

Analytics e Inteligência Artificial Data Science

Tema da aula
Análise de Redes Sociais



BUSINESS SCHOOL

Graduação, pós-graduação,
MBA, Pós- MBA, Mestrado
Profissional, Curso In
Company e EAD



CONSULTING

Consultoria personalizada
que oferece soluções
baseadas em seu
problema de negócio



RESEARCH

Atualização dos
conhecimentos e do material
didático oferecidos nas
atividades de ensino



Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais **escolas de negócio do mundo**, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 **projetos de consultorias** em organizações públicas e privadas.



Único curso de graduação em administração a receber as notas máximas



A primeira escola brasileira a ser finalista da maior competição de MBA do mundo



Única *Business School* brasileira a figurar no *ranking* LATAM



Signatária do Pacto Global da ONU



Membro fundador da ANAMBA - Associação Nacional MBAs



Credenciada pela AMBA - Association of MBAs



Credenciada ao Executive MBA Council



Filiada a AACSB - Association to Advance Collegiate Schools of Business



Filiada a EFMD - European Foundation for Management Development



Referência em cursos de MBA nas principais mídias de circulação



O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de **Big Data, Analytics** e **Inteligência Artificial**.



Profª Drª Alessandra Montini

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil. Os diretores foram professores de grandes especialistas do mercado.

- +10 anos de atuação.
- +9.000 alunos formados.

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria;
- Larga experiência de mercado na resolução de *cases*;
- Participação em congressos nacionais e internacionais;
- Professor assistente que acompanha o aluno durante todo o curso.

Estrutura

- 100% das aulas realizadas em laboratórios;
- Computadores para uso individual durante as aulas;
- 5 laboratórios de alta qualidade (investimento +R\$2MM);
- 2 unidades próximas à estação de metrô (com estacionamento).



PROFA. DRA. ALESSANDRA DE ÁVILA MONTINI

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Tem muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e Inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em Estatística Aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Parecerista da FAPESP e colunista de grandes portais de tecnologia.





PROF. ÂNGELO CHIODE, MSc

Bacharel, mestre e candidato ao PhD em Estatística (IME-USP), atua como professor de Estatística Aplicada para turmas de especialização, pós-graduação e MBA na FIA. Trabalha como consultor nas áreas de Analytics e Ciência de Dados há 13 anos, apoiando empresas na resolução de desafios de negócio nos contextos de finanças, aquisição, seguros, varejo, tecnologia, aviação, telecomunicações, entretenimento e saúde. Nos últimos 5 anos, tem atuado na gestão corporativa de times de Analytics, conduzindo projetos que envolviam análise estatística, modelagem preditiva e *machine learning*. É especializado em técnicas de visualização de dados e design da informação (Harvard) e foi indicado ao prêmio de Profissional do Ano na categoria Business Intelligence, em 2019, pela Associação Brasileira de Agentes Digitais (ABRADi).



Conteúdo Programático

6



DISCIPLINAS



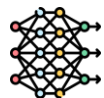
**IA E TRANSFORMAÇÃO
DIGITAL**



ANALYTICS



**INTELIGÊNCIA ARTIFICIAL:
MACHINE LEARNING**



**INTELIGÊNCIA ARTIFICIAL:
DEEP LEARNING**



**EMPREENDEDORISMO E
INOVAÇÃO**



**COMPORTAMENTO
HUMANO E SOFT SKILLS**

TEMAS: ANALYTICS E MACHINE LEARNING

ANÁLISE EXPLORATÓRIA DE DADOS

INFERÊNCIA ESTATÍSTICA

TÉCNICAS DE PROJEÇÃO

TÉCNICAS DE CLASSIFICAÇÃO

TÓPICOS DE MODELAGEM

TÉCNICAS DE SEGMENTAÇÃO

TÓPICOS DE ANALYTICS

MANIPULAÇÃO DE BASE DE DADOS

AUTO ML

TEMAS: DEEP LEARNING

REDES DENSAS

REDES CONVOLUCIONAIS

REDES RECORRENTES

MODELOS GENERATIVOS

FERRAMENTAS

LINGUAGEM R

LINGUAGEM PYTHON

DATABRICKS



Conteúdo da Aula

- 1. Introdução
- 2. Objetivo
- 3. Definições Gerais
- 4. Tipos de Redes
- 5. Métodos de Análise de Redes
- Referências Bibliográficas



1. Introdução



Case: Vendas Especializadas

1. INTRODUÇÃO | ANÁLISE DE REDES SOCIAIS

9

Exemplo:

Analisar as interações entre membros de um time de vendas especializado (coparticipação em negociações, compartilhamento de leads etc.), a fim de identificar os agentes mais colaborativos e engajados.

Aplicação:

Área comercial



Case: Relações Corporativas

1. INTRODUÇÃO | ANÁLISE DE REDES SOCIAIS

10

Exemplo:

Identificar funcionários com diferentes graus de influência e capacidade de disseminação de opiniões dentro de uma corporação, a fim de compreender as dinâmicas de relacionamento e descobrir potenciais novas lideranças.

Aplicação:

Gestão e RH



Case: Epidemiologia

1. INTRODUÇÃO | ANÁLISE DE REDES SOCIAIS

11

Exemplo:

Analisar fluxos de movimentação populacional entre regiões para mitigar a transmissão de doenças infecciosas, intensificando precauções em áreas com alta concentração de pessoas ou de alta intermediação.

Aplicação:

Saúde pública



2. Objetivo



Objetivo

2. OBJETIVO | ANÁLISE DE REDES SOCIAIS

A **análise de redes sociais** (*social network analysis*, ou SNA) é um conjunto de métodos estatísticos destinados a compreender estruturas representadas por meio de **redes/grafos**, permitindo identificar e interpretar as relações entre os seus elementos (indivíduos, entidades etc.).

A origem do SNA remonta ao uso de **sociogramas** na década de 1930, utilizados para visualizar a estrutura das interações sociais. Com o advento computacional e aumento da capacidade de processamento de dados, a técnica evoluiu e tornou-se útil em áreas como marketing, ciência política e epidemiologia.



Exemplo de sociograma

Nesta aula, vamos estudar alguns dos principais métodos para análise estatística de dados em formato de redes sociais.



3. Definições Gerais



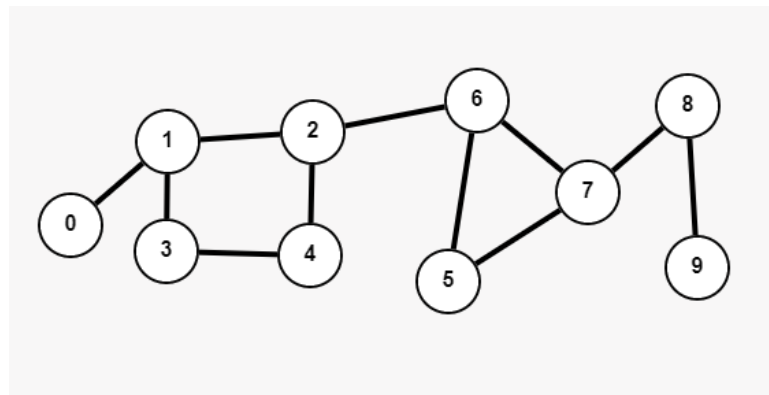
Grafos

3. DEFINIÇÕES GERAIS | ANÁLISE DE REDES SOCIAIS

Um grafo é uma estrutura matemática utilizada para representar um conjunto de **objetos** e as eventuais **relações/conexões** entre eles.

Um grafo consiste em dois componentes fundamentais:

- ✓ **nós** (ou **vértices**), que representam os elementos;
- ✓ **arestas** (ou **ligações**), que representam as conexões entre esses elementos.



Exemplo de grafo

Em SNA, os nós podem representar **pessoas, organizações, regiões** etc. Já as arestas representam relações, interações ou colaborações.

Créditos da imagem: https://pt.wikipedia.org/wiki/Ciclo_%28teoria_de_grafos%29

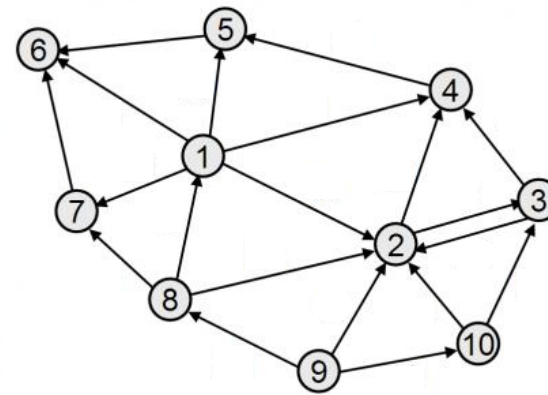
Grafos Direcionados e Não Direcionados

3. DEFINIÇÕES GERAIS | ANÁLISE DE REDES SOCIAIS

16

Em um grafo **não direcionado**, cada aresta representa uma ligação mútua entre os nós envolvidos, e é representada como um segmento de reta. O exemplo do slide anterior corresponde a este caso.

Já em um grafo **direcionado**, cada aresta denota uma **direção**, ou seja, representa uma conexão que possui um nó de partida e um nó de destino. Utiliza-se **setas** para representar tais ligações.



Exemplo de grafo direcionado

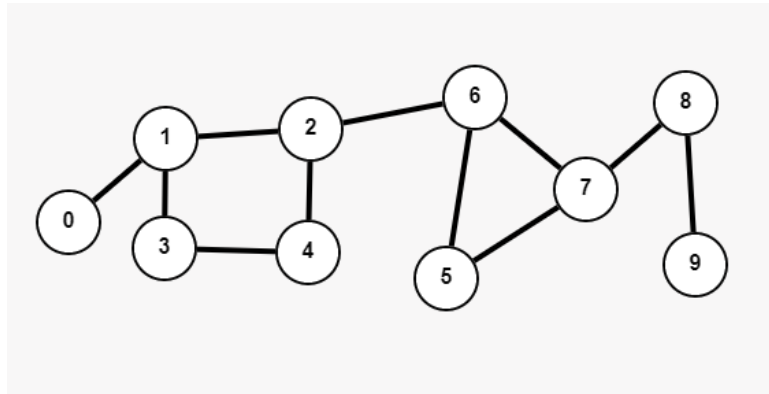
Também é possível que a ligação seja **bidirecional**, que poderia ser representada por uma seta que aponta para os dois lados.

Créditos da imagem: https://www.ime.usp.br/~pf/algoritmos_para_grafos/aulas/weightedgraphs.html

Grau de um Nó

3. DEFINIÇÕES GERAIS | ANÁLISE DE REDES SOCIAIS

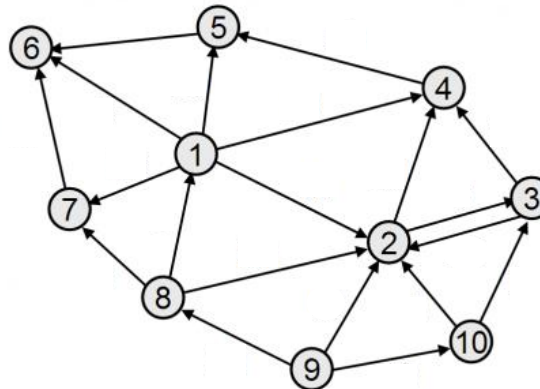
O **grau** de um nó indica quantas conexões ele possui.



Qual o grau de cada nó deste grafo?

- Nós 0 e 9: **Grau 1**
- Nós 3, 4, 5 e 8: **Grau 2**
- Nós 1, 2, 6 e 7: **Grau 3**

Em redes direcionadas, é possível distinguir entre **grau de entrada** (quantas conexões chegam ao nó) e **grau de saída** (quantas conexões partem do nó).



Quais os graus de entrada e saída de cada nó deste grafo?

Exemplos:

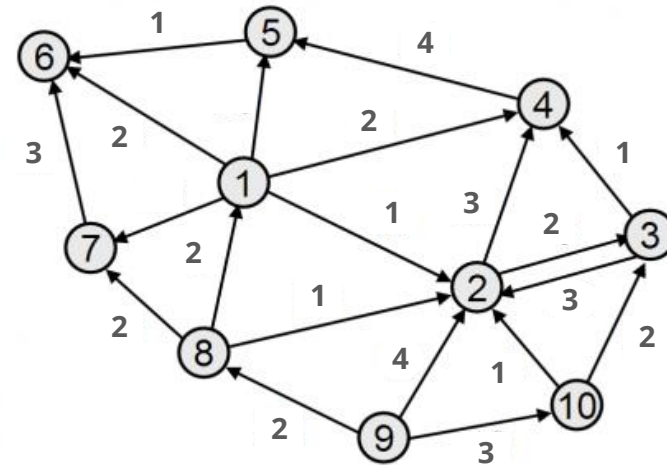
- Nó 1: **Grau de entrada 1**
Grau de saída 5
- Nó 2: **Grau de entrada 5**
Grau de saída 2

Força de uma Aresta

3. DEFINIÇÕES GERAIS | ANÁLISE DE REDES SOCIAIS

18

A **força** de uma aresta representa o **peso** da conexão existente. Trata-se de uma forma opcional de medir diferentes intensidades de conexão exercida entre os nós.



Exemplo de grafo com arestas de diferentes forças

Exemplos de interpretação:

- *Mídias digitais:* O **usuário 8** interagiu com 2 postagens do **usuário 1**.
- *Logística:* O **armazém 5** recebeu 4 cargas vindas do **armazém 4**.
- *Produção científica:* O **autor 6** foi citado por 3 autores, e não citou ninguém.

4. Tipos de Redes



Matriz de Adjacências

4. TIPOS DE REDES | ANÁLISE DE REDES SOCIAIS

Para distinguir entre os principais tipos de redes, partimos da definição de **matriz de adjacências**.

Matriz de adjacências

	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	1	1	1	1
3	1	1	0	0	1	0
4	0	1	0	0	0	0
5	0	1	1	0	0	0
6	0	1	0	0	0	0

A matriz de adjacências é **quadrada**, ou seja, a sua quantidade de **linhas** é igual à quantidade de **colunas**. Cada elemento da matriz representa a existência ou não de uma conexão (aresta) entre os dois elementos (nós) representados na linha e na coluna respectiva.

Tipos de Redes

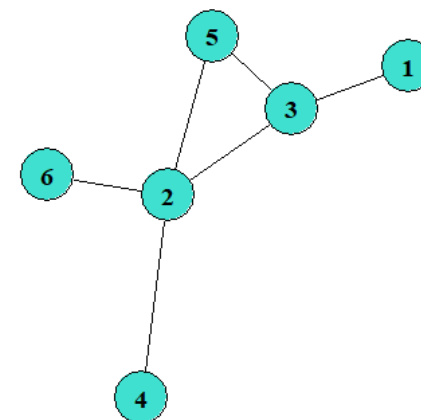
4. TIPOS DE REDES | ANÁLISE DE REDES SOCIAIS

Exemplo 1: Arestas **não direcionadas** e com **mesmo peso**

Matriz de adjacências

	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	1	1	1	1
3	1	1	0	0	1	0
4	0	1	0	0	0	0
5	0	1	1	0	0	0
6	0	1	0	0	0	0

Representação da rede



Neste tipo de rede, a matriz de adjacências contém apenas **valores 1 e 0**, que indicam a presença ou ausência de aresta entre os nós correspondentes, respectivamente.

Além disso, a matriz é **simétrica** em relação à diagonal principal, pois a existência de ligação entre dois nós X e Y **implica** ligação entre Y e X.

Tipos de Redes

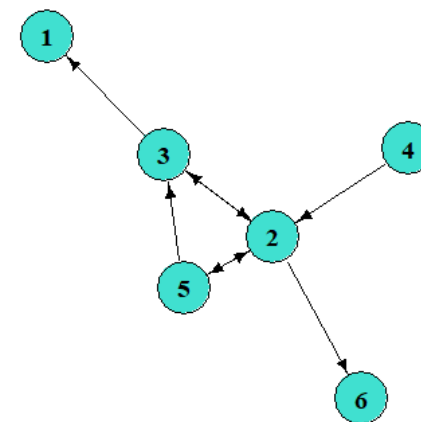
4. TIPOS DE REDES | ANÁLISE DE REDES SOCIAIS

Exemplo 2: Arestas **direcionadas** e com **mesmo peso**

Matriz de adjacências

	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	1	0	1	1
3	1	1	0	0	0	0
4	0	1	0	0	0	0
5	0	1	1	0	0	0
6	0	0	0	0	0	0

Representação da rede



Neste tipo de rede, a matriz de adjacências contém apenas **valores 1 e 0**, que indicam a presença ou ausência de aresta entre os nós correspondentes, respectivamente.

Agora, a matriz pode **não ser simétrica** em relação à diagonal principal, pois a existência de ligação entre dois nós X e Y **não implica** ligação entre Y e X.

Tipos de Redes

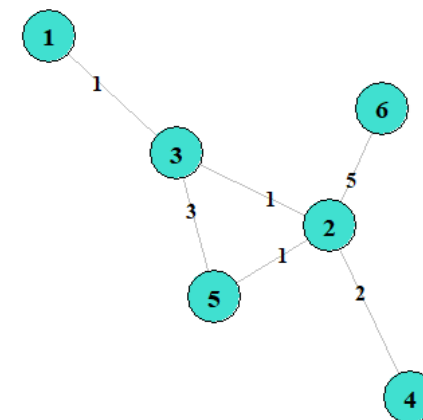
4. TIPOS DE REDES | ANÁLISE DE REDES SOCIAIS

Exemplo 3: Arestas **não direcionadas** e com **pesos diferentes**

Matriz de adjacências

	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	1	2	1	5
3	1	1	0	0	3	0
4	0	2	0	0	0	0
5	0	1	3	0	0	0
6	0	5	0	0	0	0

Representação da rede



Neste tipo de rede, a matriz de adjacências contém valores numéricos que indicam a **força da aresta** que une cada par de nós. Quando não existe aresta de ligação, a força é zero.

Além disso, a matriz é **simétrica** em relação à diagonal principal, pois a força de ligação entre dois nós X e Y é a **mesma força** de ligação entre Y e X.

Tipos de Redes

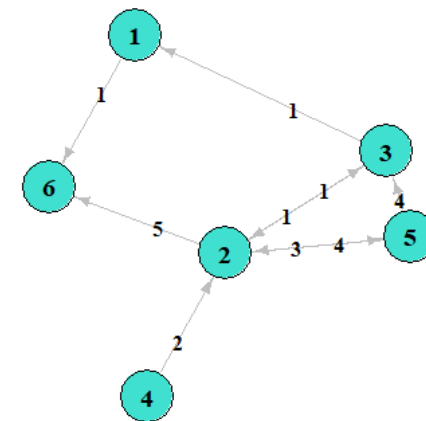
4. TIPOS DE REDES | ANÁLISE DE REDES SOCIAIS

Exemplo 4: Arestas **direcionadas** e com **pesos diferentes**

Matriz de adjacências

	1	2	3	4	5	6
1	0	0	0	0	0	1
2	0	0	1	0	3	5
3	1	1	0	0	0	0
4	0	2	0	0	0	0
5	0	4	4	0	0	0
6	0	0	0	0	0	0

Representação da rede



Neste tipo de rede, a matriz de adjacências contém valores numéricos que indicam a **força da aresta** que une cada par de nós. Quando não existe aresta de ligação, a força é zero.

Agora, a matriz pode **não ser simétrica** em relação à diagonal principal, pois a força de ligação entre dois nós X e Y pode não ser a mesma força de ligação entre Y e X.

5. Métodos de Análise de Redes



Métodos de Análise

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

26

Para analisar estruturas de redes sociais, alguns dos principais métodos são:

- **Análise visual da rede**, por meio de um grafo que representa seus nós e arestas.
- **Análise de coesão** entre os nós da rede, por meio de aspectos como distância média, diâmetro, densidade e transitividade.
- **Identificação de melhores caminhos** entre dois nós da rede, seja em termos de comprimento ou de custo associado.
- **Cálculo de medidas de centralidade**, que avaliam a importância ou influência individual dos nós da rede.
- **Detecção de comunidades**, ou seja, subgrupos de nós mais densamente conectados entre si do que com o restante da rede.



Análise Visual da Rede

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

Para obter a representação visual da rede (grafo) por meio de uma **matriz de adjacências**, pode-se utilizar a função `graph_from_adjacency_matrix` (pacote `igraph` do R).

Também é possível obter a visualização da rede por meio de uma **base de dados** construída em formato específico, usando a função `graph_from_data_frame` do mesmo pacote do R. Esse método é computacionalmente mais adequado para redes com grande quantidade de elementos e poucas conexões.

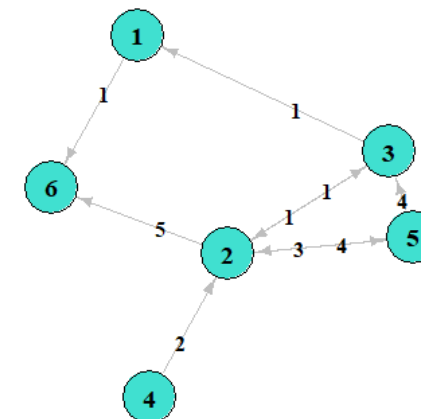
Base de dados

from	to	weight
1	6	1
2	3	1
2	5	3
2	6	5
3	1	1
3	2	1
4	2	2
5	2	4
5	3	4

Matriz de adjacências

	1	2	3	4	5	6
1	0	0	0	0	0	1
2	0	0	1	0	3	5
3	1	1	0	0	0	0
4	0	2	0	0	0	0
5	0	4	4	0	0	0
6	0	0	0	0	0	0

Representação da rede



Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

28

Uma empresa comercializadora de insumos agrícolas possui o interesse de otimizar os custos (R\$) associados às rotas dos vendedores em suas visitas aos clientes, ao longo de 15 municípios de atuação (de A reserve O). Tais custos englobam consumo de combustível, honorário fixo do vendedor, pagamento de pedágios e depreciação estimada do veículo.

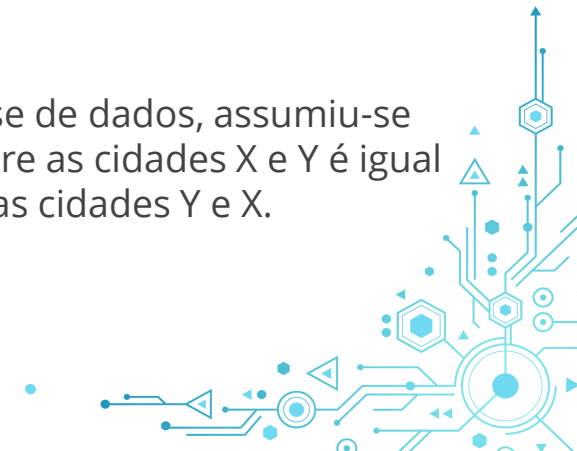


ORIGEM	DESTINO	CUSTO
A	C	14
A	D	14
A	E	35
A	F	25
B	C	29
B	E	17
B	G	37
C	D	22
C	K	16
D	K	42
...

Obs.: Nesta base de dados, assumiu-se que o custo entre as cidades X e Y é igual ao custo entre as cidades Y e X.

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.

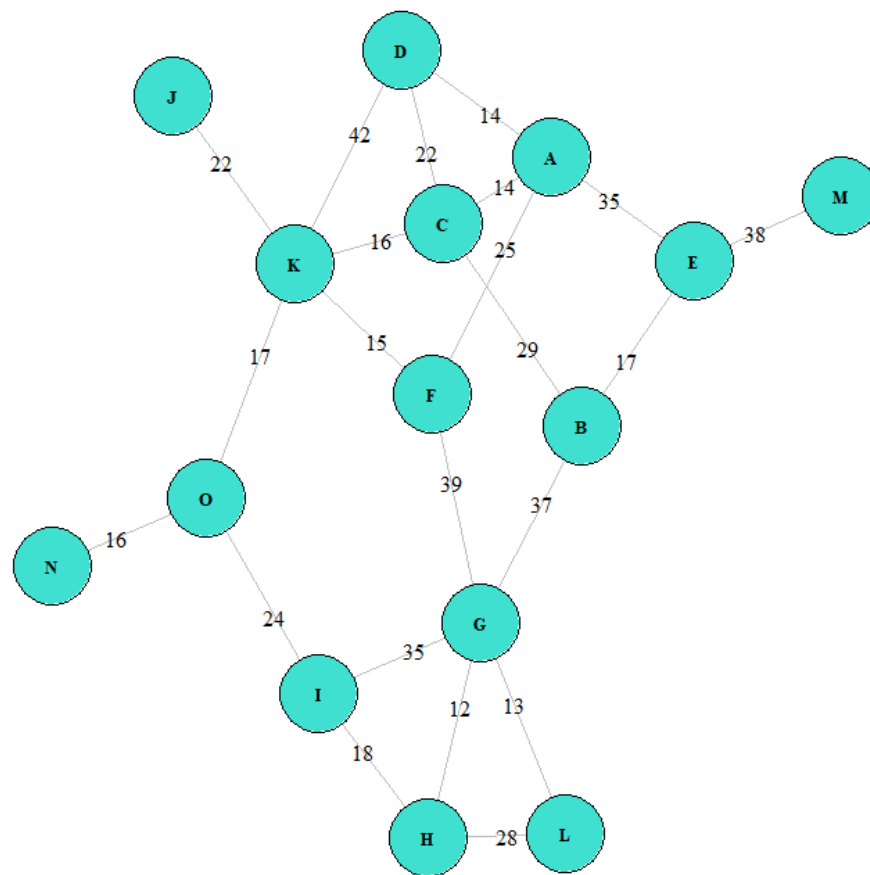


Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

29

Análise visual da rede



De qual tipo é esta rede?

Arestas **não direcionadas** e
com **pesos diferentes**

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.

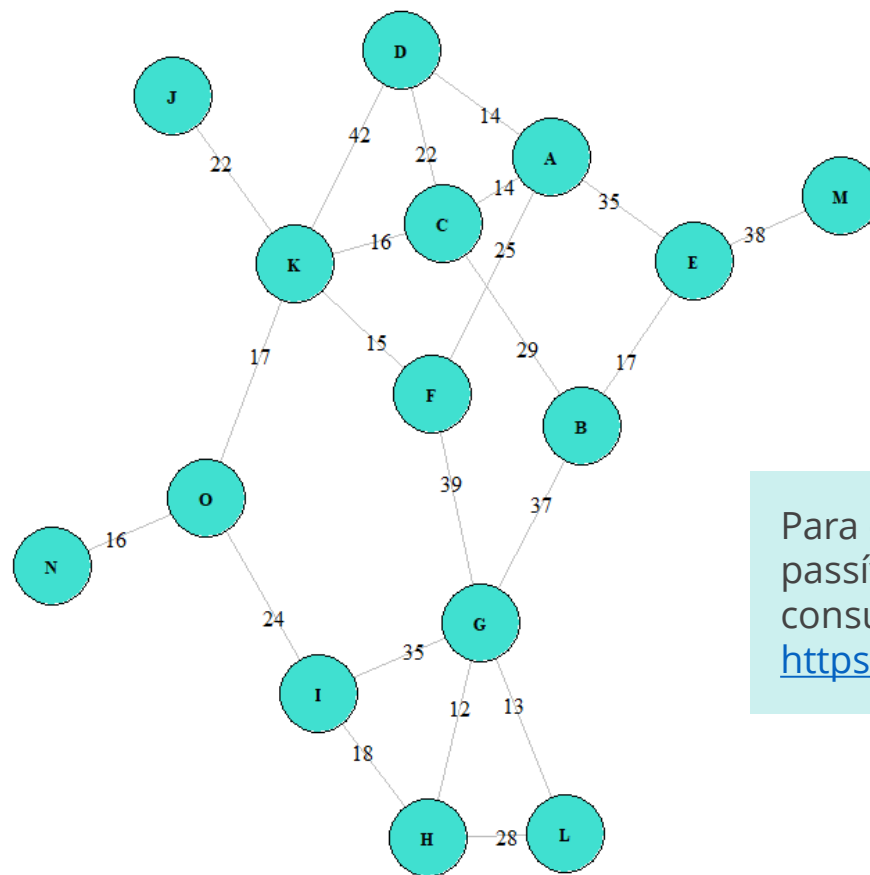


Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

30

Análise visual da rede



Para explorar todos os parâmetros gráficos passíveis de configuração no pacote *igraph*, consulte:

<https://igraph.org/r/doc/plot.common.html>

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



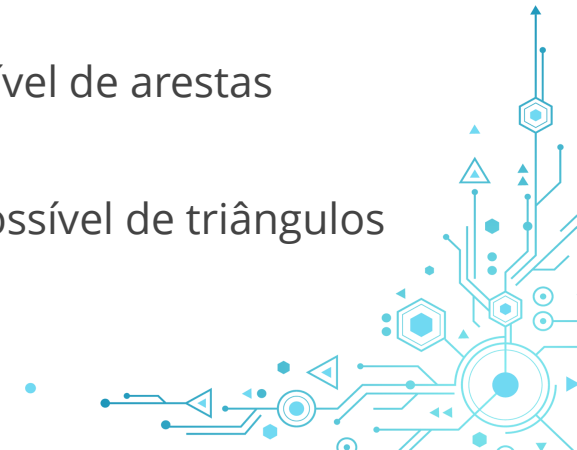
Análise de Coesão

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

A **análise de coesão** consiste numa análise descritiva da rede, por meio de medidas que representam a força e a densidade das conexões dentro de uma rede.

As principais medidas de coesão são:

- ✓ **Distância geodésica média**
Média das menores distâncias entre todos os pares de nós
- ✓ **Custo médio**
Médias dos custos associados às distâncias geodésicas entre todos os pares de nós
- ✓ **Diâmetro**
Maior distância geodésica entre todos os pares de nós
- ✓ **Custo diametral**
Maior custo associado às distâncias geodésicas entre todos os pares de nós
- ✓ **Densidade**
Proporção de arestas presentes, em relação ao total possível de arestas
- ✓ **Transitividade** (ou **triangulação**)
Proporção de triângulos presentes, em relação ao total possível de triângulos



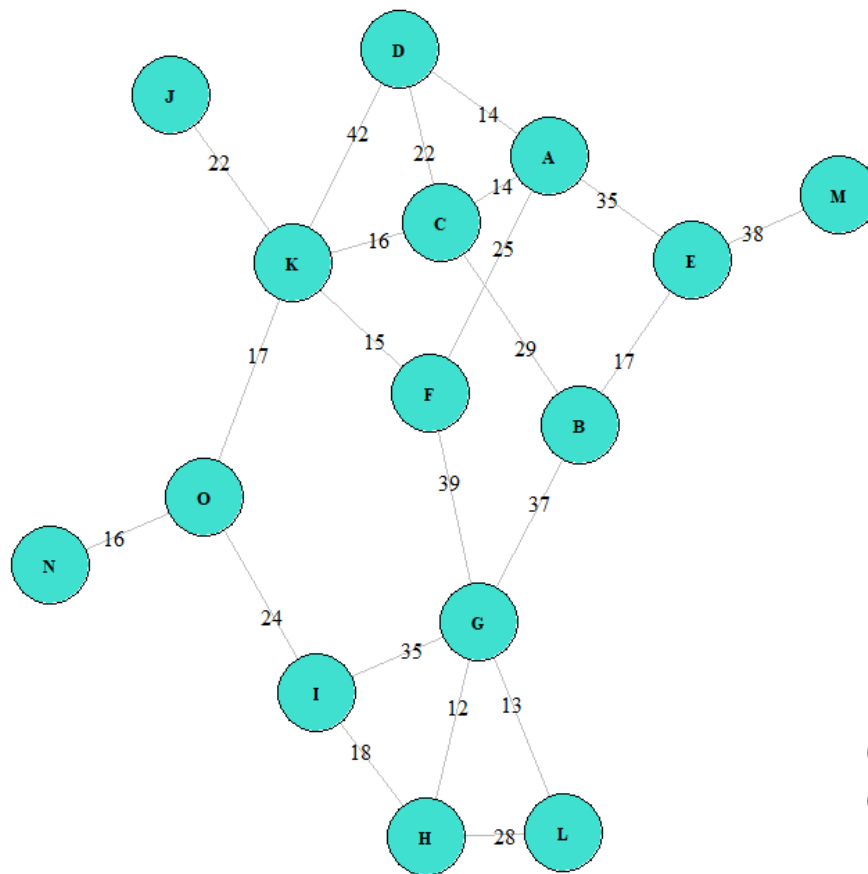
Análise de coesão

Distância geodésica média

Para ir de uma cidade a outra dentro da rede de atuação, por quantas cidades o vendedor passa, em média? (contando a de destino)

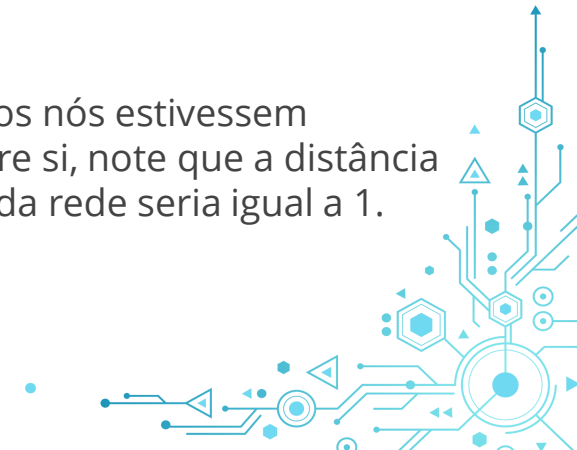
Resposta: 2,5 cidades

Função em R: *mean_distance*



Obs.: Se todos os nós estivessem conectados entre si, note que a distância média dos nós da rede seria igual a 1.

Arquivo: Otimizacao_Rotas.txt



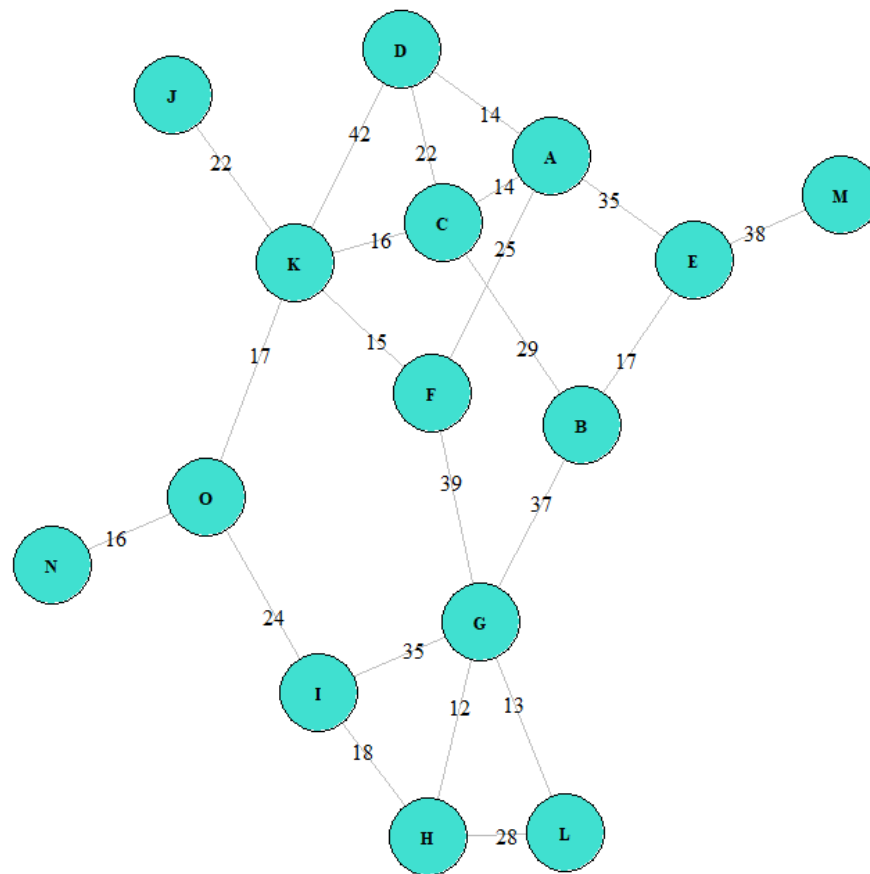
Análise de coesão

Custo médio

Para ir de uma cidade a outra dentro da rede de atuação, qual é o custo médio (R\$) da viagem para a empresa?

Resposta: R\$ 57,02

Função em R: `mean_distance(weights = ...)`



Arquivo: Otimizacao_Rotas.txt



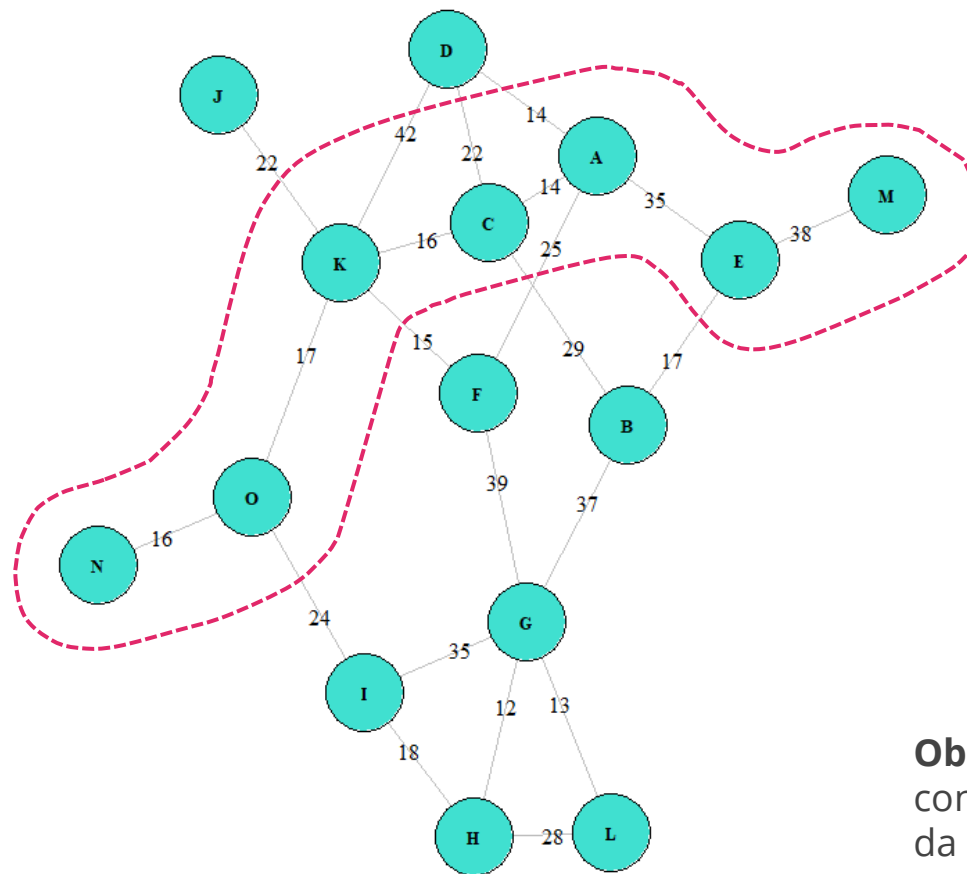
Análise de coesão

Diâmetro

Qual é a maior quantidade de cidades pelas quais o vendedor precisa passar para chegar de uma cidade a outra, dentro da rede de atuação? (contando a de destino)

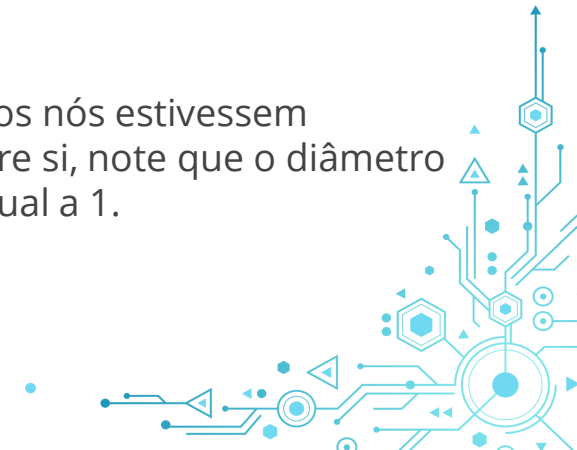
Resposta: 6 cidades (de N a M)

Função em R: *diameter*



Obs.: Se todos os nós estivessem conectados entre si, note que o diâmetro da rede seria igual a 1.

Arquivo: Otimizacao_Rotas.txt



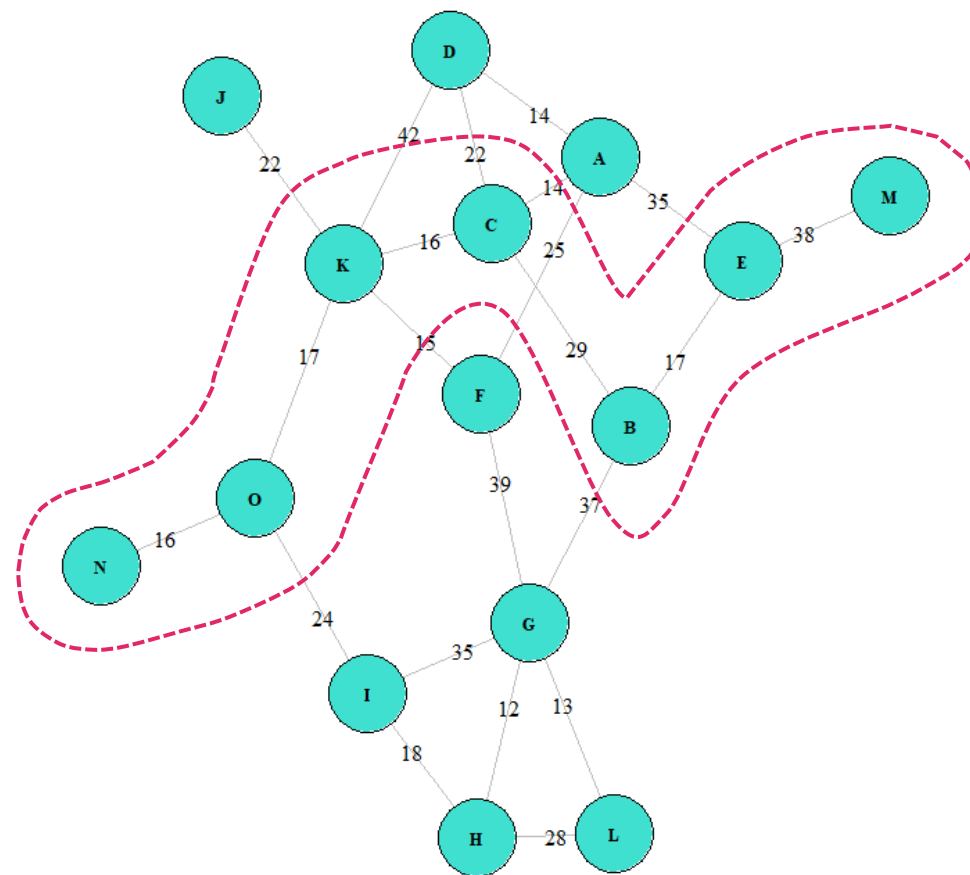
Análise de coesão

Custo diametral

Entre as cidades N e M, cuja viagem é a mais longa em termos de quantidade de cidades, qual é o menor custo (R\$) possível que a empresa pode ter?

Resposta: R\$ 133

Função em R: `diameter(weights = ...)`



Arquivo: Otimizacao_Rotas.txt



Análise de coesão

Densidade

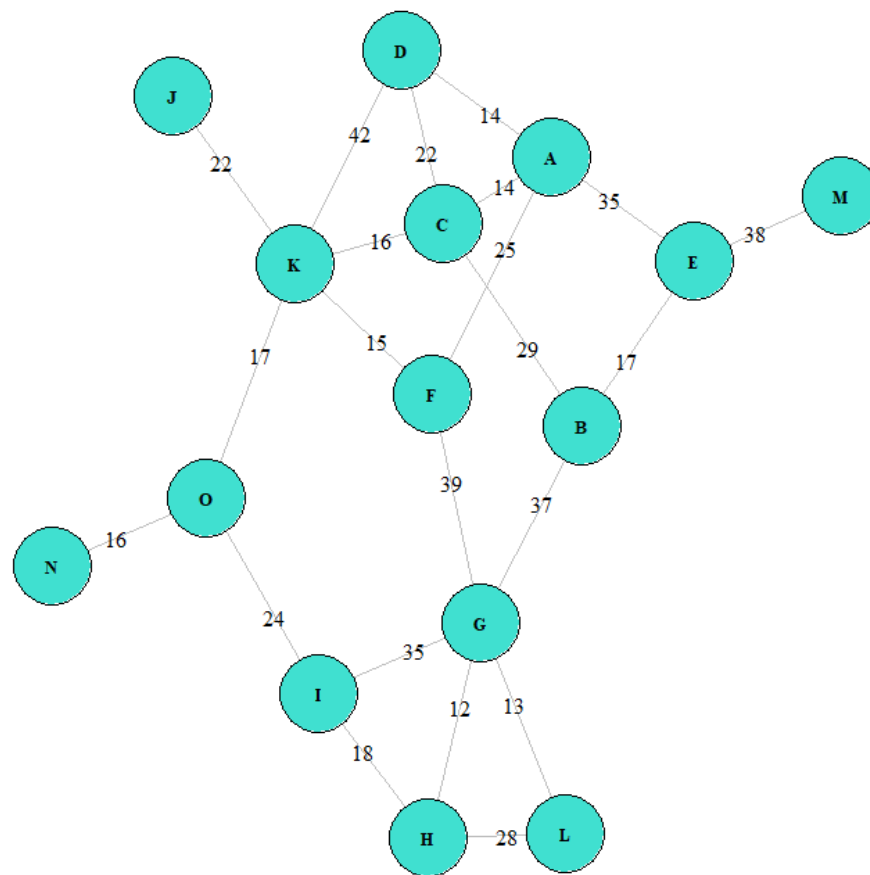
% de arestas = **0,21 = 21%**

Valor	Densidade
0% a 10%	Muito baixa
10% a 30%	Baixa
30% a 50%	Moderada
50% a 70%	Alta
70% a 100%	Muito alta

Função em R: *edge_density*

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



Análise de coesão

Transitividade

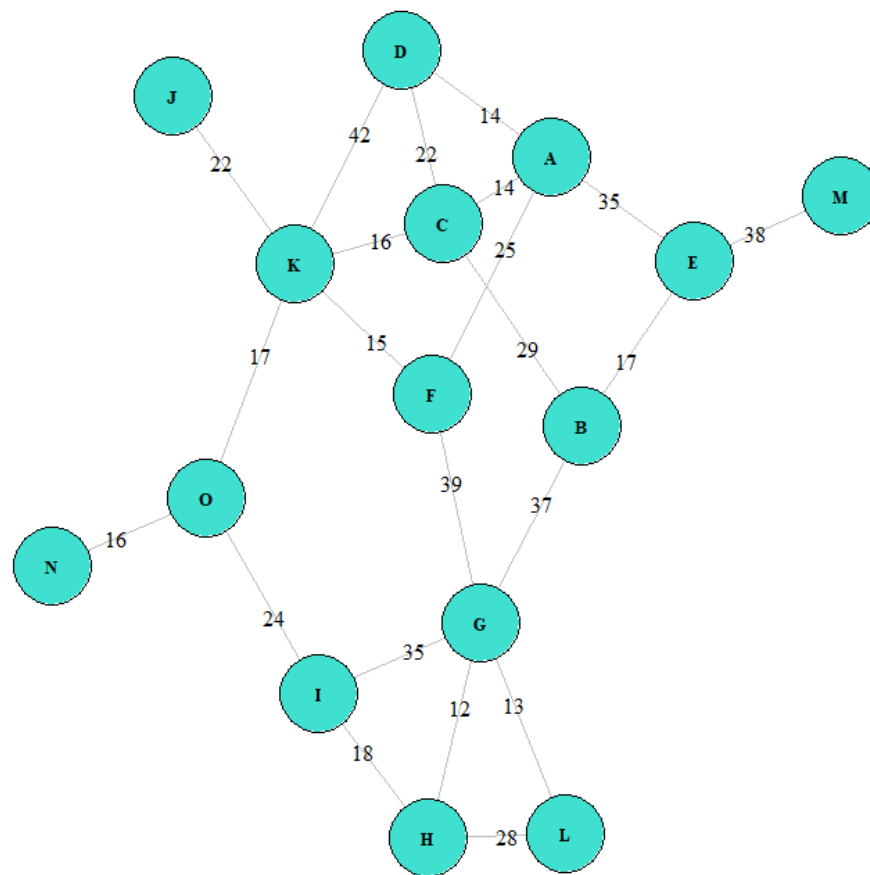
% de triângulos = **0,22 = 22%**

Valor	Transitividade
0% a 10%	Muito baixa
10% a 30%	Baixa
30% a 50%	Moderada
50% a 70%	Alta
70% a 100%	Muito alta

Função em R: `transitivity(..., type = "global")`

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.

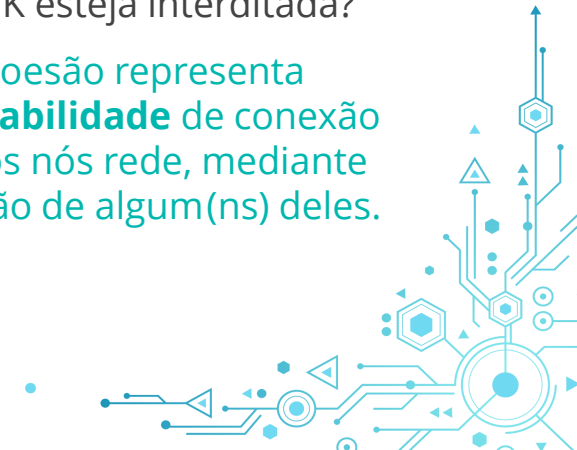


Qual a interpretação de um **baixo grau** de coesão?

Qual o impacto na conectividade entre as cidades H e F, caso a cidade G esteja interditada?

E entre as cidades O e J, caso a cidade K esteja interditada?

Baixa coesão representa **vulnerabilidade** de conexão entre os nós rede, mediante remoção de algum(ns) deles.



Identificação de Melhores Caminhos

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

Identificar visualmente o **melhor caminho** (mais curto, ou menos custoso) de conexão entre dois nós pode não ser um trabalho simples e rápido, especialmente em redes mais extensas.

Por meio da função *distances* (pacote *igraph* do R), é possível obter uma matriz que contém as **distâncias geodésicas** (menores comprimentos de caminho) entre dois nós, considerando as ligações existentes entre eles.

	A	B	C	D	E	F	G	I	J	K	L	N	M	H	O
A	0	2	1	1	1	1	2	3	3	2	3	4	2	3	3
B	2	0	1	2	1	2	1	2	3	2	2	4	2	2	3
C	1	1	0	1	2	2	2	3	2	1	3	3	3	3	2
D	1	2	1	0	2	2	3	3	2	1	4	3	3	4	2
E	1	1	2	2	0	2	2	3	4	3	3	5	1	3	4
F	1	2	2	2	2	0	1	2	2	1	2	3	3	2	2
G	2	1	2	3	2	1	0	1	3	2	1	3	3	1	2
I	3	2	3	3	3	2	1	0	3	2	2	2	4	1	1
J	3	3	2	2	4	2	3	3	0	1	4	3	5	4	2
K	2	2	1	1	3	1	2	2	1	0	3	2	4	3	1
L	3	2	3	4	3	2	1	2	4	3	0	4	4	1	3
N	4	4	3	3	5	3	3	2	3	2	4	0	6	3	1
M	2	2	3	3	1	3	3	4	5	4	4	6	0	4	5
H	3	2	3	4	3	2	1	1	4	3	1	3	4	0	2
O	3	3	2	2	4	2	2	1	2	1	3	1	5	2	0

Em destaque, o diâmetro da rede.

Identificação de Melhores Caminhos

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

A função *distances* também fornece a matriz com os **menores custos de caminho** entre dois nós, considerando as ligações existentes entre eles.

	A	B	C	D	E	F	G	I	J	K	L	N	M	H	O
A	0	43	14	14	35	25	64	71	52	30	77	63	73	76	47
B	43	0	29	51	17	60	37	67	67	45	50	78	55	49	62
C	14	29	0	22	46	31	66	57	38	16	79	49	84	75	33
D	14	51	22	0	49	39	78	79	60	38	91	71	87	90	55
E	35	17	46	49	0	60	54	84	84	62	67	95	38	66	79
F	25	60	31	39	60	0	39	56	37	15	52	48	98	51	32
G	64	37	66	78	54	39	0	30	76	54	13	70	92	12	54
I	71	67	57	79	84	56	30	0	63	41	43	40	122	18	24
J	52	67	38	60	84	37	76	63	0	22	89	55	122	81	39
K	30	45	16	38	62	15	54	41	22	0	67	33	100	59	17
L	77	50	79	91	67	52	13	43	89	67	0	83	105	25	67
N	63	78	49	71	95	48	70	40	55	33	83	0	133	58	16
M	73	55	84	87	38	98	92	122	122	100	105	133	0	104	117
H	76	49	75	90	66	51	12	18	81	59	25	58	104	0	42
O	47	62	33	55	79	32	54	24	39	17	67	16	117	42	0

Em destaque, o custo diametral da rede.

Como vimos antes, o menor custo de viagem entre as cidades N e M é de **R\$ 133**. Este é o maior custo de viagem entre dois nós da rede de atuação, portanto, é o custo diametral.

Identificação de Melhores Caminhos

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

Por meio da função *shortest_paths* (pacote *igraph* do R) é possível obter o detalhamento de qual é o **melhor caminho** entre dois nós, seja em termos de **comprimento** ou de **custo**.

Também é possível obter uma lista de **todos os caminhos** que conectam dois nós, por meio da função *all_simple_paths*.



Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

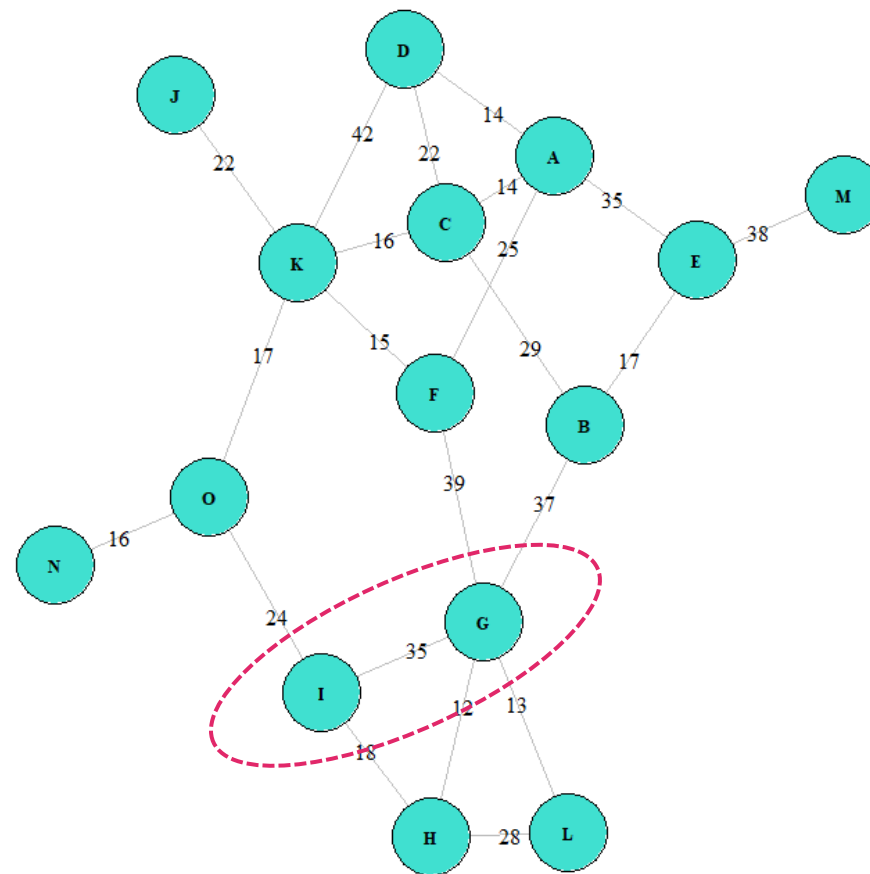
41

Identificação de melhores caminhos

Qual é o melhor caminho entre as cidades I e G, em termos de quantidade de cidades?

Resposta: I → G

Função em R: *shortest_paths*



Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

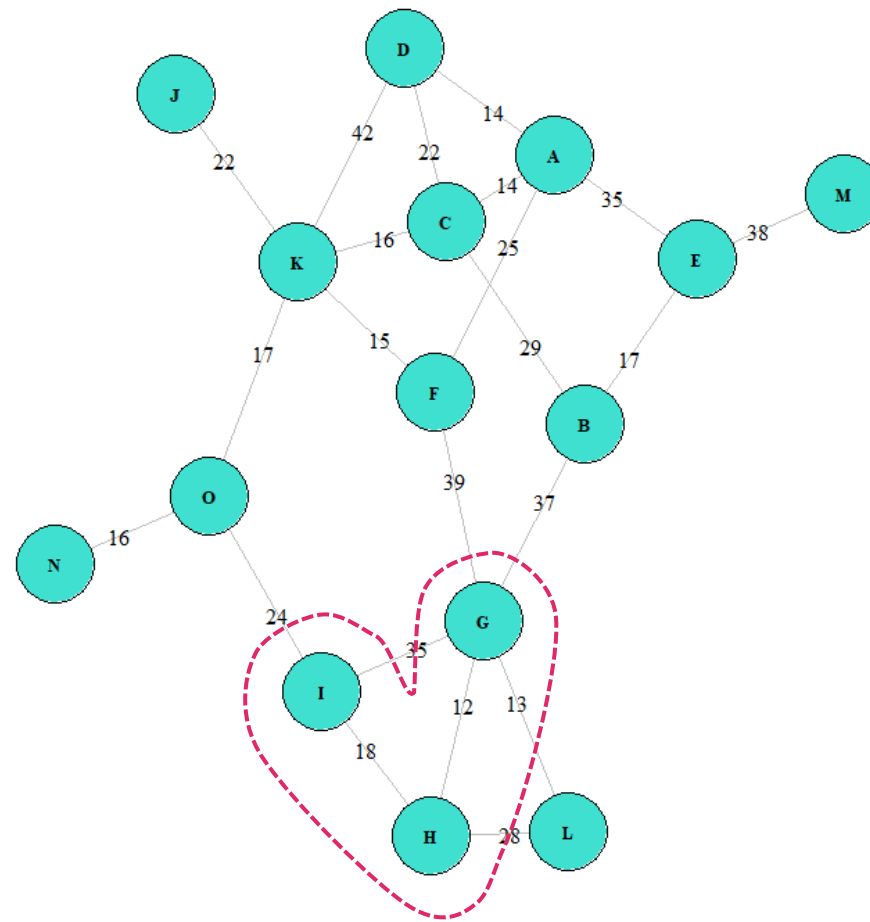
42

Identificação de melhores caminhos

Qual é o melhor caminho entre as cidades I e G, em termos de custo (R\$) para a empresa?

Resposta: I → H → G

Função em R: `shortest_paths(weights = ...)`



Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



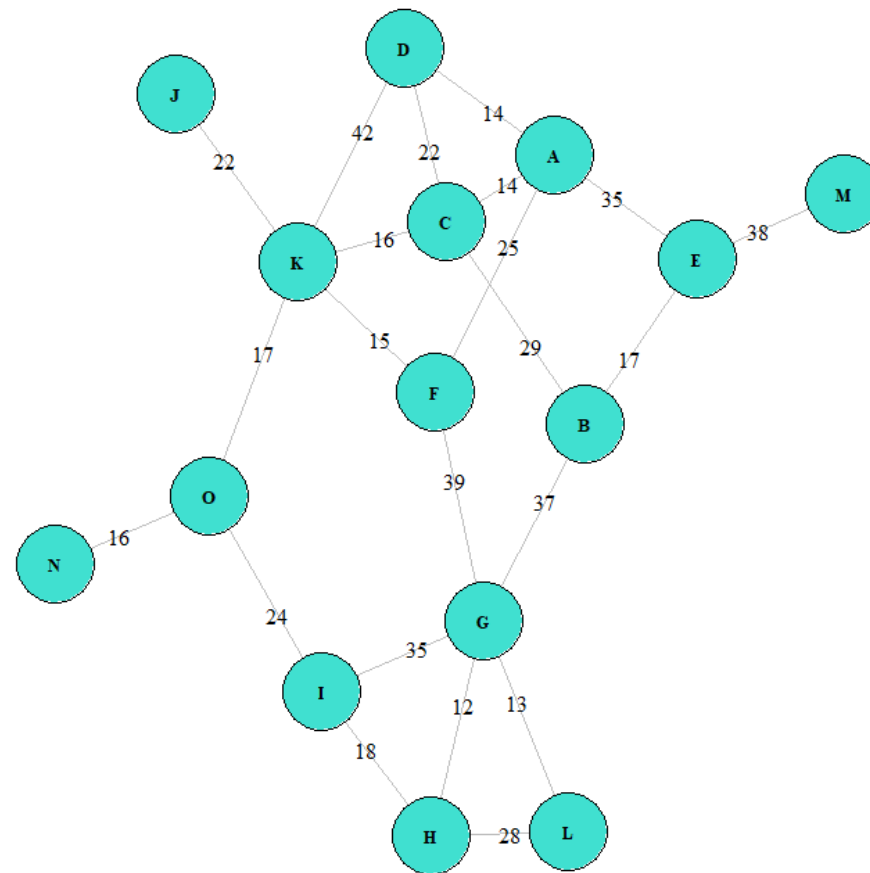
Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

43

Identificação de melhores caminhos

Exercício: Qual é o melhor caminho entre as cidades L e D, em termos de custo (R\$) para a empresa?



Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

44

Identificação de melhores caminhos

Quantos caminhos podem levar da cidade I para a cidade G? Quais são esses caminhos?

Resposta: 20 caminhos

I → G

I → H → G

I → H → L → G

I → O → K → C → A → E → B → G

I → O → K → C → A → F → G

I → O → K → C → B → E → A → F → G

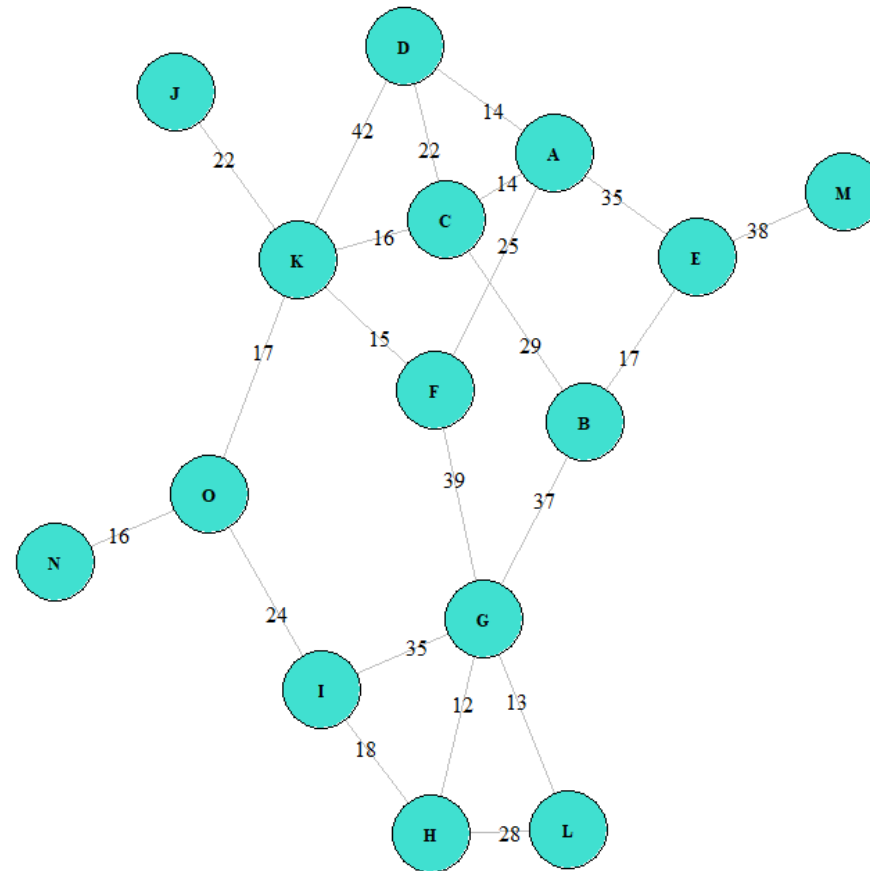
I → O → K → C → B → G

...

Função em R: *all_simple_paths*

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



Case: Otimização de Rotas

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

45

Identificação de melhores caminhos

Quantos caminhos podem levar da cidade I para a cidade G, passando por no máximo 4 cidades? (contando a de destino)

Resposta: 4 caminhos

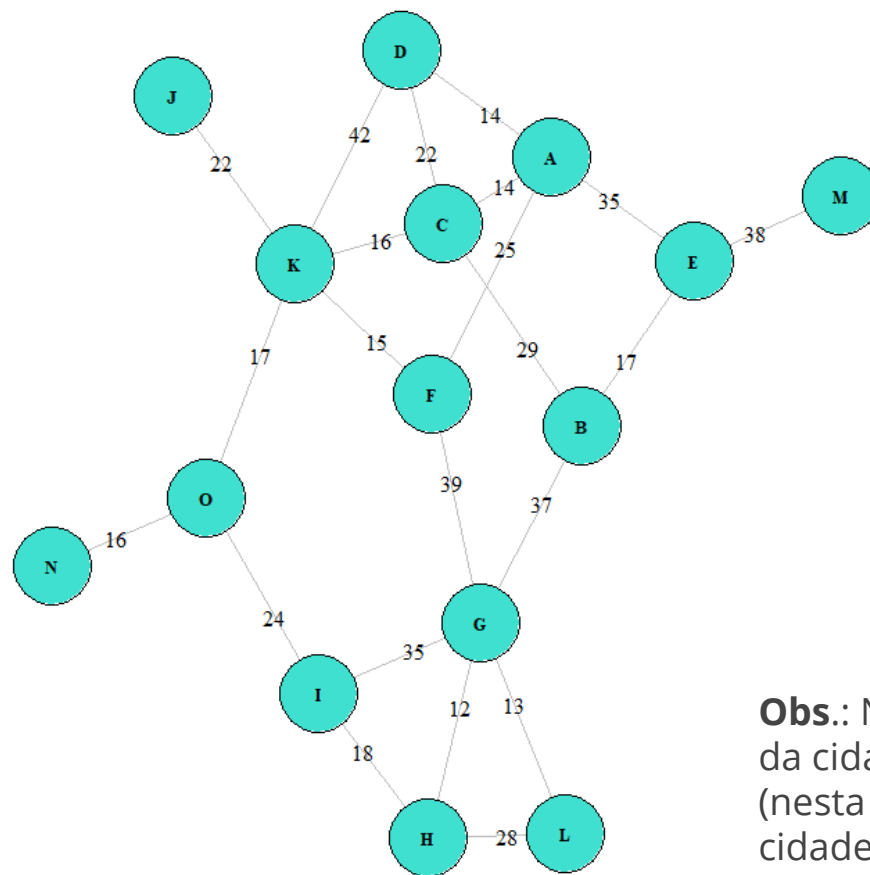
I → G

I → H → G

I → H → L → G

I → O → K → F → G

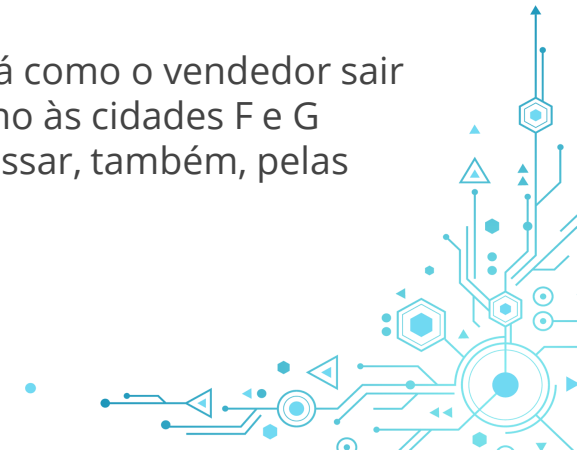
Função em R: `all_simple_paths(..., cutoff = 4)`



Obs.: Note que não há como o vendedor sair da cidade I com destino às cidades F e G (nesta ordem) sem passar, também, pelas cidades O e K.

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



Medidas de Centralidade

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

Podemos descrever os nós de uma rede a partir de **medidas de centralidade**, que ajudam a mensurar a **influência** de cada nó de uma rede, sob as seguintes perspectivas:

- ✓ **Centralidade de grau**

Quantidade de conexões diretas que um nó possui. Denota o grau direto de atividade do nó ou, em alguns contextos sociais, de popularidade do nó.

- ✓ **Centralidade de proximidade**

Inverso da soma das distâncias do nó para todos os outros nós. Denota o quão rapidamente um nó pode alcançar outros nós.

- ✓ **Centralidade de intermediação**

Quantidade de vezes que um nó aparece nos caminhos mais curtos entre outros nós. Denota atuação de nós como “pontes” ou intermediários na rede.



Medidas de centralidade

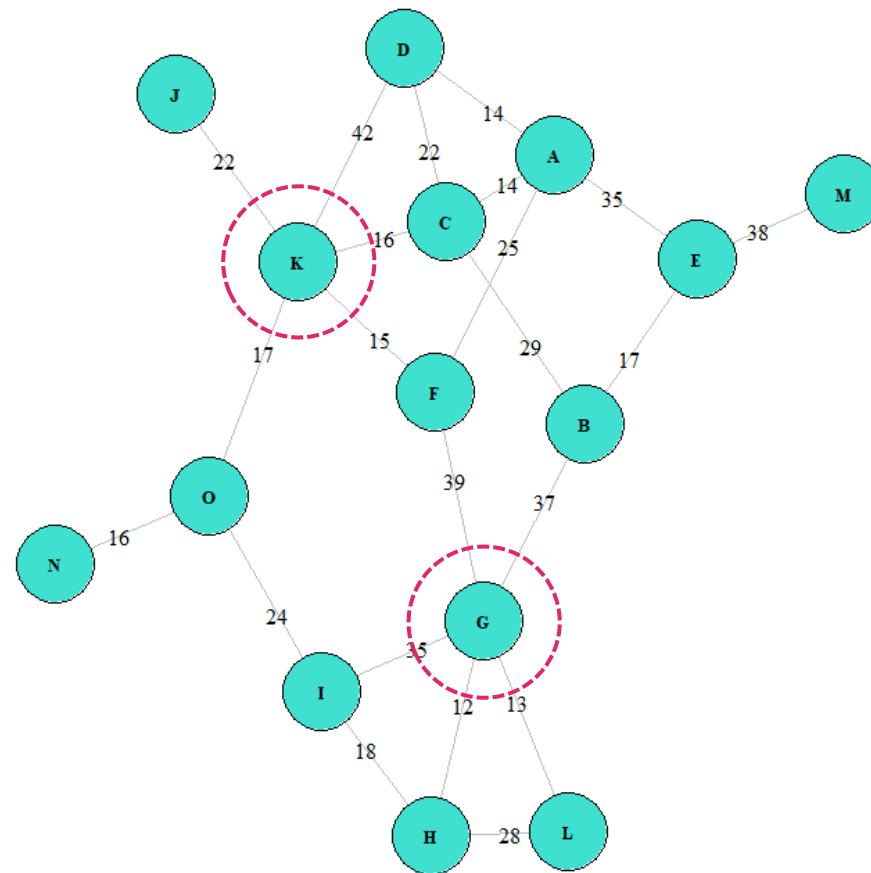
Centralidade de grau

Quais cidades possuem mais conexões com outras cidades na rede de atuação?

Resposta: **G e K**

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
4	3	4	3	3	3	5	3	3	1	5	2	1	1	3

Função em R: *degree*



Arquivo: Otimizacao_Rotas.txt



Medidas de centralidade

Centralidade de proximidade

Quais cidades possuem maior proximidade em relação a todas as demais, concomitantemente?

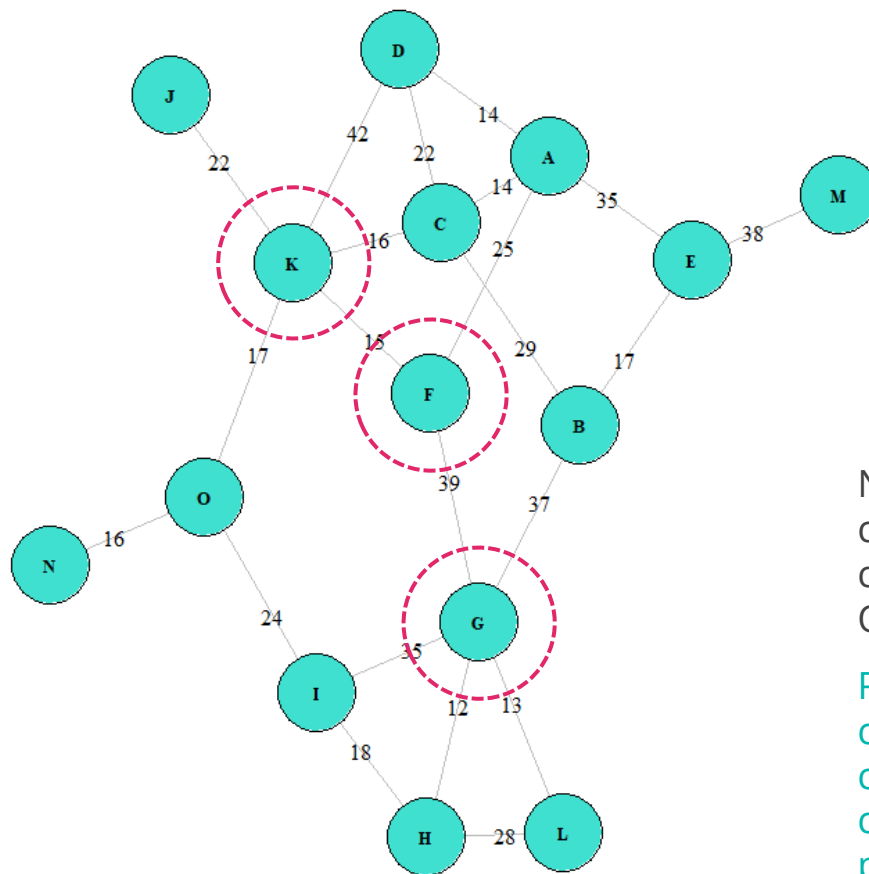
Resposta: **F, G e K**

A	B	C	D	E
0.032	0.034	0.034	0.030	0.028
F	G	H	I	J
0.037	0.037	0.028	0.031	0.024
K	L	M	N	O
0.036	0.026	0.020	0.022	0.030

Função em R: *closeness*

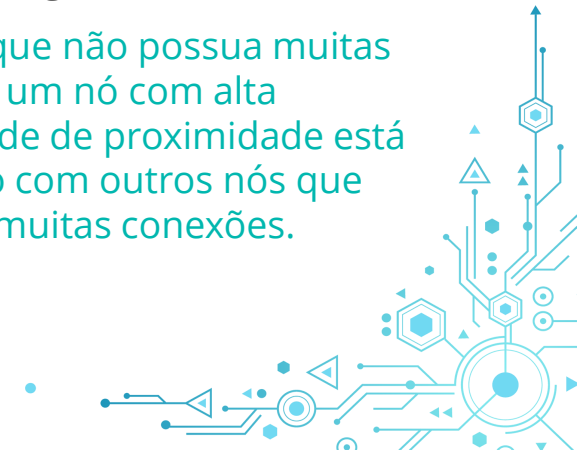
Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.



Note que a **cidade F** possui baixa centralidade de grau, mas alta centralidade de proximidade. O que isso significa?

Por mais que não possua muitas conexões, um nó com alta centralidade de proximidade está conectado com outros nós que possuem muitas conexões.



Medidas de centralidade

Centralidade de intermediação

Quais cidades atuam como intermediadoras frequentes nos caminhos entre outras cidades?

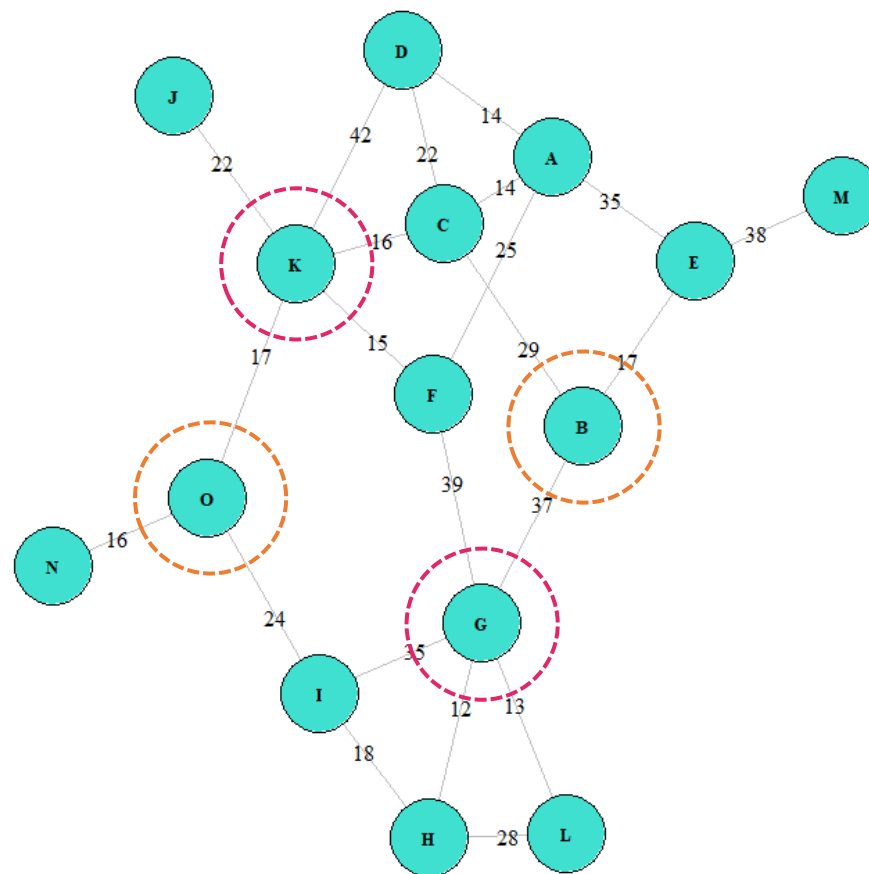
Resposta: **K e G**, seguidas de **O e B**

A	B	C	D	E
12.3	16.0	10.4	3.1	13.5
F	G	I	J	K
14.0	26.5	9.1	0.0	28.9
L	N	M	H	O
0.0	0.0	0.0	1.5	17.8

Função em R: *betweenness*

Arquivo: Otimizacao_Rotas.txt

@LABDATA FIA. Copyright all rights reserved.





Detecção de Comunidades

Por fim, os métodos de detecção de comunidades servem para **identificar subgrupos de nós** que estão mais densamente conectados entre si do que com o restante da rede.

Essa tarefa é realizada por meio de **algoritmos de clusterização** específicos para dados em formato de grafos. Os algoritmos mais comuns são:

Algoritmo	Precisão (capacidade de encontrar os clusters "reais")	Velocidade de execução
Louvain	Alta	Alta
Infomap	Moderada	Moderada
Walktrap	Moderada	Moderada
Edge Betweenness (Girvan-Newman)	Moderada	Lenta
Label Propagation	Baixa	Alta





Detecção de Comunidades

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

Por fim, os métodos de detecção de comunidades servem para **identificar subgrupos de nós** que estão mais densamente conectados entre si do que com o restante da rede.

Essa tarefa é realizada por meio de **algoritmos de clusterização** específicos para dados em formato de grafos. Os algoritmos mais comuns são:

Algoritmo	Precisão (capacidade de encontrar os clusters "reais")	Velocidade de execução
Louvain	Alta	Alta
Infomap	Moderada	Moderada
Label Propagation	Baixa	Alta

O algoritmo de Louvain possui um parâmetro chamado *"resolução"* que pode ser controlado manualmente, a fim de aumentar ou diminuir a quantidade desejada de *clusters*.



Case: Interação entre Colaboradores

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

52

O diretor de ciência de dados de uma empresa de tecnologia deseja avaliar quais são os principais grupos de interação entre os colaboradores da área. Para isso, foi realizada uma pesquisa corporativa com a questão: *Com quais membros da área de ciência de dados você colabora de forma recorrente para realização do seu trabalho?*



A área de ciência de dados desta empresa é composta por 50 funcionários, divididos nas seguintes atribuições:

- 15 **cientistas de dados** (*data scientists*) identificados como DS01, DS02, ..., DS15
- 10 **analistas de dados** (*data analysts*), identificados como DA01, DA02, ..., DA10
- 10 **engenheiros de dados** (*data engineers*), identificados como DE01, DE02, ..., DE10
- 10 **analistas de negócios** (*business analysts*), identificados como BA01, BA02, ..., BA10
- 5 **engenheiros de machine learning** (*ML engineers*), identificados como ML01, ML02, ..., ML05

Todos eles responderam à pesquisa, indicando os nomes de todos os colaboradores com quem possuem interação de forma recorrente.

Arquivo: Interacao_Colaboradores.txt

@LABDATA FIA. Copyright all rights reserved.



lab.data

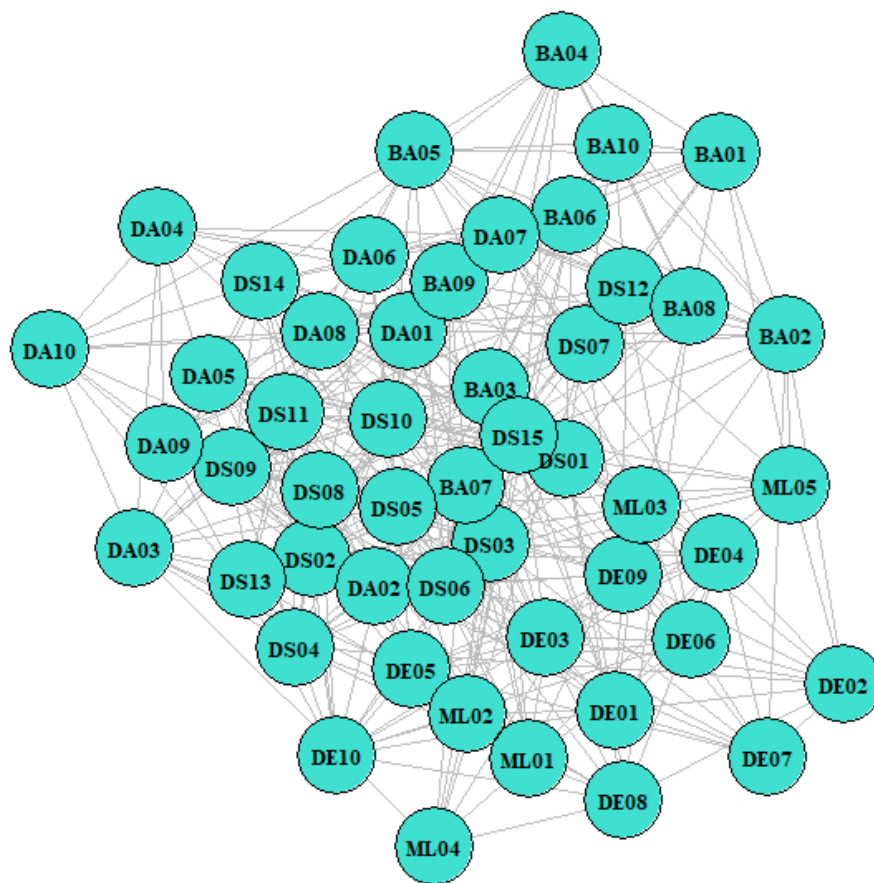


Case: Interação entre Colaboradores

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

53

Análise visual da rede



De qual tipo é esta rede?

Arestas **não direcionadas** e
com **pesos iguais**

Arquivo: Interacao_Colaboradores.txt

@LABDATA FIA. Copyright all rights reserved.

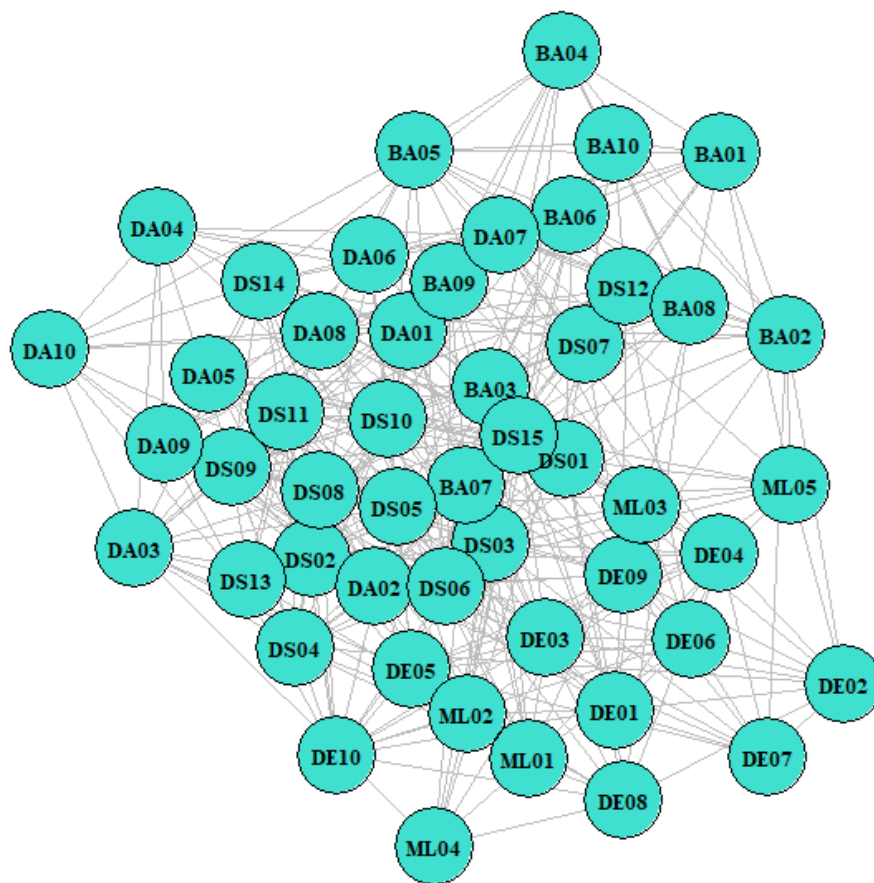


Case: Interação entre Colaboradores

5. MÉTODOS DE ANÁLISE DE REDES | ANÁLISE DE REDES SOCIAIS

54

Análise visual da rede



É difícil examinar os padrões de conexão entre nós em redes sociais mais extensas e/ou densas.

Arquivo: Interacao_Colaboradores.txt

@LABDATA FIA. Copyright all rights reserved.



Identificação de comunidades

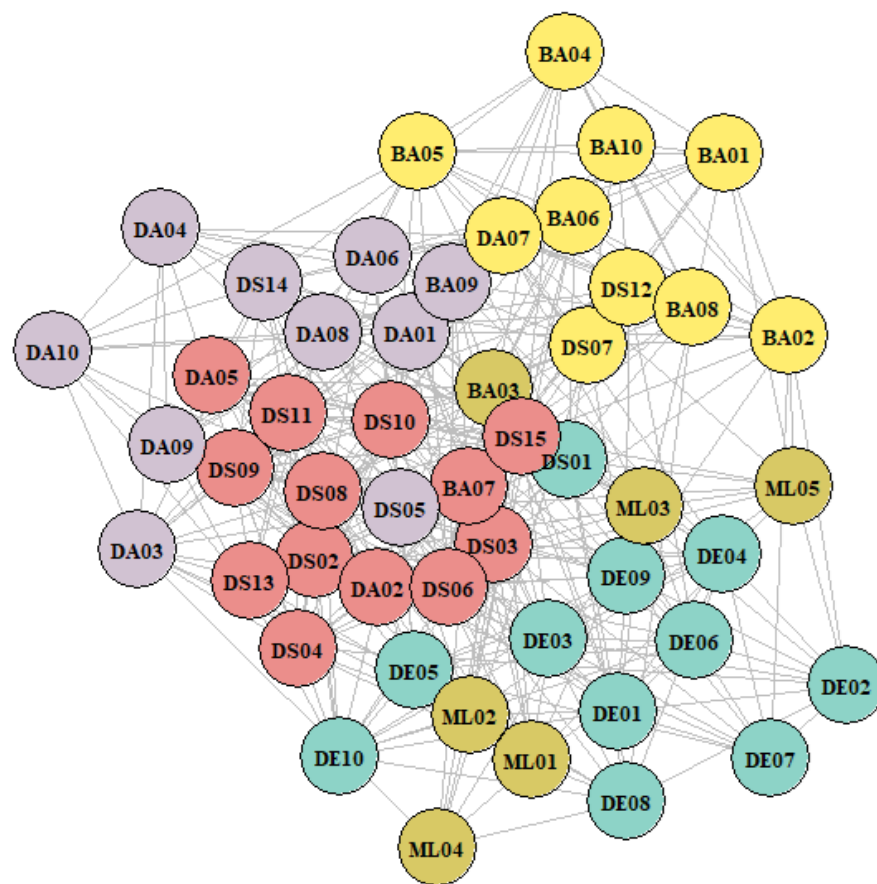
Algoritmo: Louvain

Parâmetro de resolução = **1** (*default*)

O valor padrão do parâmetro de resolução fornece uma segmentação próxima a das subdivisões de cargos da área. Porém, note que alguns funcionários estão mais próximos de outros cargos do que de seus pares.

Função em R:

```
set.seed(123)  
cluster_louvain(..., resolution = 1)
```



Arquivo: Interacao_Colaboradores.txt



Identificação de comunidades

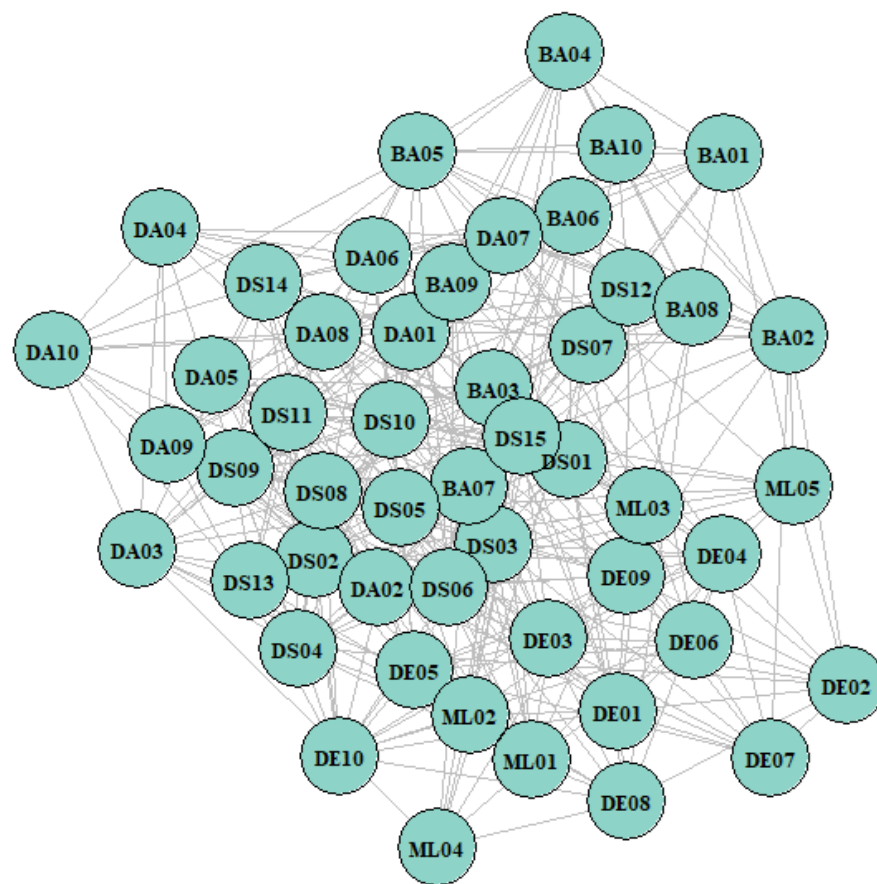
Algoritmo: Louvain

Parâmetro de resolução = **0,5**

Alterando o parâmetro de resolução para um valor menor, resultamos numa segmentação muito genérica.

Função em R:

```
set.seed(123)  
cluster_louvain(..., resolution = 0.5)
```



Arquivo: Interacao_Colaboradores.txt



Identificação de comunidades

Algoritmo: Louvain

Parâmetro de resolução = 2

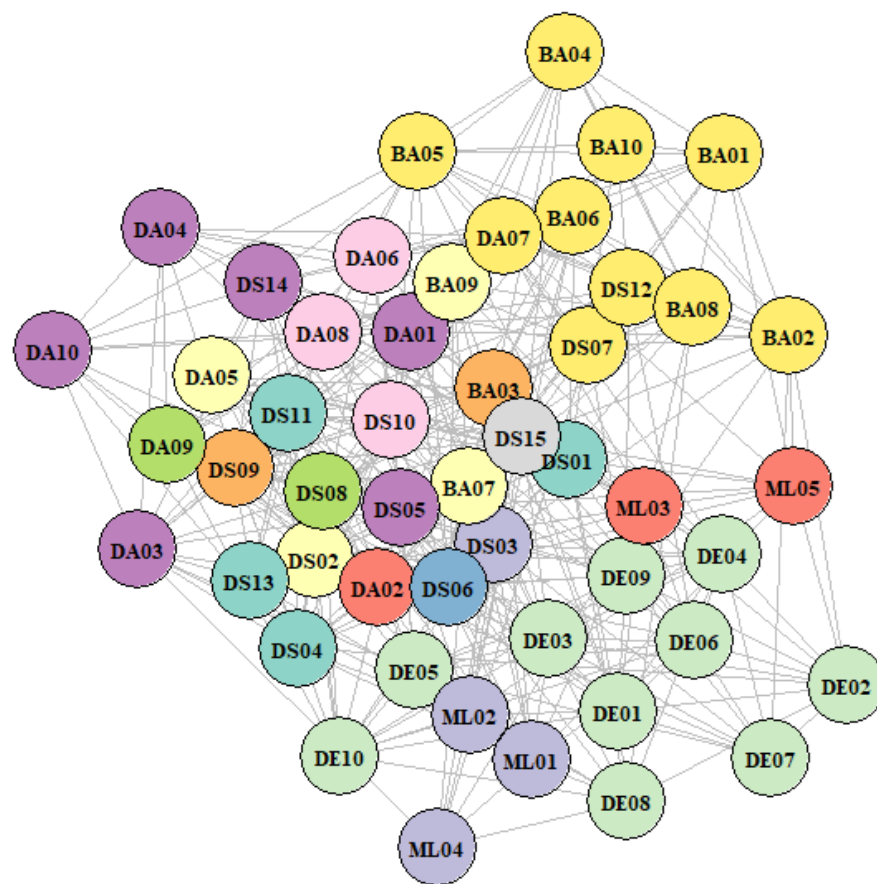
Alterando o parâmetro de resolução para um valor maior, resultamos numa segmentação mais detalhada, com subgrupos menores de colaboradores com maior interação: possíveis “squads”.

Função em R:

```
set.seed(123)  
cluster_louvain(..., resolution = 2)
```

Arquivo: Interacao_Colaboradores.txt

@LABDATA FIA. Copyright all rights reserved.

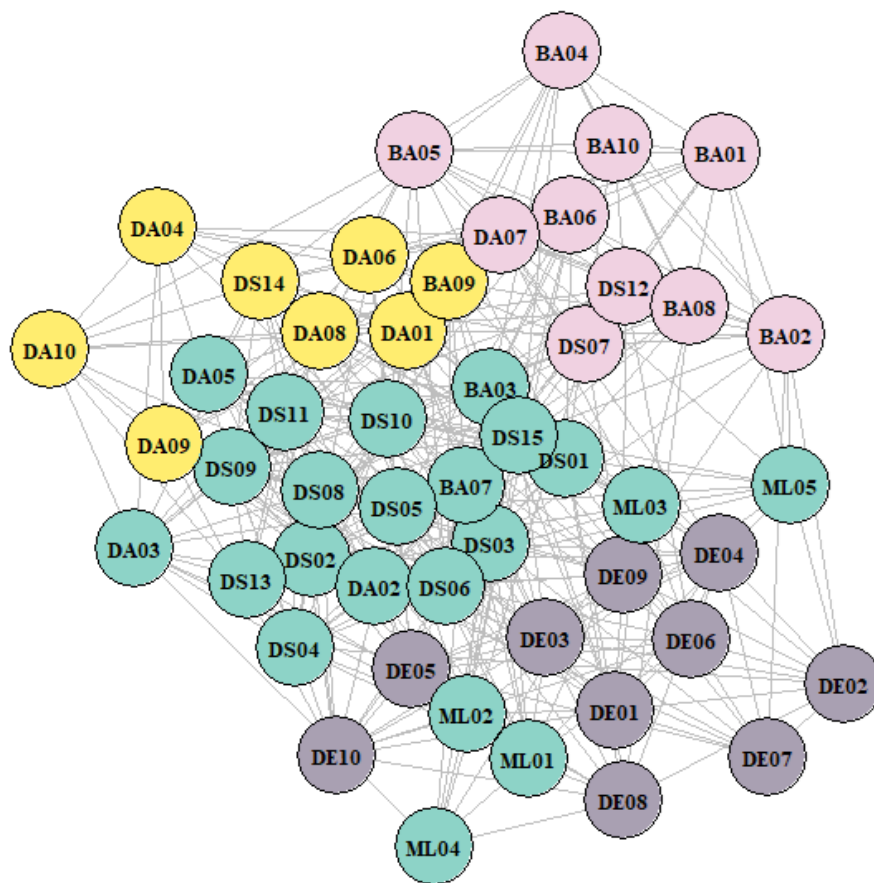


Identificação de comunidades

Algoritmo: Walktrap

Resultou em uma segmentação próxima à do algoritmo Louvain com resolução igual a 1, porém com uma comunidade a menos.

Função em R:
`cluster_walktrap`

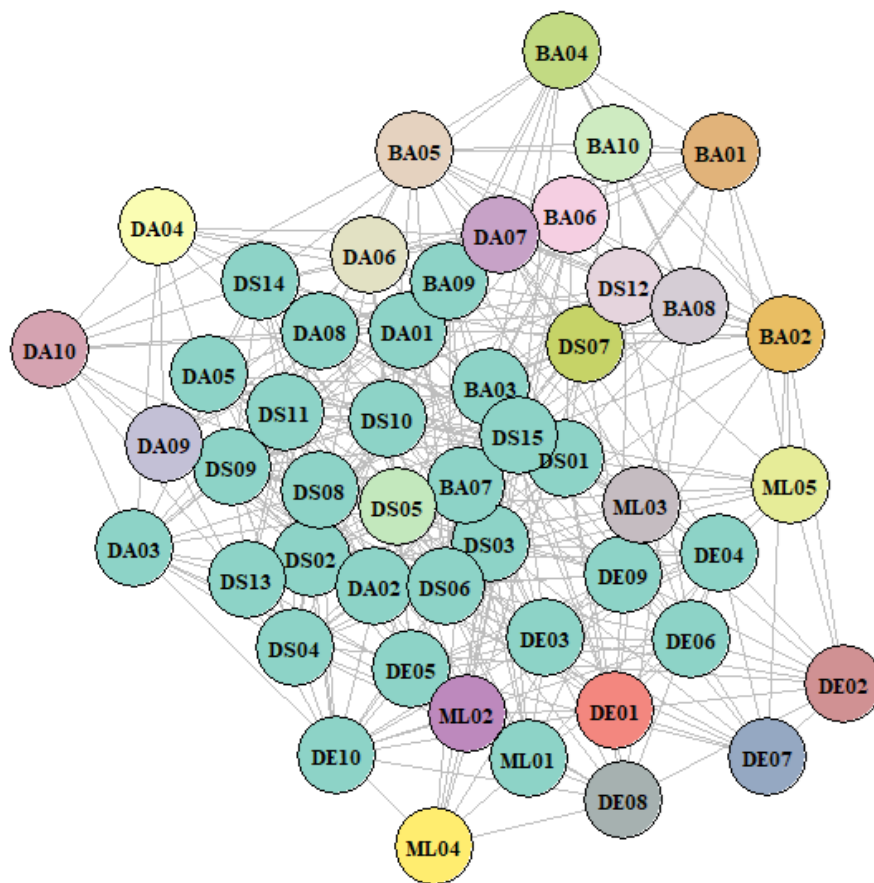


Arquivo: Interacao_Colaboradores.txt



Algoritmo: Edge Betweenness

Função em R:
cluster_edge_betweenness



@LABDATA FIA. Copyright all rights reserved.



Identificação de comunidades

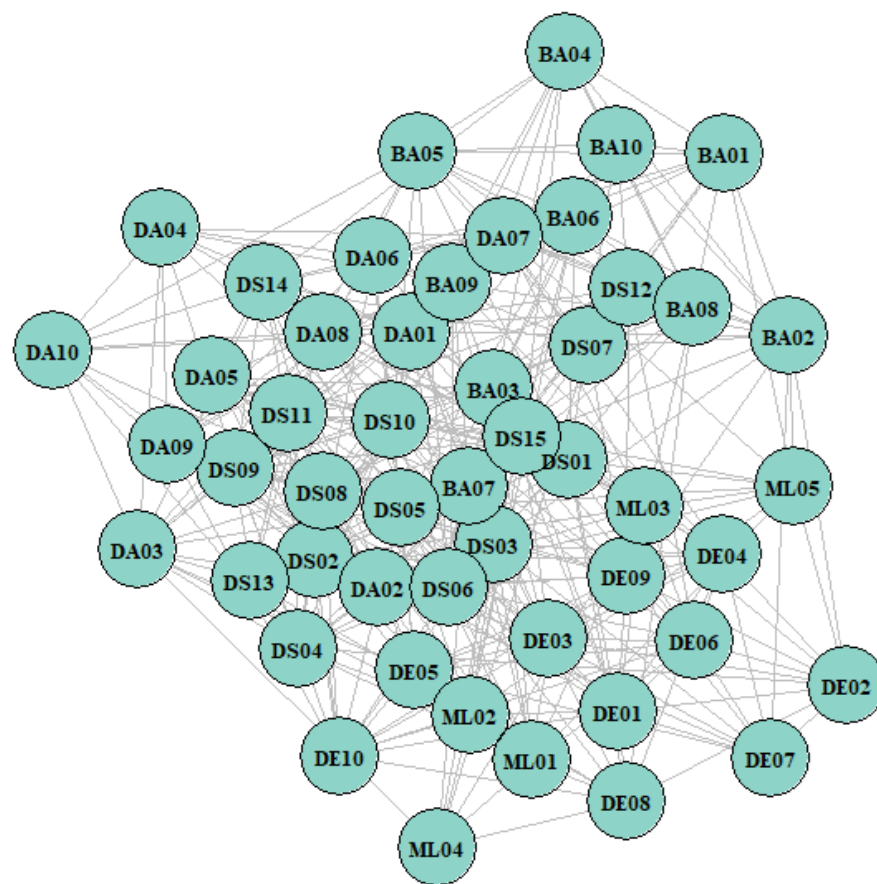
Algoritmos: Infomap e Label Propagation

Não foram capazes de identificar comunidades.

Funções em R:

`cluster_infomap`

`cluster_label_prop`



Arquivo: Interacao_Colaboradores.txt



- Wasserman, F., Faust K. *Social Network Analysis - Methods and Applications*. Cambridge, 1994.
- McCulloh, I. et al. *Social Network Analysis with Applications*. Wiley, 2013.
- Documentação do pacote *igraph* do R:
 - <https://r.igraph.org/articles/igraph.html>
- Sobre métodos de detecção de comunidades:
 - Smith N. R. et al. *A Guide for Choosing Community Detection Algorithms in Social Network Studies - The Question Alignment Approach*. American Journal of Preventive Medicine, 59 (4), 597–605, 2020.
 - <https://igraph.org/c/doc/igraph-Community.html>
 - <https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae> (em Python)





lab.data

<http://labdata.fia.com.br>
Instagram: @labdatafia
Facebook: @LabdataFIA

