant-pollinator interaction between a bee and a flower, plant-frugivore interaction between a bird and a fleshy fruit plant, and mutualistic association between fungi and algae (i.e. a lichen).

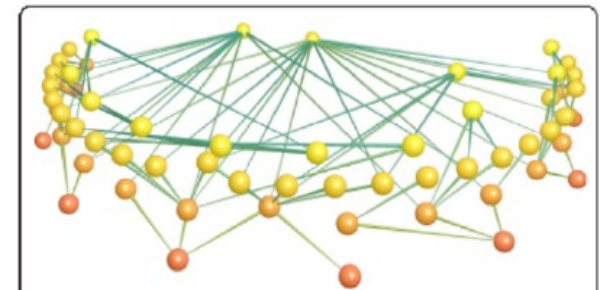


Figure 2 Example of a complex food web in a real community (grassland in the United Kingdom). Nodes and edges (in the graph) represent species and trophic interactions among them respectively. (Image produced with FoodWeb3D, written by R.J. Williams and provided by the Pacific Ecoinformatics and Computational Ecology Lab <http://www.foodwebs.org>).

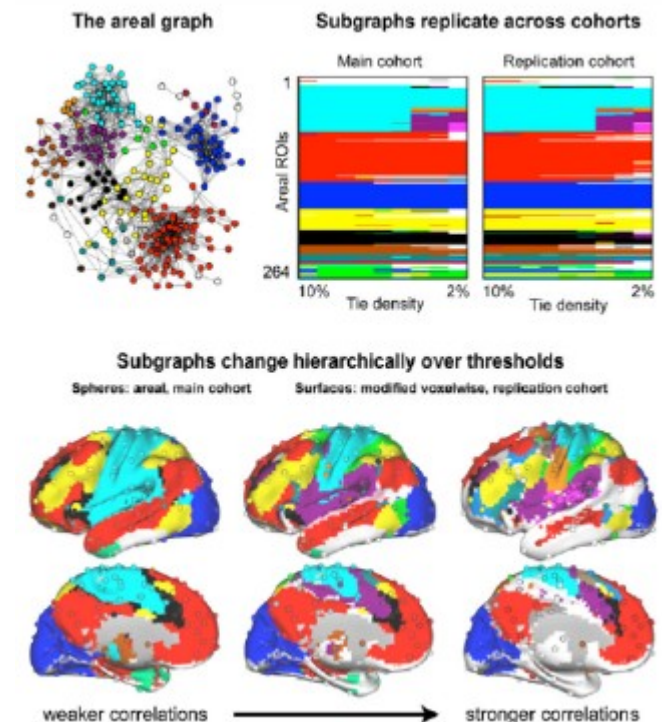
Análises de Redes Biológicas

■ Redes de Funções Cerebrais

■ Baseado em Imagens de MRI

■ Mapeamento de regiões cerebrais em grafos

■ Análise da condição cerebral de indivíduos com doenças (Alzheimer, Esquizofrenia, Esclerose múltipla)

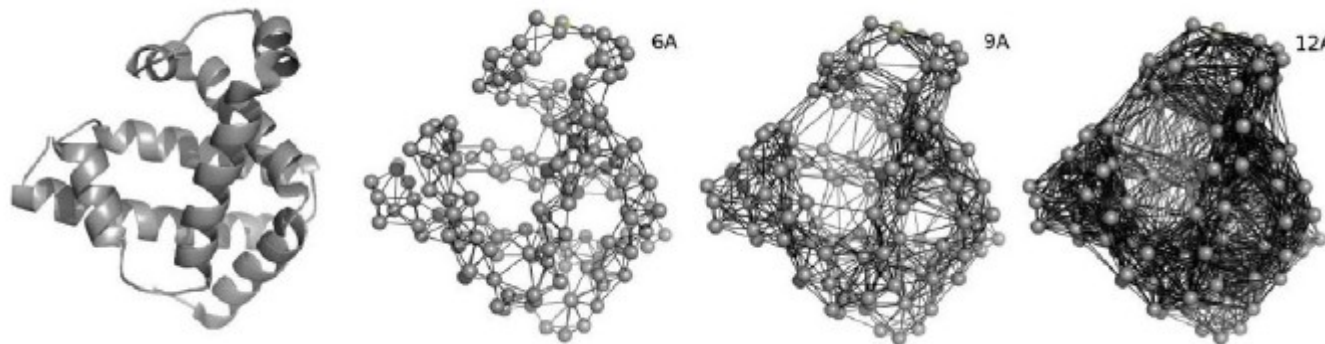


Análises de Redes Biológicas

■ Redes de Interações Atômicas

■ Classificação estrutural e predição de função por meio de contatos atômicos

■ CSM (Cutoff Scanning Matrix)



Análises de Redes Biológicas

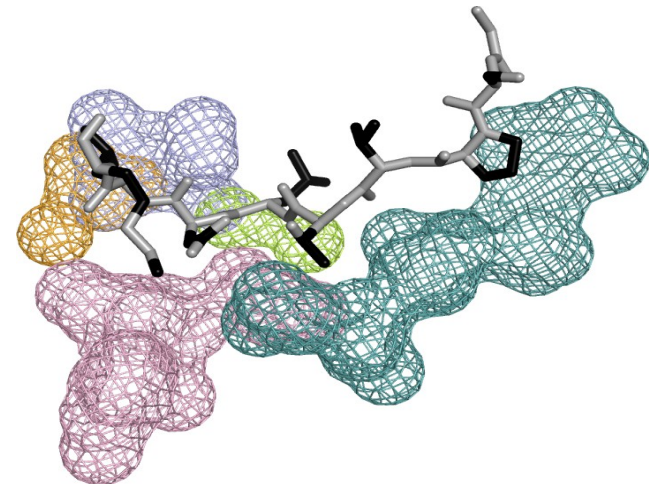
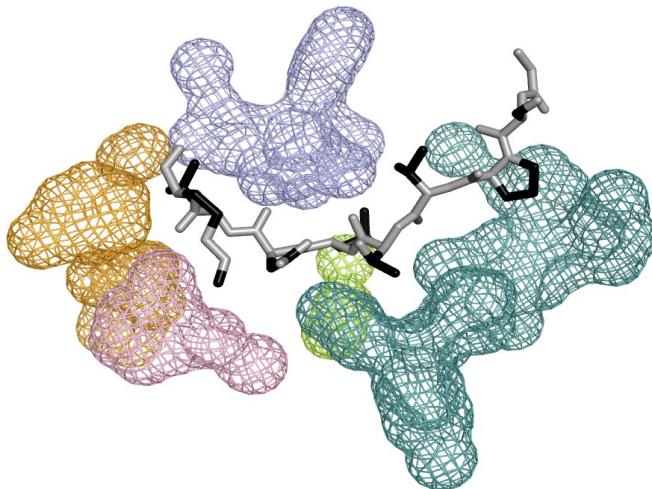
■ Redes de Interações Atômicas

■ Identificar regiões hidrofóbicas na interface em complexos proteicos

■ Busca por componentes conexos

■ Padrões frequentes em casos de inibição cruzada

■ Hydropace



Métricas em Redes Complexas

- Medidas de Centralidades

- Grau

- Closeness

- Betweenness

Métricas em Redes Complexas

Grau:

É uma medida relativa aos vértices de um grafo

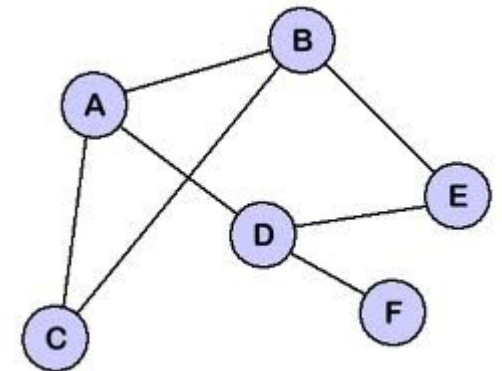
O grau de um vértice é dado pelo número de arestas que lhe são incidentes

Exemplo:

Grau 3 = A, B, D

Grau 2 = C, E

Grau 1 = F



Caminhos em Grafo

Distância natural métrica entre todos os pares de vértices.

Caminhos mais curtos podem ser definidos pela menor distância entre pares de vértices

Em um grafo sem arestas ponderados o caminho é definido pelo número de arestas de um ponto a outro

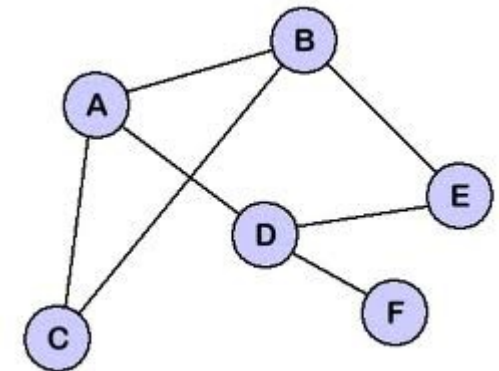
Exemplo:

A - F = (A-C-B-E-D-F) (5)
(A-B-E-D-F) (4)
(A-D-F) (2)

Um grafo pode ter vários caminhos mais curtos

Exemplo:

A - E = (A-B-E) (2)
(A-D-E) (2)



Métricas em Redes Complexas

Closeness - Proximidade:

É uma medida definida pelo comprimento de seus caminhos mais curtos

Define o quanto cada vértice está próximo dos demais

Quanto mais central é o vértice menor é a distância do seu total para todos os outros vértices

Exemplo:

$F = 11$

1-D

2-E

2-A

3-B

3-C

$D = 7$

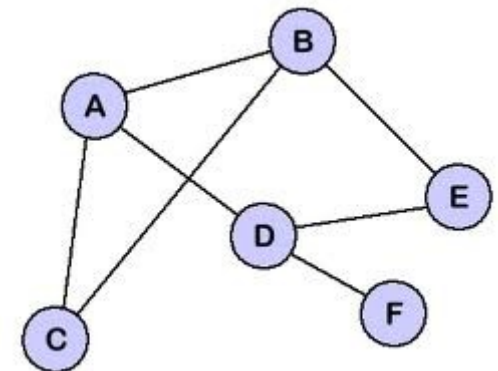
1-F

1-E

1-A

2-B

2-C



Métricas em Redes Complexas

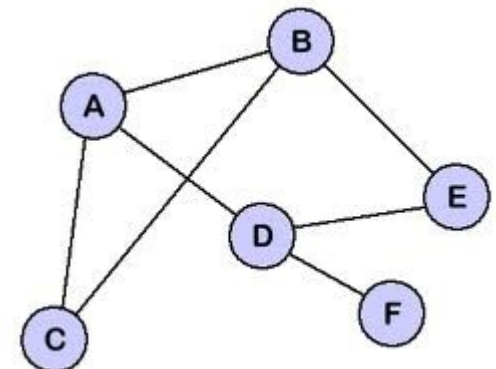
Betweenness – Intermediação (Vértice):

É uma medida que quantifica o número de vezes que **um VÉRTICE age como ponte** ao longo do caminho mais curto entre dois outros vértices.

1. Para cada par de vértices calcular os caminhos mais curtos entre eles.
2. Para cada par de vértices determinar a fração de caminhos mais curtos que passam através do vértice em questão.
3. Somar esta fração de todos os pares de vértices.

Exemplo:

D = A – E
A – F
C – F



Métricas em Redes Complexas

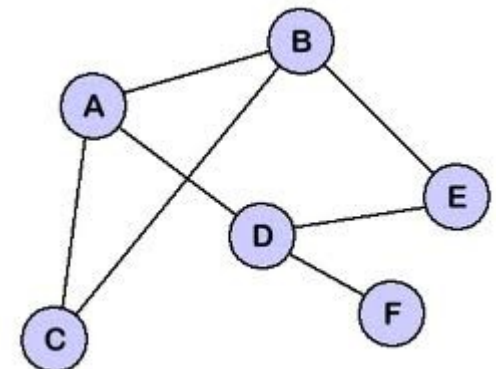
Betweenness – Intermediação (Vértice):

É uma medida que quantifica o número de vezes que **um VÉRTICE age como ponte** ao longo do caminho mais curto entre dois outros vértices.

1. Para cada par de vértices calcular os caminhos mais curtos entre eles.
2. Para cada par de vértices determinar a fração de caminhos mais curtos que passam através do vértice em questão.
3. Somar esta fração de todos os pares de vértices.

Exemplo:

D = A – E
A – F
C – F



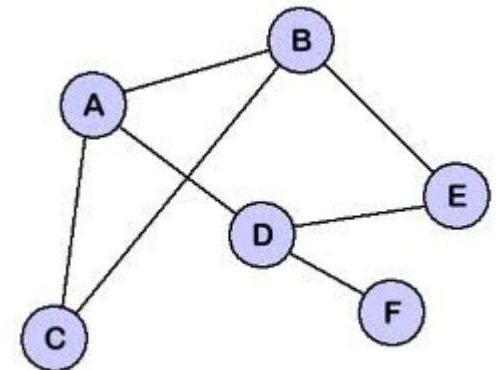
Métricas em Redes Complexas

Betweenness – Intermediação (Aresta):

É uma medida que quantifica o número de vezes que **uma ARESTA** age como **ponte** ao longo do caminho mais curto entre dois outros vértices.

Exemplo:

A - D = A - E
A - F
C - F



O que é o R?

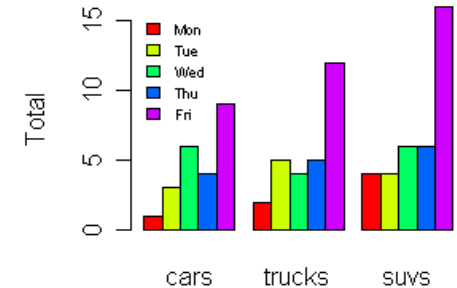
O R é um software livre para análise estatística e gráficos

Possui um repositório de pacotes ou bibliotecas chamado CRAN

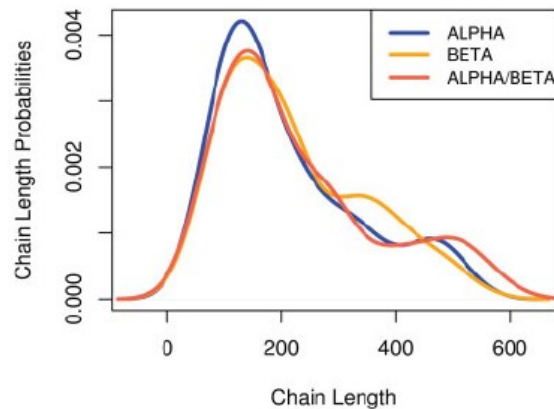
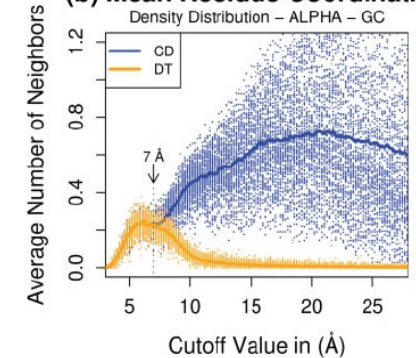
<http://www.r-project.org/>

<http://www.sr.bham.ac.uk/~ajrs/R/r-gallery.html>

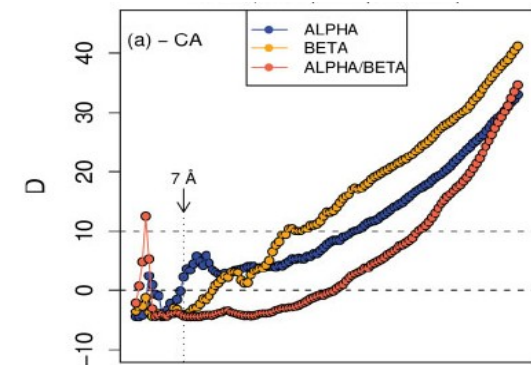
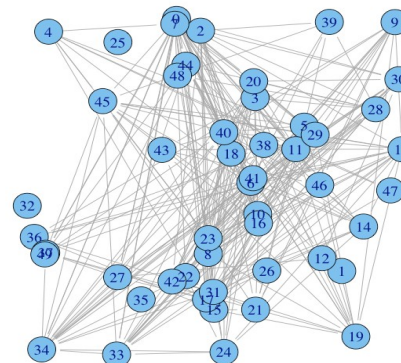
Autos



(b) Mean Residue Coordination



The bare minimum



Usando o R

Windows: <http://www.r-project.org/downloads>

Linux Ubuntu: `sudo apt-get install r-base`
`R`
`install.packages("igraph")`

www.dcc.ufmg.br/~valdete/metricas/G1.net

Manipulando Grafos no R

Formato de entrada: Arquivo no formato Pajek NET

*Vertices 6

1 "A"

2 "B"

3 "C"

4 "D"

5 "E"

6 "F"

*Edges

1 2 1.000

1 3 1.000

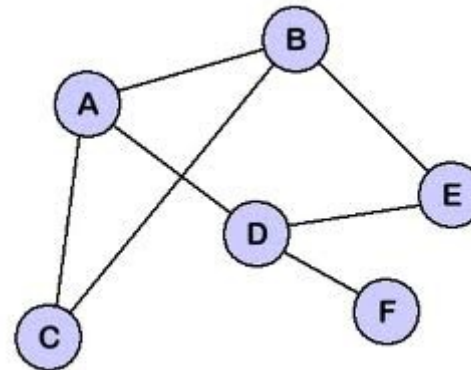
1 4 1.000

2 3 1.000

2 5 1.000

4 5 1.000

4 6 1.000



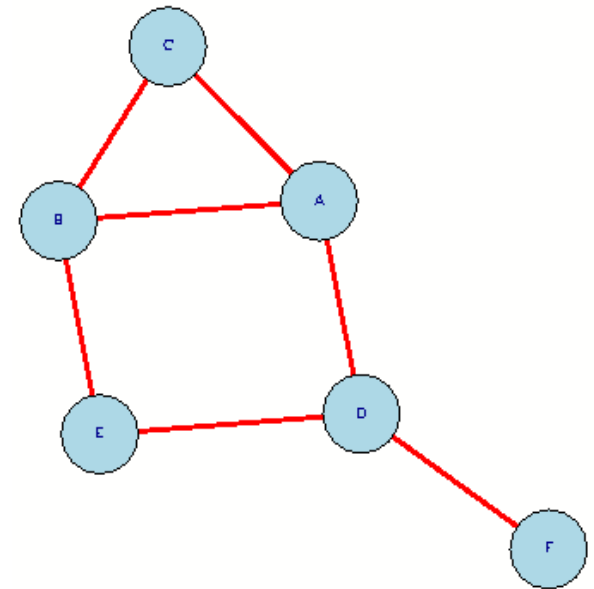
Manipulando Grafos no R

Carregar o Grafo:

```
R  
library(igraph)  
g=read.graph("G1.net", format="pajek")
```

#Plotar o Grafo:

```
tkplot(g)  
tkplot(g, vertex.size=30, vertex.color="lightblue", edge.color="red",  
vertex.label.font=2)  
  
nomes=V(g)$id  
  
tkplot(g, vertex.label=nomes, vertex.size=30, vertex.color="lightblue", edge.width=3,  
edge.color="red")
```



Manipulando Grafos no R

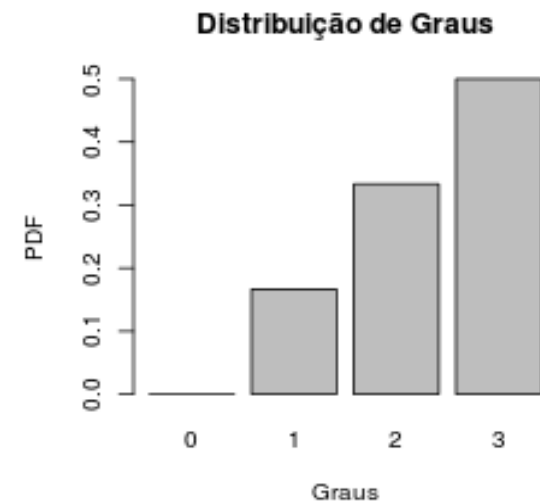
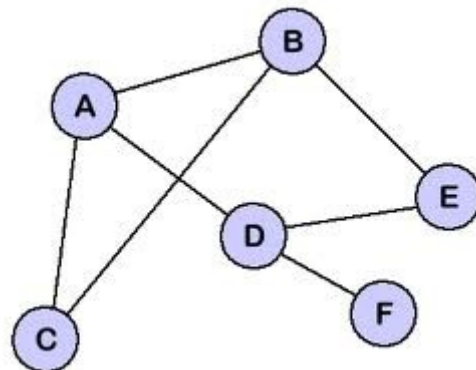
Distribuição de Graus:

```
library(igraph)
g = read.graph("G1.net",format="pajek", directed=FALSE)
d1 = degree.distribution(g, cumulative=FALSE)
len = length(d1)

png(filename="G1_graus.png", height=500, width=500, bg="white")

barplot(d1, main="Distribuição de Graus",xlab="Graus", ylab="PDF",
names.arg=c(0:(len-1)))

dev.off()
```



Manipulando Grafos no R

Closeness Vértices:

```
library(igraph)
g = read.graph("G1.net", format="Pajek")
cl = closeness(g)
cl

cl_max= max(cl)
cl_max

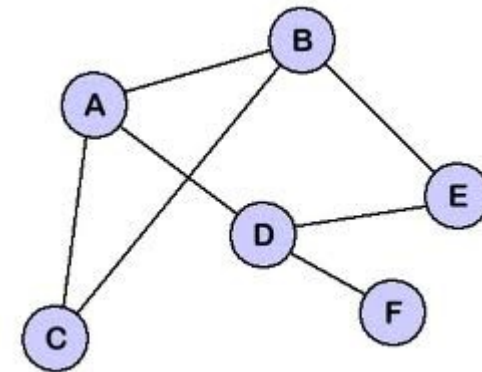
vindex_max = which(closeness(g) == cl_max)
vindex_max

vmax=V(g)[vindex_max]$id
vmax

cl_min= min(cl)
cl_min

vindex_min = which(closeness(g) == cl_min)
vindex_min

vmin=V(g)[vindex_min]$id
vmin
```



Manipulando Grafos no R

Betweenness Vértices:

```
library(igraph)
g = read.graph("G1.net", format="Pajek")
b = betweenness(g)
b

bmax=max(b)
bmax

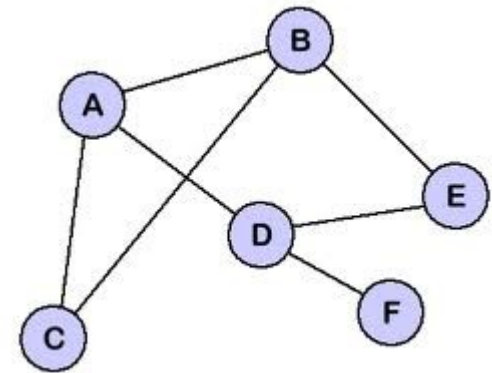
bindex_max = which(betweenness(g) == bmax)
bindex_max

bmax=V(g)[bindex_max]$id
bmax

bmin= min(b)
bmin

bindex_min = which(betweenness(g) == bmin)
bindex_min

bmin=V(g)[bindex_min]$id
bmin
```



Manipulando Grafos no R

Betweenness Arestas:

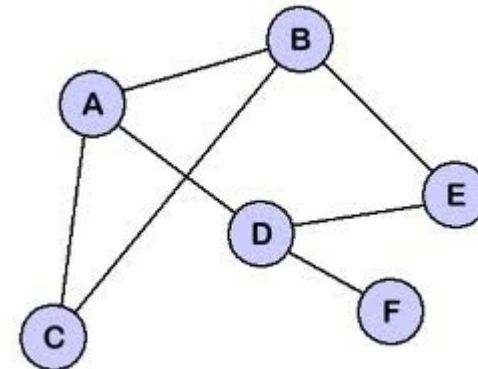
`vcount(g)`

`ecount(g)`

`E(g)`

`eb = edge.betweenness(g)`

`eb`



Referências

Site do R: <http://www.r-project.org/>

Informações sobre R(Igraph): <http://igraph.sourceforge.net/>

<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

D E V Pires, R C Melo-Minardi, , M A Santos, C H da Silveira, M M Santoro, and W Meira Junior. Cutoff Scanning Matrix (CSM): structural classification and function prediction by protein inter-residue distance patterns. BMC Genomics, 12 Suppl 4:S12, 2011.

V M Gonçalves-Almeida, D E V Pires, R C de Melo-Minardi, C H da Silveira, W Meira, and M M Santoro. Hydropace: understanding and predicting cross-inhibition in serine proteases through hydrophobic patch centroids. Bioinformatics, 28(3):342–9, February 2012.