GRAFOS

Redes Sociais e Econômicas

Prof. André Vignatti

TEORIA DOS GRAFOS E REDES SOCIAIS

grafos: permite formular propriedades de redes em uma *linguagem unificadora*

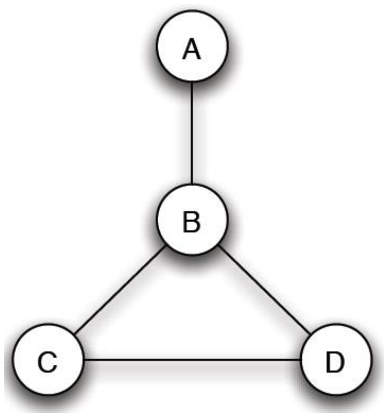
as definições centrais aqui são simples, e podemos descrevê-las de forma relativamente rápida

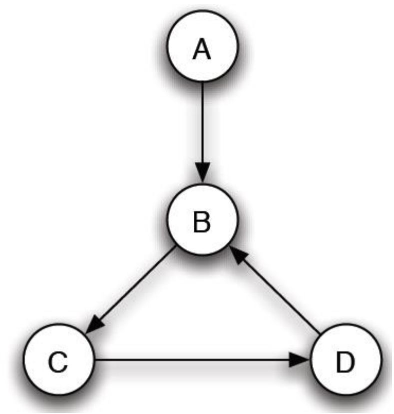
veremos algumas aplicações dessas definições

GRAFOS: DEFINIÇÕES BÁSICAS

Um **grafo** é uma forma de especificar relações entre coleção de objetos

- **grafo**: um conjunto de objetos (nós), com alguns pares de objetos conectados por links (arestas)
- **vizinhos**: nós conectados por aresta
- **grafos direcionados**: expressam relações assimétricas





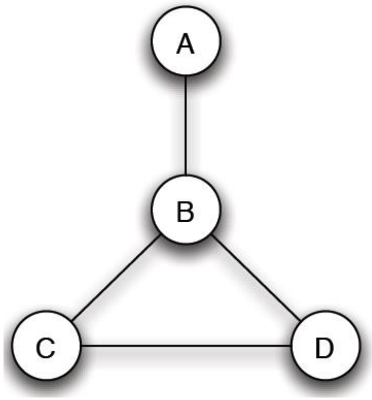
GRAFOS: DEFINIÇÕES BÁSICAS

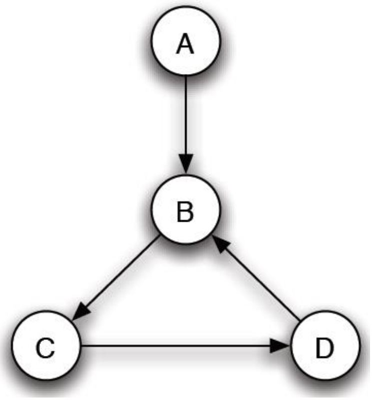
na figura da esquerda, a relação entre os vértices de uma aresta é simétrica

às vezes, queremos expressar relações assimétricas

para isso, usamos um grafo com arestas direcionadas

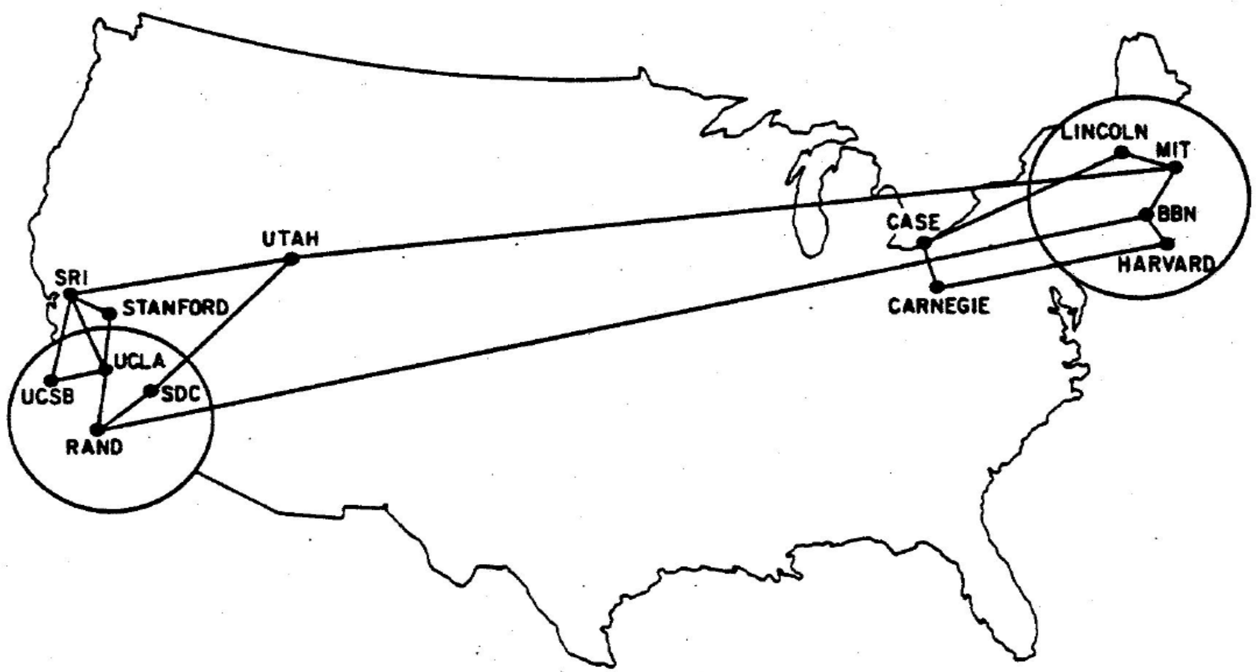
cada aresta direcionada é um link de um nó para outro, com a direção sendo importante





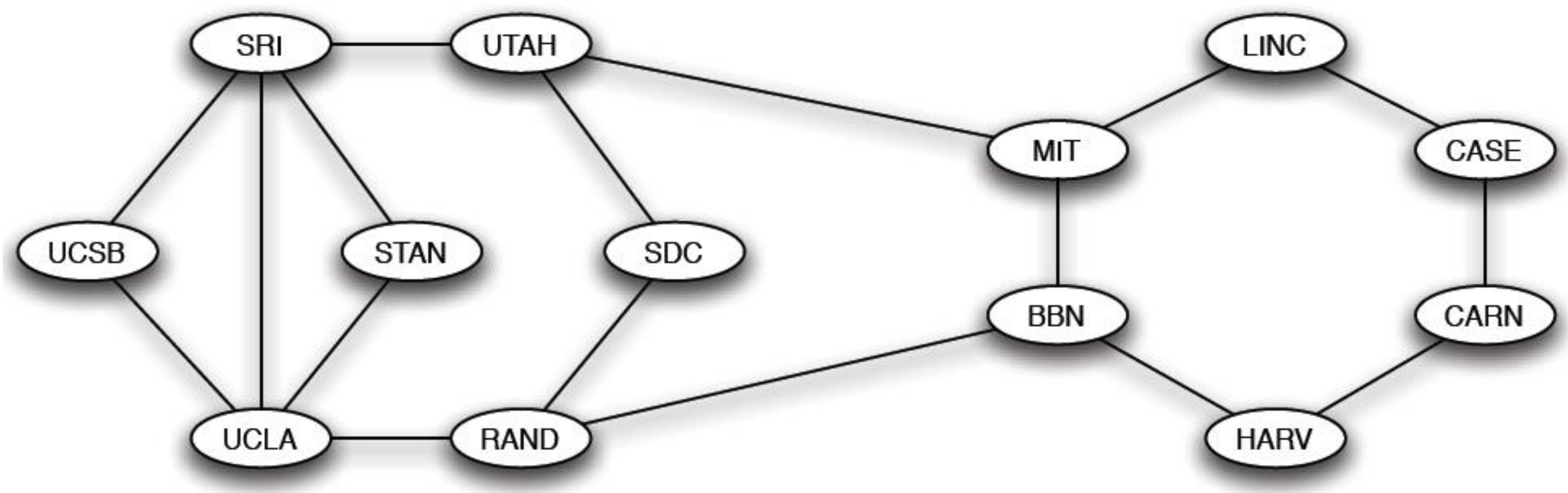
GRAFOS COMO MODELOS DE REDES

grafos são úteis porque servem como modelos matemáticos de estruturas de rede



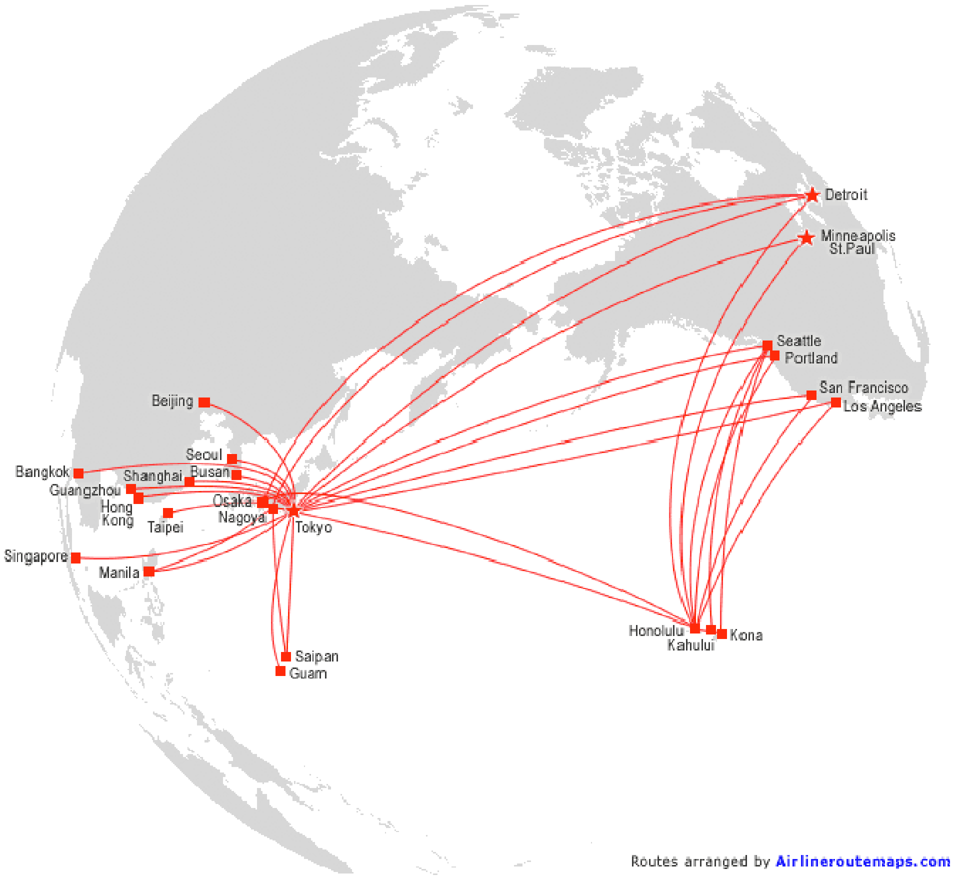
ARPANET - Internet no início, com 13 sites em 1970

REPRESENTAÇÃO DE GRAFO DA ARPANET



SITUAÇÕES PRÁTICAS ENVOLVENDO GRAFOS

redes de transporte: nós são locais, arestas ligam os locais

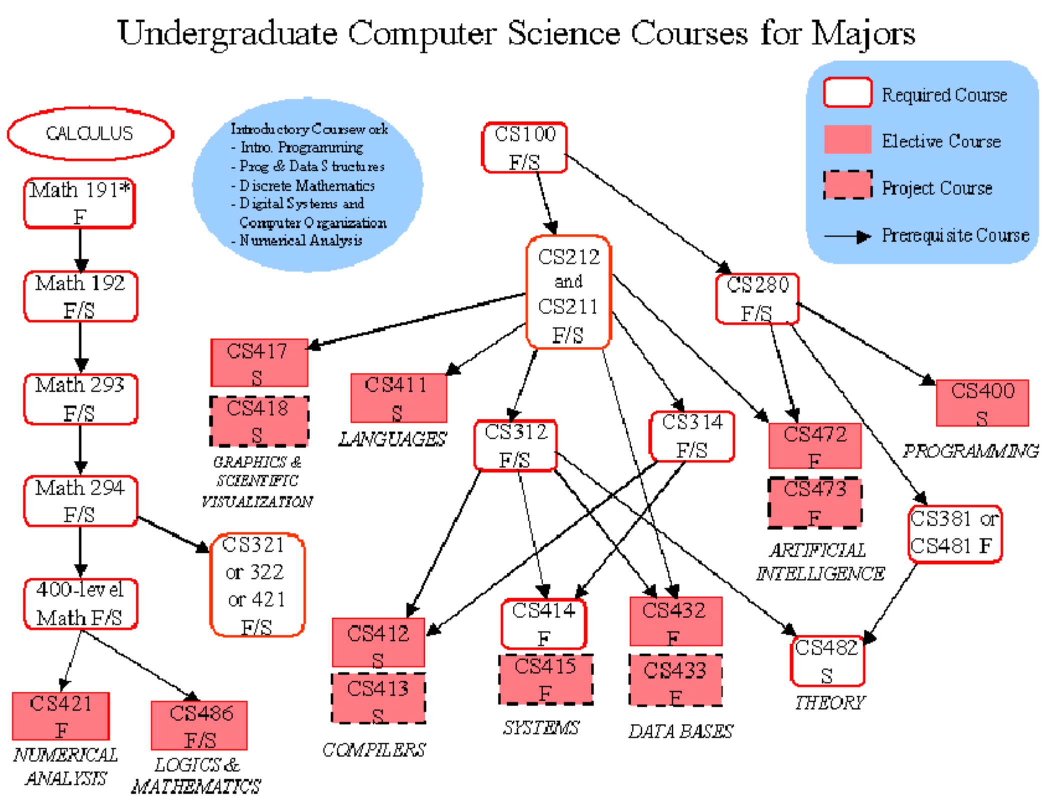


SITUAÇÕES PRÁTICAS ENVOLVENDO GRAFOS

redes de transporte: nós são locais, arestas ligam os locais

SITUAÇÕES PRÁTICAS ENVOLVENDO GRAFOS

redes de dependência (DAG): nós são tarefas, arestas direcionadas indicam que uma tarefa deve ser feita antes da outra



SITUAÇÕES PRÁTICAS ENVOLVENDO GRAFOS

rede estrutural: nós são junções, arestas são ligações físicas

- (teoria de rigidez – geometria, engenharia mecânica)

CAMINHOS E CONECTIVIDADE

- **caminho**: sequência de nós onde cada par consecutivo na seqüência é ligado por uma aresta

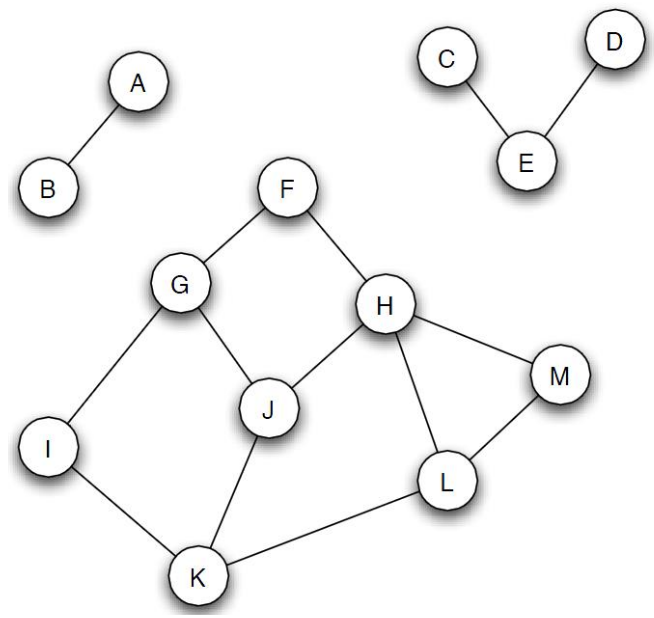
- **ciclo**: caminho com pelo menos três arestas, onde o primeiro e o último nós são os mesmos, mas todos os outros nós são distintos

dado um grafo, é natural perguntar se cada nó pode chegar a qualquer outro nó por um caminho

um grafo é **conexo** se, para cada par de nós, existe um caminho entre eles

COMPONENTES

se um grafo não é conexo, ele se divide em **componentes**



componente gigante: considere a rede social do **mundo inteiro,** com uma arestas entre duas pessoas se são amigas

esta rede de amizade globais é **conexa**?

COMPONENTE GIGANTE

provavelmente não: o comportamento de um único nó pode desfazer a conexidade

mesmo que a rede de amizade global não seja conexa, o componente que você faz parte é muito grande

de fato, você "alcança", mesmo indiretamente, pessoas das mais diversas origens

parece de fato provável que exista um **componente** que contém uma *fração significativa da população mundial*

COMPONENTE GIGANTE

isto é verdade quando se olha dados sobre redes grandes

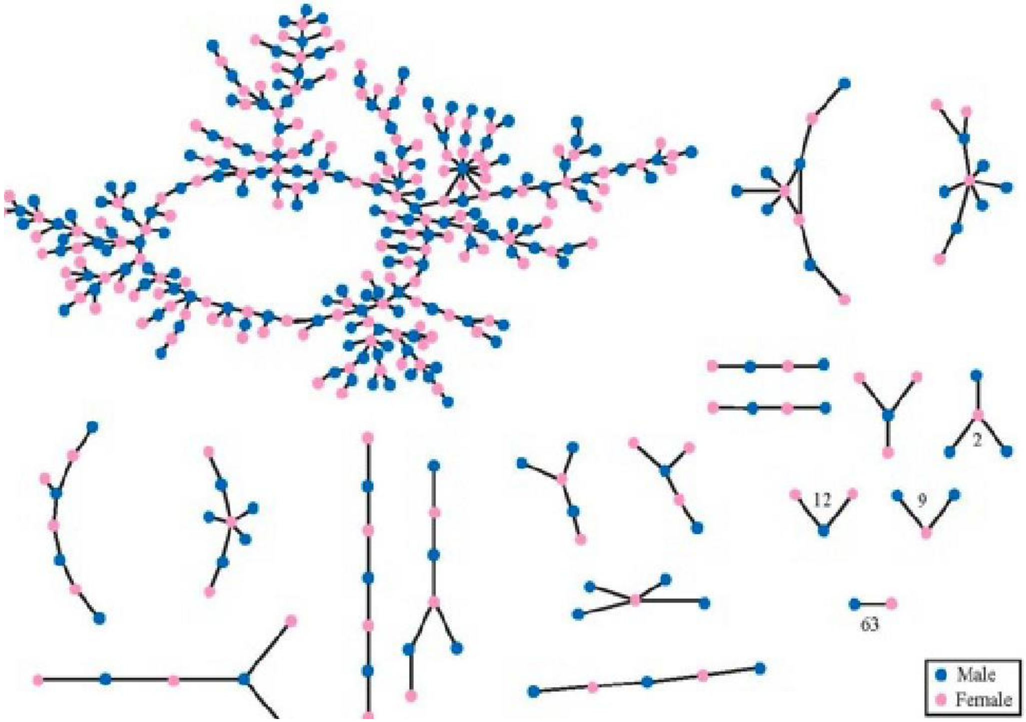
tais redes muitas vezes têm o que se chama de **componente gigante**

componente gigante é um termo informal para um componente conexo que contém uma fração significativa de todos os nós

quando uma rede contém um componente gigante, **quase sempre contém apenas um**

COMPONENTE GIGANTE

figura: relacionamentos românticos em uma escola americana em 18 meses



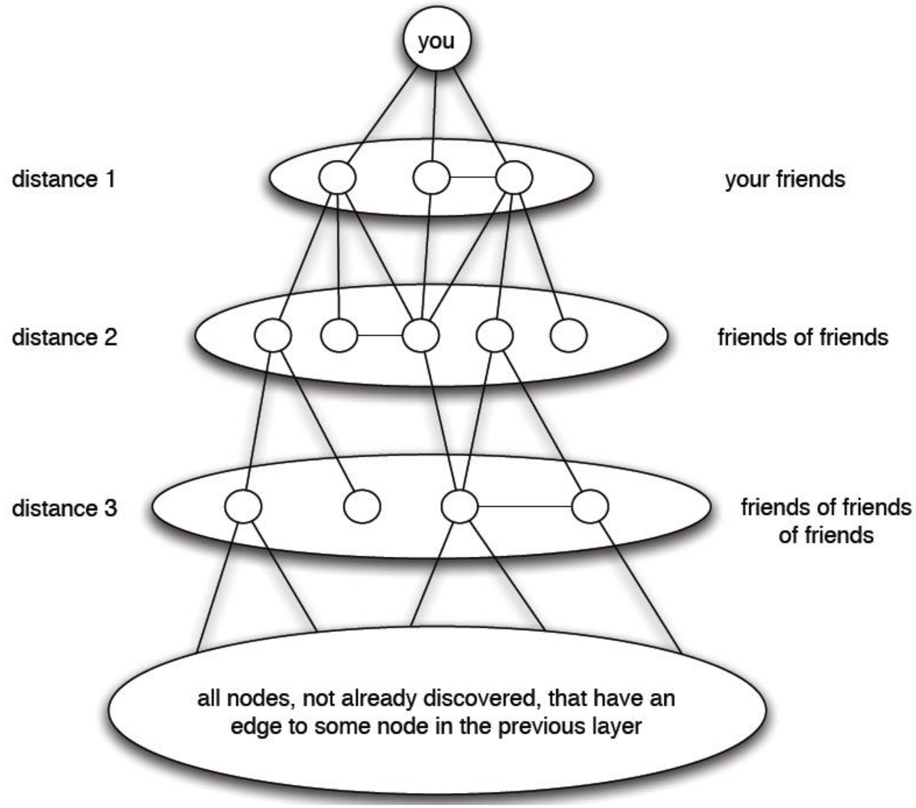
- é possível ver um componente gigante
- até estudantes com apenas uma relação estão no componente gigante

BUSCA EM LARGURA

comprimento de um caminho: número de arestas que ele contém do começo ao fim

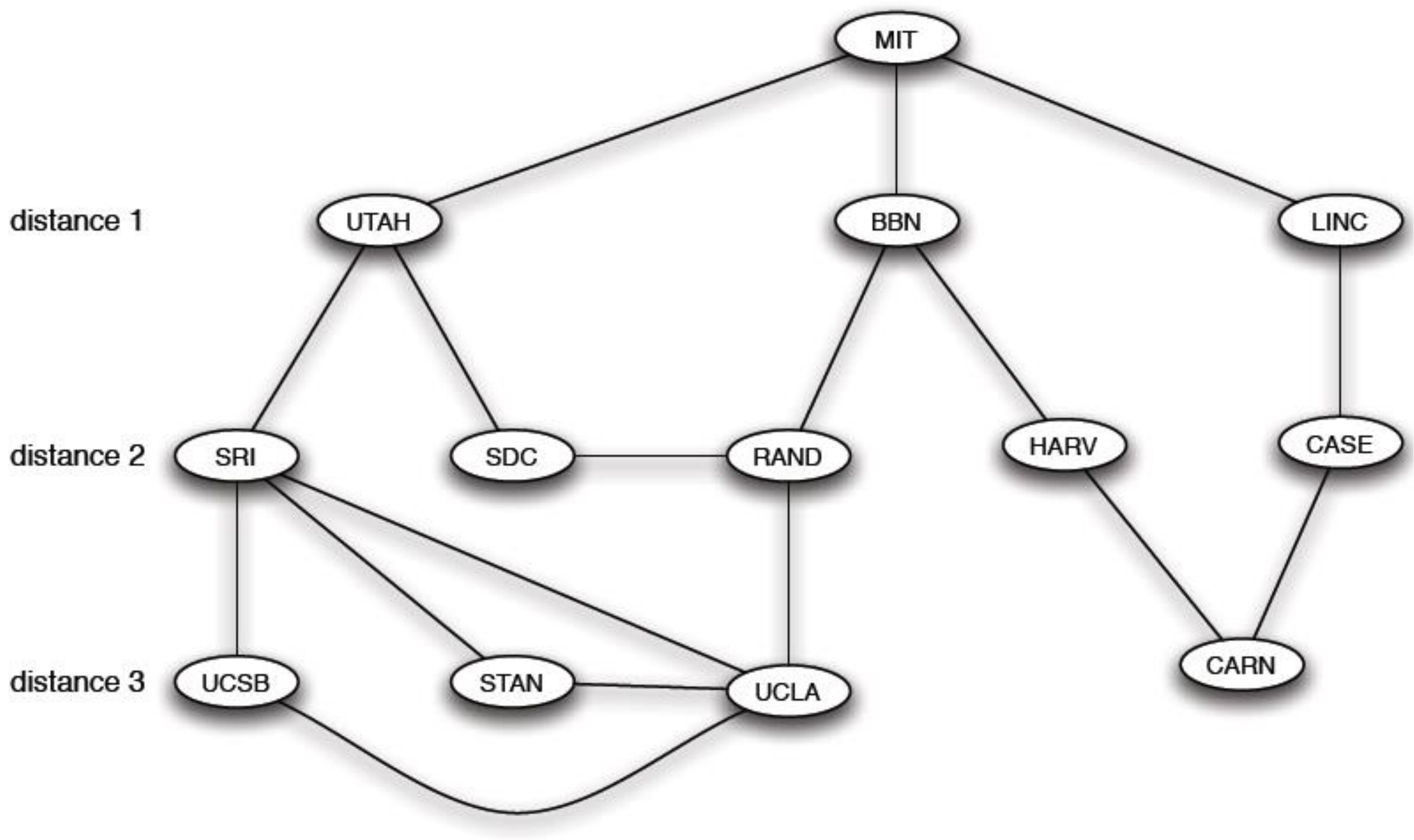
distância entre dois nós: o comprimento do caminho mais curto entre eles

como descobrir a distância entre dois nós: **busca em largura**



BUSCA EM LARGURA: EXEMPLO

busca em largura no grafo da Arpanet, começando do MIT



BUSCA EM LARGURA

1) Declare que todos os seus **amigos** estão a **distância 1**

2) Encontrar os **amigos deles** (sem contar os que já são amigos de você), e declarar que estes estão à **distância 2**

3) Em seguida, você encontra todos **amigos deles** (novamente, sem contar as pessoas que já encontrou a distâncias 1 e 2) declarem que estes estão à **distância 3**

4) (...)

BUSCA EM LARGURA

(...) Continuando desta maneira, você busca em **camadas sucessivas**, cada uma representando a próxima distância a partir do início

cada nova camada é construída a partir de todos os nós que

- (i) não tenham sido descobertos em camadas anteriores,
- (ii) têm uma aresta para algum nó na camada anterior

conclusão: determina distâncias e organiza os nós com base em suas distâncias a partir de um nó inicial

FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

voltando aos experimentos da rede global de amizades: a existência de um componente gigante na verdade implica em **algo mais forte**

além de caminhos ligando uma grande fração da população mundial, esses caminhos são **surpreendentemente curtos**

esta ideia é chamada de **fenômeno do mundo pequeno**

- o mundo parece "pequeno" pois o caminho entre você e quase qualquer pessoa é curto
- também conhecida como **seis graus de separação**

FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

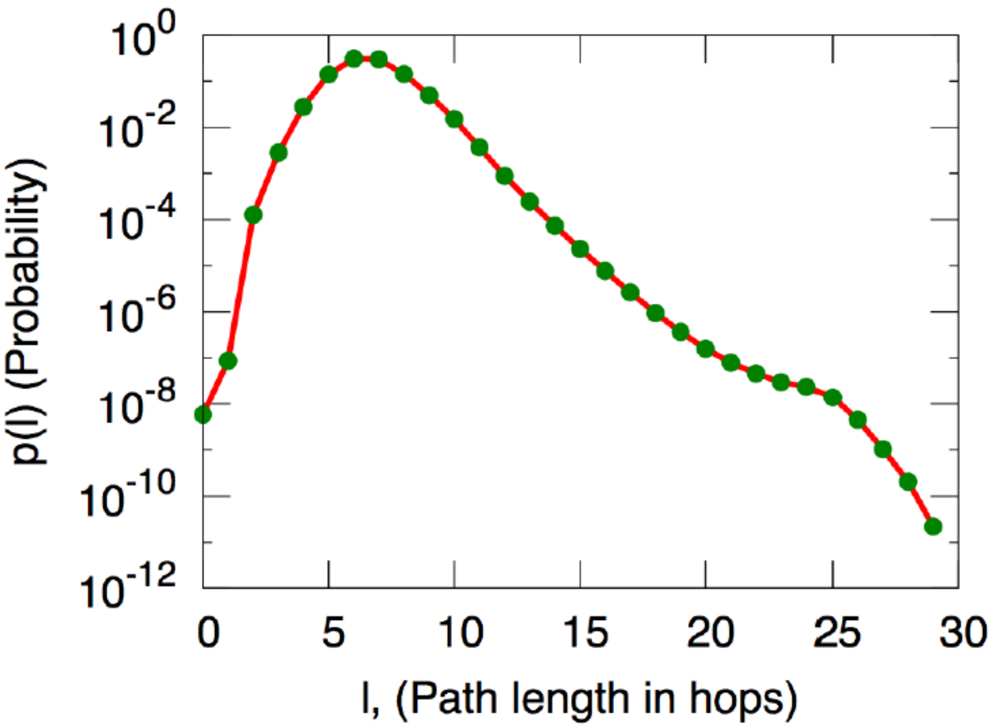
experimento: Microsoft Instant Messaging

- analisaram 240 milhões de contas no Microsoft Instant Messenger
- construíram um grafo: nó corresponde a um usuário
- aresta entre dois usuários: se eles conversaram durante o período de um mês de observação
- o grafo mostrou um **componente gigante** contendo quase todos os nós
- as distâncias dentro do componente gigante eram muito pequenas: distância média de 6,6

FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

Figura: distribuição de distâncias de uma amostra aleatória de 1.000 usuários

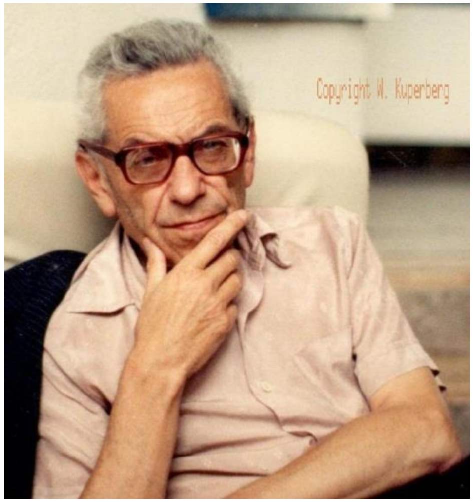
a busca em largura foi realizada separadamente para cada um desses mil usuários



FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

outra experiência em menor escala (milhares ao invés de milhões)

- um dos matemáticos mais importantes: **Paul Erdos**
- publicou cerca de 1000 artigos em sua carreira
- Erdos foi considerado um ponto central na estrutura de colaboração dos artigos



Os **nós** correspondem aos matemáticos, e **arestas** que conectam pessoas que publicaram um artigo juntas

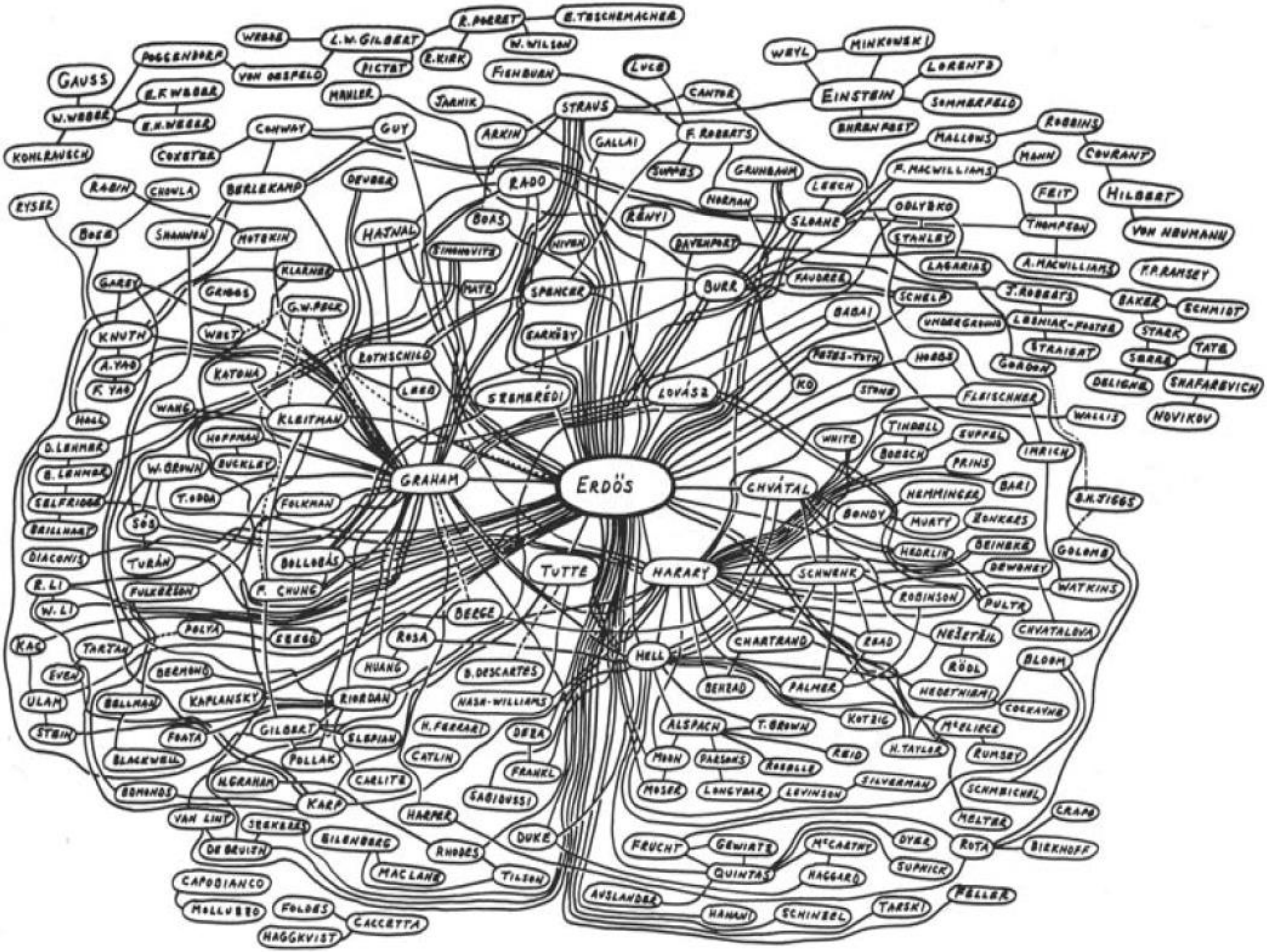
FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

O número de Erdos é a distância entre um matemático para Erdos neste grafo

A maioria dos matemáticos têm números de Erdos no máximo 4 ou 5

Exemplos:

- Albert Einstein tem número 2
- Enrico Fermi, Jon Kleinberg e Dennis Ritchie é 3
- Noam Chomsky, John Nash e Linus Pauling é 4
- Francis Crick é 5 e James Watson é 6

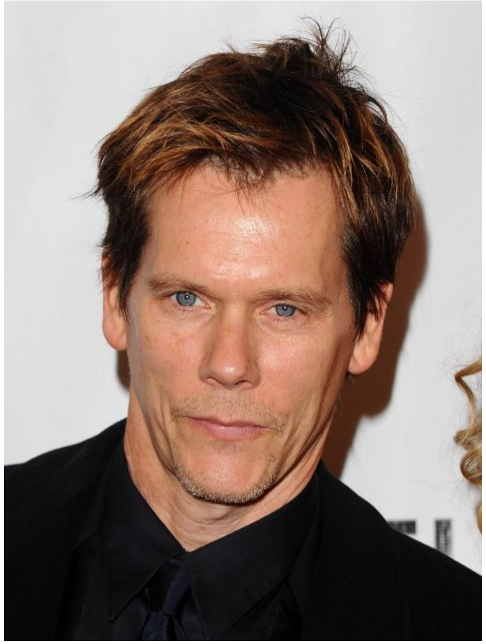


R. PORRET
E. TESCHEMACHER
WEBER
L.W. GILBERT
W. WILSON
POGGENDORF
R. KIRK
PICTET
VON OESFELD
GAUSS
E.F. WEBER
W. WEBER
E.H. WEBER
KOHLRAUSCH
MAHLER
JARNIK
FISHBURN
LUCE
CANTOR
STRAUS
WEYL
MINKOWSKI
LORENTZ
EINSTEIN
SOMMERFELD
EHRENFEST
CONWAY
GUY
COXETER
ARKIN
GALLAI
F. ROBERTS
ROBBINS
MALLOWS
MANN
COURANT
SUPPES
GRUNBAUM
F. MACWILLIAMS
HILBERT
RABIN
CHOWLA
BERLEKAMP
DEUBER
RADO
LEECH
FEIT
RYSER
NORMAN
ODLYZKO
BOSE
SHANNON
MOTZKIN
BOAS
HAJNAL
RÉNYI
SLOANE
THOMPSON
VON NEUMANN
SIMONOVITS
NIVEN
DAVENPORT
STANLEY
KLARNER
LAGARIAS
A. MACWILLIAMS
F.P. RAMSEY
GAREY
GRIGGS
G.W. PECK
MATE
SPENCER
BURR
FAUDREE
SCHELP
J. ROBERTS
BAKER
SCHMIDT
BABAI
UNDERGROUND
LESNIAK-FOSTER
STARK
KNUTH
WEST
SARKÖZY
STRAIGHT
TATE
SERRE
A. YAO
ROTHSCHILD
FEJES-TOTH
HOBBS
GORDON
DELIGNE
SHAFAREVICH
KATONA
F. YAO
LEEB
SZEMERÉDI
LOVÁSZ
KO
STONE
FLEISCHNER
NOVIKOV
WANG
KLEITMAN
WALLIS
HALL
TINDELL
HOFFMAN
WHITE
SUFFEL
D. LEHMER
BOESCH
IMRICH
W. BROWN
BUCKLEY
PRINS
E. LEHMER
GRAHAM
ERDŐS
CHVÁTAL
BARI
HEMMINGER
SELFRIDGE
T. ODDA
FOLKMAN
B.H. JIGGS
BONDY
MURTY
ZONKRES
SÓS
BRILLHART
HEDRLIN
BEINEKE
GOLOMB
DIACONIS
TURÁN
BOLLOBÁS
TUTTE
HARARY
SCHWENK
DEWDNEY
R. LI
FULKERSON
F. CHUNG
WATKINS
ROBINSON
W. LI
BERGE
PULTR
POLYA
NEŠETŘIL
CHVATALOVA
KAC
TARJAN
SZEGŐ
ROSA
CHARTRAND
ZBAD
HELL
BLOOM
RÖDL
EVEN
BERMOND
HUANG
B. DESCARTES
BEHZAD
PALMER
HEDETNIEMI
ULAM
BELLMAN
KAPLANSKY
RIORDAN
NASH-WILLIAMS
COCKAYNE
KOTZIG
ALSPACH
T. BROWN
McELIECE
STEIN
FOATA
GILBERT
SLEPIAN
H. FERRARI
DEZA
H. TAYLOR
RUMSEY
REID
BLACKWELL
POLLAK
CATLIN
FRANKL
MOON
PARSONS
ROSELLE
CARLITZ
SCHMEICHEL
EDMONDS
H. GRAHAM
SILVERMAN
MOSER
LONGYEAR
LEVINSON
SABIDUSSI
KARP
HARPER
MELTER
CRAPO
VAN LINT
SZEKERES
DUKE
EILENBERG
FRUCHT
GEWIRTZ
McCARTHY
DYER
ROTA
BIRKHOFF
DE BRUIJN
RHODES
MACLANE
TILSON
QUINTAS
HAGGARD
SUPNICK
CAPOBIANCO
AUSLANDER
FELLER
MOLLUZZO
FOLDES
CACCETTA
HANANI
SCHINZEL
TARSKI
HAGGKVIST

FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

outra experiência bem conhecida é o grafo de colaboração entre atores e atrizes de cinema

- Os nós são os atores do filme
- Uma aresta conecta dois atores se eles já apareceram juntos em um filme
- O **número de Bacon** é a distância a o ator a *Kevin Bacon* neste grafo



FENÔMENO DO MUNDO PEQUENO OU SEIS GRAUS DE SEPARAÇÃO

usando as listas de elenco do Internet Movie Database (IMDB), é possível calcular os números de Bacon para todos os artistas usando busca em largura

assim como na colaboração dos matemáticos, aqui também temos um mundo pequeno

o número de Bacon médio, sobre todos os artistas no IMDB, é de aproximadamente **2,9**

é um desafio encontrar um que é maior que 5

DADOS DE REDES

para realizar pesquisas em redes pequenas (e.g. clube de karatê), é fácil: é só fazer entrevistas!

para ***redes grandes***, devemos automatizar isso

as principais redes grandes onde são feitas pesquisas:

- **Grafos de Colaboração**: quem trabalha com quem num ambiente. Ex: atores, matemáticos, Fortune 500, colaboradores da Wikipedia (trabalharam no mesmo artigo), World of Warcraft (se participaram da mesma missão)
- **Redes Quem-fala-com-Quem**: Microsoft IM, emails dentro de uma empresa ou universidade, ligações telefônicas, e até cara-a-cara (usando dados de proximidade de telefones celulares)

DADOS DE REDES

- **Grafos de ligações entre informações**: páginas Web, blogs, artigos da Wikipedia, citações de artigos acadêmicos
- **Redes tecnológicas**: entre computadores na Internet, entre estações de força para redes de energia
- **Redes do mundo natural**: cadeia alimentar, conexões neurais em cérebros, metabolismo de células (nós são compostos, arestas são interações químicas)

DADOS DE REDES

Exemplo: networkrepository.com

Data & Network Collections. Find and interactively VISUALIZE and EXPLORE hundreds of network data

ANIMAL SOCIAL NETWORKS	816	INTERACTION NETWORKS	29	SCIENTIFIC COMPUTING	11
BIOLOGICAL NETWORKS	37	INFRASTRUCTURE NETWORKS	8	SOCIAL NETWORKS	77
BRAIN NETWORKS	116	LABELED NETWORKS	105	FACEBOOK NETWORKS	114
COLLABORATION NETWORKS	20	MASSIVE NETWORK DATA	21	TECHNOLOGICAL NETWORKS	12
CHEMINFORMATICS	646	MISCELLANEOUS NETWORKS	2668	WEB GRAPHS	36
CITATION NETWORKS	4	POWER NETWORKS	8	DYNAMIC NETWORKS	115
ECOLOGY NETWORKS	6	PROXIMITY NETWORKS	13	TEMPORAL REACHABILITY	38
ECONOMIC NETWORKS	16	GENERATED GRAPHS	221	BHOSLIB	36
EMAIL NETWORKS	6	RECOMMENDATION NETWORKS	36	DIMACS	78
GRAPH 500	8	ROAD NETWORKS	15	DIMACS10	84
HETEROGENEOUS NETWORKS	15	RETWEET NETWORKS	34	NON-RELATIONAL ML DATA	211

Exemplo: networkrepository.com

Summary of notation.

$\|V\|$	Number of nodes
$\|E\|$	Number of edges
d_{max}	Maximum degree
d_{avg}	Average degree
r	Assort. Coeff.
$\|T\|$	Number of triangles (3-clique)
$\|T\|_{avg}$	Average triangles formed by a edge
$\|T\|_{max}$	Maximum number of triangles formed by a edge
κ_{avg}	Average local clustering coefficient
κ	Global clustering coefficient
K_{max}	Maximum k-core number
ω_{lb}	Lower bound on the size of the maximum clique

Graph Name	$\|V\|$	$\|E\|$	d_{max}	d_{avg}	r	$\|T\|$	T_{avg}	T_{max}	κ_{avg}	κ	K	ω_{heu}	Size	Download
bn-mouse-visual-cortex-1	29	44	9	3	-0.47	3	-	1	0.05	0.02	3	2	2 KB	Download
bn-cat-mixed-species-brain-1	65	1K	72	35	0.01	28K	433	1K	0.71	0.60	27	11	6 KB	Download
bn-macaque-rhesus-brain-2	91	628	96	13	-0.71	10K	109	792	1.63	0.37	16	12	4 KB	Download
bn-macaque-rhesus-cerebral-cortex-1	91	2K	248	43	-0.34	86K	945	4K	1.30	0.56	35	21	9 KB	Download
bn-macaque-rhesus-interareal-cortical-network-2	93	3K	120	57	-0.53	174K	2K	4K	1.56	0.75	56	31	12 KB	Download
bn-mouse-visual-cortex-2	193	214	31	2	-0.84	12	-	2	0.02	0.00	3	2	2 KB	Download
bn-mouse-brain-1	213	22K	314	203	-0.04	4M	17K	31K	0.80	0.77	143	35	53 KB	Download
bn-macaque-rhesus-brain-1	242	4K	164	33	-0.05	78K	322	3K	0.53	0.36	26	8	20 KB	Download
bn-mouse-kasthuri-graph-v4	1K	2K	153	3	-0.22	-	-	-	-	-	6	7	6 KB	Download
bn-mouse-retina-1	1K	577K	8K	1K	0.02	777M	722K	8M	1.55	0.61	1K	42	225 KB	Download
bn-fly-drosophila-medulla-1	2K	34K	16K	37	-0.33	11M	6K	737K	3.87	0.08	329	6	63 KB	Download
bn-human-BNU-1-0025890-session-1	178K	16M	5K	176	-0.13	2B	11K	875K	0.46	0.17	387	72	40 MB	Download
bn-human-Jung2015-M87101234	277K	64M	12K	464	0.24	25B	89K	6M	0.49	0.26	952	171	163 MB	Download

DADOS DE REDES

http://snap.stanford.edu/data/

https://gephi.org/

https://networks.skewed.de/