Comunicação e Redes

Msc. Carlos Reynaldo Portocarrero Tovar
Prof. Fabrício Olivetti de França

Agosto 2020

Sumário

1 Introdução **4**
1.1 Comunicação e Redes 7
1.2 Objetivos do Curso 8
1.3 Conceitos Básicos 8
1.3.1 Redes 9
1.3.2 Teoria dos Grafos 11
1.3.3 Ciência das Redes 12
1.3.4 Redes 12
1.3.5 Exemplo de Redes Reais : Famílias de Florença – Séc. XV 14

2 Teoria dos Grafos **16**
2.1 As Sete Pontes de Königsberg 16
2.2 Grafo Simples 16
2.2.1 Grau de um vértice 20
2.3 Grafo Direcionado 22
2.3.1 Grau de um vértice em Grafos Direcionados 24
2.4 Caminhos em uma Rede 24
2.4.1 Caminhos em Redes Direcionadas 25
2.4.2 Conetividade da Rede 26
2.4.3 Conexão 26
2.5 Modelagem Computacional 28
2.5.1 Estudo das Redes 28
2.6 Problemas reais onde as redes são aplicadas 29
2.6.1 Problema do Caixeiro Viajante 29
2.6.2 Modelar usando Redes 31
2.7 Representação Computacional 32
2.7.1 Rede como matriz 33
2.7.2 Representação: Matriz de Adjacência 33
2.7.3 Rede como lista de Arestas 36

3 Percurso em Grafos **38**
3.1 Distância entre os nós . . . 38
3.2 Distancia em Redes não Ponderadas . . . 39
3.3 Árvore . . . 44
3.4 Ciclos . . . 45
3.5 Conetividade . . . 45
3.6 Distância em Redes Ponderadas . . . 47
3.7 O algoritmo Dijkstra . . . 48
3.8 Distância em Redes Reais . . . 52

4 Centralidade **55**
4.1 Importância dos nós . . . 55
4.2 Grau . . . 56
4.3 Proximidade . . . 56
4.4 Betweenness . . . 60
4.5 PageRank . . . 63
4.5.1 Centralidade de Autovetor . . . 63
4.5.2 Autovalor e Autovetor . . . 64
4.5.3 Algoritmo PageRank . . . 64
4.6 Centralidade nas Redes Reais . . . 68
4.7 Agrupamentos e Pontes . . . 71
4.8 Coeficiente de Agrupamento . . . 72
4.9 Agrupamento em Redes Reais . . . 77

5 Redes Aleatórias **78**
5.1 Modelo genérico de Rede . . . 78
5.2 Redes Aleatórias: Erdõs-Rényi . . . 79
5.3 Redes Aleatórias . . . 79

6 Redes Mundo Pequeno **88**
6.1 Cadeias (*Chains*) . . . 88
6.2 Experimentos de Mundo Pequeno . . . 89
6.2.1 Experimento de Milgram . . . 89
6.2.2 Seis graus de Erdos . . . 91
6.3 Modelo Mundo Pequeno . . . 92

7 Padrões de Distribuição **100**
7.1 Distribuição dos Parâmetros da Rede . . . 100
7.1.1 Distribuição de Dados . . . 100
7.2 Medidas de Distribuição . . . 100
7.3 Lei da Potência . . . 103
7.4 Invariância em Escala . . . 103
7.5 Lei de Potência em Redes . . . 105
7.6 Cidades como Redes . . . 107
7.7 Padrões de Exponente . . . 108
7.8 Equações de Crescimento . . . 109

7.9 Modelos de Rede Livres de Escala 112
7.9.1 Redes sem Escala 113
7.9.2 Ligação Preferencial 114
7.10 Modelo Barabási-Albert 115
7.10.1 Alterações no modelo Barabási-Albert 119
7.10.2 Função de Afinidade 119

8 Assortatividade e Comunidade 121
8.1 Assortatividade 121
8.1.1 Assortatividade em Redes Reais 122
8.2 Comunidades 130
8.2.1 Particionamento de Redes 130
8.2.2 Métodos de Particionamento 133
8.2.3 Girvan-Newman 135
8.3 Particionamento Espectral 135
8.4 Modularidade 142

9 Redes de Computadores 146
9.1 Roteamento 153
9.1.1 Redes Desconhecidas 153
9.1.2 Situação Hipotética 155
9.1.3 Algoritmo de Roteamento 155

10 Aplicações 162
10.1 Análise da rede de interações dos personagens de Breaking Bad . 162

11 Difusão de Informação 168
11.1 Modelo SIR 168
11.2 Modelo SIR em uma rede 171
11.3 Informação e Contágio 173
11.4 Percolação 175
11.5 Ponto Crítico de uma Rede 176
11.6 Percolação e Tolerância a Falhas 180

1 Introdução

"Tudo está conectado, assim como o sistema solar", é uma afirmação forte, provocativa, mas consistente com o que será passado nesse curso. Essa imagem (Figura 1) mostra o sistema solar e a força gravitacional exercida pelo sol nos planetas. A órbita de um planeta pode ser afetada bem pouco pela força gravitacional de um planeta vizinho, embora esse efeito é bem pequeno, ele é mensurável. Antes de Netuno ser descoberto, astrônomos do século XIX notaram irregularidades na órbita de Urano e perceberam que era resultado da força gravitacional de um planeta além dele.

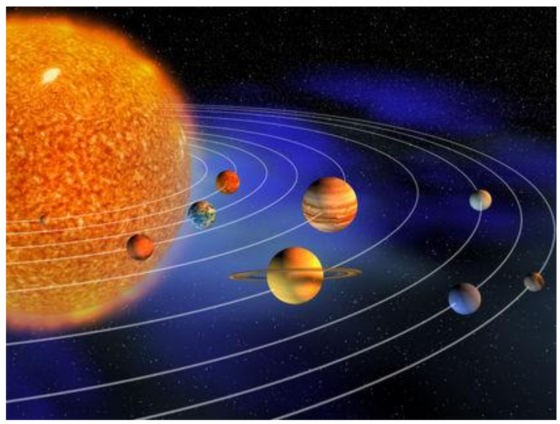


Figura 1: Sistema Solar

O universo também esta conectado (Figura 2). A grande motivação dessa disciplina é que iremos observar que tudo no universo está conectado através de interações, as quais são observadas macro e microscopicamente.

Os seres vivos estão conectados, um exemplo e a Cadeia Alimentar (Figura 3). Para sobreviver, os seres vivos interagem entre si através de um modelo de presa-predador, em que um ser vivo serve de fonte de energia para outro.

As pessoas estão conectadas através do Convívio social (Figura 4). Nós fazemos parte de diversos círculos sociais, familiares, trabalho, faculdade, academia..., por causa disso a sociedade consegue evoluir colaborativamente, também por conta disso doenças contagiosas conseguem se espalhar de forma eficiente.

Podemos fazer alguns experimentos simples para verificar nossa conectividade em nossa rede social restrita a nossa turma: Se eu sortear dois alunos, qual a probabilidade de eles se conhecerem?. Caso eles não se conheçam, qual a chance de eles terem pelo menos um amigo em comum?. Se verificarmos os amigos dos amigos, eventualmente encontraremos uma ligação entre eles? (Figura 5).

E por que isso importa?. Ter pleno conhecimento do funcionamento dessas interações podem nos ajudar a compreender o mundo e tomar decisões:

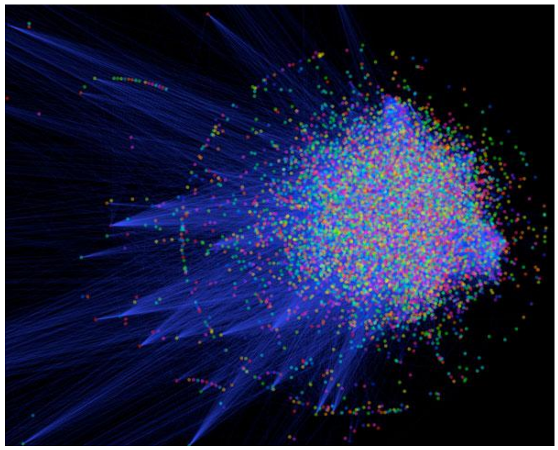


Figura 2: Universo

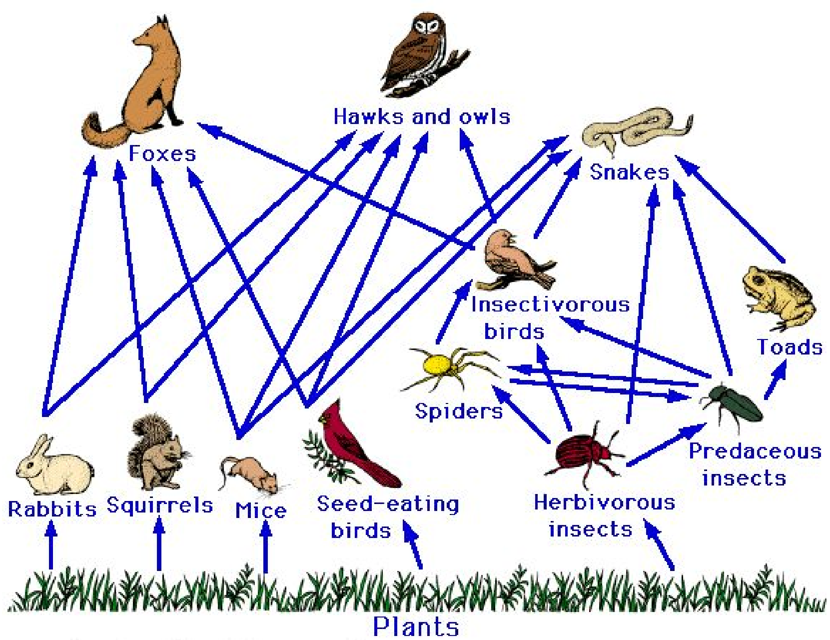


Figura 3: Cadeia Alimentar, que mostra a relação entre presas e depredadores.

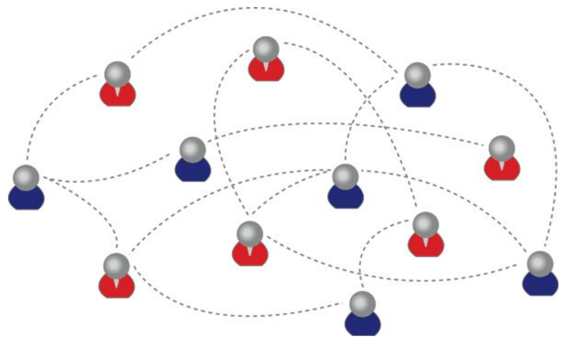


Figura 4: Interações entre os seres humanos.

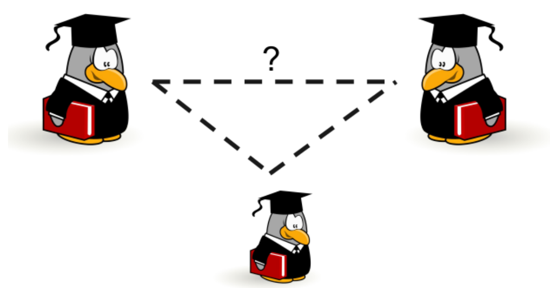


Figura 5: Qual é a probabilidade de dois alunos selecionados aleatoriamente se conhecerem?

- Estratégicas;
- Econômicas;
- Ambientais;
- Saúde Pública;
- dentre outras

Como exemplos temos: estratégias de negócio, militares, como criar um modelo econômico robusto, promover o equilíbrio ambiental, verificar dinâmica de queimadas, planejar campanhas de vacinação que proteja a população a um custo mínimo.

1.1 Comunicação e Redes

Essa interação e conexão entre objetos de estudo é chamado de REDES. A informação transmitida nas redes é a COMUNICAÇÃO. Na rede de Cadeia alimentar os objetos de estudo são os seres vivos e a comunicação é a energia que é transmitida da presa para o predador (Figura 6).

Figura 6: Fluxo de energia entre presa e depredador na Cadeia Alimentar.

Em uma rede de transporte os objetos são os pontos que servem de origem e destino (ex.: aeroportos) e a comunicação é o produto que está sendo transportado de um ponto à outro. (Figura 7).

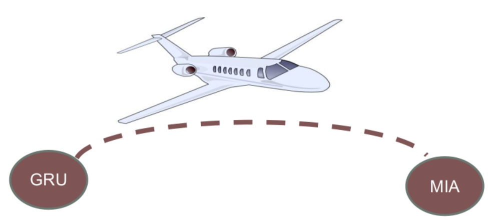


Figura 7: Voo de São Paulo (GRU) para Miami (MIA).

Já na rede social os objetos são as pessoas e a comunicação pode ser desde uma conversa até a transmissão de vírus (Figura 8).

Figura 8: Comunicação entre as pessoas.

1.2 Objetivos do Curso

Apresentar conceitos básicos de comunicação em redes:

- Modelos matemáticos;
- Métodos de análise;
- Ferramentas de apoio;

Mostrando o uso desses conceitos de forma multidisciplinar, envolvendo:

- Computação;
- Engenharia;
- Sociologia e;
- Biologia.

1.3 Conceitos Básicos

O conceito de Comunicação refere-se à transmissão de informação de um ponto A até um ponto B (Figura 9).

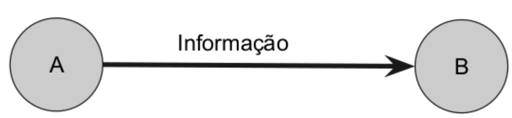


Figura 9: Comunicação

A "informação" deve ser entendida genericamente como o objetivo da transmissão, exemplos:

- Telefonar para seu colega para estudarem juntos;
- Transportar produtos em um navio;
- Transmitir gripe para outras pessoas.

A transmissão de uma doença ocorre quando um indivíduo infectado entra em contato com uma pessoa saudável (Figura 10). A doença se espalha conforme as pessoas infectadas interagem em seu ambiente social(Figura 11, 12). Cada pessoa tem uma capacidade específica de transmissão (Figura 13). Eventualmente o grupo todo acaba infectado (Figura 14).

Figura 10: Transmissão doenças.

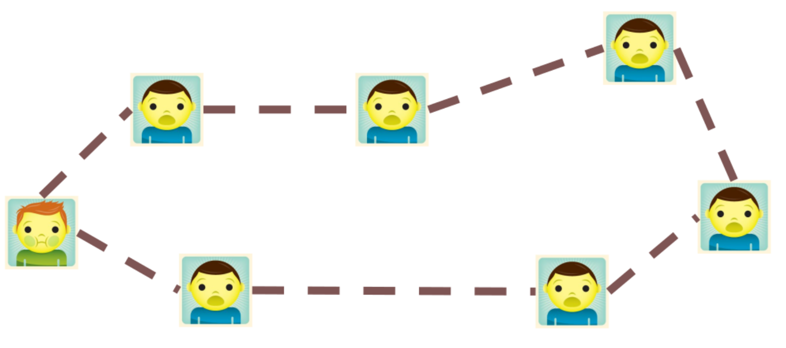


Figura 11: Transmissão doenças, uma pessoa infectada.

1.3.1 Redes

Essas estruturas interconectadas são denominadas REDES. Na matemática uma rede também é chamada de GRAFO. Cada ponto da rede é chamado de NÓ, também conhecidos como atores, vértices e objetos. A ligação que serve de caminho para o envio de informação entre dois nós é chamado de ARESTA

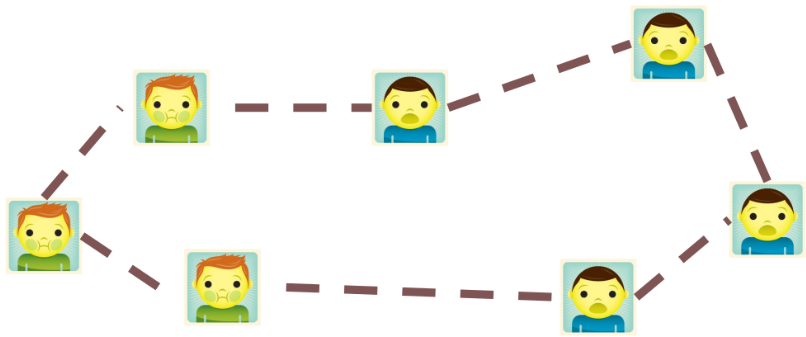


Figura 12: Transmissão doenças, poucas pessoas infectadas.

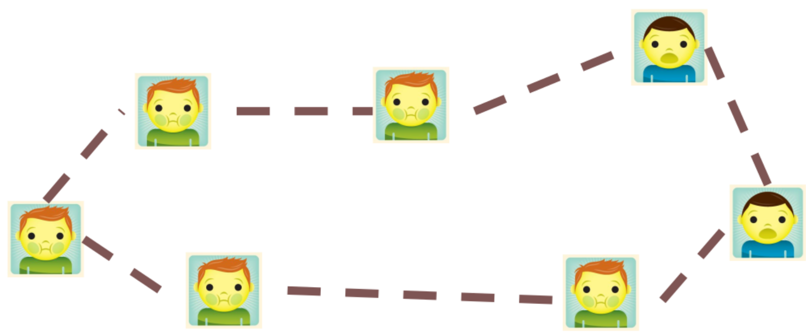


Figura 13: Transmissão doenças, muitas pessoas infectadas.

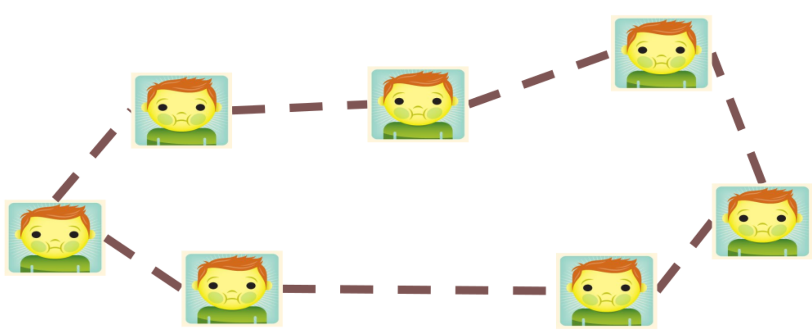


Figura 14: Transmissão doenças, todas as pessoas infectadas.

(Figura 15). Para facilitar a visualização de uma rede os nós são representados por círculos e as arestas por traços cheios (Figura 17).

Figura 15: Duas pessoas infetadas com relação entre elas pode-se representar como uma aresta entre dois vértices.

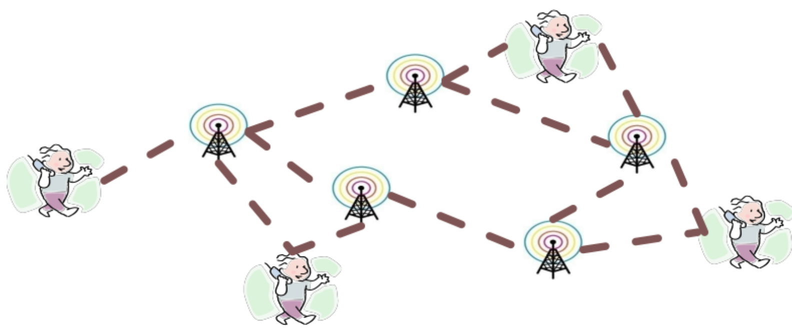


Figura 16: As Redes de Telecomunicações são um exemplo de Grafos.

1.3.2 Teoria dos Grafos

É a área da matemática que estuda essas redes. Na Teoria dos Grafos, um grafo G é um par ordenado $G(V,E)$ composto por um conjunto V contendo n vértices ou nós e um conjunto de arestas E contendo pares não ordenados de nós explicitando a relação entre eles.

$$G = (V, A)$$
$$V = \{v_1, v_2, \dots, v_n\}$$
$$E = \{(x, y)|x, y \in V, x \neq y\}$$

Cada nó costuma ser identificado por um nome (Figura 18):

$$V = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

As arestas passam a ser o conjunto de pares de nós que tem uma ligação:

$$E = \{(1,2), (2,3), (2,4), (2,5), (3,4), (4,8), (5,6), (5,7), (6,7), (7,8), (7,9), (8,9)\}$$

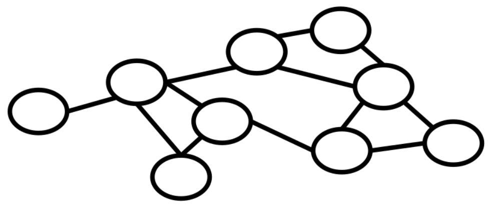


Figura 17: Grafo com círculos como vértices e linhas entre eles representando as arestas.

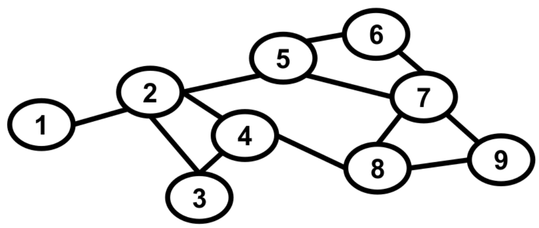


Figura 18: Grafo com números nos vértices para identifica-os.

1.3.3 Ciência das Redes

As redes reais apresentam características distintas que contém informações úteis, um objetivo do estudo das redes é extrair tais informações (Figura 19). Por exemplo, existe alguma espécie predadora que está causando desiquilíbrio ecológico?, alguma espécie ameaçada?, que grupos de pessoas devem ser vacinadas prioritariamente, qual o melhor trajeto para o transporte de uma mercadoria?

Em uma rede pequena aparentemente é fácil descobrir essas relações. Mas e as redes que temos na prática?, por exemplo:

- Uma rede de portos marítimos tem mais de 700 nós!;
- Uma rede de cadeia alimentar pode contar com mais de 8 milhões de seres vivos!;
- Uma rede epidêmica envolve milhares ou até milhões de pessoas em contato todos os dias!

1.3.4 Redes

O estudo de comunicação e redes se envolve:

- Determinar e conhecer o melhor caminho da transmissão;

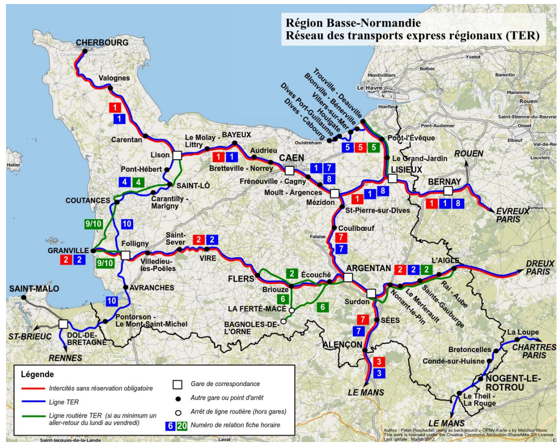


Figura 19: Rede do Transporte expresso regional em França. `https://commons.wikimedia.org/w/index.php?curid=18745016`

- Entender como difundir ou restringir acesso a informação;
- Explicar o crescimento dessa rede e suas implicações.

Um ponto importante no estudo das redes é a classificação, pois existem diversos tipos de redes reais com características distintas:

- Redes Sociais;
- Redes de Transporte;
- Redes de Informação;
- Redes Biológicas.

As Redes Sociais estudam as relações entre indivíduos, grupos e organizações. Por exemplo:

- rede de amizade;
- rede de epidemias;
- rede de empresas (Fornecedoras <-> distribuidoras).

As Redes de Transporte são construídas de forma otimizada para o transporte de algum produto ou serviço. Por exemplo:

- rede de transporte coletivo;
- rede de vasos sanguíneos;
- rede de distribuição de água, rios.

As Redes de Informação interconectam conhecimentos formalizados. Por exemplo:

- Wikipédia;
- rede de citação de artigos;
- redes de palavras.

As Redes Biológicas estudam a interação nos sistemas biológicos. Por exemplo:

- redes neurais;
- redes de proteínas;
- cadeia alimentar.

1.3.5 Exemplo de Redes Reais : Famílias de Florença – Séc. XV

Essa rede mostra a conexão entre diversas famílias poderosas de florença no século 15 Cada nó dessa rede representa uma família e as arestas indicam que houve pelo menos um casamento entre membro das duas famílias conectadas (Figura 20).

Exceto pela Pucci, toda família pode entrar em contato com qualquer outra, através de, no máximo, CINCO intermediários. Mas na média elas estão bem próximas, sendo necessários apenas 2,5 intermediários.

As famílias Medici, Guadagni e Strozzi são as mais bem posicionadas estratégicamente a Medici foi a mais influente em sua época. . . , as Strozzi e Guadagni foram as mais ricas. Notamos também a existência de grupos de famílias ilustrado pelas cores dos nós. A formação de grupos ajuda a proteger os negócios.

Outro fato a se notar é a existência de "triângulos". Esses triângulos representam o quão fechada as famílias são entre si. Com essas informações podemos responder algumas perguntas. Qual melhor forma de me aproximar da família Strozzi? Se faço parte da família Lambeteschi, a melhor forma é ser apresentado a família Bischeri, através dos Guadagnis!

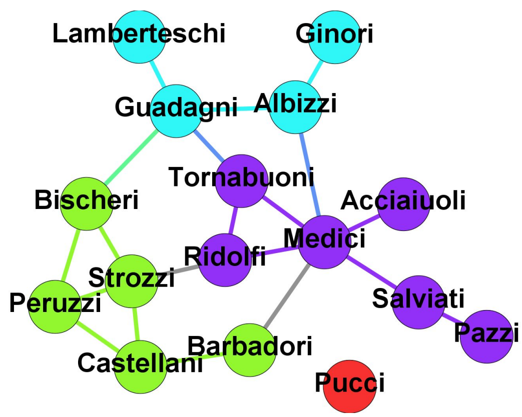


Figura 20: Rede das Famílias Florência representadas por um grafo onde os cores representam grupos de famílias.

2 Teoria dos Grafos

2.1 As Sete Pontes de Königsberg

Na cidade de Köningsberg localizada na Prussia (agora chamada de Kaliningrad na Russia), eles tinham sete pontes cruzando o rio Prególia que dividia a cidade em 4 partes (Figura 21). As pessoas dessa cidade costumavam caminhar por ela todo domingo. Para tornar o passeio mais divertido elas decidiram criar uma brincadeira, alguém consegue fazer sua caminhada dominical de tal forma que, você passe por todas as 4 regiões da cidade e atravesse todas as pontes uma única vez?.

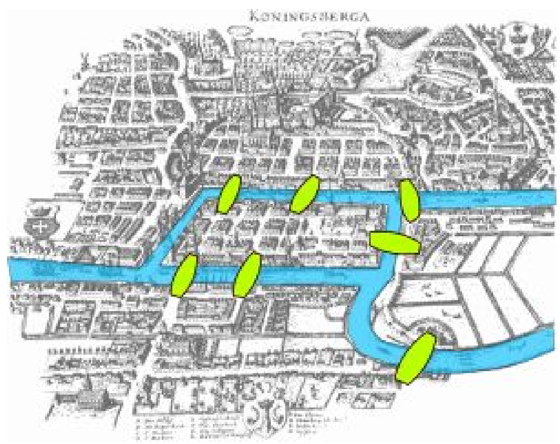


Figura 21: As Sete Pontes de Königsberg

As pessoas começaram a anotar as combinações que elas testavam de tal forma a enumerar todas elas até encontrar uma solução!. Então, Euler apareceu e disse "tentar resolver isso dessa forma, uma solução de cada vez, não vai nos trazer nada de novo. Pode ser que encontrem uma solução, mas isso não fará com que entendamos o problema!". Então ele disse "Vamos nos concentrar nas partes da cidade, não nas pontes. Vamos dar nomes para cada parte da cidade" (Figura 22). Uma explicação interessante pode-se encontrar no `https://www.youtube.com/watch?v=W18FDEA1jRQ`.

2.2 Grafo Simples

Um grafo simples G e um par $G(V, E)$, onde V é um conjunto não vazio de objetos denominados vértices (ou nós) e E (do inglês *edges* - arestas) é um subconjunto de pares não ordenados de V, nos quais não se tem uma direção, é dizer o fluxo de dados é bidirecional. O problema das pontes de Königsberg pode ser representado utilizando um grafo simples, onde as áreas representam os vértices e os pontes representam as arestas (Figura 23).

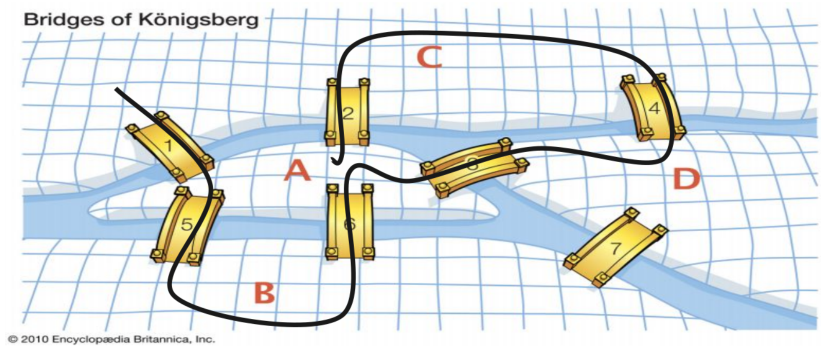


Figura 22: Mapa com as partes da cidade nomeadas e com uma tentativa de recorrer todas as pontes.

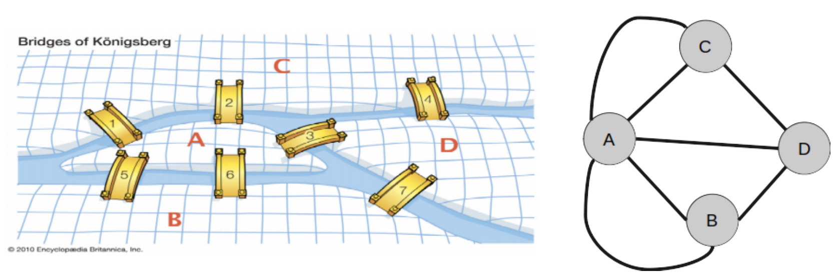


Figura 23: Grafo com a representação das áreas e pontes de Königsberg

Se pensamos em apenas três pedaços da cidade e interconectá-las com duas pontes, se tem essa representação. Nessa situação eu posso fazer o caminho $B \to D \to C$ ou $C \to D \to B$(Figura 24).

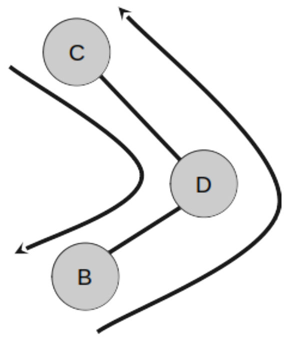


Figura 24: Grafo representando 3 áreas e 2 pontes. As zetas representam a direção com que pode se fazer o percurso.

Com quatro pontes eu tenho $B \to D \to C \to D \to B$ ou $C \to D \to B \to D \to C$ ou $D \to C \to D \to B \to D$. Podemos começar de qualquer ponto e conseguimos uma solução (Figura 25).

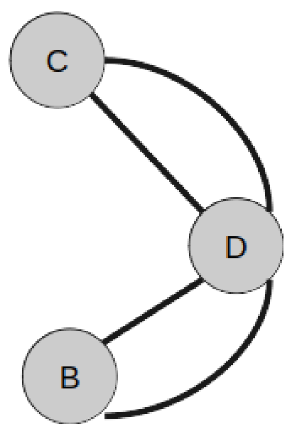


Figura 25: Grafo representando 3 áreas e 4 pontes.

Com três pontes eu devo tomar cuidado! Não posso fazer $B \to D \to C$ pois faltara percorrer uma das pontes, nem $C \to D$. Porém, eu posso fazer $D \to B \to D \to C$ ou $C \to D \to B \to D$(Figura 26). O que essas rotas possuem em comum?, elas começam ou terminam em C ou em D!. Euler percebeu que, se eu tiver um número ímpar de pontes eu devo começar ou terminar meu trajeto no local que possui acesso a um número ímpar das pontes.

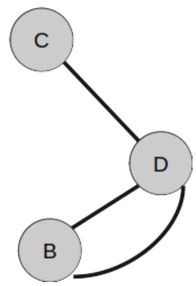


Figura 26: Grafo representando 3 áreas e 3 pontes.

Uma outra situação observada por Euler é quando cada região possui um número par de pontes, você pode começar de qualquer ponto e terminar nele mesmo! Isso implica que podemos eliminar esse ciclo do problema para nos concentrarmos em uma solução!! (Figura 27).

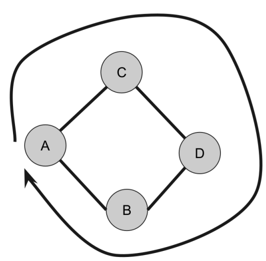


Figura 27: Se cada vértice tem numero par de vértices, se pode fazer um ciclo que comece e termine no mesmo ponto.

Vamos ver o que aprendemos até então sobre o problema! Sempre que chegamos em uma parte de terra passando por uma ponte, se esse não for a região final, precisamos necessariamente ter outra ponte para sairmos de lá!. Em outra palavras:

- Se o vértice não for o ponto inicial e nem final, ele necessariamente precisa ter um número par de arestas.
- Os vértices iniciais e finais podem ter número ímpar, porém ambos devem possuir a mesma paridade.

Na situação que um certo nó tem um número ímpar de arestas, um outro nó também precisa ter um número ímpar, fazendo com que esses dois nós sejam os pontos de partida e chegada. Então temos a outra regra que é: Os vértices iniciais e finais podem ter número ímpar, porém ambos devem possuir a mesma paridade. Ou seja, se o nó inicial for par, o final também deve ser; se o inicial for ímpar, o final também deve ser.

Após a segunda guerra mundial, 2 das pontes foram destruídas (Figura 28). E agora temos um problema diferente com 5 pontes!. Dessa forma o vértice C e B possuem 2 arestas, e A e D possuem 3! (Figura 29). Partimos fazendo o ciclo, e sobra apenas uma aresta para percorrermos chegando em uma solução!(Figura 30).

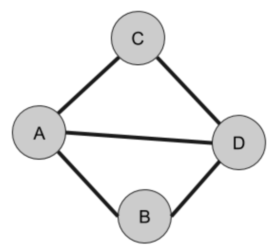


Figura 28: Grafo dos pontos após a segunda guerra mundial.

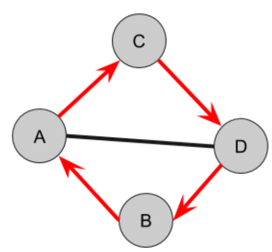


Figura 29: Grafo dos pontos após a segunda guerra mundial.

2.2.1 Grau de um vértice

Em grafos, o número de arestas que parte ou chega em um nó é chamado de GRAU (Figura 31). Vamos relacionar cada nó da Rede dos Pontes de Königsberg com o grau correspondente (Figura 32).

O grau de um vértice nas redes sociais representa a importância/popularidade. Em uma rede social, o grau de um nó nos ajuda a medir o quanto aquela determinada pessoa é importante ou popular. Em uma rede de amizades, por exemplo o Facebook, o grau de uma pessoa é a quantidade de amigos que ela possui. Se ela possui um grau muito acima da média de seu grupo, ela pode ser considerada popular. O grau na rede do Instagram mede a soma dos seguidores e de quem ela segue. Caso esse número seja alto, ela pode ser uma pessoa útil

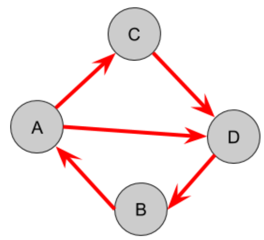


Figura 30: Grafo dos pontos após a segunda guerra mundial.

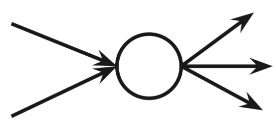


Figura 31: Vértice de um grafo com arestas de entrada e de saída.

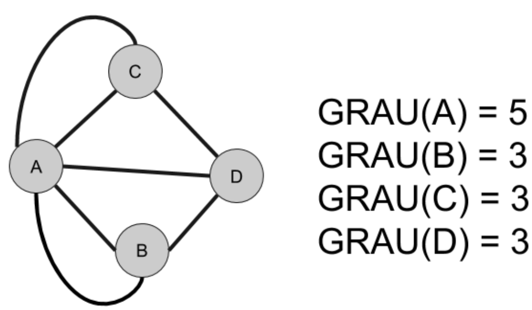


Figura 32: Graus dos nós da rede dos pontes de Königsberg.

para disseminar informação para toda a rede, por exemplo, fazer propaganda para alguma marca (Figura 33).

Figura 33: Grau de uma pessoa numa rede social.

2.3 Grafo Direcionado

Um grafo orientado, grafo dirigido, grafo direcionado ou diágrafo é um par $G = (V, E)$ onde V é o conjunto de vértices e E é o conjunto de arestas direcionadas, também chamadas setas, isso quer dizer que o fluxo da informação é bidirecional (Figura 34).

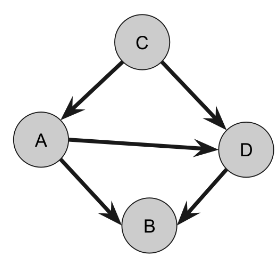


Figura 34: Grafo Direcionado com 4 vértices.

A Cadeia alimentar representa um grafo direcionado, pois as arestas possuem direção (Figura 35). Nessa rede uma raposa recebe energia do coelho, mas o coelho não recebe da raposa. Podemos pensar em três tipos importantes de nós:

- Transmissores, aqueles que transmitem energia para várias espécies

- Receptores, Os que recebem energia de várias espécies
- Intermediários, Nós que enviam e recebem de diversas espécies

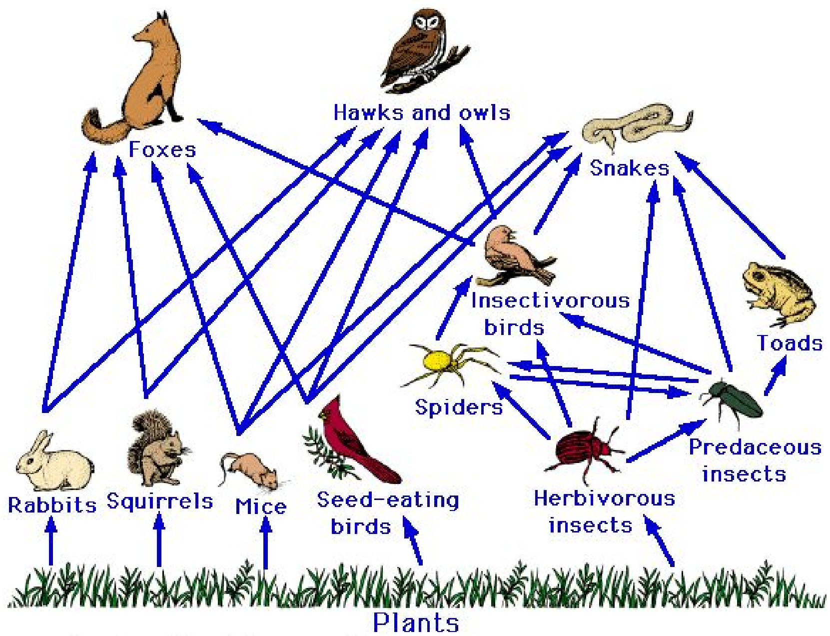


Figura 35: Cadeia Alimentar

Nós que transmitem energia para várias espécies são caracterizadas por aquelas que conseguem gerar energia do sol (plantas). Como elas não consomem energia de nenhuma espécie e, diversas espécies consomem energia através dela, as plantas tem várias arestas apontando parar fora dela. Essa quantidade é mensurada como GRAU DE SAÍDA, que é o número de arestas saindo do nó.

Nós que recebem energia de várias espécies são caracterizadas por aquelas que são grandes predadores e não são presas de outras espécies. Nesse caso, essas espécies contêm arestas apenas apontando para elas, ou seja, chegando nelas. Essa quantidade é mensurada como GRAU DE ENTRADA, que é o número de arestas chegando em um nó. Já as espécies intermediárias que são predadoras de diversas espécies mas também servem de presas tem tanto arestas chegando nelas como saindo delas. Essas espécies podem ser classificadas pelo GRAU DE ENTRADA, pelo GRAU DE SAÍDA e pela soma desses dois que dá o GRAU TOTAL do nó.

2.3.1 Grau de um vértice em Grafos Direcionados

São utilizados trés graus: Grau de Entrada G_{ent} (conjunto de arestas que chegam ao vértice), Grau de Saída G_{saida} conjunto de arestas que saem do vértice), Grau Total $G(v_i)$ (soma das arestas que saem e entram ao vértice).

$$G_{ent} = |\{(v_j, v_i)|(v_j, v_i) \in E\}|$$
$$G_{saida} = |\{(v_i, v_j)|(v_i, v_j) \in E\}|$$
$$G(v_i) = G_{ent} + G_{saida}$$

2.4 Caminhos em uma Rede

Frequentemente é possível utilizar mais de um caminho para transmitir uma informação entre dois nós. Um caminho de um grafo G é a sequência de nós $(v_1, v_2, \ldots, v_k)$ tal que dois vértices consecutivos v_i e v_{i+1} sejam adjacentes. Apresentamos um exemplo (Figura 36) de dois caminhos para um mesmo par de nós 2 e 7.

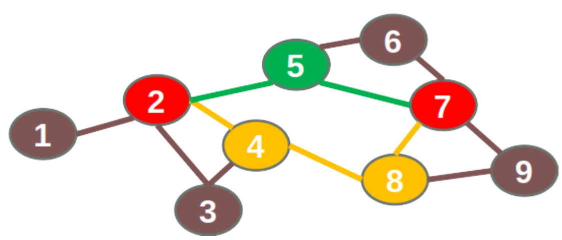


Figura 36: Um grafo simples com dois caminhos entre os nos 2 e 7, um caminho de cor verde e outro amarelo.

Alguns dos caminhos possíveis são:

- Caminho 1 : $(2, 4, 8, 7)$;
- Caminho 2 : $(2, 5, 7)$;
- Caminho 3 : $(2, 1, 2, 5, 7)$;
- Caminho 4 : $(2, 4, 7)$.

Já que os caminhos tem que ser adjacentes, o Caminho 4 não forma um caminho válido. A existência de múltiplos caminhos faz com que exista a necessidade em encontrarmos os melhores caminhos para transmitir a informação! Por exemplo, se estou levando um produto de um ponto à outro, quero passar pelo menor número de nós.

O primeiro fato a ser observado é que caminhos que tem repetição de nós não podem ser os melhores, pois prolonga o caminho de forma redundante.

Esse critério faz que o Caminho 3 não seja ótimo. Levando em conta apenas o número de nós em um caminho, o Caminho 2 é o melhor!, pois o Caminho 1 possui tamanho 3 e o Caminho 2 possui tamanho 2.

O comprimento de um caminho é o número de arestas desse caminho. Mas em certas redes cada aresta tem um valor numérico que representa o custo da transmissão de informação. Esse custo pode representar:

- capacidade de receber o sinal de uma torre de transmissão (ex.: distância entre elas);
- combustível gasto para percorrer tal aresta;
- energia gasta para consumir a presa menos o ganho de energia ao comê-la, etc.

As redes com arestas que apresentam custo são denominadas Redes Ponderadas (Figura 37).

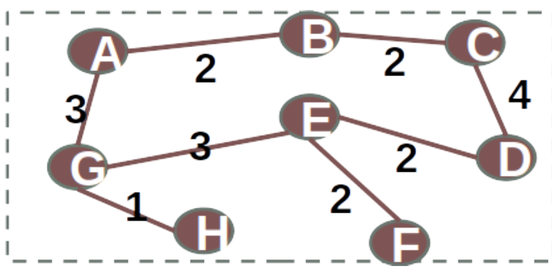


Figura 37: Exemplo de uma Rede Ponderada.

Se somarmos os custos para cada caminho (Figura 38), percebemos que agora o Caminho 1 é o melhor!.

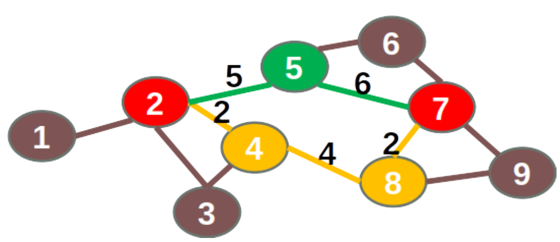


Figura 38: Um grafo simples com caminhos, com pesos nas arestas.

2.4.1 Caminhos em Redes Direcionadas

Em redes direcionadas, os caminhos devem respeitar a direção das setas. Se as arestas representam vias de mão única, os carros que trafegam nela somente podem seguir em uma única direção (Figura 39).

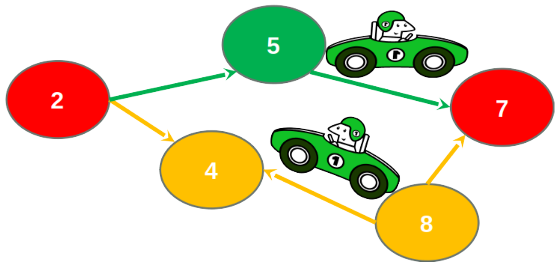


Figura 39: Ruta de automóveis apresentado como um grafo dirigido.

Note que com as arestas direcionadas o Caminho 1 se torna impraticável (Figura 40).

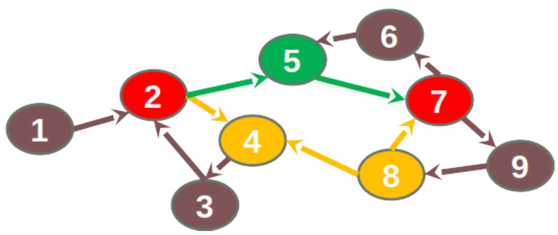


Figura 40: Um grafo dirigido representando caminhos.

2.4.2 Conetividade da Rede

Em algumas redes nem todos os nós estão conectados. Pode existir um grupo de nós que não são acessíveis (Figura 41), p. ex. coletamos apenas parte das informações de usuários do twitter. Isso ocorre quando:

- Fazemos uma aquisição incompleta dos dados da rede;
- Ocorrem falhas pontuais em certos nós;
- A rede ainda está em formação.

2.4.3 Conexão

Uma rede é dita CONEXA ou CONECTADA se existe pelo menos um caminho conectando quaisquer dois nós $\forall v_i, v_j \in V, \exists$ caminho(v_i, v_j), caso contrário ela é dita DESCONEXA ou DESCONECTADA $\forall v_i, v_j \in V$, $\nexists$ caminho(v_i, v_j).

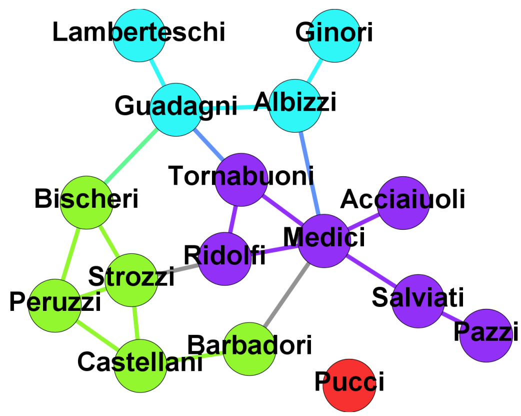


Figura 41: Um grafo dirigido com dois caminhos entre os nós 2 e 7, um caminho de cor verde e outro amarelo.

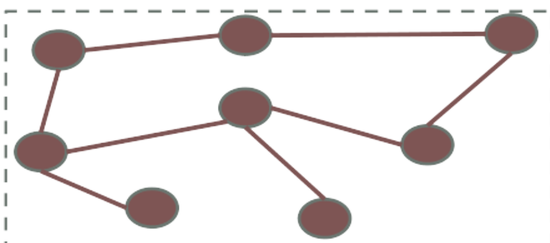


Figura 42: Exemplo rede conectada.

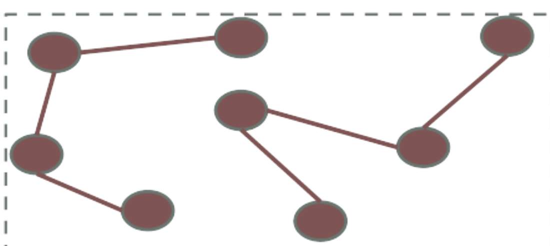


Figura 43: Exemplo rede desconectada.

2.5 Modelagem Computacional

O primeiro estudo de redes complexas surgiu em 1934, por Jacob L. Moreno, em seu livro Who Shall Survive?, com um estudo quantificando o papel de cada indivíduo na sociedade na qual reside (Figura 44) [9] .

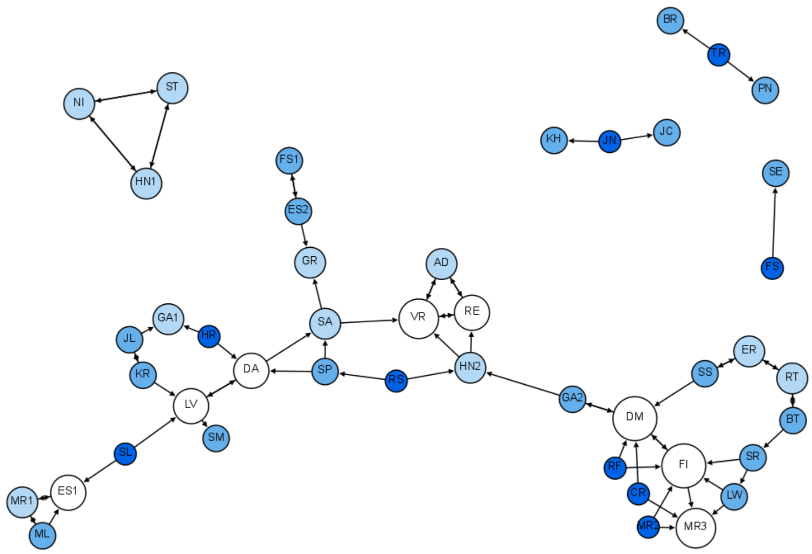


Figura 44: Grafo do estudo de redes complexas de Jacob L. Moreno 1934. http://www.martingrandjean.ch/social-network-analysis-visualization-morenos-sociograms-revisited/.

Esse estudo foi feito em pequenos grupos sociais como, por exemplo, uma classe de 5ª. série de uma escola. Jacob fez um estudo visual e através de contagem dos padrões observados (p.ex. quantas formações de triângulos foram encontrados em uma rede). Isso só foi possível pela dimensão reduzida de sua coleta de dados. Porém, com a redução dos dados não é confiar em certas propriedades estatística das redes construídas.

2.5.1 Estudo das Redes

Com o aumento de recursos computacionais o estudo das redes ganhou força recentemente. Temos computadores com capacidade de coleta de grande volume de dados. Por exemplo, 1 dia de coleta de dados do Twitter sobre um tema específico ocupa cerca de 0.1% de um tamanho popular de HD.

Além disso, é possível calcular estatísticas de redes com milhares, milhões e até bilhões de arestas em um tempo razoável. Com isso podemos focar nas propriedades estatísticas dos grafos ao invés de apenas alguns nós individuais.

Capacidade de armazenamento + capacidade de processamento = estatísticas gerais das redes, busca por padrões de interesse, análises da dinâmica da informação.

2.6 Problemas reais onde as redes são aplicadas

Além da representação e estudo das relações, existem muitos problemas e aplicações do mundo real que podem (e devem) ser modelados em forma de redes. Ao modelar um problema em forma de rede temos como vantagens:

- Visualizar melhor o problema;
- Aplicar diversos algoritmos já existente;
- Analisar as propriedades da rede.

2.6.1 Problema do Caixeiro Viajante

Imagine um vendedor ambulante que deseja encontrar o caminho mais curto que passe por todas as cidades de seu país. Um problemas clássico é o do CAIXEIRO VIAJANTE. No exemplo da Figura 45, vemos o caminho mais curto que passa por diversas cidades da Alemanha. Já vimos a definição de caminho, mais adiante aprenderemos alguns algoritmos sobre o assunto.

Figura 45: Mapa da Alemanha com uma ruta possível entre as cidades.

Uma possível aplicação desse problema é na construção de um tour de shows de uma banda internacional que deve passar por diversos países. Na Figura 46 temos um exemplo da Euro Tour da banda Kiss. Esse parece o melhor trajeto possível? Vale notar que alguns casos apresentam restrições extras como impossibilidade de fazer certo trajeto, restrição de conflito de agenda para certos países, etc.

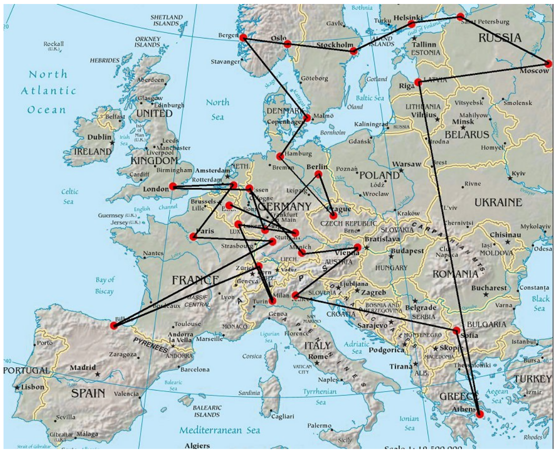


Figura 46: Mapa Kiss Tour `http://www.kissinuk.com/bb/viewtopic.php?f=1&t=3458`

Um outro exemplo é a definição da rota de entrega de correspondências, dado que o carteiro deve entregar cartas em um certo conjunto de residências, qual a melhor ordem para cobrir toda sua região andando o menos possível? E se o planejamento levar em conta a parada para almoço? Qual trajeto passa por algum restaurante no horário apropriado? (Figura 47).

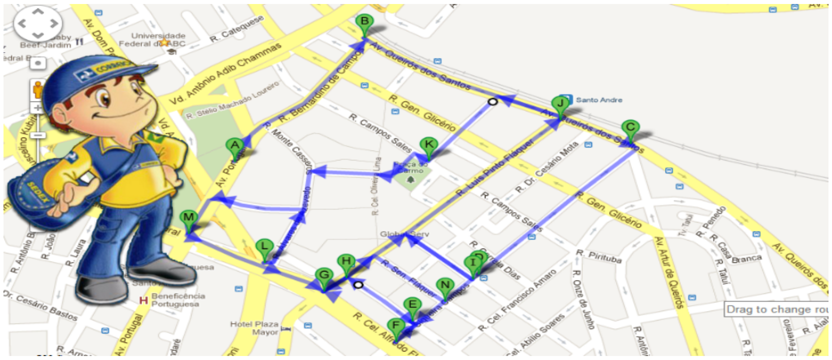


Figura 47: Mapa de um caminho de correio.

Perfuração de placas de circuitos elétricos. Qual a sequência de furos que corresponde à um menor número de movimentos do perfurador? Isso pode reduzir o tempo de fabricação dos circuitos além de minimizar o desgaste do perfurador (Figura 48).

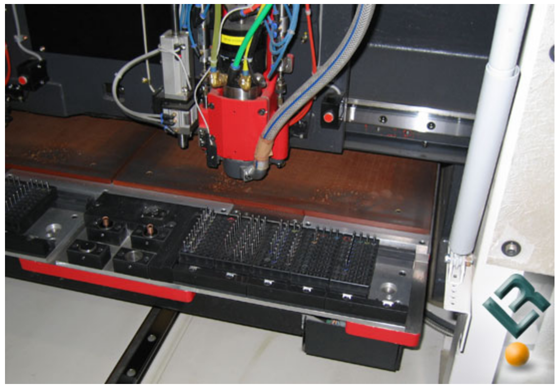


Figura 48: Placa de circuito elétrico `http://www.legitreviews.com/article/525/2/`.

2.6.2 Modelar usando Redes

Uma empresa que realiza entregas na Grande São Paulo possui um centro de distribuição e um caminhão (Figura 49). Qual caminho o caminhão deve percorrer de modo a realizar todas as entregas com a menor quilometragem?

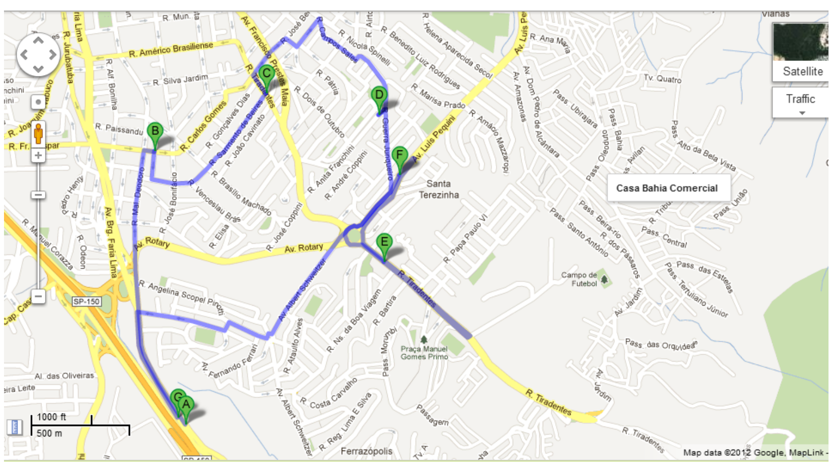


Figura 49: Rota do caminhão de entregas.

Uma maneira de resolver o problema é modelar-o como uma rede ponderada. Cada cruzamento é representado por um vértice e cada rua por uma aresta ($V =$

cruzamentos E = ruas). O peso pode-se representar como Tempo Gasto / Custo Total. Mas podemos também minimizar:

- Tempo gasto na viagem (peso = fluxo)
- Custo total da viagem (peso = pedágios + gasolina)

E se pudermos comprar mais de um caminhão?, teríamos então 3 problemas:

- Quantos caminhões minimizam o custo?
- Quais pontos cada caminhão deve cobrir?
- Quais rotas cada caminhão deve fazer?

Para uma solução exata um computador capaz de testar 1 milhão de rotas por seg. tomaria os tempos descritos na Tabela 1.

Nós	Rotas Possíveis	Tempo
5	24	$\sim$0
10	362.880	$<$ 1 seg.
15	87 bilhões	$\sim$24 horas
20	1,2 x 10_{17}	$>$92 mil anos
25	6,2 x 10_{23}	$\sim$4,7 x 10_{11} anos
30	8,8 x 10_{30}	$\sim$6,7 x 10_{18} anos
35	2,95 x 10_{38}	$\sim$2,2 x 10_{26} anos

Tabela 1: Dados de tempo em que um computador capaz de testar um milhão de rotas por seg. consegui resolver o problema, $\sim$4,7 x 10_{11} anos é maior do que a idade do universo.

Todos esses problemas pertencem a classe de problemas NP-Completo. O tempo necessário para obter a solução exata demanda que a gente teste todas as permutações de rotas possíveis, ou seja, para n nós precisamos calcular $n!$ possibilidades. Precisamos de formas mais inteligentes de encontrar soluções para esses problemas...

2.7 Representação Computacional

Como representar uma rede? Redes complexas reais geralmente contêm muitos nós e arestas sendo impossível visualiza-la por completo.Para trabalhar com essas redes é necessário utilizarmos recursos computacionais. Com isso vem a necessidade de adotarmos uma forma de representar a rede numericamente para a leitura e processamento da mesma.

2.7.1 Rede como matriz

Uma forma de representar as redes é em forma matricial. A matriz A, chamada de matriz de adjacência, de dimensão $|V|x|V|$ tem um valor igual à 1 no elemento a_{ij} se existe uma aresta ligando o nó i ao nó j, e 0 caso contrário. Onde se tem a $|V|$ como o numero de nós.

2.7.2 Representação: Matriz de Adjacência

Nas Redes não direcionadas (Figura 50), somando os valores de cada linha temos o grau do nó correspondente!. Já que as arestas permitem um intercambio de informação bidirecional, sua Matriz de adjacência é simétrica.

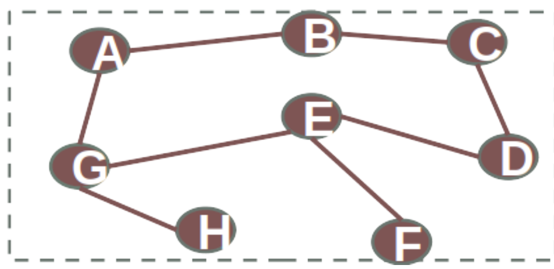


Figura 50: Rede não direcionada.

	A	B	C	D	E	F	G	H
A	0	1	0	0	0	0	1	0
B	1	0	1	0	0	0	0	0
C	0	1	0	1	0	0	0	0
D	0	0	1	0	1	0	0	0
E	0	0	0	1	0	1	1	0
F	0	0	0	0	1	0	0	0
G	1	0	0	0	1	0	0	1
H	0	0	0	0	0	0	1	0

Figura 51: Representação de uma rede não direcionada mediate uma matriz de adjacência.

Nas Redes Direcionadas temos os graus de entrada e saída. Somando os valores de cada coluna temos o grau de entrada! Somando os valores de cada linha temos o grau de saída!. Não se tem simetria na Matriz de adjacência. (Figura 47).

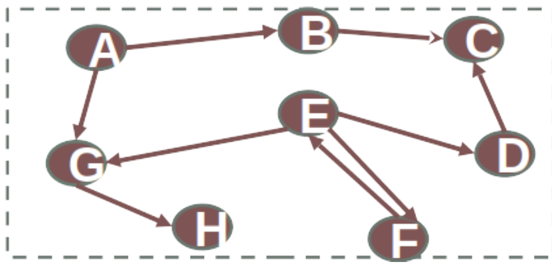


Figura 52: Exemplo rede direcionada.

	A	B	C	D	E	F	G	H
A	0	**1**	0	0	0	0	**1**	0
B	0	0	**1**	0	0	0	0	0
C	0	0	0	0	0	0	0	0
D	0	0	**1**	0	0	0	0	0
E	0	0	0	**1**	0	**1**	**1**	0
F	0	0	0	0	**1**	0	0	0
G	0	0	0	0	0	0	0	**1**
H	0	0	0	0	0	0	0	0

Figura 53: Representação de uma rede direcionada mediate uma matriz de adjacência.

Nas Redes Ponderadas podemos colocar os pesos diretamente na matriz de adjacência (Figura 47). Mas é necessário o uso de um valor especial representando a ausência de ligações.

Ao realizar a operação Adj^2 na matriz não ponderada, obtemos um resultado interessante. A matriz original nos indica a existência de um caminho entre dois nós de tamanho 1. Essa matriz ao quadrado, nos mostra quantos caminhos de tamanho 2 existem entre dois nós. Por exemplo, o caminho $A - C$ vai obter o

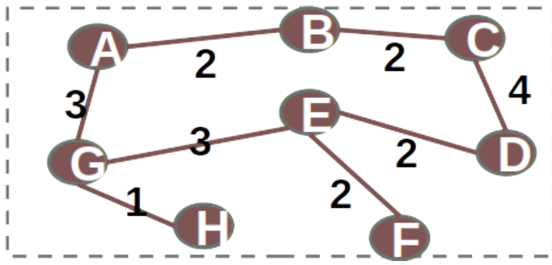


Figura 54: Exemplo rede ponderada.

	A	B	C	D	E	F	G	H
A	0	2	0	0	0	0	3	0
B	2	0	2	0	0	0	0	0
C	0	2	0	4	0	0	0	0
D	0	0	4	0	2	0	0	0
E	0	0	0	2	0	2	3	0
F	0	0	0	0	2	0	0	0
G	3	0	0	0	3	0	0	1
H	0	0	0	0	0	0	1	0

Figura 55: Representação de uma rede ponderada mediate uma matriz de adjacência.

valor 1 pois $A - B$ e $B - C$ tem valores 1. Consequentemente, Adj, nos indica quantos caminhos existem entre dois nós com tamanho n. (Figura 57).

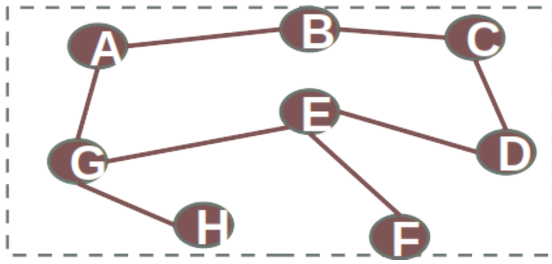


Figura 56: Rede não direcionada

	A	B	C	D	E	F	G	H
A	2	0	1	0	1	0	0	1
B	0	2	0	1	0	0	1	0
C	1	0	2	0	1	0	0	0
D	0	1	0	2	0	1	1	0
E	1	0	1	0	3	0	0	1
F	0	0	0	1	0	1	1	0
G	0	1	0	1	0	1	3	0
H	1	0	0	0	1	0	0	1

Figura 57: Matriz de adjacência depois de aplicar a potencia.

2.7.3 Rede como lista de Arestas

Alguns programas de análise de redes utilizam a representação de lista de arestas para o processamento. Essa lista geralmente pode conter atributos para os nós e para as arestas (Figura 58).

$$V = \{A, B, C, D, E, F, G, H\}$$

$$E = \{(A, B), (B, C), (C, D), (D, E)\,(E, F), (E, G), (G, H), (A, G)\}$$

Por exemplo, em redes sociais os atributos podem ser:

- Pessoas (nós): idade, sexo, nacionalidade, renda, escolaridade.
- Relacionamento (arestas): tipo de relação, grau da relação, tempo da relação.

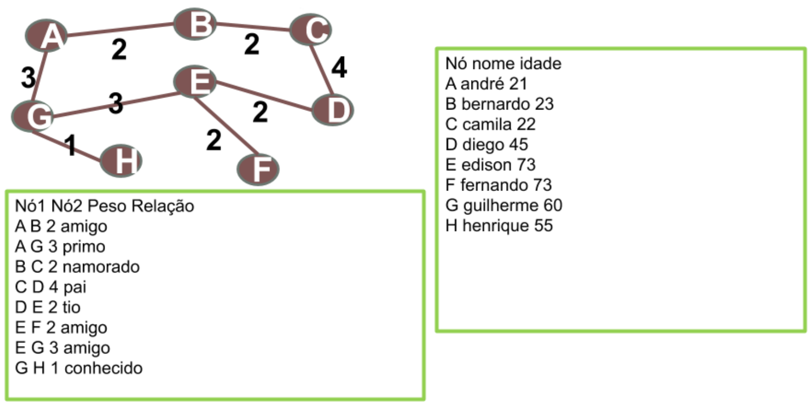


Figura 58: Exemplos de uma rede com as arestas representando atributos.

3 Percurso em Grafos

3.1 Distância entre os nós

Se eu escolher duas famílias aleatórias nessa rede (Figura 59), qual o tamanho do menor caminho entre elas que eu devo esperar? Tomem como exemplo a família Ginori e a Peruzzi. O menor caminho entre elas é passando por Albizzi, Guadagni e Bischeri. Esse caminho tem tamanho 4 pois tem 4 arestas entre elas. Em uma rede muito grande é interessante termos o valor esperado desse tamanho.

Figura 59: Rede de Famílias Italianas.

Vamos definir a distância entre dois nós, v_i e v_j, como o tamanho do menor caminho entre esses dois nós: $d(v_i) =$ Tamanho do menor caminho entre v_i e v_j. A distância média da rede é calculada como a média da distância entre todos os pares de nós diferentes entre si. Esse valor reflete o tamanho esperado entre dois vértices quaisquer.

$$\bar{d} = \frac{1}{n \cdot (n-1))} \sum_{v_i, v_j \in V, v_i \neq v_j} d(v_i, v_j)$$

Uma outra medida interessante é o diâmetro que indica a distância máxima que pode ser observada na rede. Podemos pensar nesse valor como um pior caso das distâncias.

$$diam = max_{v_i, v_j} d(v_i, v_j)$$

Alguns exemplos donde é importante a medição entre dois nós são:

- O tempo médio que um produto é entregue a um consumidor;

- Quão rápido um boato será difundido em uma rede social;
- Estimar a velocidade que um vírus se espalha.

A importância de conhecer o melhor percurso entre dois nós em uma rede transcende a simples economia de custos e recursos. Em muitos casos a própria estrutura da rede é determinada pela própria otimização do caminho. Uma rede com distância média reduzida é associada com eficiência de transporte. Pense em uma rede metroviária, construindo essa rede com uma distância média reduzida implica em um menor tempo de viagem para seus usuários (Figura 60). Em uma rede de distribuição de energia, uma rede com menor distância média tende a ter menor probabilidade de quedas de energia (Figura 61).

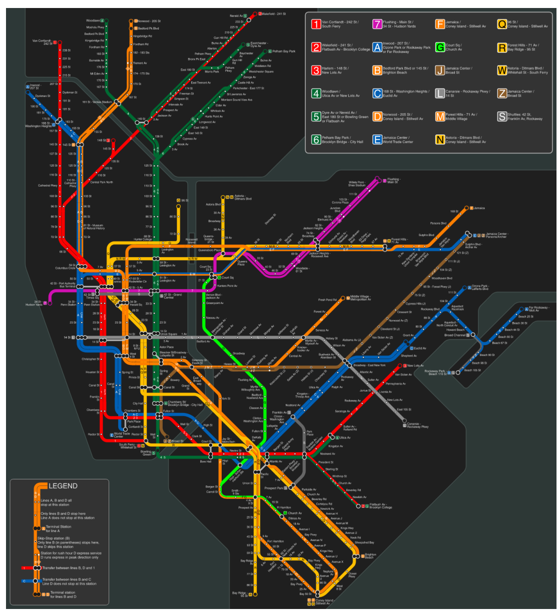


Figura 60: Subway New York. "File:New York Subway Map Alargule.svg"by Alargule is licensed under CC BY-SA 3.0

3.2 Distancia em Redes não Ponderadas

Em uma rede social os nós são representados pelas pessoas pertencentes à rede e as arestas indicam que duas pessoas tem alguma relação de contato entre si. Por exemplo, no Facebook os nós são os usuários cadastrados e as arestas indicam que duas pessoas fazem parte da lista de amigos uma da outra (Figura 62).

Certo usuário resolve publicar em seu mural uma propaganda de um produto desenvolvido por ele. Como essa propaganda se espalha pela rede? Supondo

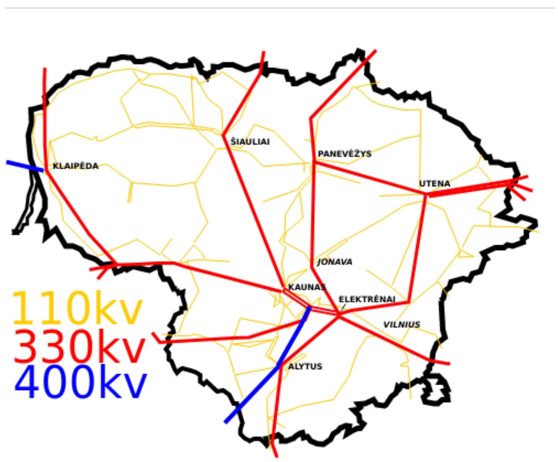


Figura 61: Rede de Energia Lituânia. "File:Lithuania High Voltage Electricity Links.svg"by Bearas is licensed under CC BY-SA 4.0

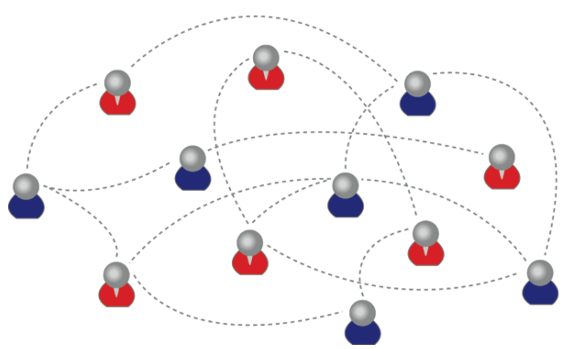


Figura 62: Rede Social

que todos que olham a propaganda compartilham em seu mural, quantos passos serão necessários para atingir a rede toda? Vamos supor que a informação começa a se espalhar pelo nó A (Figura 63). Do nó A, a informação consegue atingir os nós B e C (Figura 64). Cada nó, por sua vez, espalha a informação para os nós vizinhos (Figura 65, 66, 67,68).

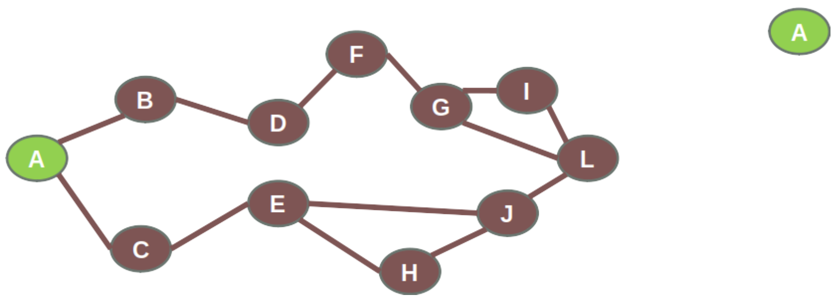


Figura 63: Exemplo percurso em um grafo.

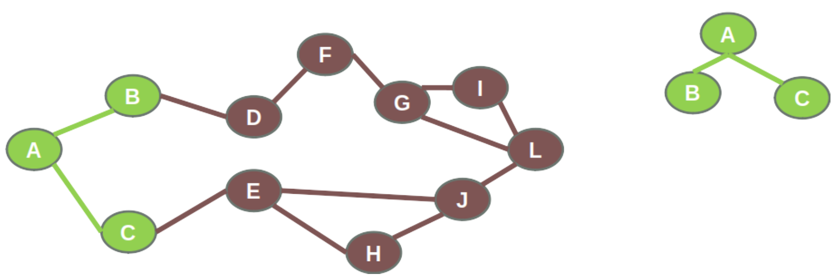


Figura 64: Exemplo percurso em um grafo.

Esse procedimento é conhecido como busca em largura. Com ele conseguimos determinar, partindo de um determinado nó, qual o menor número de passos (*hops*) é necessário para atingir qualquer outro nó da rede. Esse procedimento mostra também a dinâmica da informação. É possível ver como a informação tende a caminhar (Figura 69).

As distâncias do nó A até os outros nós são:

$$d(A, -) = [1, 1, 2, 2, 3, 3, 3, 4, 4, 5]$$

Repetindo o mesmo procedimento para os outros nós e somando os valores de cada lista temos:

$$d(A, -) = [1, 1, 2, 2, 3, 3, 3, 4, 4, 5] = 28$$

$$d(B, -) = [1, 2, 2, 3, 3, 4, 4, 4, 4] = 27$$

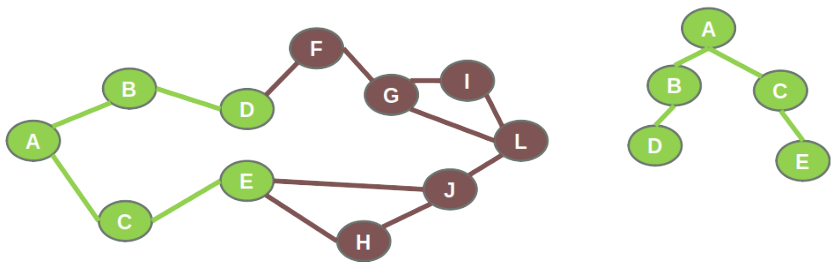


Figura 65: Exemplo percurso em um grafo.

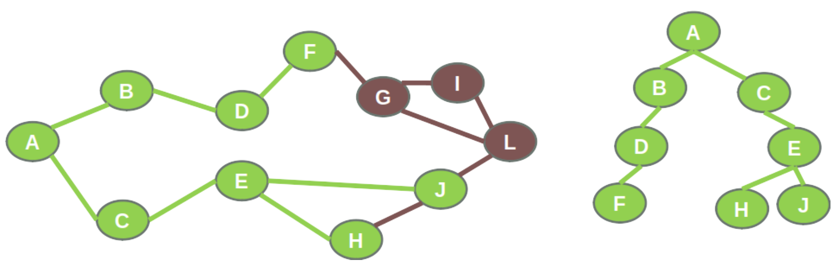


Figura 66: Exemplo percurso em um grafo.

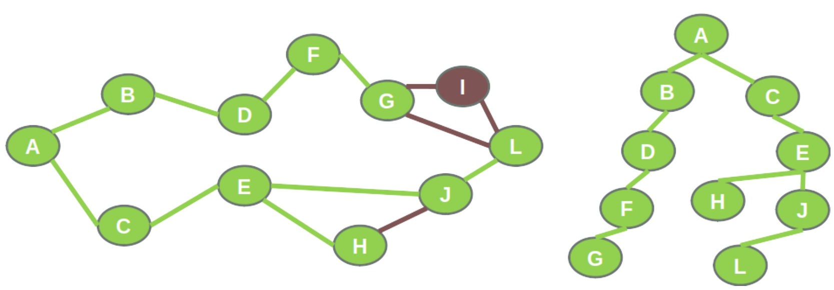


Figura 67: Exemplo percurso em um grafo.

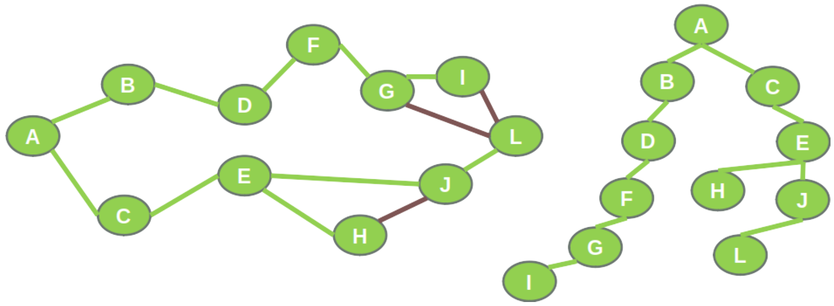


Figura 68: Exemplo percurso em um grafo.

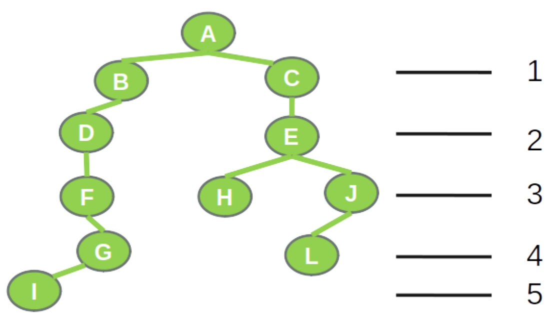


Figura 69: Grafo resultado de uma busca em largura, com as distancias.

$$d(C,-) = [2,2,2,3,3,4,4,4] = 24$$
$$d(D,-) = [1,2,3,3,4,4,5] = 22$$
$$d(E,-) = [1,1,2,3,3,4] = 14$$
$$d(F,-) = [1,2,2,3,4] = 12$$
$$d(G,-) = [1,1,2,3] = 7$$
$$d(H,-) = [1,2,3] = 6$$
$$d(I,-) = [1,2] = 3$$
$$d(J,-) = [1] = 1$$

Em seguida, somamos esses totais e dividimos pela quantidade de valores do agregado dessas listas.

$$\frac{28+27+24+22+14+12+7+6+3+1}{55} = 2,62$$

O diâmetro dessa rede é o maior valor observado nessa lista, no nosso caso, 5.

3.3 Árvore

Essa estrutura resultante do procedimento, que mostra o caminho da informação pelos nós, é conhecida como ÁRVORE. Um ÁRVORE é uma rede que não contém CICLOS! (Figura 70).

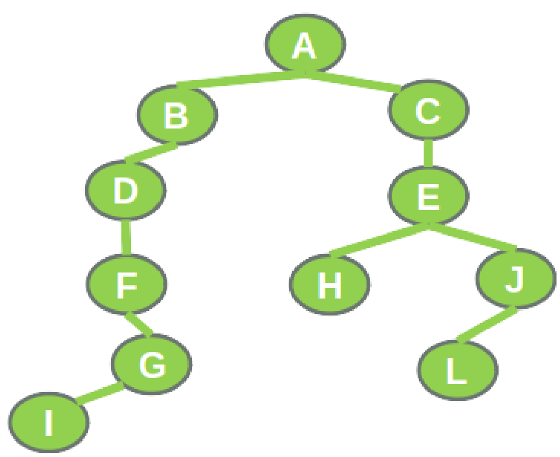


Figura 70: Arvore resultante da busca em largura.

3.4 Ciclos

Um CICLO é um caminho em uma rede em que o nó final é o mesmo da origem. Por exemplo o Caminho $[A, B, C, D, E, G, A]$ forma um ciclo (Figura 71).

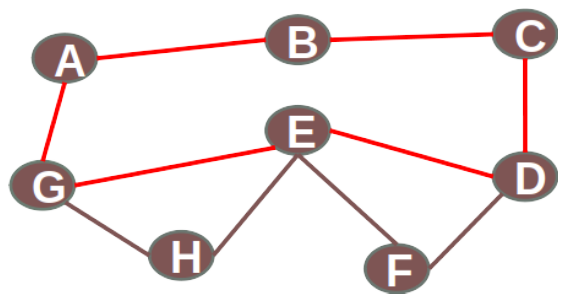


Figura 71: Exemplo de um grafo simples, presentado um ciclo formado pelas arestas em vermelho.

3.5 Conetividade

Um outro fato interessante em redes reais é que, embora tudo pareça estar conectado, pode existir alguns nós das redes que formam um grupo separado da rede (Figura 72). Em uma rede social, imagine que existam moradores de uma ilha deserta que não tem contato algum com o resto do mundo.

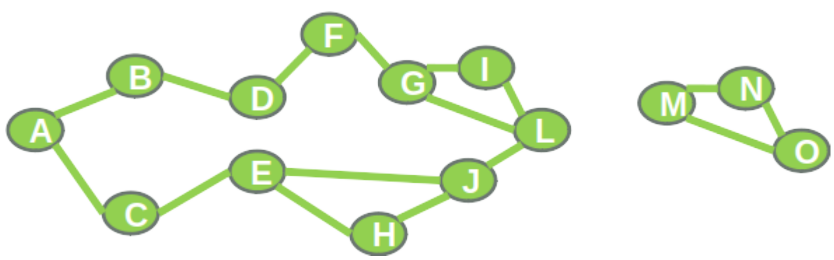


Figura 72: Grafo não conexo.

Eles formariam uma rede sem qualquer ligação com a rede principal. Quando em uma rede existem dois nós tal que não exista um caminho entre eles, ela é denominada Desconectada ou Desconexa (Figura 73).

Em contrapartida, uma rede em que existe pelo menos um caminho entre todos os nós é chamada de Conetada ou Conexa (Figura 73).

A rede formada por um subconjunto de nós e arestas da rede original forma uma SUBREDE. Toda SUBREDE contendo o conjunto de nós e arestas formando uma rede conexa são componentes conexos. (Figura 73).

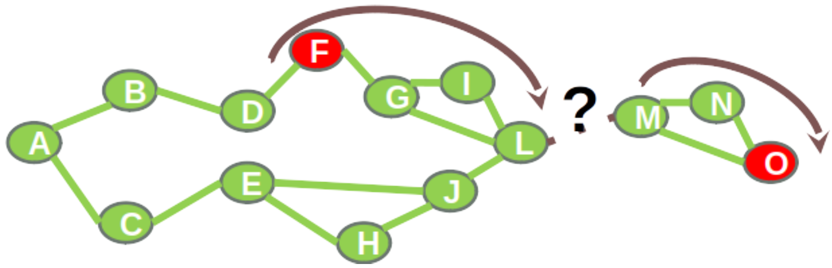


Figura 73: Rede Desconectada ou Desconexa, com dois grupos de vértices que não tem aresta entre eles.

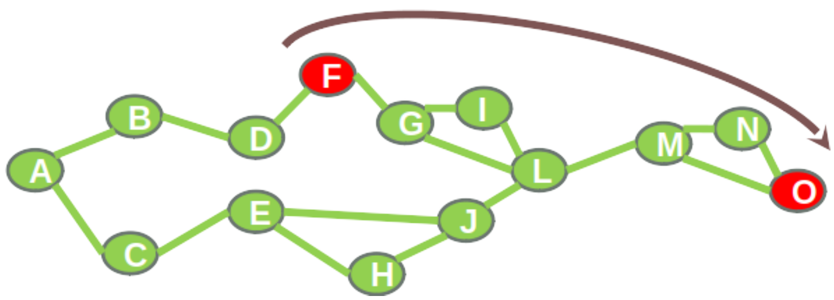


Figura 74: Rede Conetada ou Conexa, na qual existe um caminho entre cada par de vértices .

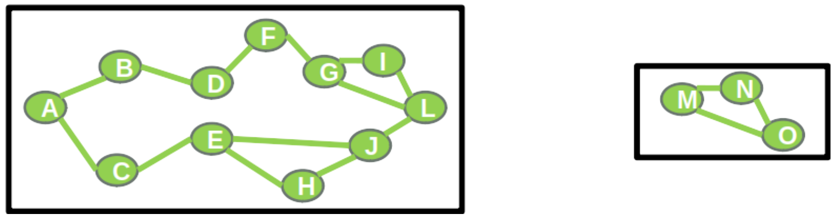


Figura 75: Rede Desconectada ou Desconexa, formada por duas sub-redes conexas denominadas componentes conexos.

Se utilizarmos a BUSCA EM LARGURA, podemos verificar se a rede é CONECTADA e se existe um COMPONENTE GIGANTE (Figura 76). .

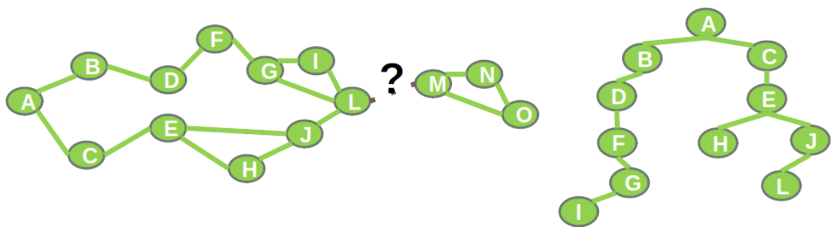


Figura 76: Rede Desconectada ou Desconexa, com uma busca em largura do lado que so tem nós da componente gigante.

3.6 Distância em Redes Ponderadas

Uma rede urbana de ruas é planejada de acordo com o crescimento populacional e comercial. Empresas de logística necessitam determinar a melhor rota de entrega dado uma lista de clientes. Nesse caso a rede já tem sua estrutura determinada por outros fatores e é necessário encontrar caminhos ótimos para propagar a informação (entrega de produtos). Os nós da rede de entrega são os pontos que devem receber os produtos e as arestas são as ruas que interligam eles (Figura 77). .

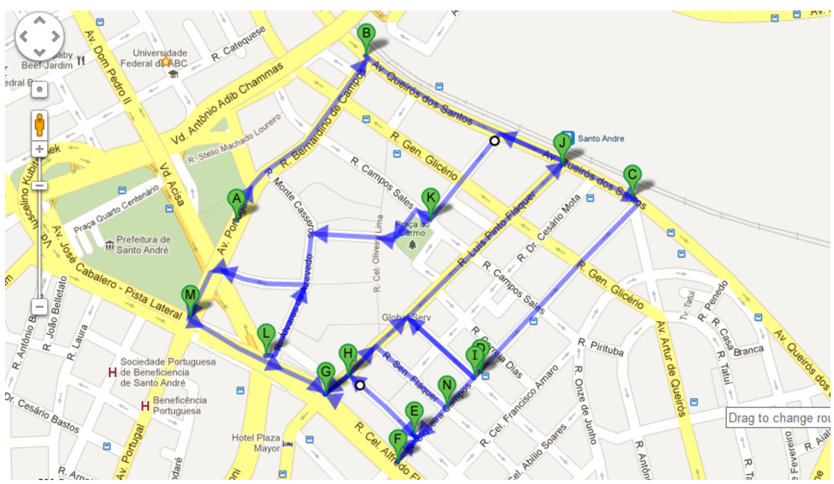


Figura 77: Ruta de entrega de produtos, sinalizada com pontos verdes para os lugares de entrega.

Essa rede tem um diferencial em relação ao exemplo anterior pois as arestas não representam sempre o mesmo valor de unidade de distância. Por exemplo, a aresta interligando os pontos A e B tem uma distância muito maior que a aresta que interliga J e C. Essa rede é, portanto, PONDERADA. Ou seja, cada aresta tem um valor associando o custo em percorrer tal trecho entre dois nós. Reparem que em um grafo ponderado o melhor caminho não necessariamente é o que percorre menos arestas. Vamos verificar o melhor caminho entre os nós A e F (Figura 78). Nesse caso, o melhor caminho entre dois pontos não pode ser obtido através da busca em largura. .

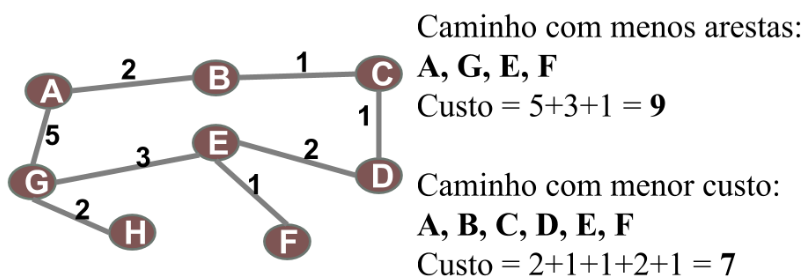


Figura 78: Grafo ponderado com o calculo do caminho com menos arestas e o caminho com menor custo.

3.7 O algoritmo Dijkstra

Um algoritmo para encontrar o caminho mínimo em um grafo foi criado por Edsger Dijkstra em 1956. Esse algoritmo se assemelha à BUSCA EM LARGURA, porém toma as decisões sobre quais nós percorrer em seguida utilizando um procedimento GULOSO.

Para encontrarmos todos os menores caminhos partindo do nó A, iniciamos indicando que a distância de A até ele mesmo é 0, e para todos os outros nós é INFINITO (INF)(Figura 79). Na busca em largura o nó A enviaria a informação paralelamente para os nós B e G (Figura 80). Expandimos tais nós e atualizamos a distância até eles somando o valor do nó A às arestas correspondentes (Figura 81). Ao invés de expandir paralelamente os nós B e G, vamos escolher o nó com menor distância para expandir. No nosso caso esse é o nó B (Figura 82).

Do nó B podemos expandir apenas o nó C, com distância igual à $2 + 1 = 3$ (Figura 83). Na sequência expandimos o nó C, que tem apenas o nó D com valor igual à $3+1 = 4$ (Figura 84). Do nó D expandimos o nó E com o valor $4+2 = 6$ (Figura 85). Como o nó G agora tem um valor menor, expandimos à partir dele, o nó H (o E já foi utilizado). O valor ficará $5 + 2 = 7$ (Figura 86). Finalmente expandimos o nó E, que leva ao nó F com valor $6+1 = 7$ (Figura 87). Não tendo mais nenhum nó para expandirmos a busca é encerrada, formando a seguinte árvore (Figura 88).

A distância média para essa rede é $4,36$ e o diâmetro é igual a 9. Para

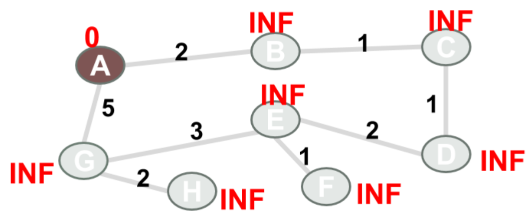


Figura 79: Exemplo do Algoritmo Dijkstra.

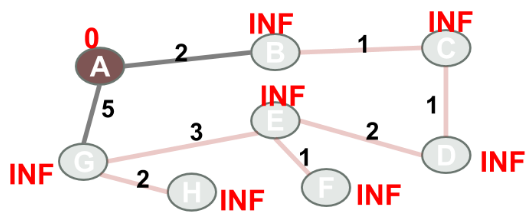


Figura 80: Exemplo do Algoritmo Dijkstra.

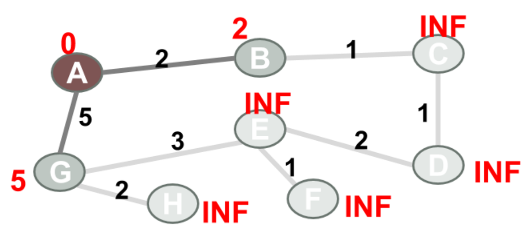


Figura 81: Exemplo do Algoritmo Dijkstra.

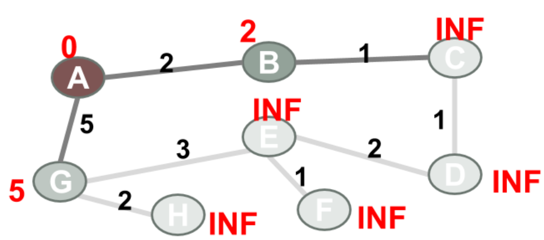


Figura 82: Exemplo do Algoritmo Dijkstra.

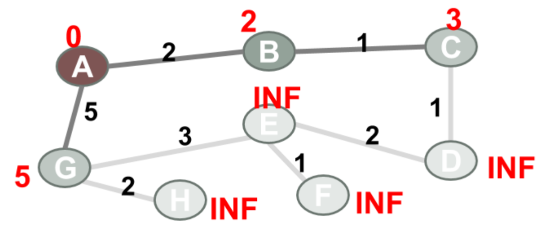


Figura 83: Exemplo do Algoritmo Dijkstra.

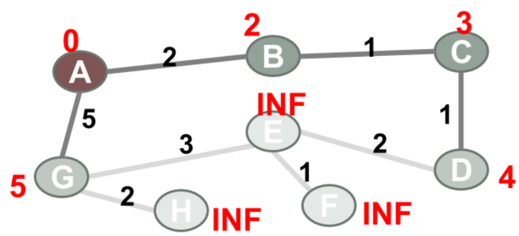


Figura 84: Exemplo do Algoritmo Dijkstra.

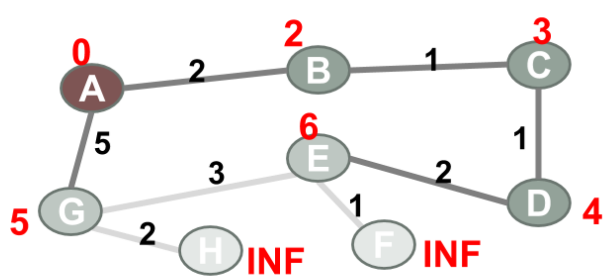


Figura 85: Exemplo do Algoritmo Dijkstra.

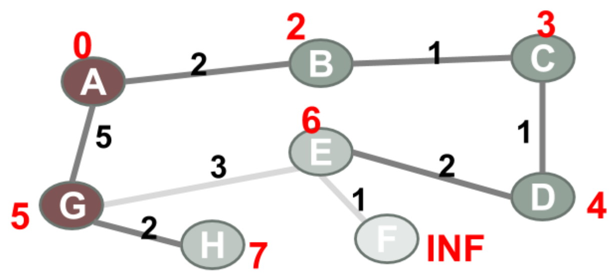


Figura 86: Exemplo do Algoritmo Dijkstra.

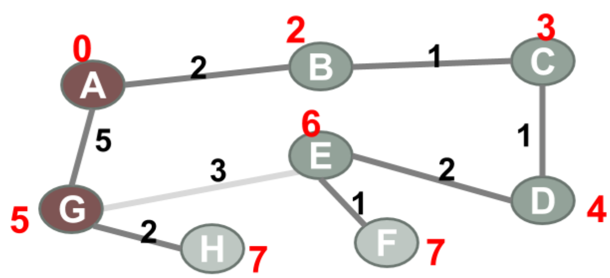


Figura 87: Exemplo do Algoritmo Dijkstra.

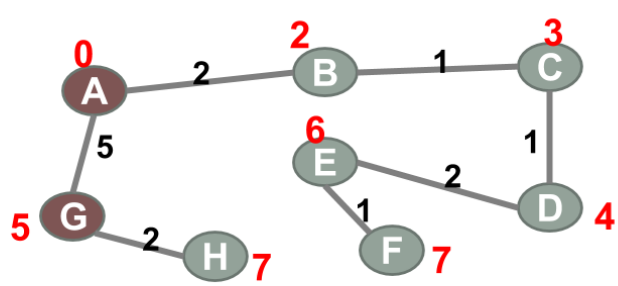


Figura 88: Exemplo do Algoritmo Dijkstra.

calcular a distância média e o diâmetro, repetimos esse procedimento para cada nó, agregamos a informação e calculamos os valores.

3.8 Distância em Redes Reais

Citamos alguns exemplos:

- Em 2012 o Facebook contava com uma rede da ordem de 721 milhões de usuários (nós) e 69 bilhões de amizades (arestas). Nessa época a distância média entre os nós era de 4,74 enquanto o diâmetro era de 41 passos. 92% dos pares de nós tem uma distância menor que 5! [3].
- A rede de transmissão de energia da parte leste dos Estados Unidos conta com 49.597 nós de transmissão e 62.985 linhas interligando esses nós [7].
- A rede de reações metabólicas da bactéria E. Coli conta com 906 metabolitos (nós) e 1,230 mapeamentos atômicos (arestas) [2] (Figura 89).
- A rede do co-autoria da área de Saúde no Brasil conta com 114.169 autores representando os nós e 659,332 relações de co-autorias (arestas) [8] (Figura 90).

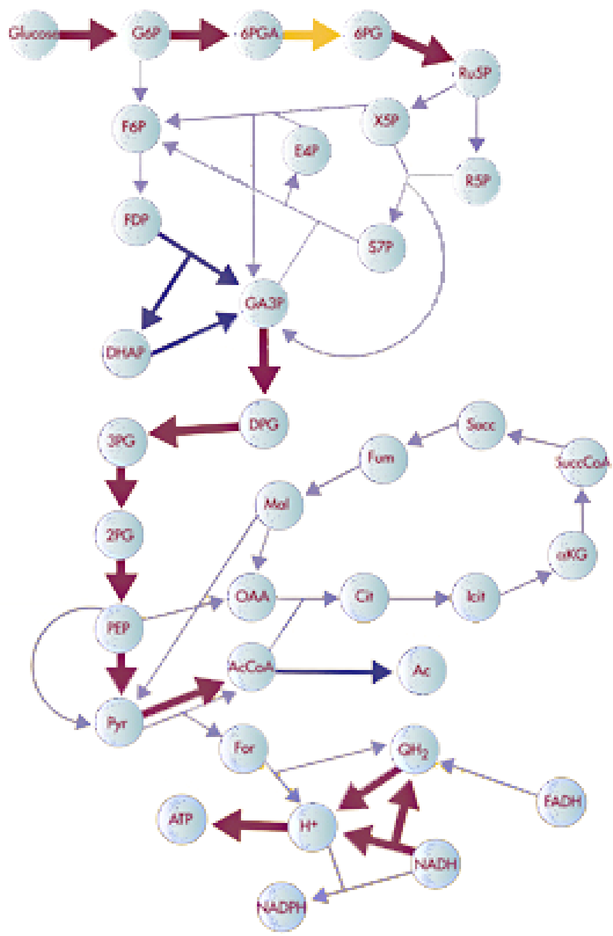


Figura 89: Rede de reações metabólicas da bactéria E. Coli. [2].

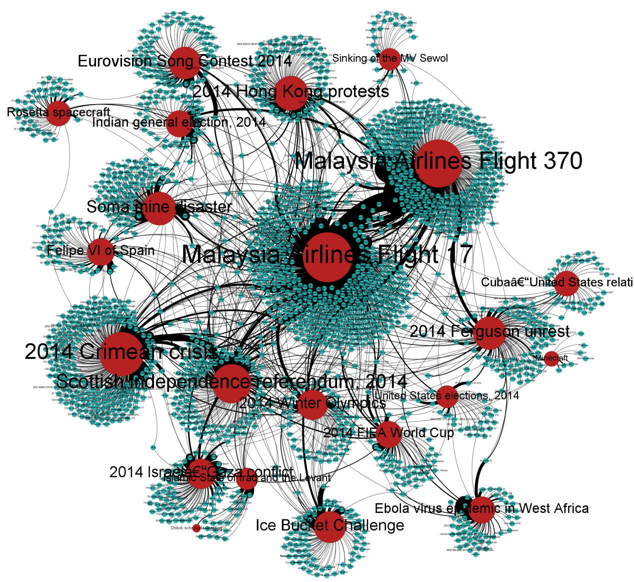


Figura 90: Rede do co-autoria da área de Saúde no Brasil [8].

4 Centralidade

4.1 Importância dos nós

Até então vimos que cada nó da rede cumpre seu papel na transmissão de informação. Sem eles a informação poderia demorar mais para chegar ao destino ou nem mesmo chegar. Porém, alguns nós são mais importantes do que outros nessa transmissão! (Figura 91).

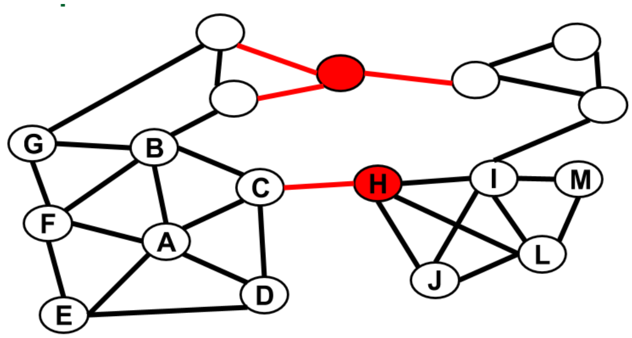


Figura 91: Grafo com com dois nós em vermelho que juntam a maioria dos vértices mostrando a importância da centralidade.

Essa importância é mensurada através da CENTRALIDADE do nó, e ela indica:

- Influência de uma pessoa em uma rede social;
- Importância de certas avenidas para o fluxo de veículos;
- Qual a ordem de importância de páginas web para serem apresentadas no resultado de busca;

Dá para perceber que identificar esses nós de maior importância pode ser vital para a estabilidade ou controle da sua rede. Também notem que a forma com que essa importância é mensurada pode ser diferente dependendo da aplicação. A centralidade tem origem em estudos de sociologia, veremos os conceitos mais conhecidos, as centralidades de:

- Grau;
- Proximidade;
- Betweenness;
- Pagerank (utilizado pelo Google).

Em uma rede social o nó central é aquele que tem mais conexões, pois influencia mais pessoas ao mesmo tempo. Em uma rede de transporte o nó importante é aquele pelo qual mais pessoas passam para ir de um local ao outro.

4.2 Grau

A ideia mais simples de um nó central é aquele que está mais conectado que os demais, ou seja, aquele que tem um maior grau. Se ele tem mais conexões, significa que:

- Ele deve ser importante (para os outros nós "desejarem" se conectar a ele);
- Uma informação partindo dele é transmitida diretamente para muito mais nós.

A centralidade de grau foi a primeira medida de centralidade já utilizada (e é também a mais simples). Simplesmente é calculada como a relação do grau do nó com o valor máximo de grau que tal nó poderia ter:

$$C_g(v_i) = \frac{g(v_i)}{|V| - 1}$$

Grau máximo possível: número total de nós menos um (exceto quando é possível conectá-lo a ele mesmo). Em resumo, essa medida pode ser interpretada como "o quão importante os outros nós julgam que ele é" e/ou "por quantas vias o nó pode receber ou emitir informação". Algumas aplicações são:

- Quantas pessoas a opinião de uma certa pessoa atinge, de acordo com seu número de amigos;
- Quantas pessoas alguém doente vai infectar;
- Qual roteador um hacker deve atacar para parar um serviço.

A centralidade pode se entender melhor com um exemplo, na figura 92 é apresentado um grafo com nós coloridos em função de seu grau. O que acontece se removermos os nós de maior centralidade de grau dessa rede? (Figura 93). Se escolhermos atacar somente o nó azul, também desconectamos a rede, mas ainda permanece alguns grupos conectados (Figura 94). O envio de informação através do nó azul é, em média, mais rápida que a partir dos nós rosas. Talvez a melhor forma de lidar com a transmissão de uma doença seja isolando o nó azul (Figura 95). Por outro lado, os nós rosas tem vantagem de escolha de múltiplos caminhos para seu objetivo, pode ser um concentrador de riquezas!

Apresentamos outro exemplo da Centralidade de Grau (Figura 96), onde a remoção do nó central não desconectou a rede! (Figura 97). Porém, a remoção dele pode causar um congestionamento de informação, pois ele era um nó "centralizador", provavelmente com maior capacidade de receber e transmitir informação.

4.3 Proximidade

Uma outra forma de pensar em centralidade de um nó é imaginando que a rede representa um espaço físico e o nó mais central é aquele que se localiza no

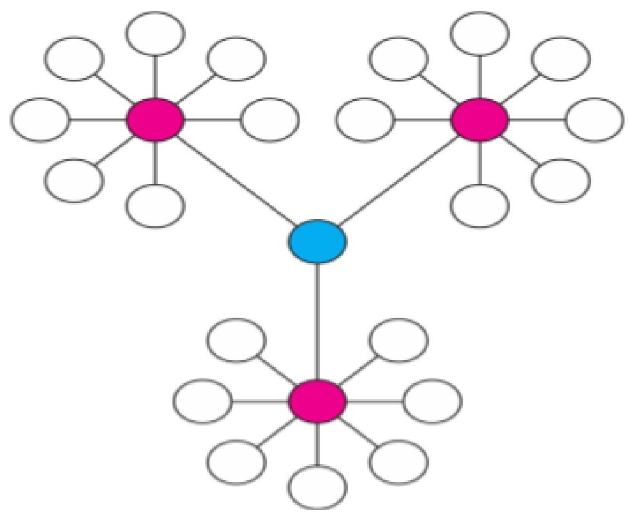


Figura 92: Grafo que mostra a centralidade com cores nos nós: azul com um $C_g = 0.125$; rosa com um $C_g = 0.333$; branco com um $C_g = 0.042$.

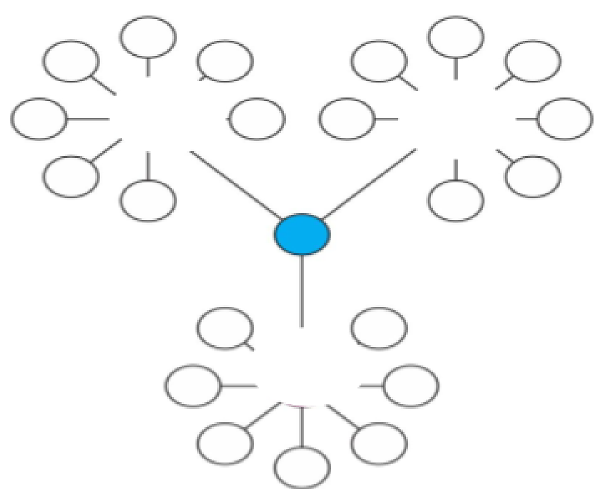


Figura 93: Grafo que mostra a centralidade com cores nos nós, foram retirados os nós de cor rosa com um $C_g = 0.333$.

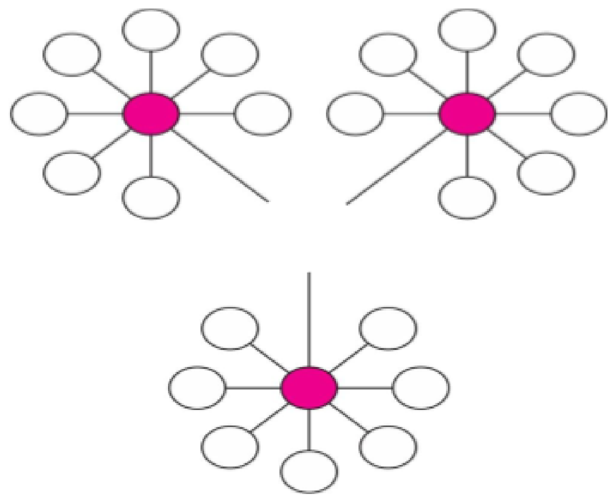


Figura 94: Grafo que mostra a centralidade com cores nos nós, foi retirado o nó de cor azul com um $C_g = 0.0.125$.

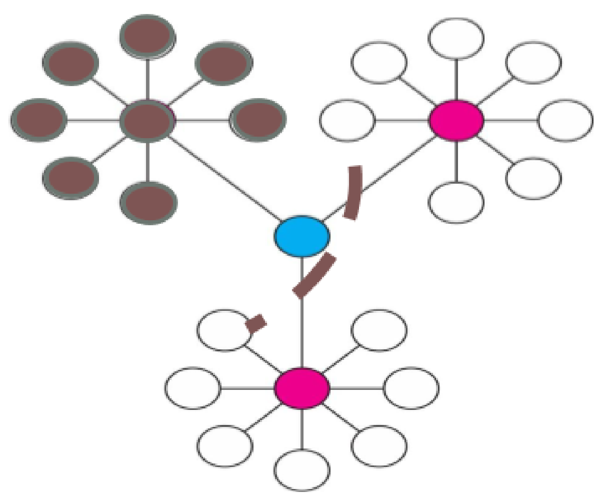


Figura 95: Grafo que mostra a centralidade com cores nos nós. Foi isolado o nó de cor azul com um $C_g = 0.0.125$.

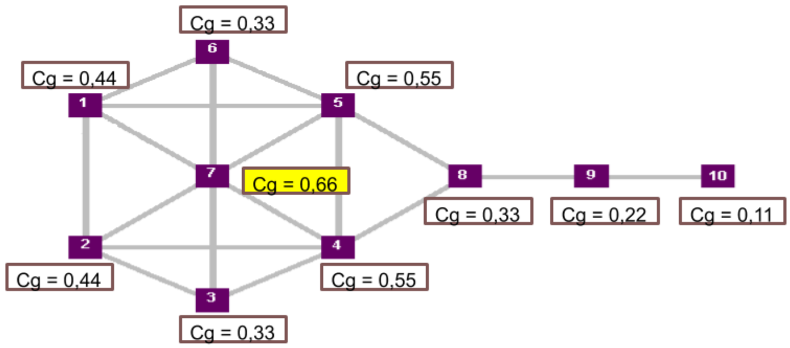


Figura 96: Exemplo da Centralidade de grau.

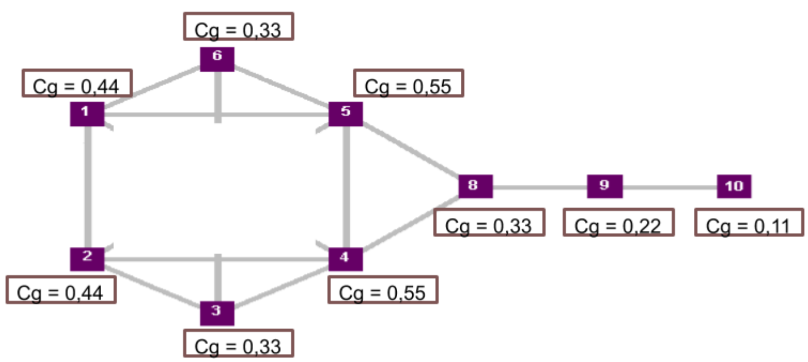


Figura 97: Exemplo da Centralidade de grau.

centro da rede . Em um espaço Euclidiano o ponto central é aquele que minimiza a distância média entre todos os outros pontos (medóide) (Figura 98). Em uma rede podemos pensar que o nó central é aquele nó que minimiza a distância média entre todos os outros nós. Um nó que tem a distância média menor pode espalhar uma informação mais rapidamente e de forma mais eficiente para todos os nós do que um nó que simplesmente tem muitas conexões.

Figura 98: Exemplo da importância de um medóide (em amarelo).

Em uma rede a distância entre dois nós pode ser calculada pelo caminho mais curto entre eles. Para medir o quão distante um nó se encontra no grafo podemos somar a distância do caminho mais curto entre esse nó e todos os outros. Logo, para verificar a proximidade do nó, basta calcular a inversa dessa distância. A centralidade de proximidade de um nó é geralmente calculada como o inverso da média das distâncias de todo nó v_j ao nó v_i. Ela pode ser interpretada como o tempo que leva para um informação partindo de v_i se propagar por um certo nó aleatório na rede. Inverso da média dos caminhos mais curtos:

$$C_p(v_i) = \frac{n-1}{\sum_{v_j}}$$

Isso pode ser relevante para ataques em redes, como selecionar pessoas para espalhara fake news, divulgar produtos e também na defesa da rede, como campanha de vacinação.

4.4 Betweenness

Essa medida foi introduzida para verificar o controle/influência de uma pessoa para transmitir uma informação entre grupos sociais diferentes [6].Em alguns estudos o nó que temos interesse, e consideramos importante, é aquele que mais contribui para o funcionamento da rede. Queremos saber qual nó causará maior impacto caso seja removido da rede. Essa medida foi introduzida para verificar o controle/influência de uma pessoa para transmitir uma informação entre grupos sociais diferentes.

Essa medida, conhecida como centralidade betweenness, mede a "carga" de informação que um nó tende a receber. Ela é calculada como a quantidade

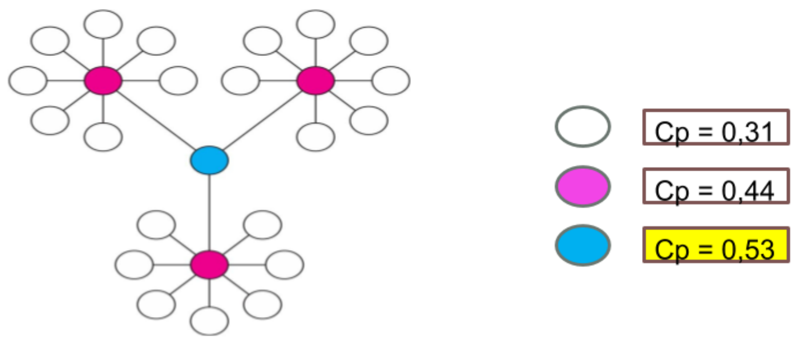


Figura 99: Centralidade Cor Proximidade

de caminhos mais curtos da rede do qual esse nó faz parte. Se a informação sempre viaja pelo caminho mais curto, quanto esse nó servirá de intermediário em diversas situações. Ela é calculada como a somatória do número de caminhos mínimos entre dois nós v_j e v_k que passam por vi dividido pelo total desses caminhos.

$$C_b(v_i) = \sum_{v_j \neq v_i \neq v_k} \frac{\sigma_{v_j,v_k}(v_i)}{\sigma_{v_j,v_k}}$$

Medida normalizada para comparação entre redes:

$$Cb'(v_i) = \frac{Cb(v_i)}{(n-1)(n-2)}$$

Grau de betweenness do nó mais central em uma rede em formato estrela. Se não for direcionada, dividir por 2 :

$$(n-1)(n-2)$$

Em redes com ORDEM e TAMANHO elevados, o valor dessa centralidade é estimado utilizando uma amostragem dos caminhos (Figura 100). Essa métrica pode ser "custosa" para calcular pois necessita conhecer todos os possíveis caminhos mais curtos entre cada par de nós.

Apresentamos um exemplo com os trés tipos de centralidade visto ate agora (Figura 102). O nó 7 tem o maior número de conexões, é o mais "popular" levando em conta a centralidade de grau, pois conhece mais pessoas. Porém, não é ele que está mais bem localizado na rede. O nó 8, que tem maior grau de centralidade betweenness, está em um ponto vital da rede. Ele interliga as duas redes formadas (uma centrada em 7 e outra em 9). Sem ele, a informação que parte do nó 1 nunca chegará ao nó 10. Os nós 4 e 5 tem menos conexões que o nó 7, mas a posição na rede em que eles se encontram é privilegiada. Eles conseguem ver tudo que acontece de forma mais rápida.

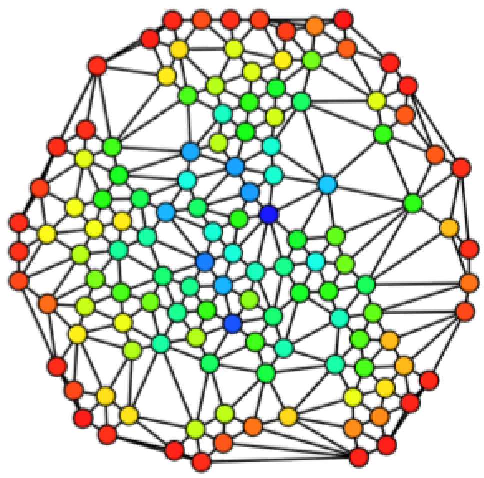


Figura 100: Rede com ordem e tamanho elevado, vermelho = 0, azul = max.

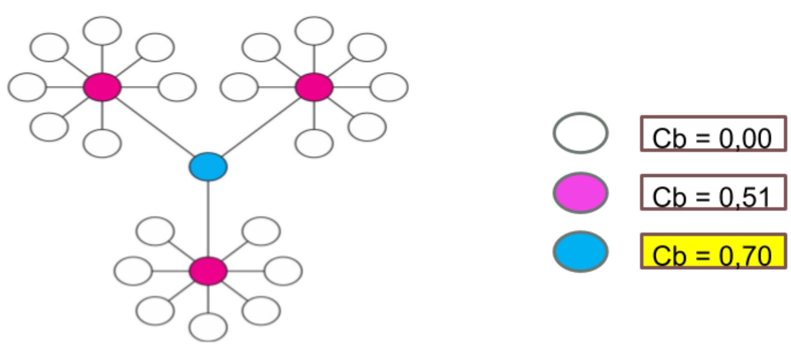


Figura 101: centralidade betweenness

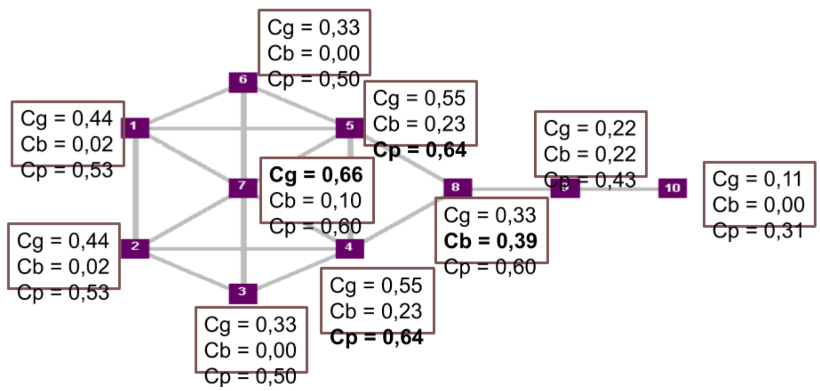


Figura 102: Exemplo de Centralidade: Cg = centralidade de grau, Cb centralidade betweenness e Cp = centralidade de proximidade.

4.5 PageRank

4.5.1 Centralidade de Autovetor

A centralidade de grau permite calcular o quão conectado é um nó. No contexto de rede social isso pode ser um determinante de influência. Porém, analisemos o caso apresentado na figura 102. Apesar dos dois nós em amarelo apresentarem um grau de influência 2, eles não apresentam efetivamente a mesma influência. Um deles está conectado a pessoas mais importantes (com grau 6 e 7), enquanto o outro com pessoas menos importante (outros de grau 2).

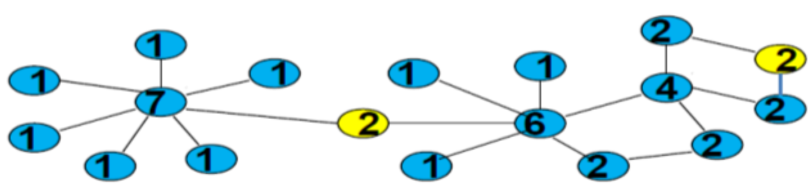


Figura 103: Grafo simples com etiquetas nos vértices que representam seu grau.

Uma forma de melhorar esse conceito foi definido com a centralidade de autovetor. Essa centralidade calcula a importância do vértice baseado em sua influência na rede. Ela é proporcional à centralidade dos vizinhos do nó i, com $a_{ij} = 1$ se os nós i e j são adjacentes e lambda é uma constante de importância.

$$C_i = \frac{1}{\lambda} \sum_j a_{i,j} C_j$$

Podemos reformular a equação passando o lambda para o outro lado e colocando na forma matricial. Para resolver essa equação basta calcularmos os autovalores e autovetores!

$$\lambda C = \sum_i \sum_j a_{i,j} C_j = AC$$

4.5.2 Autovalor e Autovetor

Dados uma matriz A, um vetor x e um escalar lambda λ, dizemos que x é um autovetor de A e lambda λ o autovalor correspondente se a seguinte equação for verdadeira.

$$Ax = \lambda x$$

Ou seja, multiplicando A por x, a direção do vetor permanece a mesma, mudando apenas sua intensidade. Obviamente temos a solução trivial quando $x = 0$, porém é interessante encontrar soluções diferentes. Dada certas condições temos um sistema com n equações e múltiplas soluções.

4.5.3 Algoritmo PageRank

Partindo da ideia da centralidade de autovetor, pesquisadores do Google pensaram em um algoritmo para determinar as páginas principais para serem mostradas em um resultado de busca. Esse algoritmo, denominado PageRank, determina a importância de um nó (página web) da seguinte forma:

- Uma página web A é importante se existem várias páginas apontando para ela e;
- Cada página B que aponta para ela incrementa o grau de centralidade proporcionalmente à quantas páginas apontam para B.

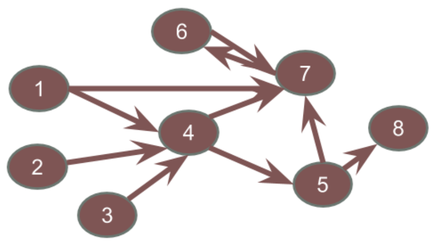
A centralidade do PageRank pode ser calculada como a somatória da centralidade de cada nó v_j que envia informação a v_i dividido pelo seu grau de saída. Note que para calcular a centralidade de v_i, precisamos conhecer as centralidades dos v_js.

$$C_{pr}(v_i) = \sum v_j | (v_j, v_i) \in E \frac{C_{pr}(v_j)}{dg_{out}(v_j)}$$

Vamos reescrever a equação na forma matricial, N é uma matriz em que o elemento n_{ij} é igual ao inverso do grau de saída do nó i, caso i, j formem uma aresta, e 0 caso contrário. Esse valor representa a probabilidade de você sair do nó i e chegar em qualquer nó adjacente a ele.

$$C_{pr} = C_{pr} N$$

Para calcular o valor de PageRank, iniciamos o vetor C_{pr} com valores quaisquer (Figura 104). Em seguida, aplicamos a equação de C_{pr} repetidamente até que os valores não se alterem (ou alterem muito pouco) (Figura 105, 106, 107, 108, 109, 110). Por causa do nó 6 e do nó 7, o cálculo entra em um laço infinito, pois do 6 só existe possibilidade de ir para o 7 e do 7 apenas para o 6 (Figura 111).



			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

1	1	1	1	1	1	1	1

Figura 104: Calculo do valor de Page Rank.

1	1	1	1	1	1	1	1

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	2,5	0,5	1	2,5	0,5

Figura 105: Calculo do valor de Page Rank.

Para evitar esse ciclo adicionamos uma pequena probabilidade de pular de um nó para outro qualquer sem necessidade de serem adjacentes. Isso é feito construindo uma matriz P na forma sendo E uma matriz com valores iguais a $1/n$ em todos os seus elementos. α é definido pelo usuário, geralmente é utilizado um valor de 0.85 e serve para amortizar o cálculo.

$$P = \alpha N + (1 - \alpha)E$$

Para essa rede a matriz P será a representada na Figura 112.Com isso alteramos nossa equação para e repetimos o procedimento.

$$C_{pr} = \frac{C_{pr}P}{||C_{pr}P||}$$

$$\begin{bmatrix} 0 & 0 & 0 & 2{,}5 & 0{,}5 & 1 & 2{,}5 & 0{,}5 \end{bmatrix} \times \begin{bmatrix} & & & 0{,}5 & & & 0{,}5 & \\ & & & 1 & & & & \\ & & & 1 & & & & \\ & & & & 0{,}5 & & 0{,}5 & \\ & & & & & & 0{,}5 & 0{,}5 \\ & & & & & & 1 & \\ & & & & & 1 & & \\ & & & & & & & \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 1{,}25 & 2{,}5 & 2{,}5 & 0{,}25 \end{bmatrix}$$

Figura 106: Calculo do valor de Page Rank.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1{,}25 & 2{,}5 & 2{,}5 & 0{,}25 \end{bmatrix} \times \begin{bmatrix} & & & 0{,}5 & & & 0{,}5 & \\ & & & 1 & & & & \\ & & & 1 & & & & \\ & & & & 0{,}5 & & 0{,}5 & \\ & & & & & & 0{,}5 & 0{,}5 \\ & & & & & & 1 & \\ & & & & & 1 & & \\ & & & & & & & \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 2{,}5 & 3{,}125 & 0{,}625 \end{bmatrix}$$

Figura 107: Calculo do valor de Page Rank.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 2{,}5 & 3{,}125 & 0{,}625 \end{bmatrix} \times \begin{bmatrix} & & & 0{,}5 & & & 0{,}5 & \\ & & & 1 & & & & \\ & & & 1 & & & & \\ & & & & 0{,}5 & & 0{,}5 & \\ & & & & & & 0{,}5 & 0{,}5 \\ & & & & & & 1 & \\ & & & & & 1 & & \\ & & & & & & & \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 3{,}125 & 2{,}5 & 0 \end{bmatrix}$$

Figura 108: Calculo do valor de Page Rank.

0	0	0	0	0	3,125	2,5	0

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	0	0	2,5	3,125	0

Figura 109: Calculo do valor de Page Rank.

0	0	0	0	0	2,5	3,125	0

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	0	0	3,125	2,5	0

Figura 110: Calculo do valor de Page Rank.

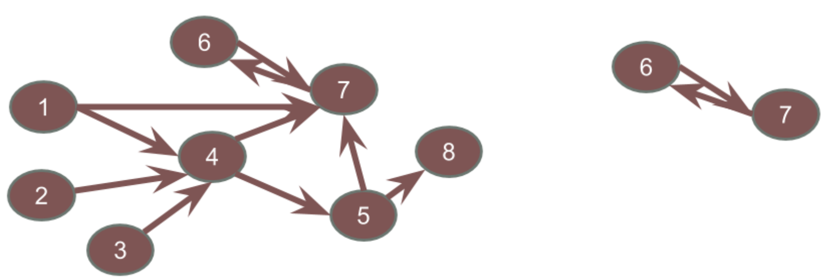
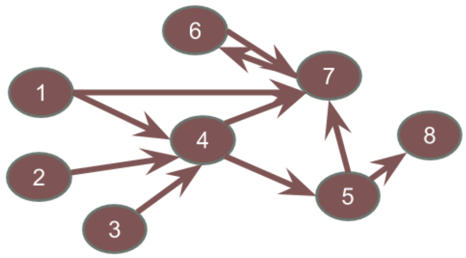


Figura 111: Calculo do valor de Page Rank.



0.02	0.02	0.02	0.44	0.02	0.02	0.44	0.02
0.02	0.02	0.02	0.87	0.02	0.02	0.02	0.02
0.02	0.02	0.02	0.87	0.02	0.02	0.02	0.02
0.02	0.02	0.02	0.02	0.44	0.02	0.44	0.02
0.02	0.02	0.02	0.02	0.02	0.02	0.44	0.44
0.02	0.02	0.02	0.02	0.02	0.02	0.87	0.02
0.02	0.02	0.02	0.02	0.02	0.87	0.02	0.02
0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Figura 112: Page Rank Pular

Essa tabela (Figura 113) mostra a sequência de aplicações desse método até chegar na convergência. Aumentando o tamanho dos nóś com os valores da última linha temos. Que é igual a ilustração original do PageRank.

0.04	0.04	0.04	0.65	0.17	0.29	0.65	0.17
0.04	0.04	0.04	0.14	0.33	0.62	0.68	0.11
0.04	0.04	0.04	0.12	0.09	0.6	0.76	0.17
0.04	0.04	0.04	0.12	0.09	0.71	0.68	0.08
0.03	0.03	0.03	0.11	0.09	0.63	0.76	0.07
0.03	0.03	0.03	0.11	0.08	0.71	0.69	0.07
0.03	0.03	0.03	0.11	0.08	0.64	0.75	0.07
0.03	0.03	0.03	0.11	0.08	0.7	0.7	0.07
0.03	0.03	0.03	0.11	0.08	0.65	0.75	0.07
0.03	0.03	0.03	0.11	0.08	0.69	0.7	0.07
0.03	0.03	0.03	0.11	0.08	0.65	0.74	0.07

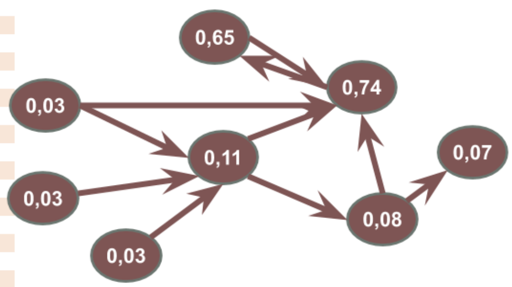


Figura 113: Page Rank Convergencia

4.6 Centralidade nas Redes Reais

Em uma rede de estilos musicais em que a relação $A \rightarrow B$ indica que o estilo A recebeu influência de B podemos perceber alguns padrões interessantes nas métricas de centralidade.

Punk Rock possui o maior grau de entrada, ou seja, mais estilos apontando para ele indicando que foi um estilo musical que influenciou muitos outros, contém diversas derivações.(Figura 114).

Já o maior grau de saída ficou com o estilo Girl Group, bandas formadas exclusivamente por mulheres, que receberam influência de diferentes gêneros

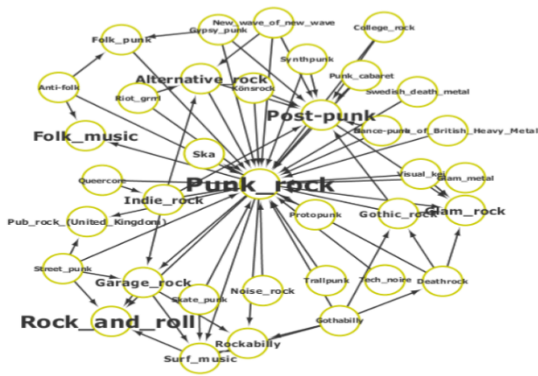


Figura 114: Rede de influencias musicais do Punk Rock.

como dance, disco, rock, pop, soul, rythm and blues, dentre outros (Figura 115).

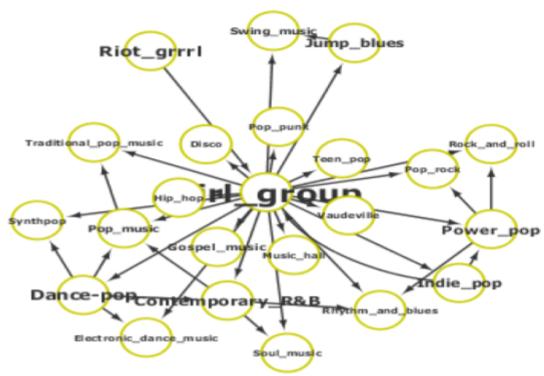


Figura 115: Rede de influencias musicais do Girl Group.

Os estilos musicais com maior grau de proximidade são o Rock, que é utilizado direta ou indiretamente como influência de muitos e o Samba Rock que mistura dois gêneros bastante influentes: Rock e Bossa Nova (Figura 116).

Punk Rock novamente aparece como o mais importante, dessa vez na centralidade de betweenness Isso ocorre por ele ser influência de muitos outros gêneros, fazendo com que exista um caminho curto entre dois gêneros passando por ele (Figura 117).

E, pela métrica de Pagerank, os três gêneros mais influentes são Rock, Punk Rock e Jazz pois esses influenciaram outros gêneros também importantes (Figura 118).

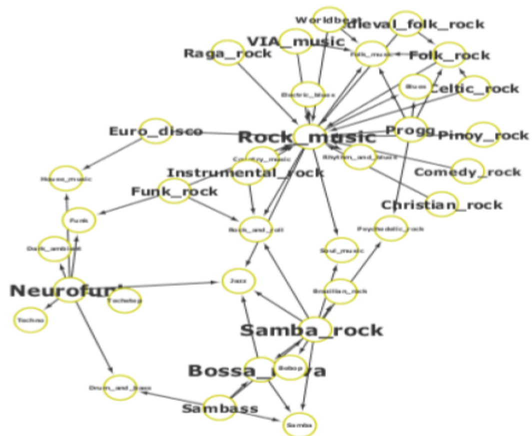


Figura 116: Rede de influencias musicais do Rock music.

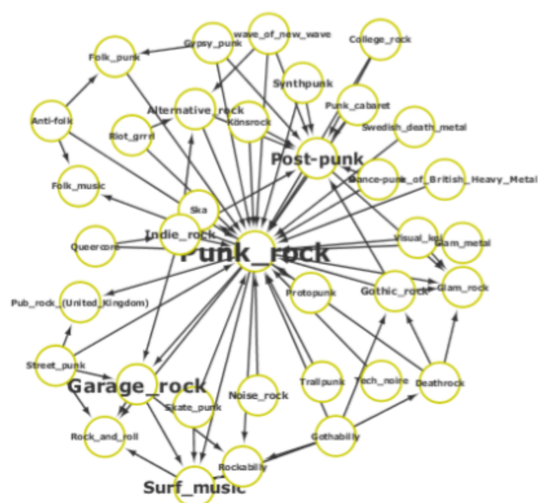


Figura 117: Rede de influencias musicais do Punk Rock.

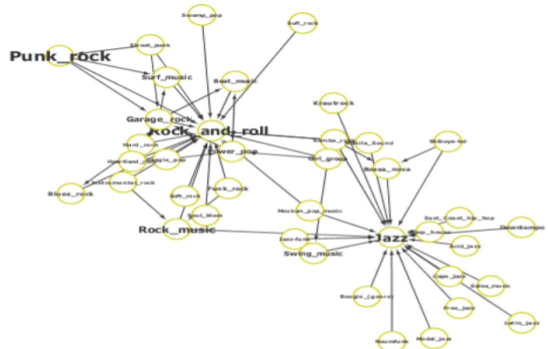


Figura 118: Rede de influencias musicais do Jazz Rock.

4.7 Agrupamentos e Pontes

Em um estudo feito nos anos 60 por Mark Granovetter, pesquisou a forma como as pessoas que acabaram de mudar de emprego ficaram sabendo de seus novos empregos (Figura 119). As respostas dos participantes indicavam que a informação partiu de alguém em sua lista de contatos pessoais. Porém, o mais surpreendente e contra intuitivo é que esses contatos foram descritos como “conhecidos” e não “amigos próximos”. Era de se esperar que tivessem a maior motivação em te ajudar! Mas então por que os “conhecidos” tem um papel tão crucial na busca de um novo emprego? (Figura 120).

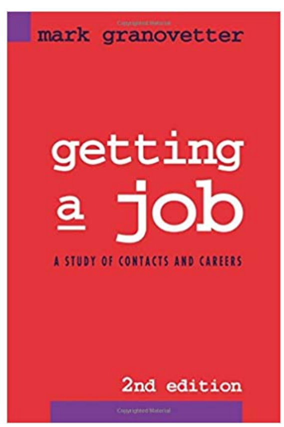


Figura 119: Portada do livro "Getting a Job"de Mark Granovetter.

Para generalizar um pouco vamos substituir a busca por emprego pela busca por informação!. Vamos começar a análise com a seguinte rede social de círculo

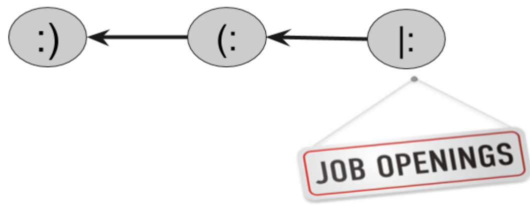


Figura 120: Segun o experimento de Mark Granovetter, as pessoas recebem informação de emprego de seus "conhecidos". "Job Openings"by rustybrick is licensed under CC BY-NC 2.0

de amizade na Figura 121. Se observarmos a relação entre os nós A, B e C podemos dizer que existe boas chances de, em certo momento da rede, dos nós B e C formarem um vínculo de amizade (Figura 122).

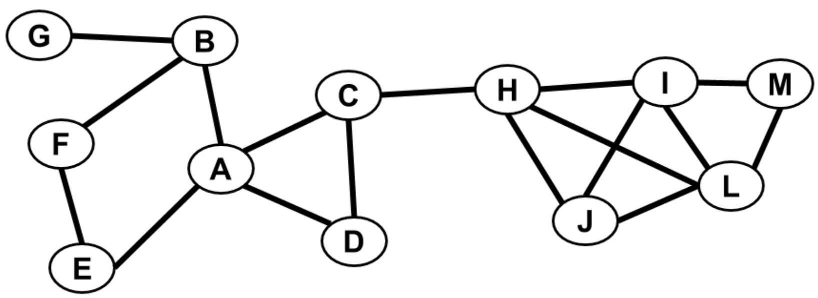


Figura 121: Exemplo de agrupamentos e pontes.

No contexto de redes sociais podemos pensar que se A é amigo de B e de C, fatalmente B e C irão se encontrar em diversas ocasiões formando, futuramente, um vínculo (Figura 123). E isso poderá ocorrer com diversos outros nós! Quanto mais amigos em comum duas pessoas tem maiores as chances de se tornarem amigas(Figura 124). Esse princípio é conhecido como FECHAMENTO TRIÁDICO. Podemos observar isso em redes representando páginas da web linkando outras páginas. Se uma página A faz referência às páginas B e C assume-se que essas páginas tratam do mesmo tema. Logo, existe uma chance de B fazer referência a C e vice-versa. Para verificar o quão próximos são os nós vizinhos a um certo nó, foi criado o COEFICIENTE DE AGRUPAMENTO.

4.8 Coeficiente de Agrupamento

Esse coeficiente mede a tendência com que os nós de uma rede se agrupam (compartilharem ligações com um mesmo nó). É baseado no fechamento

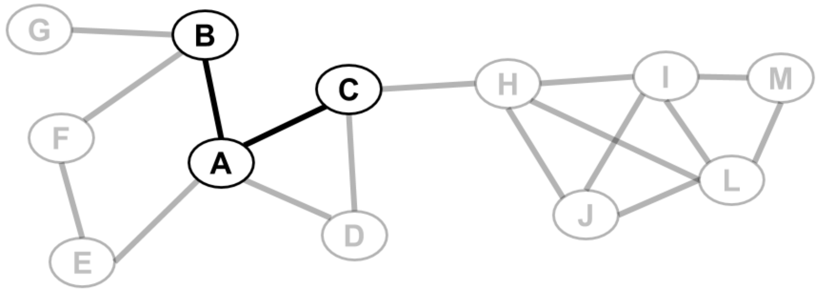


Figura 122: Exemplo de agrupamentos e pontes.

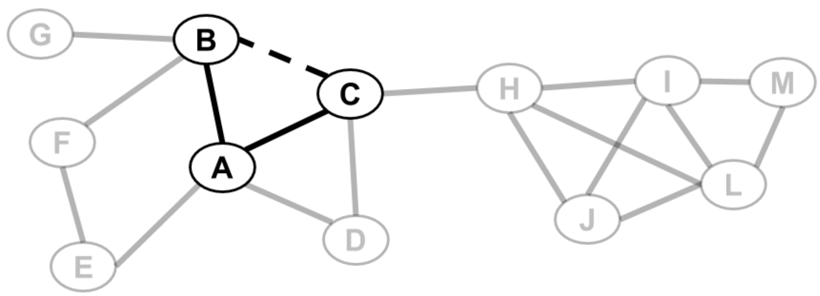


Figura 123: Exemplo de agrupamentos e pontes.

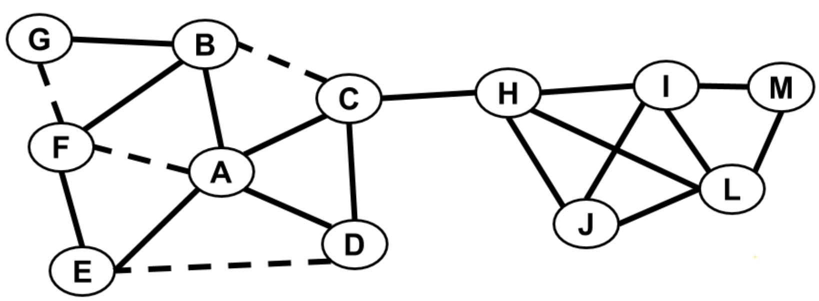


Figura 124: Exemplo de agrupamentos e pontes.

triádico:

$$C_i = \frac{|\{(j,k) \in E | j,k \in N(i)\}|}{|N(i)|(|N(i)|-1)}$$

O coeficiente de agrupamento de um nó para redes direcionadas é calculado como a fração de vizinhos do nó que tem ligações entre si com todas as possibilidades (Figura 125). Para redes não-direcionadas essa fórmula é multiplicada por 2, pois existem apenas metade das possíveis ligações (Figura 126).

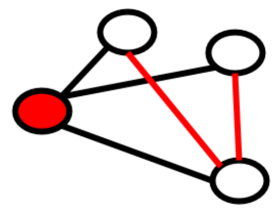


Figura 125: Coeficiente agrupamento em uma Rede não direcionada, Grau do nó = 3, Possíveis ligações = $3 * 2 = 6$, Dos 3 vizinhos existem 2 ligações entre si, Como a rede não é direcionada, multiplicamos por 2, $C = 2 * 2/6 = 2/3$

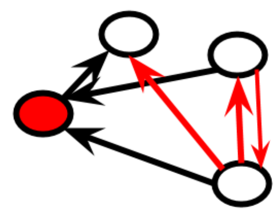


Figura 126: Coeficiente agrupamento em uma Rede direcionada, Grau total do nó = 3, Possíveis ligações = 3.2 = 6, Dos 3 vizinhos existem 3 ligações entre si, $C = 3/6 = 1/2$

Vamos calcular o coeficiente de agrupamento para cada nó da nossa rede! (Figura 127). O coeficiente de agrupamento médio da rede é a média dos coeficientes de agrupamento de todos os nós da rede. Esse valor nos indica qual a probabilidade de, ao sorteamos dois vizinhos de um certo nó, eles estarem interligados. Ou...a probabilidade da existência de triângulos nessa rede.

Pensando novamente na busca por emprego (ou informação), é de se supor que um emprego é algo escasso. Então alguém que te dá informação sobre um possível emprego conseguiu isso de uma fonte que você não tem! Considerem o nó C e seus amigos! Podemos perceber que a ligação de C com os nós A, B e D remete para uma ligação de um grupo muito bem definido de amigos. Porém, se observarmos a ligação entre C e H, percebemos que nenhum outro amigo de H é amigo de C, portanto ele pertence a um grupo distinto de amigos.

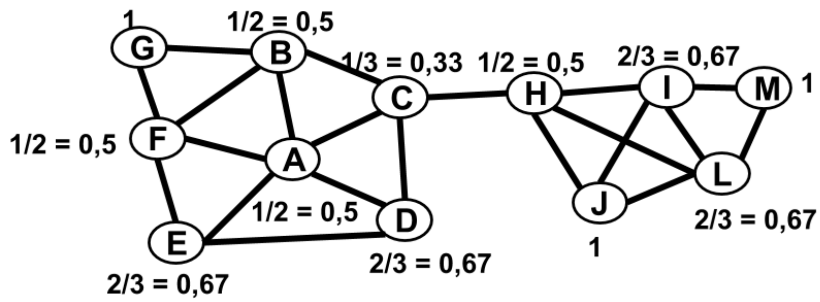


Figura 127: Grafo com os coeficientes de agrupamento em cada vértice.

O nó H sendo apenas um conhecido tem acesso às informações que o nó C não tem! Reparem que se eliminarmos a ligação entre C e H, os nós de A a G perdem comunicação com os nós de H a M. A aresta que, quando removida, desconecta a rede, é denominada PONTE (Figura 128). Mas, em redes reais, dificilmente temos uma única aresta que desconecta a rede (Figura 129).

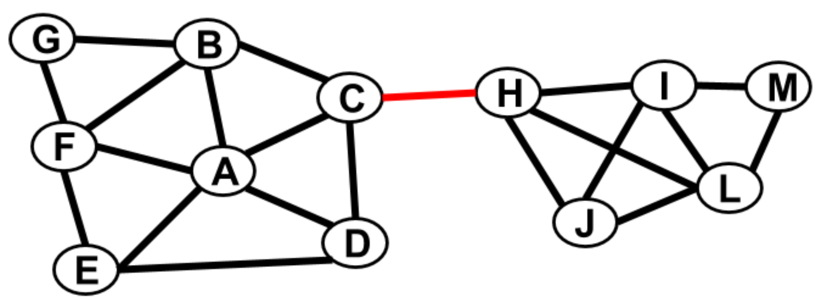


Figura 128: Grafo com uma aresta ponte em vermelho.

O menor número de arestas que desconectam a rede é denominado CONECTIVIDADE DE ARESTA da rede. Temos então que definir pontes de outra maneira! Por isso foi pensando no conceito de PONTES LOCAIS que são as arestas que, quando removidas, aumenta o caminho mínimo entre os dois nós em mais do que uma unidade. Repare que se removermos a aresta que liga os nós C ao D, a distância entre eles aumenta em apenas uma unidade! Mas ao remover a aresta $C - H$, o caminho entre esses nós aumenta em 6 unidades! Percebe-se então como o nó H, mesmo sendo apenas um conhecido, é capaz de passar informações melhores ao nó C do que seus melhores amigos (Figura 130).

Esses conceitos vistos até agora são importantes em diversas redes, pois nos ajuda a perceber:

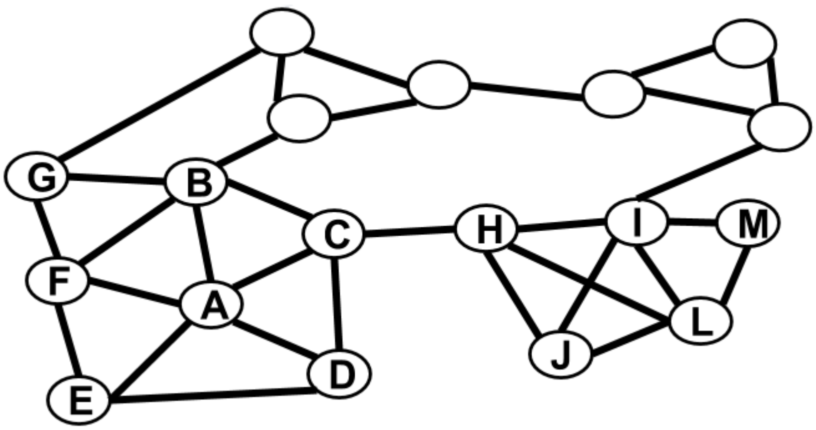


Figura 129: Nas redes reais se tem mais de um ponte.

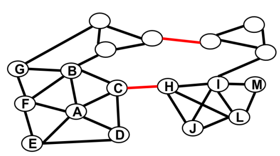


Figura 130: Grafo com duas arestas pontes em vermelho.

- Os nós e arestas críticas da rede
- Como comunidades são formadas
- Porque certos nós se juntam

4.9 Agrupamento em Redes Reais

Nas Redes Sociais, o coeficiente de agrupamento no Facebook é cerca de $0,27$. Existe uma chance de 30% de dois amigos de alguém também serem amigos...ou 70% de conhecer alguém novo através de um amigo [3].

Nas Redes Tecnológicas, a rede de transmissão de energia da parte leste dos Estados Unidos tem um coeficiente de agrupamento de $0,07$. A malha de transmissão foi planejada com 7% de redundância nos pontos mais críticos para ser tolerante a falhas[7].

Nas Redes Biológicas, a rede de reações metabólicas da bactéria E. Coli tem coeficiente de agrupamento de 0,59. As reações ocorrem em grupos, apenas algumas reações ocorrem como intermediário de um grupo para outro [2].

Redes de Informação, a rede do co-autoria da área de Saúde no Brasil tem coeficiente de agrupamento de 0,24. A transitividade entre autores é equivalente a uma rede social, as publicações em co-autoria são centradas em autores multidisciplinares [8].

5 Redes Aleatórias

5.1 Modelo genérico de Rede

Antigamente era impossível obter informações de todos os nós e arestas de algumas redes por limitações computacionais. Mesmo hoje, é impossível obter todas as interações entre proteínas de um organismo (Figura 130).

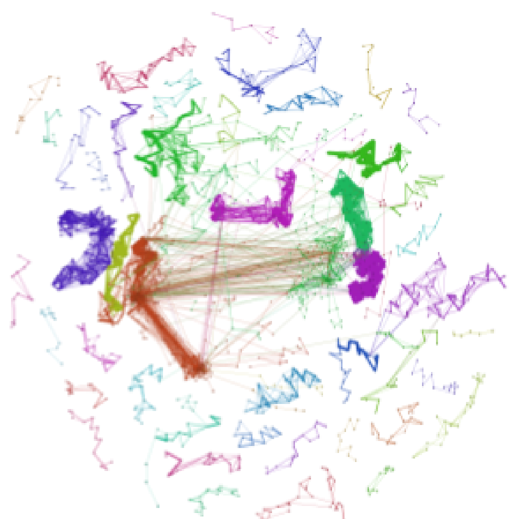


Figura 131: Interações nas Proteínas.

Imagine uma rede de co-ocorrência de palavras em um livro. Feito manualmente, mesmo com apenas algumas páginas, o trabalho poderia se tornar inviável e com alta taxa de erros. Hoje em dia ainda existem redes que não conhecemos por completo! E algumas redes ainda estão se formando!:

- Como estimar a estrutura de uma rede?;
- Como predizer os próximos nós e arestas dessa rede em construção?;
- Quais os próximos nós e arestas do planejamento de um transporte público?;
- Quais arestas irão surgir ou sumir em uma rede de relações comerciais?;
- Como uma rede social é formada?

Por que criar um modelo genérico de rede?

- Próximas conexões de transporte público?;

- Conexões serão criadas ou desfeitas em relações comerciais?;
- Qual a dinâmica de uma rede social?

Precisamos conhecer as propriedades de redes de interesse e modelar matematicamente para melhor compreensão. Além disso, como testar e validar algoritmos de redes se não temos redes de diferentes tamanhos e propriedades? Como verificar se diferentes redes reais seguem um mesmo modelo? Para isso precisamos identificar as propriedades de determinadas redes e tentar criar um modelo matemático que consegue definir uma rede com as mesmas propriedades.

5.2 Redes Aleatórias: Erdõs-Rényi

Paul Erdõs e Alfred Rényi propuseram um dos primeiros modelos para o estudo de análise de redes (Figura 132). Nesse estudo foi explicado o modelo de Redes Aleatórias, sugerindo que, muitas redes complexas na Natureza poderiam seguir um padrão aleatório de formação. Com o modelo proposto de Redes Aleatórias eles tentaram verificar como as redes sociais eram formadas.

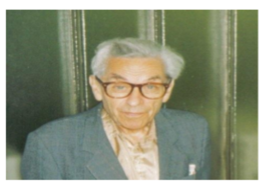



Figura 132: Fotos: na esquerda Paul Edrõs, na direita Alfred Rényi.

Para exemplificar eles pensaram na rede social formada em uma festa. Eles perceberam que bastava apenas uma conexão entre cada um dos convidados para que, ao final da festa, todos se tornassem conectados. Dessa forma concluiu-se que uma festa é um conjunto de agrupamentos de pessoas que se interligavam através de uma ou mais arestas (pontes) (Figura 133).

5.3 Redes Aleatórias

Em uma rede com n nós, criar arestas entre pares de nós de forma aleatória (por sorteio) e sem repetição. Inicialmente temos uma rede DESCONEXA (Figura 134). Quando o número de conexões aumenta, os COMPONENTES CONEXOS são interligados, formando grupos cada vez maiores. Quando o GRAU médio dos nós atinge um valor acima de 1, o número de COMPONENTES DESCONEXOS diminui exponencialmente. Rapidamente a REDE se torna CONECTADA (Figura 135).

Como uma analogia imagine uma festa onde cada cada convidado conhece de 2 a 3 pessoas (Figura 136).Os convidados formam então rodas de conversa e novas arestas são formadas (Figura 137). Existe a troca de rodas de conversa,

Figura 133: Em uma festa só é preciso uma conexão entre cada um dos convidados para se tornassem conectados. `http://greveufabc.wordpress.com/2012/08/08/fotos-da-festa-greve-olimpica/`

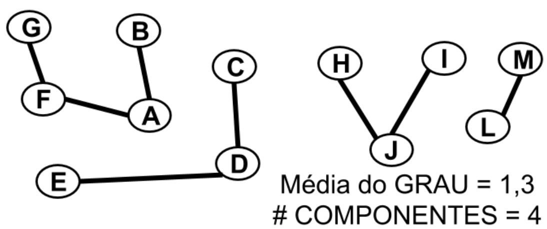


Figura 134: Rede aleatória com elementos desconexos.

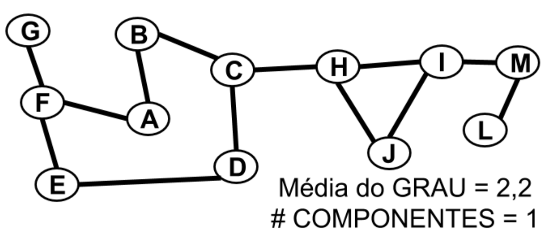


Figura 135: Redes aleatória conexa.

gerando mais novas arestas (Figura 138). E, eventualmente, desconhecidos também passam a se conhecer(Figura 139). Forma-se, ao final, uma rede conectada e com diâmetro pequeno (Figura 140). Porém com um coeficiente de agrupamento não tão alto! (Figura 141).

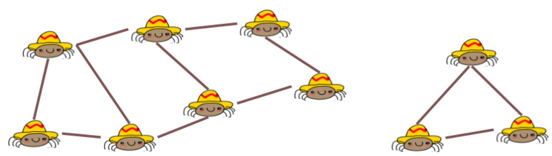


Figura 136: Analogia de uma Rede Aleatória com um Festa.

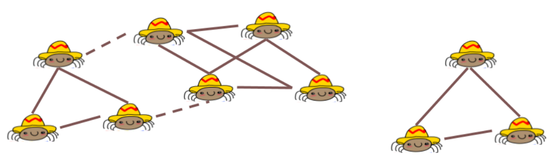


Figura 137: Analogia de uma Rede Aleatória com um Festa.

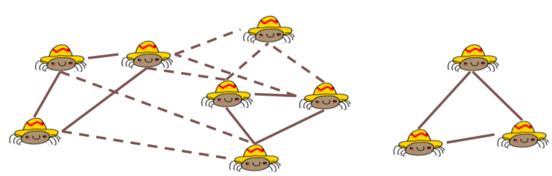


Figura 138: Analogia de uma Rede Aleatória com um Festa.

Criando uma Rede Aleatória, O modelo para a criação de uma rede aleatória de Erdõs-Rényi é representado por $G(n, p)$.

$$P = \frac{|E|}{n(n-1)}$$

Como ilustração vamos utilizar p=0.3 no modelo a seguir (Figura 142). Para cada aresta, sorteia um número aleatório uniforme entre 0 e 1 (Figura 143). Se o número for menor ou igual a 0,3 a aresta é criada (Figura 144). Caso contrário, a aresta é ignorada (Figura 145, Figura 146). Das possíveis arestas, cerca de 13 serão sorteadas (Figura 147). O coeficiente de agrupamento médio é igual a 0,44 (Figura 148).

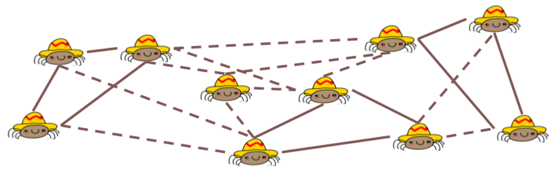


Figura 139: Analogia de uma Rede Aleatória com um Festa.

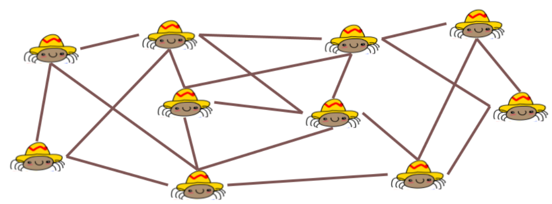


Figura 140: Analogia de uma Rede Aleatória com um Festa.

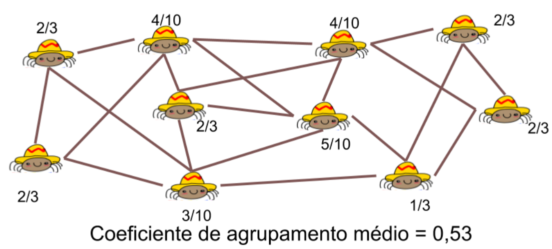


Figura 141: Analogia de uma Rede Aleatória com um Festa.

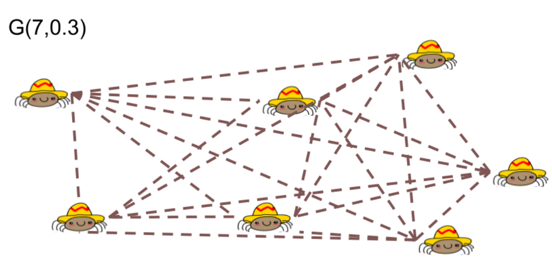


Figura 142: Geração de uma Rede aleatória.

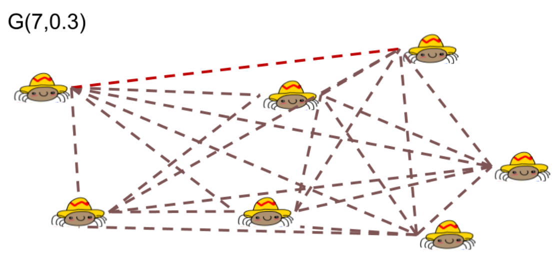


Figura 143: Geração de uma Rede aleatória.

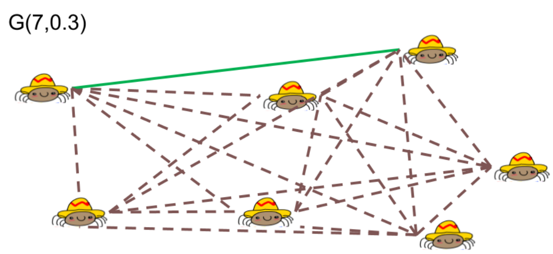


Figura 144: Geração de uma Rede aleatória.

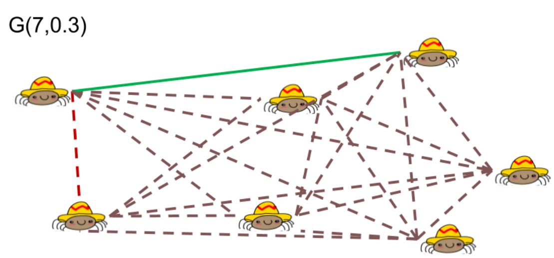


Figura 145: Geração de uma Rede aleatória.

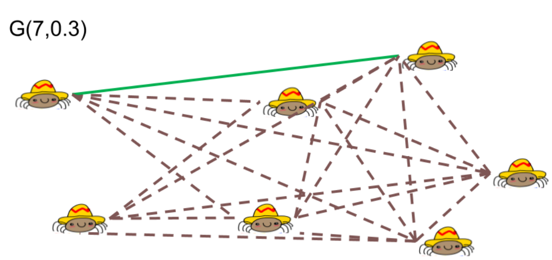


Figura 146: Geração de uma Rede aleatória.

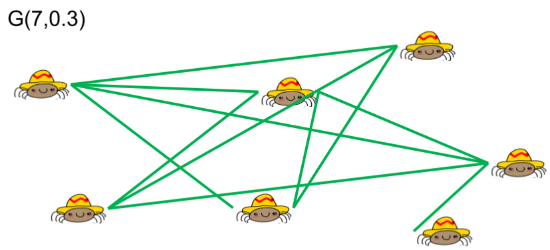


Figura 147: Geração de uma Rede aleatória.

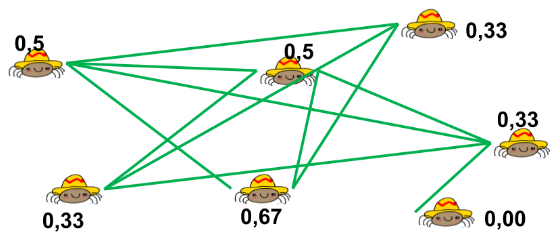


Figura 148: Geração de uma Rede aleatória.

Embora com esse modelo possamos gerar redes aleatórias de qualquer ORDEM, logo foi identificado que ela apresenta algumas características que diferem de redes reais, tais como:

- Tendência de geração de nós com coeficiente de agrupamento baixo e baixa ocorrência de fechamentos triádicos;
- A distribuição dos graus dos nós segue uma forma diferente de redes reais (Figura 149).

Nesse modelo $G(n, p)$, a quantidade de arestas esperada ao final do processo é calculada pela probabilidade de criação de uma aresta multiplicado por quantas arestas poderiam ser criadas. Aqui estamos levando em conta um grafo não direcionado É dizer no modelo $G(n, p)$ se espera obter $p * n * (n - 1)$ ou $E(arestas) = \binom{n}{2}p$. A pergunta que queremos fazer é, qual a probabilidade de nesse modelo um nó v possuir grau k?, $P(grau(v) = k) =?$

Essa probabilidade é a probabilidade de a inserção de aresta ocorrer k vezes e não ocorrer n-1-k vezes, ou seja, se posso inserir 10 arestas e estou interessado em grau 3, eu devo sortear um valor maior que p 3 vezes e menor por 7 vezes. E isso pode ocorrer de diversas maneiras, o total de combinações de arestas é calculada pelo binomial de n-1 nós que podem se ligar a esse nó e o número de nós que quero que se liguem, que é k.

$$P(grau(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

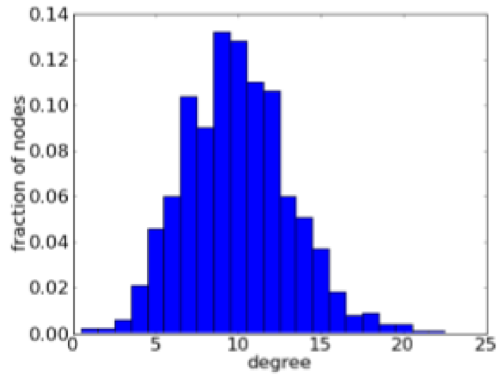


Figura 149: Distribuição dos Graus em redes Aleatórias. `https://mathinsight.org/image/network_ER_1000_0_01_undir_degree_distribution`

Essa formula pode ser aproximada para uma exponencial quando n tende a infinito.

$$P(grau(v) = k) \rightarrow \frac{(nk)^k \epsilon^{-np}}{k!}, n \rightarrow \infty$$

Esse modelo tem uma tendência de geração de nós com coeficiente de agrupamento baixo e baixa ocorrência de fechamentos triádicos (Figura 150).

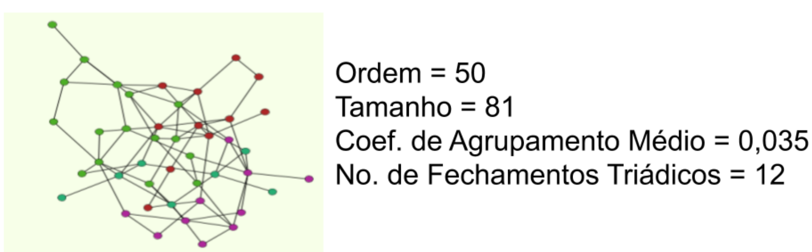


Figura 150: Exemplo de Rede aleatória com coeficiente de agrupamento baixo.

O formato da distribuição binomial para esse tipo de rede não corresponde ao que observamos na prática (Figura 151).

Nas redes reais temos essa distribuição, com uma cauda *long*, que aprenderemos mais adiante (Figura 151).

A grande utilidade desse modelo é que ela possibilita o estudo da difusão da informação na rede. Na aula 2 aprendemos o procedimento de Busca em Largura que indica como uma informação percorre a rede caso ela se propague igualmente para todos os vizinhos. Porém, na realidade, a informação percorre uma aresta

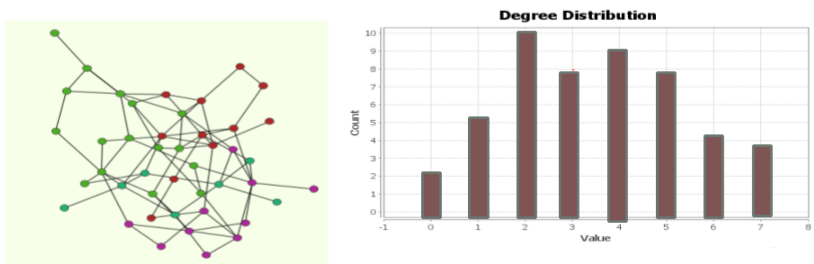


Figura 151: Rede aleatória com sua distribuição dos graus.

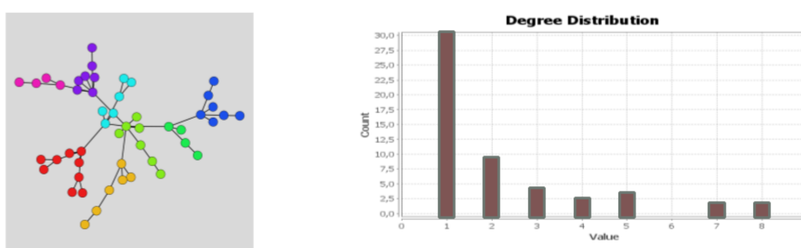


Figura 152: Rede real com sua distribuição dos graus diferente a uma de redes aleatórias.

com certa probabilidade!. Com isso, temos que a informação se propaga da mesma forma como uma rede aleatória Erdõs-Rényi é gerada!(Figura 153).

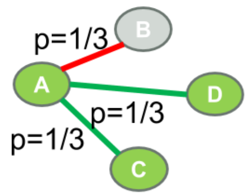


Figura 153: Rede aleatória com probabilidades nas arestas.

6 Redes Mundo Pequeno

Vimos que uma característica das redes reais é as propriedades de mundo pequeno: distância média pequena e coeficiente de agrupamento alto. Dados os nós A, B, C, D tal que (A,B) e (A,C) formam uma aresta. Em uma rede aleatória de erdos a probabilidade de ser criada a aresta (B,C) e (C,D) é a mesma!!. Mas em uma rede mundo pequeno o fato de A estar conectado tanto a B quanto a C faz com que a probabilidade de existir (B,C) seja maior do que (C,D). Lembrem-se do FECHAMENTO TRIÁDICO (Figura 154).

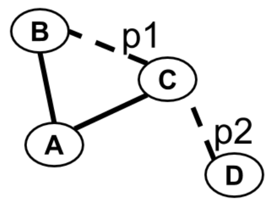


Figura 154: Na rede aleatória á probabilidade de criar uma aresta é a mesma $p1 = p2$, ma numa rede mundo pequeno depende do Fechamento Triádico $p1 > p2$.

6.1 Cadeias (*Chains*)

Após a segunda guerra mundial os governos e a sociedade voltavam a atenção em como reconstruir as cidades de forma otimizada levando em conta:

- Trânsito
- Vizinhança
- Distribuição demográfica

A importância desses estudos foram reforçados pelo autor húngaro Frigyes Karinthy em uma série de histórias curtas intituladas "Tudo é diferente" (Everything is Different). Uma dessas histórias tinha o nome de "Cadeias" (*Chains*) que explorava ideias que seriam futuramente alvo de interesse e estudo de diversos cientistas do campo de teoria das redes (Figura 155).

Ele comentava que o mundo moderno estava ENCOLHENDO assim como a conectividade entre os seres humanos através dos meios de comunicação e de transporte. Na história ele afirmava que quaisquer dois indivíduos estavam interligados através de, no máximo, 5 conhecidos. Um dos personagens desse conto propôs o seguinte jogo: "Vamos selecionar qualquer pessoa entre o 1,5 bilhão de habitantes da Terra, qualquer um em qualquer lugar. Aposto que usando não mais do que 5 pessoas, uma delas sendo um conhecido dele, eu conseguirei contactar a pessoa escolhida.".

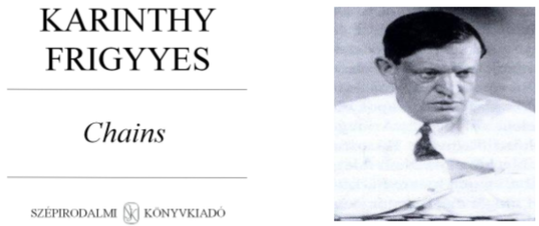


Figura 155: Esquerda: Historia Curta "Cadeias" (*Chains*), Direita : Frigyes Karinthy (autor)

6.2 Experimentos de Mundo Pequeno

Pool & Kochen, em 1958 fizeram diversos estudos em redes sociais tentando responder as seguintes questões:

- Qual a probabilidade de que duas pessoas escolhidas aleatoriamente de uma população tenham um amigo em comum?
- Quão longe as pessoas estão uma das outras através em sua rede de contatos?

Em experimentos feitos utilizando simulação de Monte Carlo com dados sociais adquiridos por Michael Gurevich, eles chegaram em um grau de separação de 3 pessoas na população americana.

6.2.1 Experimento de Milgram

Mas em 1967, Stanley Milgram deu continuidade aos experimentos de Pool e Kochen e propôs o seguinte experimento: Enviar várias cartas partindo de Nebraska/Kansas com destino a uma pessoa em Boston, porém através de intermediários . Foram enviadas 160 cartas com o seguinte conteúdo: Se você não conhece pessoalmente a pessoa destino, não tente contatá-la diretamente. Envie essa carta para um conhecido pessoal que tem mais chances do que você de conhecer a pessoa destino (Figura 156).

Das 160 cartas preparadas, 42 retornaram. O menor caminho foi de UMA conexão e o mais longo de DEZ. O valor MÉDIO foi de 5,5 conexões, o que foi surpreendentemente baixo. Arredondando os 5,5 temos a famosa frase SEIS GRAUS DE SEPARAÇÃO (Figura 157). O termo SEIS GRAUS DE SEPARAÇÃO foi cunhada e popularizada pelo autor John Guare que escreveu uma peça teatral com este título para a Broadway em 1990, que se tornou filme em 1993. O termo MUNDO PEQUENO diz respeito ao fato de que muitas redes reais, embora tenham um número grande de nós, existe um CAMINHO DE TAMANHO PEQUENO entre quaisquer dois nós.

Figura 156: Mapa de EEUU sinalizando o fluxo de cartas entre os estados no experimento de Milgram.

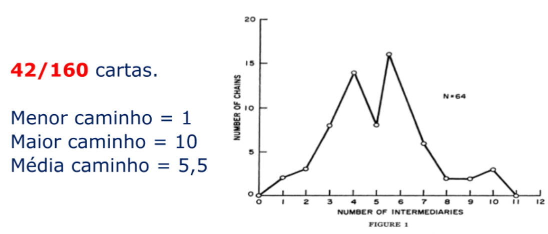


Figura 157: Relação entre o numero de cadeias e intermediários.

6.2.2 Seis graus de Erdos

Em 1995, Grossman & Ion coletaram dados de uma rede de co-autoria de artigos matemáticos tendo como foco central Paul Erdõs. Erdõs foi co-autor de artigos com milhares de outros autores, porém tiverem alguns autores em particular que ele escreveu mais artigos (esse fenômeno será visto mais adiante). Dos dados coletados foi criado o número de Erdõs:

- Um autor que escreveu um artigo em co-autoria com Erdõs tem um número de Erdõs igual a 1.
- Um autor que não escreveu nenhum artigo em co-autoria com ele, mas escreveu em co-autoria com alguem com valor 1 teria um número de Erdõs igual a 2.
- E assim por diante...

O número de Erdõs varia entre 0 (ele mesmo) até 15, tendo uma média de 4,65, não muito distante dos seis graus de separação (Figura 158). Por conta da interdisciplinaridade que temos atualmente no meio científico, muitos cientistas de diversas áreas possuem um número de Erdõs finito. O número de Erdõs pode ser verificado em `http://www.ams.org/mathscinet/collaborationDistance.html`.

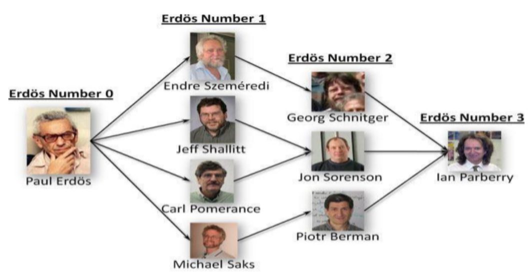


Figura 158: Rede de pesquisadores com o número de Erdõs. `http://larc.unt.edu/ian/numbers.html`

Similar ao número de Erdõs, no jogo dos Seis Graus de Bacon procura-se verificar o grau de separação entre o ator Kevin Bacon e todos os outros atores de Hollywood no site `http://oracleofbacon.org/` (Figura 159). Utilizou-se a base de dados disponível no site `http://www.imdb.com/interfaces/`. Algumas caraterística de este experimento são:

- Kevin Bacon obviamente tem grau 0.
- Um ator tem grau de bacon 1 se ele atuou em um mesmo filme que Kevin Bacon.

- Um ator tem grau de bacon 2 se ele nunca contracenou com Kevin Bacon mas atuou em um filme com um ator que tem grau 1.
- O número médio atual de Bacon é 2,989.
- O maior grau de Bacon até o momento é 8.
- Aproximadamente 12% dos atores não conseguem se conectar com Bacon.
- Dennis Hopper é um melhor centro para esse jogo, tendo um grau médio de separação de 2,835 em relação aos outros atores.

Existem pessoas com número de Bacon E número de Erdõs!!!, `http://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93Bacon_number`. Bruce Reznick, professor da Universidade de Illinois tem um número de Erdős igual a 1 e um número de Bacon igual a 2, pois foi um extra em Pretty Maids All in a Row com Roddy McDowall `http://en.wikipedia.org/wiki/Pretty_Maids_All_in_a_Row` .

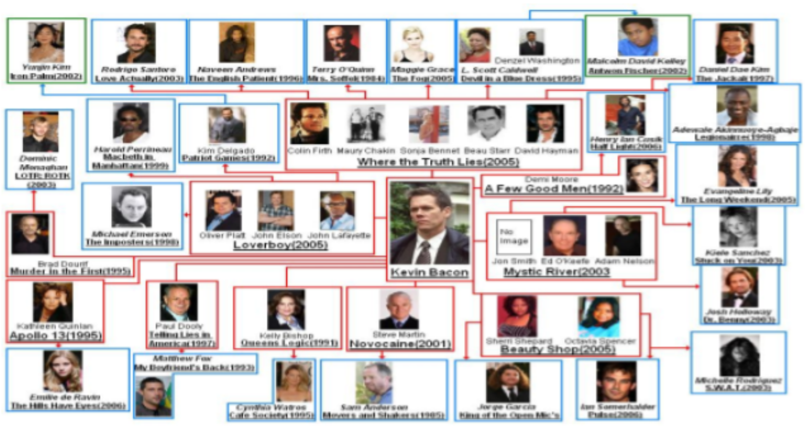


Figura 159: Rede de atores no jogo dos "Seis Graus de Bacon" `http://blog.globalpatentsolutions.com`.

6.3 Modelo Mundo Pequeno

Em 1998, Watts e Strogatz definiram um modelo matemático para gerar uma rede aleatória com características de um rede mundo pequeno (Figura 160). Esse modelo simplesmente utiliza o modelo Erdõs-Rényi partindo de uma rede com estrutura em ANEL. Rede em ANEL é quando as arestas interligam os k nós vizinhos de tal forma que eles possam ser organizados em forma de uma circunferência (Figura 161). O modelo parte de uma rede regular em anel, respeitando o número n de nós e cada nó interligado aos k nós vizinhos mais próximos por meio de arestas (Figura 162).

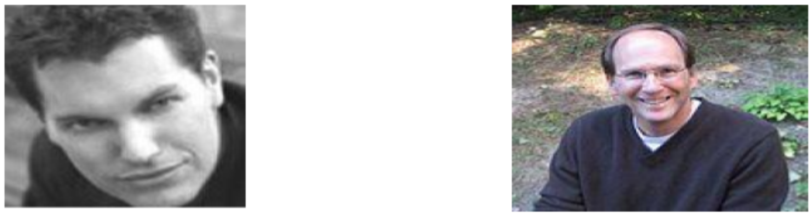


Figura 160: Na esquerda Watts e na direita Strogatz.

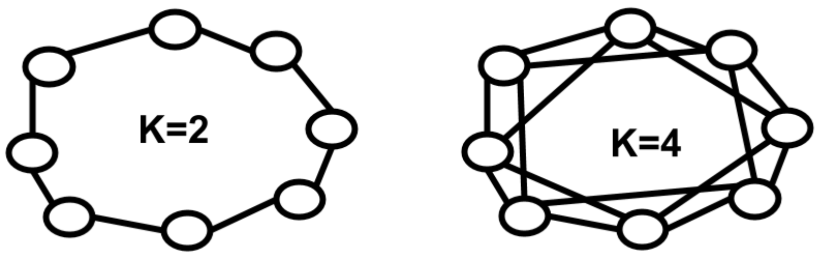


Figura 161: Redes Pequeno mundo em forma de anel, de onde k é o numero de arestas.

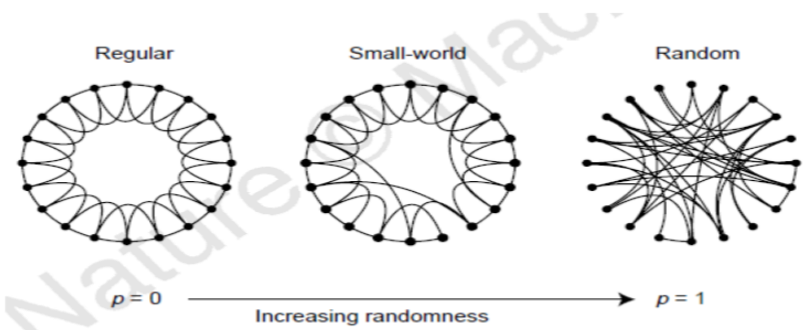


Figura 162: Variação dos tipos de redes em função a sua aleatoriedade [10].

Vamos fazer um exemplo, o modelo $G(n, k, p)$, cria uma rede com n nós, cada nó com grau inicial igual a k e probabilidade de reconexão igual a p. Para cada nó n_i, conectá-lo aos nós n_j tal que $j = i - k/2..i + k/2$ (definindo o grau de n_i como k). Para cada nó n_i e cada aresta (n_i, n_j) desse nó, reconecta essa aresta, criando uma aresta (n_i, n_k) com um nó escolhido aleatoriamente com probabilidade uniforme e $k \neq i$ dada uma probabilidade p (Figura 163).

Para cada aresta, sorteia um número entre 0 e 1 e, se for menor ou igual a $0,3$, reconecta a aresta (Figura 164) (Figura 165) (Figura 166). Com esses parâmetros, cerca de 4 arestas vão ser reconectadas ao todo (Figura 167). E o coeficiente de agrupamento médio igual a $0,84$ (Figura 168). Um coeficiente de $0,84$ para uma rede é grande ou pequeno? Para comparar vamos calcular a DENSIDADE da rede, que é a razão entre o número de arestas e quantas arestas ela poderia ter (Figura 169). A densidade da rede é a probabilidade de que, se pegarmos dois nós aleatoriamente, eles estejam conectados (Figura 170). Então se a média do coeficiente de agrupamento é muito maior que p:

$$p = \frac{2E}{n(n-1)}$$

No nosso caso, $p = 12/(6 * 5/2) = 12/15 = 0,8$. E nosso $C = 0,84$; note que a diferença é menor pois temos uma rede com poucos nós, em redes maiores a diferença tende a ser maior (Figura 171). As redes criadas pelo modelo de Watts-Strogatz tem as seguintes propriedades:

- O grau de separação médio varia rapidamente entre $n/2k$ e $\ln(n)/\ln(k)$ quando a probabilidade p varia entre 0 e 1;
- O coeficiente de agrupamento fica em torno de $3(1-p)3/4$;
- A distribuição do grau continua se aproximando de Poisson.

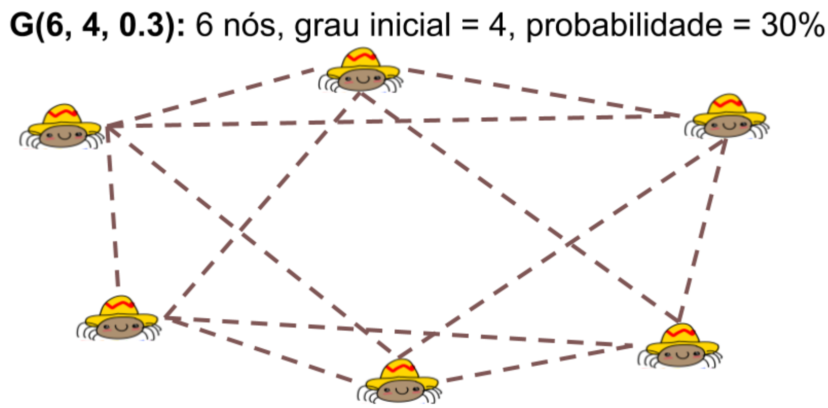


Figura 163: Exemplo rede Pequeno Mundo.

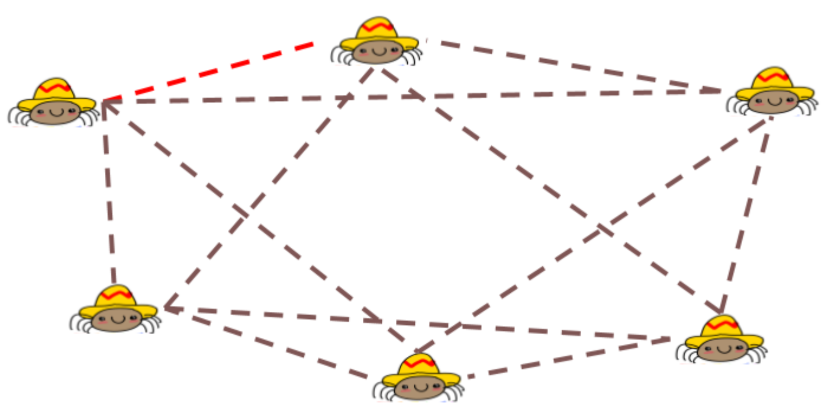


Figura 164: Exemplo rede Pequeno Mundo.

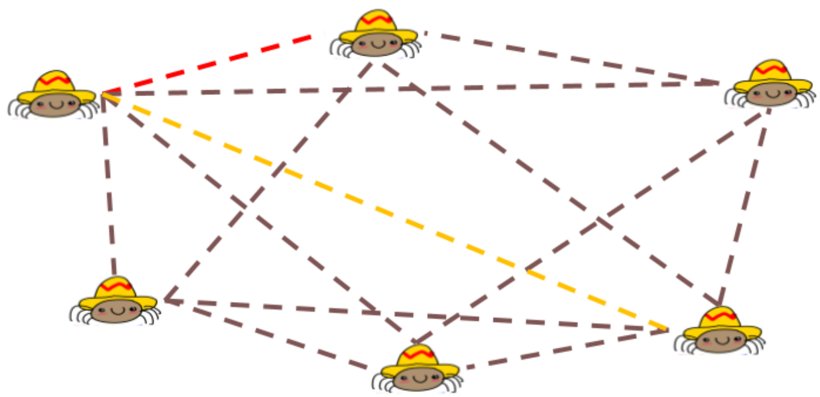


Figura 165: Exemplo rede Pequeno Mundo.

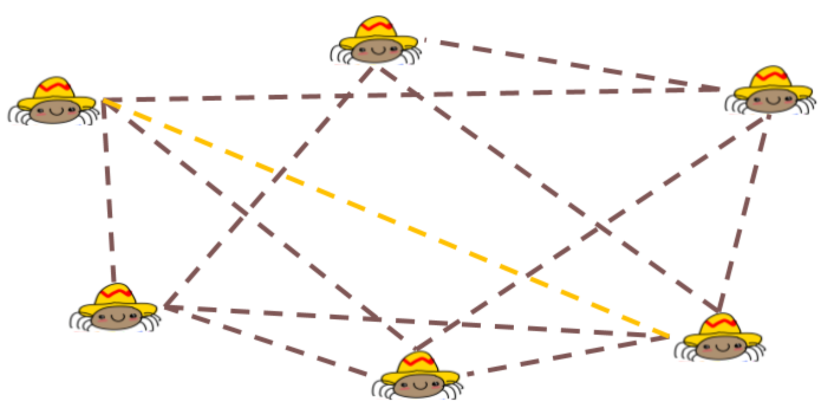


Figura 166: Exemplo rede Pequeno Mundo.

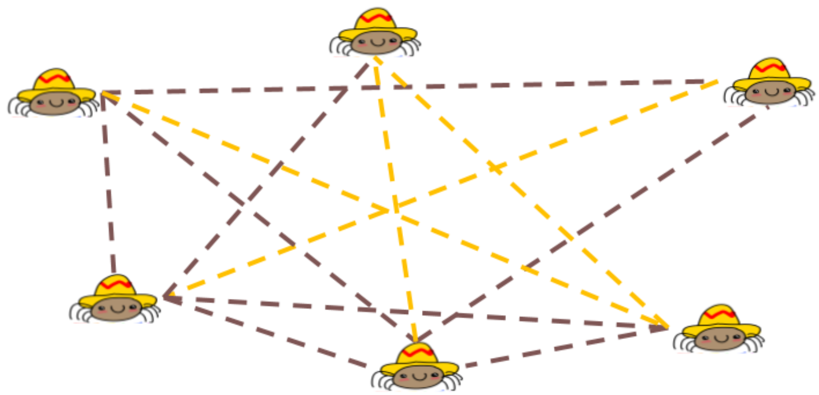


Figura 167: Exemplo rede Pequeno Mundo.

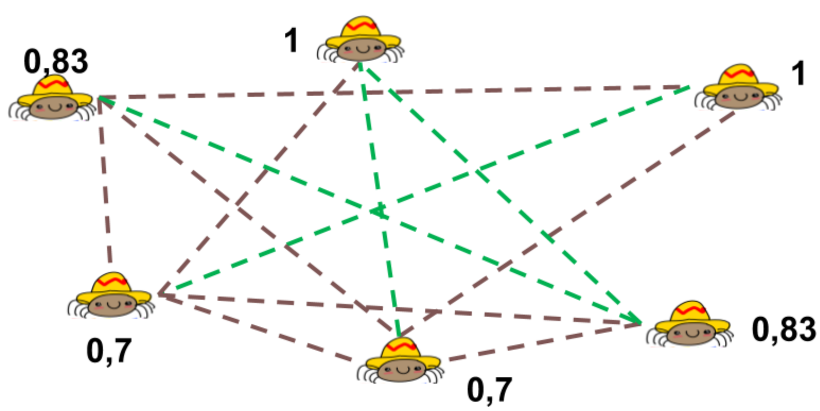


Figura 168: Exemplo rede Pequeno Mundo.

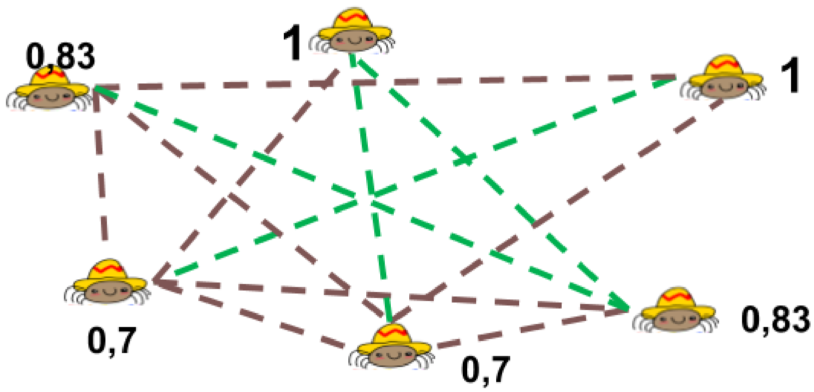


Figura 169: Exemplo rede Pequeno Mundo.

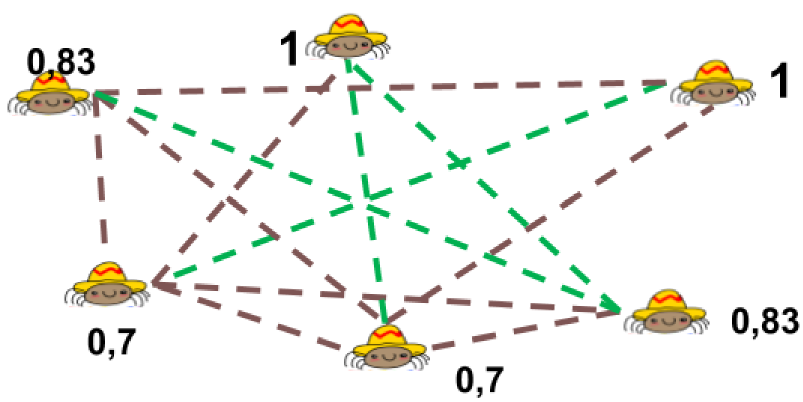


Figura 170: Exemplo rede Pequeno Mundo.

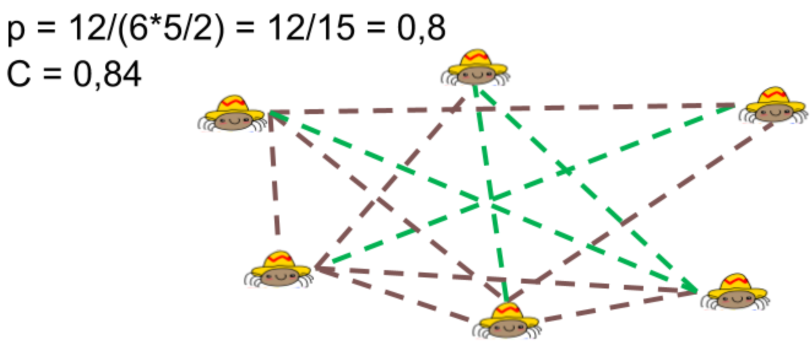


Figura 171: Exemplo rede Pequeno Mundo.

7 Padrões de Distribuição

7.1 Distribuição dos Parâmetros da Rede

7.1.1 Distribuição de Dados

Ao realizar medições de diversas naturezas, observamos uma distribuição normal (Gaussiana, curva-sino) dos valores mensurados. Um exemplo disso é a Altura dos goleiros do campeonato brasileiro: a maioria dos goleiros tem uma altura em torno de 1,88 a 1,91 metros. Porém temos uma minoria em torno de 1,82m e outra minoria em torno de 1,97m (extremos) (Figura 172).

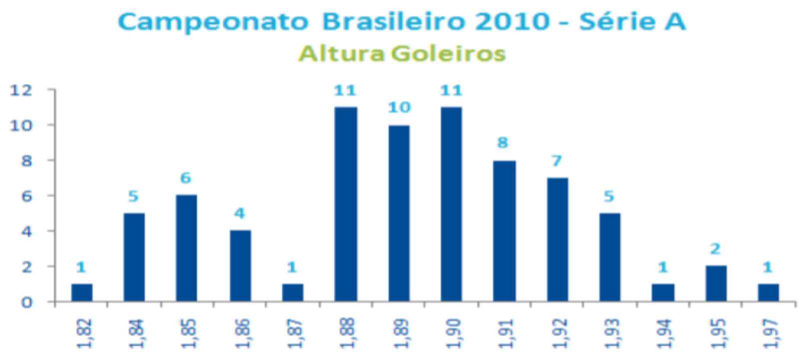


Figura 172: Distribuição da quantidade por altura dos goleiros.

Outro exemplo é a Velocidade média de carros em rodovias: em rodovias como a Imigrantes, por exemplo, a velocidade média da maioria dos carros fica em torno de 80Km/h e 100Km/h, tendo uma minoria em torno de 50Km/h e outra minoria em torno de 120Km/h (Figura 173). Porém, muitas observações possuem uma distribuição diferente, seguindo uma função de potência. Levando em conta a rodovia bandeirantes e o sistema anchieta-imigrantes, a maioria dos carros percorrem cerca de 100Km (Santos – SP), enquanto uma minoria percorre trechos muito maiores (viajantes ocasionais) (Figura 174).

7.2 Medidas de Distribuição

Esse tipo de distribuição foi observada pelo economista italiano Vilfredo Pareto quando estudava a distribuição da riqueza na Itália (Figura 175). Ele percebeu que:

- Grande parte das terras pertenciam a minoria da população;
- Boa parte do lucro de uma empresa chegava no bolso de pouquíssimos funcionários.

Essa distribuição foi observada em outras situações, como:

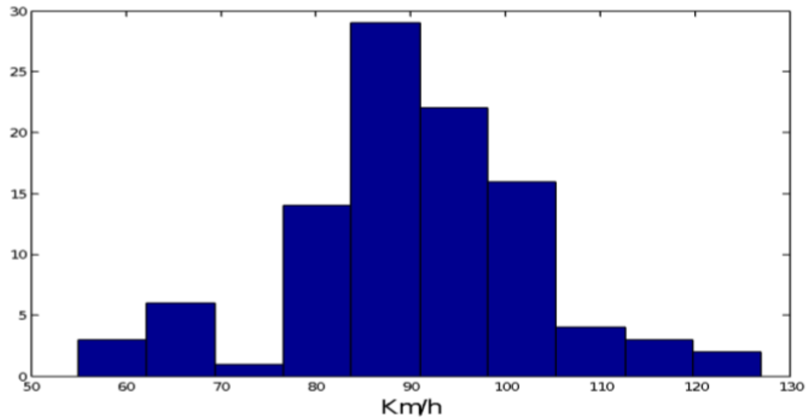


Figura 173: Distribuição velocidade numa rodovia.

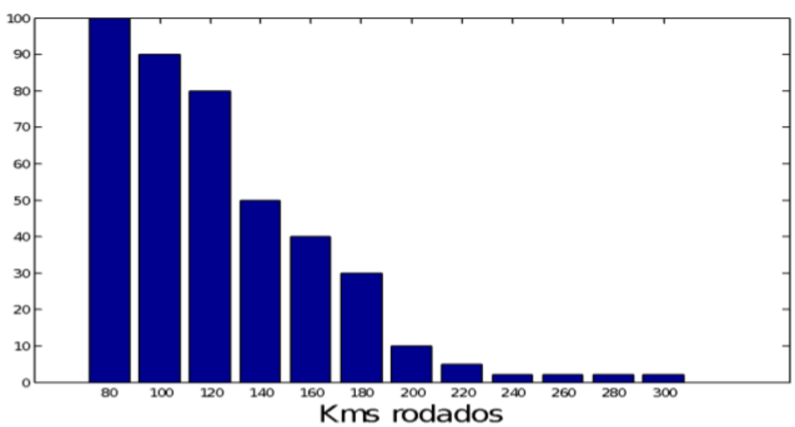


Figura 174: Distribuição do percurso em base da rodovia bandeirantes e o sistema anchieta-imigrantes.

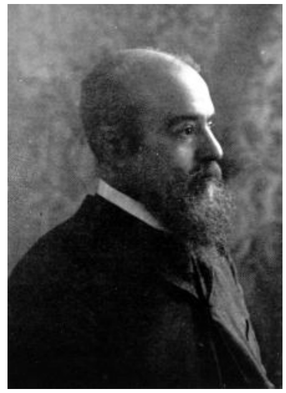


Figura 175: Vilfredo Pareto.

- Poucas cidades concentram maior parte da população;
- Maior parte dos crimes são cometidos pela menor parte dos criminosos;
- Uma minoria das palavras de um idioma aparecem com uma frequência muito acima da maioria das palavras.

Vejam, por exemplo, a distribuição de cidades de acordo com o tamanho de sua população (Figura 176). Plotando esse mesmo gráfico em escala logarítmica ($\lg(x) * \lg(F(x))$), temos que a distribuição se torna aproximadamente linear (Figura 177).

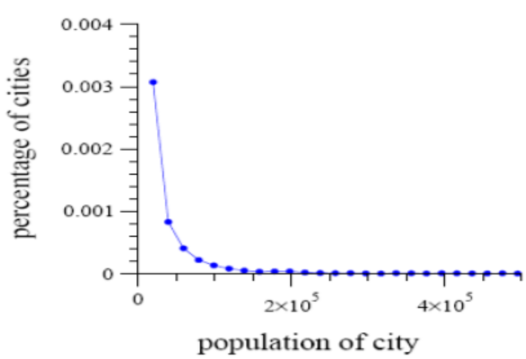


Figura 176: Histograma de ás populações de todas as cidades dos US com uma população de 10000 ou mais.

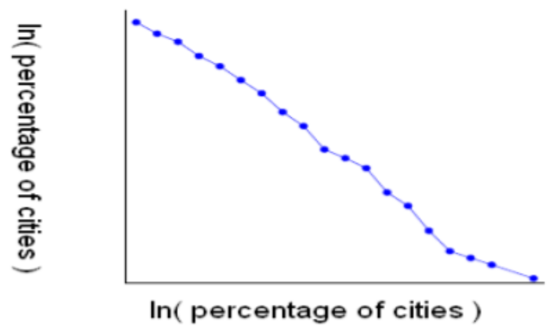


Figura 177: Outro Histograma com a mesma data, mais desenhada nas escalas logarítmicas.

7.3 Lei da Potência

Esse modelo linear pode ser descrito como:

$$
\begin{aligned}
\lg(y) &= -\alpha \lg(x) + b \\
y &= \epsilon^{-\alpha \lg(x) + b} \\
y &= \epsilon^{\lg(x^{-\alpha}) + b} \\
y &= \epsilon^{\lg(x^{-\alpha})} \epsilon^{b} \\
y &= c \cdot x^{-\alpha}
\end{aligned}
$$

O tipo de distribuição regida pela função $f(x) = c \cdot x^{-\alpha}$ é chamada distribuição de lei de potência. Essa distribuição implica que pequenas ocorrências são extremamente comuns, enquanto grandes quantidades são raras. Outra característica da Lei da Potência é que ela é invariante em escala. Se pegarmos uma amostra da nossa distribuição ela terá o mesmo formato da distribuição completa. Multiplicar a variável da função de potência simplesmente aumenta ou reduz o valor da função em um fator constante.

7.4 Invariância em Escala

Dada a relação $f(x) = ax^k$, escalonando o argumento x por um fator constante causa apenas um escalonamento proporcional da própria função. Isto é:

$$
\begin{aligned}
f(x) &= ax^k \\
f(c \cdot x) &= a(c \cdot x)^k = a \cdot c^k x^k \inf f(x)
\end{aligned}
$$

Um exemplo de invariância a escala são os fractais, os fractais são imagens auto-similares em que observando os detalhes microscópicos obtemos uma imagem similar ao macroscópico. Esse tipo de padrão é frequentemente encontrado na natureza, principalmente na estrutura das plantas (Figura 179).

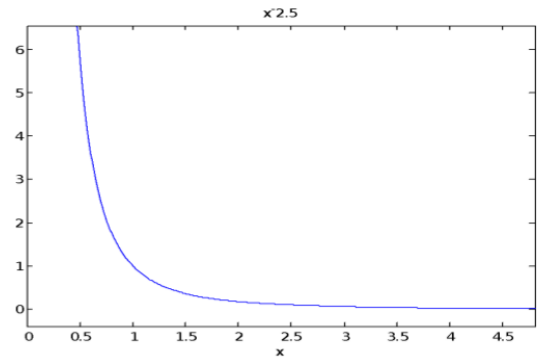


Figura 178: Distribuição função potencia

Figura 179: Brócoli formado por fractais. `https://en.wikipedia.org/wiki/File:Fractal_Broccoli.jpg`

7.5 Lei de Potência em Redes

A distribuição de Pareto e a lei de potência em redes é geralmente verificada através de dados estatísticos das propriedades dos nós e da rede. Grau: verifica-se a função de frequência $f(k)$ de nós com grau k. A invariância da rede vem do fato que a proporção dos nós com grau $2k$ e dos nós com grau k é uma constante para qualquer k. Dessa forma não há distinção da estrutura global e local da a rede (Figura 180).

$$\frac{f(2k)}{f(k)} = \frac{c(2k)^{-\alpha}}{ck^{-\alpha}} = 2^{-\alpha}$$

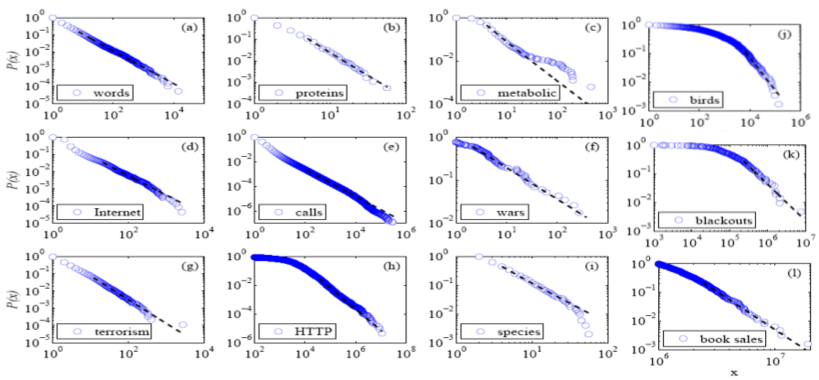


Figura 180: Correlação da Lei de Potência nas redes.

Propriedade: A distância média entre os nós cresce de forma logarítmica em função do tamanho da rede ($log(N)$).

$$\bar{d} \infty \lg n$$

Redes com essa característica tem uma menor chance de se tornar desconectada quando sofre um ataque aleatório.Exemplos de ataques em rede:

- Hackers atacando redes de internet
- Falha em um nó de uma malha de distribuição de energia
- Surgimento de um vírus
- Incêndio em uma floresta

Porém, um ataque direcionado aos nós centrais pode facilmente desconectar a rede, dividindo a rede em vários COMPONENTES CONEXOS. Em redes que tem uma capacidade de transmissão de informação especificada para cada nó e aresta, o ataque a um nó central pode sobrecarregar outros nós centrais, gerando uma desconexão em cascata.

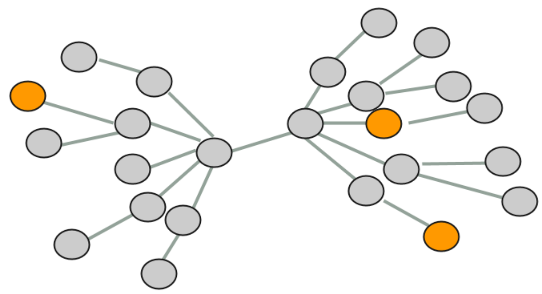


Figura 181: Invariância em Redes.

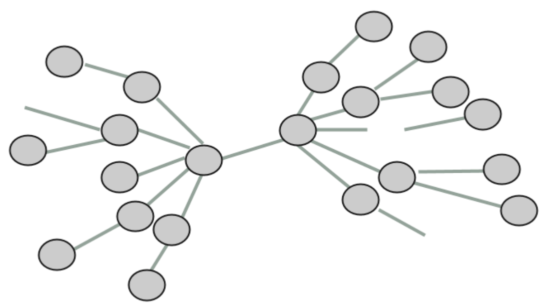


Figura 182: Invariância em Redes.

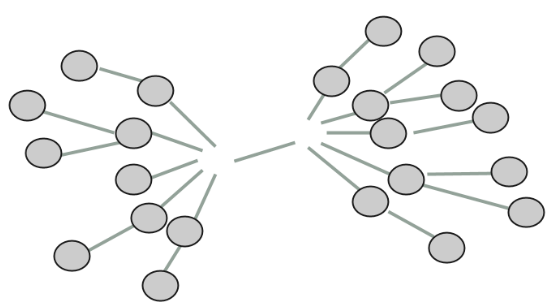


Figura 183: Invariância em Redes.

7.6 Cidades como Redes

As cidades urbanas podem ser encaradas como um ser vivo. Na verdade a vida de uma cidade é composta pela rede formada pelos seus habitantes. Diversas características importantes emergem por conta disso.

Supondo que o crescimento de uma cidade segue a lei de potência, podemos definir $N(t)$ como o número de habitantes no instante t e $Y(t)$ uma propriedade mensurável dessa cidade nesse instante. Com isso temos a equação:

$$Y(t) = Y_0 N(t)^\alpha$$

Podemos citar como exemplos de propriedades mensuráveis: :

- Energia consumida;
- Riqueza;
- Patentes;
- Poluição;
- Crimes;
- Número de mercados;
- etc.

Nessa tabela (Figura 184), beta β representa nosso coeficiente α. Aqui podemos ver os diferentes expoentes encontrado para diferentes propriedades. O R^2 indica quão acertado é o uso desse expoente para os dados utilizados, mais próximo de 1 melhor.

Y	β	95% CI	Adj-R^2	Observations	Country–year
New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
"Supercreative" employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
GDP	1.15	[1.06,1.23]	0.96	295	China 2002
GDP	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003

Figura 184: Tabela 1 `https://www.pnas.org/content/pnas/104/17/7301.full.pdf`

Continuando a tabela anterior, percebemos alguns expoentes menores do que 1. A escala em lei de potência das cidades, difere do padrão encontrado nos seres vivos em que os expoentes eram múltiplos de 1/4 (Figura 185). Veremos mais adiante que esse expoente tem relação com a dimensão do sistema

Y	β	95% CI	Adj-R^2	Observations	Country–year
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003
Total housing	1.00	[0.99,1.01]	0.99	316	U.S. 1990
Total employment	1.01	[0.99,1.02]	0.98	331	U.S. 2001
Household electrical consumption	1.00	[0.94,1.06]	0.88	377	Germany 2002
Household electrical consumption	1.05	[0.89,1.22]	0.91	295	China 2002
Household water consumption	1.01	[0.89,1.11]	0.96	295	China 2002
Gasoline stations	0.77	[0.74,0.81]	0.93	318	U.S. 2001
Gasoline sales	0.79	[0.73,0.80]	0.94	318	U.S. 2001
Length of electrical cables	0.87	[0.82,0.92]	0.75	380	Germany 2002
Road surface	0.83	[0.74,0.92]	0.87	29	Germany 2002

Data sources are shown in *SI Text*. CI, confidence interval; Adj-R^2, adjusted R^2; GDP, gross domestic product.

Figura 185: Tabela 2 `https://www.pnas.org/content/pnas/104/17/7301.full.pdf`

7.7 Padrões de Exponente

O expoente nas cidades podem apresentar três características distintas:

- $\alpha = 1$: comportamento linear, a quantidade mensurada cresce igual ao número de novos habitantes. associado com necessidades individuais (emprego, casa, consumo de água)
- $\alpha < 1$: comportamento sublinear, a quantidade mensurada cresce em uma taxa menor do que a população. quantidades materiais e infraestrutura (postos de combustível, extensão de cabos elétricos).
- $\alpha > 1$: comportamento super-linear, a quantidade mensurada cresce em uma taxa maior do que a população. contrapartidas sociais (informação, crime, saúde).

O fato das contrapartidas sociais apresentarem um fator de escala super-linear, indica que, ao contrário dos sistemas biológicos, a vida urbana apresenta uma maior velocidade quanto mais ela cresce e se desenvolve (Figura 186).

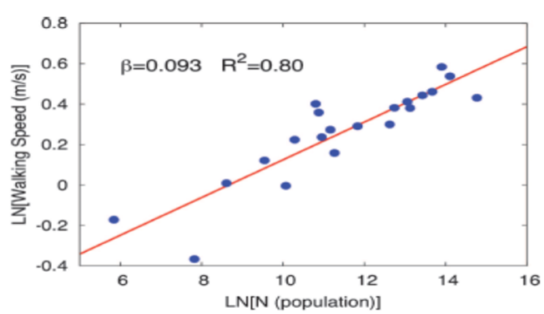


Figura 186: A velocidade de caminhada aumenta linearmente com a quantidade de população.

7.8 Equações de Crescimento

Denotamos Y como a quantidade do recurso mensurável que estamos estudando. Se R é a quantidade de recurso por unidade de tempo para manter um indivíduo na população e E a quantidade necessária para adicionar um indivíduo. Também definimos N como a quantidade de indivíduos na população e dN/dt a taxa de crescimento dela.

Temos então que a quantidade de um certo recurso é igual a quanto desse recurso preciso manter pra cada indivíduo, multiplicado pela população e temos que somar também quanto recurso eu preciso adicionar multiplicado pela taxa de crescimento do indivíduo. Isso é uma equação diferencial, podemos rearranjar para chegarmos em uma equação da taxa de crescimento da população.

$$Y = R \cdot N + E \cdot \frac{dN}{dt}$$

$$\frac{dN}{dt} = \frac{Y}{E} - \frac{R}{E} \cdot N$$

Lembrando que Y segue uma lei de potência, temos que a taxa:

$$Y = Y_0 N(t)^\alpha$$

$$\frac{dN}{dt} = \frac{Y}{E} - \frac{R}{E} \cdot N \longrightarrow \frac{dN(t)}{dt} = \frac{Y_0}{E} - \frac{R}{E} \cdot N$$

A solução dessa equação diferencial fica. Ela descreve a equação que rege o crescimento da população, ou seja, em um certo instante t, qual o tamanho da população? Mas não se assustem com essa equação, vamos fazer algumas análises em função de α se ele está caminhando para 1, se ele é acima de 1, ou se é abaixo de 1.

$$N(t) = \left[\frac{Y_0}{R} + \left(N(0)^{1-\alpha} - \frac{Y_0}{R}\epsilon^{-\frac{R}{E}(1-\alpha)t}\right)\right]^{\frac{1}{1-\alpha}}$$

Quando α se aproxima de 1, temos um crescimento exponencial da população! Ou seja, quando falamos de crescimento linear, como recurso de residências, empregos, isso favorece o crescimento exponencial da população, é possível relacionar a quantidade E/R como o tempo que um indivíduo médio atinge maturidade e passe a produzir recurso, então podemos ler essa equação como, dada uma população inicial $N(0)$, em certo momento eles podem produzir recurso (p.ex., novos empregos) que permitem um aumento da população a quantidade de recursos que eles criam é suficiente para sustentar esse crescimento exponencial (Figura 187).

Quando $\alpha < 1$, observamos um crescimento logístico. Em certo momento ocorre uma estagnação por causa dos recursos limitados! Os recursos com α menor que 1 são aqueles relacionados a serviço, por exemplo, mercados, postos de gasolina, esse tipo de recurso, pensando regionalmente, consegue atender um número limitado de pessoas isso impõe esse limite no crescimento em seu

$$N(t)_{\alpha \to 1} = N(0)e^{\frac{(Y_0 - R)t}{E}}$$

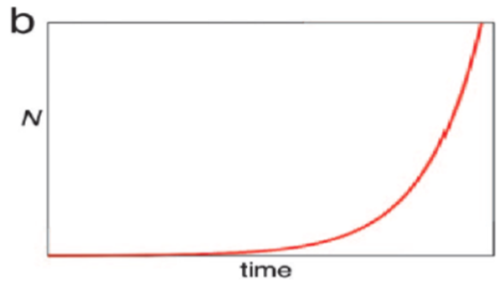


Figura 187: Equações Crescimento.

$$\alpha < 1$$

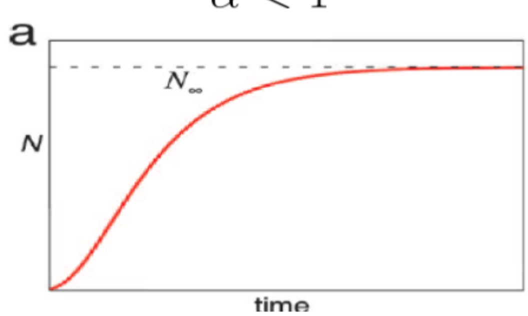


Figura 188: Equações Crescimento.

entorno. Em certo momento, para sustentar mais pessoas precisamos construir novos postos de serviço, porém em outra região (Figura 188).

Quando α é maior que 1, observamos um crescimento explosivo, isso é guiado pela inovação e riqueza. Esse gráfico implica uma população infinita em um tempo t_c finito Esse tipo de recurso é associado com riqueza, por exemplo, novas patentes, investimento em pesquisa, a criação desse tipo de recurso causa um impacto no ambiente urbano do entorno atraindo um crescimento gigantesco na população pensem aqui mesmo na cidade de São Paulo, quantas pessoas são atraídas para cá por conta da riqueza da cidade, das inúmeras oportunidades Esse tipo de crescimento é insustentável e causa todo tipo de problema como necessidade de aumento no fornecimento de residências, produtos de consumo (Figura 189).

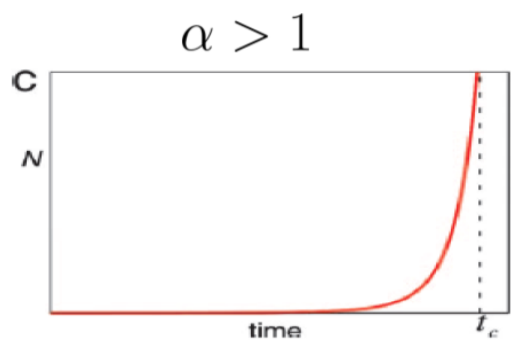


Figura 189: Equações Crescimento.

E, como esses recursos não podem ser infinitos o que realmente acontece após t_c é que o sistema entra em colapso!(Figura 190).

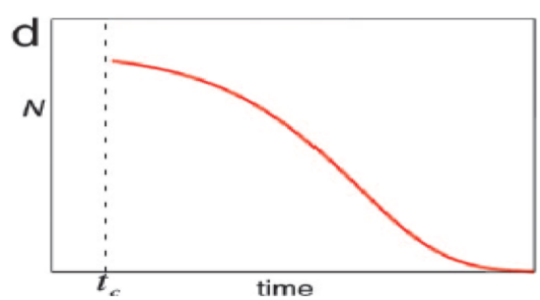


Figura 190: Equações Crescimento.

Na prática o que acontece é que em épocas de crise, surgem novas formas de uso e criação de recursos que dá novo "fôlego" para o crescimento. Mas o tempo entre crises se torna cada vez menor (Figura 191).

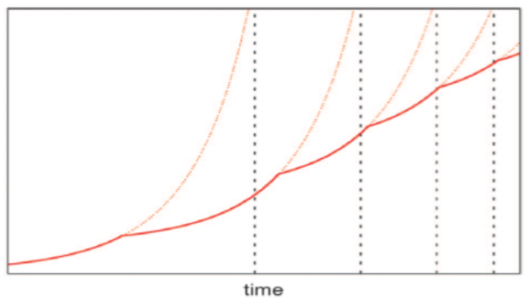


Figura 191: Equações Crescimento.

Em resumo, temos que a auto-organização da cidade pode ser dividida em três tipos de recurso, aqueles relacionados com produtos de consumo que tem como objetivo a otimização de seu uso, eles são sustentáveis e implicam em um crescimento sigmoidal temos aqueles relacionados a riqueza a criação de novos recursos que podem levar a um colapso do crescimento e finalmente aqueles de manutenção individual que são recursos consumidos por um único indivíduo ou família e implica no crescimento exponencial (Figura 192).

Expoente	Objetivo	Organização	Crescimento
$\alpha < 1$	Otimização e eficiência	Sustentável	Sigmoidal
$\alpha > 1$	Riqueza, recursos, informação	Agregadora	Explosão/Colapso
$\alpha = 1$	Manutenção Individual	Individual	Exponencial

Figura 192: Escala Auto-organização.

7.9 Modelos de Rede Livres de Escala

Seguindo os modelos de redes aleatória e de mundo pequeno vistos anteriormente e, com a percepção da característica de lei de potência de redes reais, é desejável criar modelos que consigam reproduzir a criação de diversas redes reais e suas características. O primeiro passo para definir tal(is) modelo(s) é investigar como o padrão invariante em escala é construído. Os fractais que mencionamos anteriormente são produzidos através de uma construção iterativa de sua figura. Basicamente pegamos o estado atual, replicamos e rearranjamos (Figura 193).

Figura 193: Triangulo de Sierpinski.

7.9.1 Redes sem Escala

As redes sem escala também podem ser vistas como um processo iterativo. Esse processo inicia com um único nó, ou uma única aresta e, para cada novo nó inserido escolhe-se, seguindo um modelo, quais nós ele se conectará. Percebe-se que a rede cresce dinamicamente tendo como base o modelo das escolhas de ligações de cada novo nó. Isso leva a um outro tema interessante a ser estudado em redes reais: como elas crescem?(Figura 194).

Basicamente existem dois pontos importantes no crescimento de uma rede sem escala:

- Taxa de crescimento: em que taxa novos nós e arestas surgem?
- Ligação preferencial: com quais nós os novos nós irão se conectar?

Cada novo nó n escolhe m outros nós para se conectar criando m novas arestas.

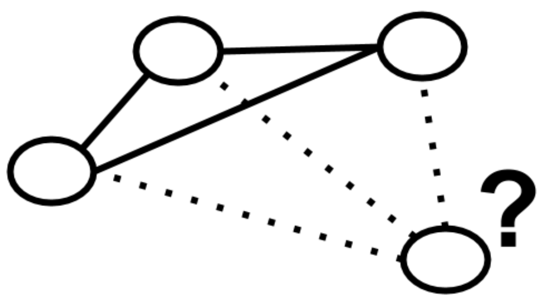


Figura 194: Rede Sem Escala, como crecem?

Redes reais têm certas propriedades observáveis em seu crescimento:

- em uma rede social, novos nós preferem se conectar aos nós mais populares;
- em certas redes os nós mais antigos param de receber novas ligações (atores se aposentam, pessoas morrem, informações se tornam defasadas);
- em uma cadeia alimentar certos animais são presas mais fáceis.

7.9.2 Ligação Preferencial

Um modelo de ligação preferencial que aparenta representar a maioria das redes sem escala reais é o "Processo Yule", também conhecido como "Efeito Matthew", "Ricos ficam mais ricos" e muitos outros nomes. Basicamente o modelo indica que um novo nó de uma rede tem preferência para se conectar com os nós melhores conectados, ou seja, de maior grau.

Tomando como exemplo a evolução das espécies e partindo do princípio que todos os seres vivos tem um ancestral comum, temos que inicialmente existe apenas uma única espécie, a qual pertence a um determinado gênero (Figura 195).

Figura 195: Exemplo de Ligação preferencial aplicado á relação Gênero-Especie.

.

A única espécie que temos sofre mutação e gera uma nova espécie. Como a mutação é pequena, a variação genética é pequena e, consequentemente, a nova espécie ainda pertence ao mesmo gênero (Figura 196).

Figura 196: GExemplo de Ligação preferencial aplicado á relação Gênero-Especie.

Com o passar do tempo e com sucessivas mutações, eventualmente uma espécie se distanciará do gênero de onde foi mutado, determinando um novo gênero (Figura 197).

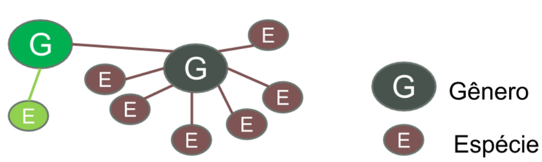


Figura 197: Exemplo de Ligação preferencial aplicado á relação Gênero-Especie.

Logo, a especiação se torna mais provável nos gêneros com maior número de espécies, pois a taxa de mutação é constante (Figura 198).

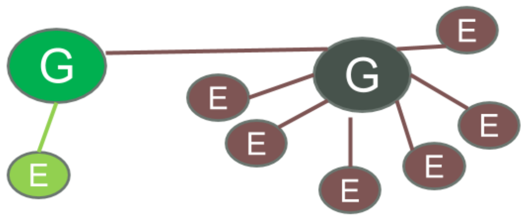


Figura 198: Exemplo de Ligação preferencial aplicado á relação Gênero-Especie.

Esse mesmo efeito pode ser observado na economia: a concentração de riqueza. Quem é rico, tem maiores chances de ficar mais rico. É observável que, mesmo em países desenvolvidos, cerca de 90% da riqueza está concentrada em 5% da população. Se analisarmos um sistema econômico em seu início, mesmo que toda população ganhe exatamente o mesmo salário, alguns tem predisposição a gastar menos do que os outros, levando a concentração de riqueza nesses ao longo do tempo (Figura 199).

Figura 199: "Ricos ficam mais ricos", https://www.flickr.com/photos/ennor/4973546393/

.

Analogamente, o crescimento populacional nas cidades sofre o mesmo efeito. Uma cidade com melhores empregos atraem mais pessoas, o crescimento no número de sua população causa surgimento de mais empregos, atraindo mais pessoas e sucessivamente...

7.10 Modelo Barabási-Albert

O modelo Barabsi-Albert cria redes sem escala utilizando o modelo de ligação preferencial, os passos são:

- Inicialização: crie uma rede com alguns nós.
- Crescimento: insira outros nós a cada instante de tempo.
- Ligação Preferencial: Para cada novo nó, crie algumas arestas com probabilidade:

$$P(k_i) = \frac{k_i}{\sum_{j=1}^{n} k_j}$$

Apresentamos um exemplo do modelo, vamos iniciar com um conjunto de 3 nós (Figura 200). Em cada passo, um novo nó é inserido e 2 arestas são desenhadas. Inicialmente com probabilidade uniforme (Figura 201). Esse novo nó tem probabilidade 2/10 de ser escolhido no próximo passo (Figura 202). Cada novo nó inserido as probabilidades se alteram (Figura 203). Os nós com maior grau, começam a captar cada vez mais ligações (Figura 204). E as probabilidades cada vez mais reflete a concentração de riquezas (Figura 205). E as probabilidades cada vez mais reflete a concentração de riquezas (Figura 206) (Figura 207). A distância média é igual a $1,67$ e o coeficiente de agrupamento médio é igual a $0,724$ (Figura 208). Vamos ver a distribuição dos graus: Verificamos que ele tende a seguir uma lei de potência (Figura 209). Vamos ver a distribuição dos coef. de agrupamento: Verificamos que ele também segue uma lei de potência (Figura 210).

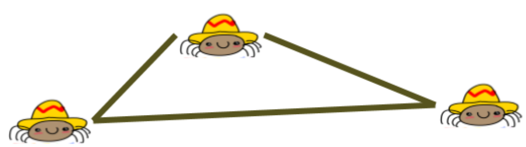


Figura 200: Exemplo modelo Barabási-Albert

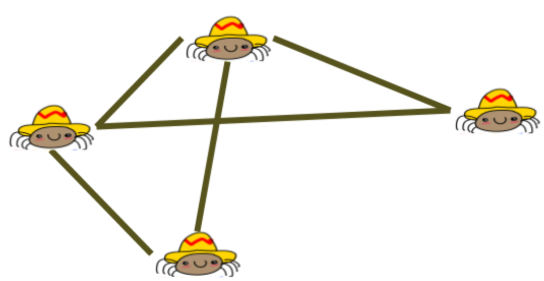


Figura 201: Exemplo modelo Barabási-Albert.

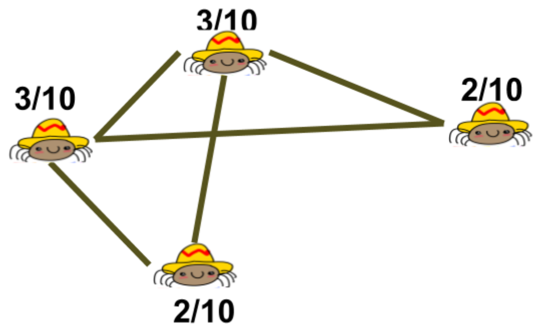


Figura 202: Exemplo modelo Barabási-Albert.

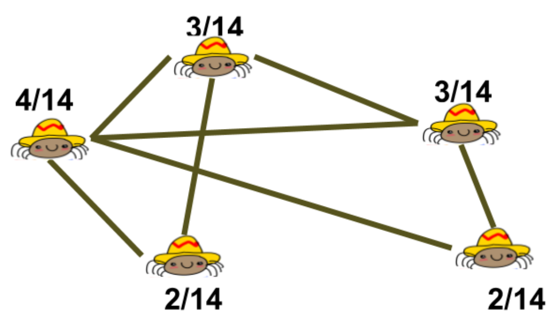


Figura 203: Exemplo modelo Barabási-Albert.

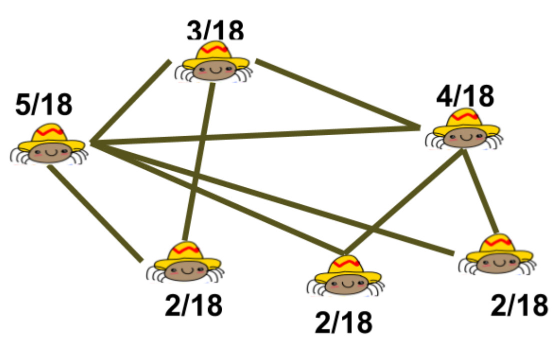


Figura 204: Exemplo modelo Barabási-Albert.

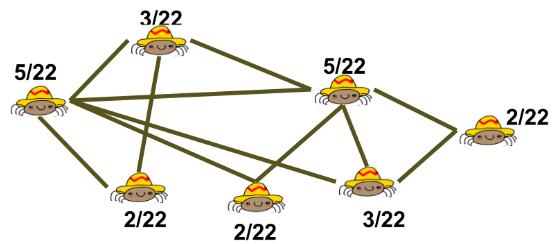


Figura 205: Exemplo modelo Barabási-Albert.

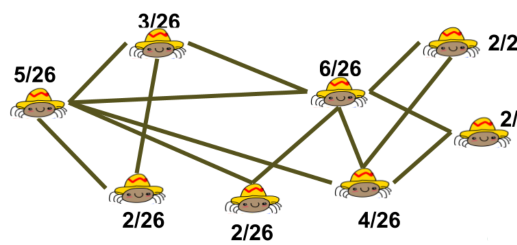


Figura 206: Exemplo modelo Barabási-Albert.

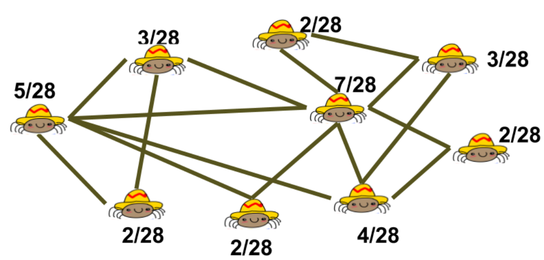


Figura 207: Exemplo modelo Barabási-Albert.

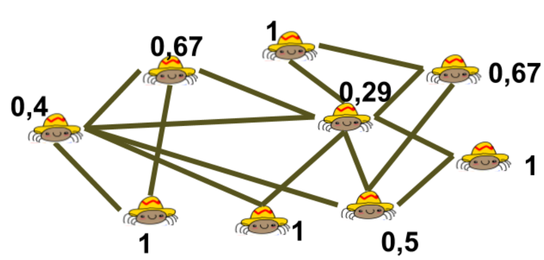


Figura 208: Exemplo modelo Barabási-Albert.

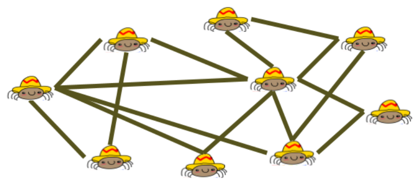


Figura 209: Exemplo modelo Barabási-Albert.

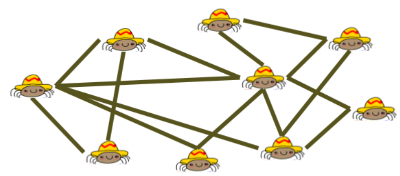


Figura 210: Exemplo modelo Barabási-Albert.

Esse modelo gera uma rede sem escala com fator de potência $\alpha = 3$. O diâmetro da rede cresce de forma logarítmica em função do tamanho da rede, ou seja, ela possui um crescimento lento. O coeficiente de agrupamento também segue uma lei de potência com $\alpha = -3/4$.

$$P(k) \approx k^{-3}$$

$$D \approx \frac{\ln(n)}{\ln(\ln(n))} P(c) \approx c^{-\frac{3}{4}}$$

7.10.1 Alterações no modelo Barabási-Albert

Algumas modificações foram propostas para tornar o modelo mais próximo das redes reais:

- Criação de função de afinidade, indicando que um determinado nó é mais propenso a receber ligações;
- Crescimento não apenas dos nós, mas também das arestas
- Envelhecimento dos nós.

7.10.2 Função de Afinidade

Se na rede estudada, cada nó tem uma importância bem definida, é possível usar essa mensuração para influenciar na probabilidade de novas conexões:

- Uma estação de metrô com bastante demanda;

- Uma pessoa muito famosa;
- Um animal carnívoro que se alimenta de muitos outros.

Com isso temos a probabilidade de inserir uma nova conexão proporcional a:

$$P(k_i, \eta_i) = \frac{k_i \eta_i}{\sum_{j=1}^{n} k_j \eta_j}$$

8 Assortatividade e Comunidade

8.1 Assortatividade

Significado geral:

- Semelhança decorrente de uma mesma origem ou de um ancestral comum.
- Tendência de um pessoa se associar a outra com propriedades similares.

Significado (em Grafos): Tendência de um nó se conectar a outro com propriedades similares.

Uma propriedade comum em muitas redes complexas é seu crescimento seguir uma lei de potência. Isso ocorre pois certos nós com maiores recursos podem atrair mais recurso ainda: ricos ficam mais ricos, populares ficam mais populares. Esse fenômeno é conhecido como ligação preferencial e diz que ao surgir um novo nó na rede, este tem uma preferência em se conectar com outros nós com determinada propriedade. Ex.: se conectar com nós de maior grau. Duas situações podem ser observadas em redes reais:

- Conexão assortativa ou homofílica, os nós de maior grau preferem se conectar com nós similares ;
- Conexão disassortativa ou heterofílica, Os nós de maior grau preferem se conectar com nós diferentes.

Em redes reais pode-se pensar em:

- Pessoas em uma rede social se conectam a pessoas com mesma popularidade;
- Portais agregadores na Internet contém links para sites com poucos links (geradores de conteúdo).

A Assortatividade r mede justamente se uma rede tem preferência por determinadas conexões ou não. Esse valor de r varia de -1 a 1 e representa:

- $r > 0$: os nós tendem a se conectarem com outros nós de grau similar.
- $r < 0$: os nós tendem a se conectarem com outros nós de graus diferentes (grau alto com grau baixo).
- $r \sim 0$: os nós não tem preferência.

O coeficiente de Assortatividade de uma rede é calculada como a correlação de Pearson entre os graus dos pares de nós conectados entre si:

$$r = \frac{1}{\sigma_0 \sigma_d} \sum_{k_1, k_2} k_1 k_2 (P(k_1, k_2) - P(k_1) P(k_2))$$

$P(k_1, k_2)$ representa quantas vezes ocorre uma aresta em que o nó de origem tem $grau = k_1$ e o nó de destino tem $grau = k_2$ divido pelo número de arestas (ou $2 * |E|$ caso o grafo seja não direcionado). $P(k_1)$ representa quantas arestas tem nó de origem com grau igual a k_1 divido pelo número de arestas. $P(k_2)$ representa quantas arestas tem nó de destino com grau igual a k_2 divido pelo número de arestas:

$$P(k_1, k_2) = \frac{|\{(v,u)|(v,u) \in E, g(v) = k_1, g(u) = k_2\}|}{|E|}$$

$$P(k_1) = \frac{|\{(v,u)|(v,u) \in E, g(v) = k_1\}|}{|E|}$$

$$P(k_2) = \frac{|\{(v,u)|(v,u) \in E, g(u) = k_2\}|}{|E|}$$

São os desvios padrão da distribuição do grau de entrada e saída dos nós:

$$\sigma_o = \sqrt{\sum_{k_1} k_1^2 P(k_1) - \left(\sum_{k_1} k_1 P(k_1)\right)^2}$$

$$\sigma_d = \sqrt{\sum_{k_2} k_2^2 P(k_2) - \left(\sum_{k_2} k_2 P(k_2)\right)^2}$$

Vamos trabalhar com um exemplo, vamos calcular a assortatividade dessa rede (Figura 211). Os graus de entrada de cada nó são (Figura 212). Os graus de saída de cada nó são (Figura 213). Enumerando, temos nós com grau igual a $2, 3, 4, 5$. Vamos relacionar quantas arestas interligam o grau x ao grau y. Temos um total de 34 arestas ($2 * 17$, pois é não-direcionado) (Figura 214). Dividimos os valores da tabela por 34 para obtermos $P(k_1, k_2)$ (Figura 215). A soma das linhas resulta em $P(k_1)$ e a soma das colunas resulta em $P(k_2)$ (Figura 216). Com isso podemos calcular $R(k_1, k_2) = k_1 k_2 (P(k_1, k_2) - P(k_1)P(k_2))$ (Figura 217). Aplicando aos valores se tem $\sum_{k_1,k_2} R(k_1, k_2) = -0.54$ (Figura 218). O desvio-padrão deve ser calculado em relação aos graus e quantas vezes eles ocorrem (Figura 219,220, 221). Finalmente, calculamos:

$$r = \frac{1}{\sigma_0 \sigma_d} \sum_{k_1,k_2} = \frac{-0.54}{1.0034 \cdot 1.0034} = -0.5363$$

Esse valor é apenas aproximado, pois arredondamos os cálculos para duas casas decimais. O valor correto da assortatividade dessa rede é $-0,28$; que ainda indica uma disassortatividade, ou heterofilia.

8.1.1 Assortatividade em Redes Reais

A assortatividade também pode ser mensurada em relação a outros aspectos da rede, não apenas grau. Em um estudo verificou-se que a rede de conversas no twitter (respostas de um usuário para outro) tem assortatividade positiva em relação aos tweets “alegres” [4].

Em um artigo um pesquisador fez o seguinte experimento, gerando:

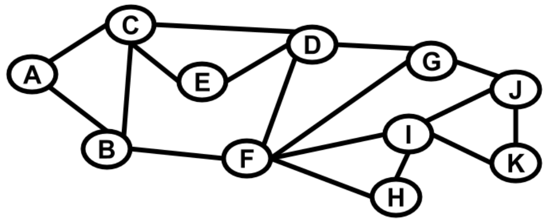


Figura 211: Exemplo Assortividade.

A	2
B	3
C	4
D	4
E	2
F	5
G	3
H	2
I	4
J	3
K	2

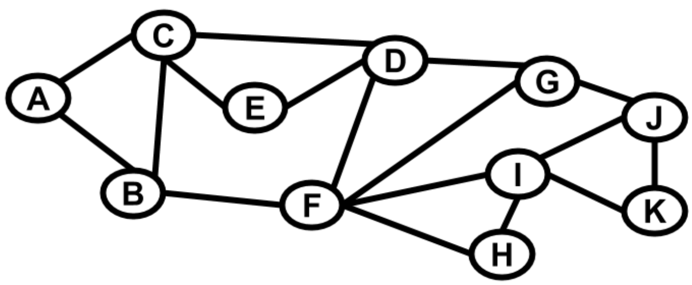


Figura 212: Exemplo Assortividade.

A	2
B	3
C	4
D	4
E	2
F	5
G	3
H	2
I	4
J	3
K	2

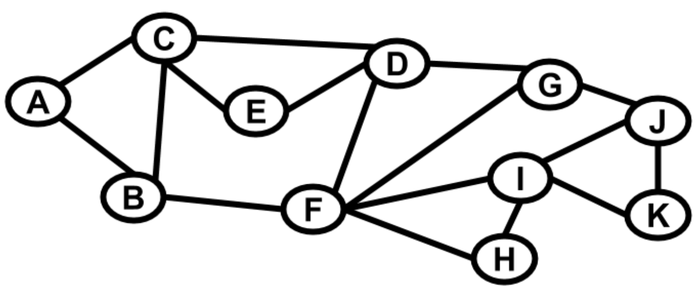


Figura 213: Exemplo Assortividade.

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

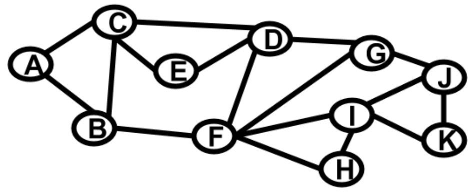


Figura 214: Exemplo Assortividade.

P(k1,k2)	2	3	4	5
2	0,00	0,06	0,15	0,03
3	0,06	0,06	0,09	0,06
4	0,15	0,09	0,06	0,06
5	0,03	0,06	0,06	0,00

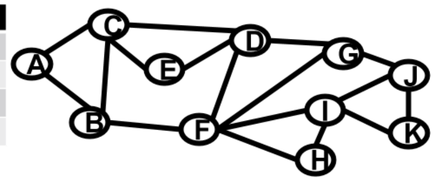


Figura 215: Exemplo Assortividade.

P(k1,k2)	2	3	4	5	P(k1)
2	0,00	0,06	0,15	0,03	0,24
3	0,06	0,06	0,09	0,06	0,27
4	0,15	0,09	0,06	0,06	0,36
5	0,03	0,06	0,06	0,00	0,15
P(k2)	0,24	0,27	0,36	0,15	1

Figura 216: Exemplo Assortividade.

P(k1,k2)	2	3	4	5	P(k1)
2	0,00	0,06	0,15	0,03	0,24
3	0,06	0,06	0,09	0,06	0,27
4	0,15	0,09	0,06	0,06	0,36
5	0,03	0,06	0,06	0,00	0,15
P(k2)	0,24	0,27	0,36	0,15	1

R(k1,k2)	2	3	4	5
2	-0,23	-0,03	0,51	-0,06
3	-0,03	-0,12	-0,09	0,29
4	0,51	-0,09	-1,11	0,12
5	-0,06	0,29	0,12	-0,56

Figura 217: Exemplo Assortividade.

R(k1,k2)	2	3	4	5
2	-0,23	-0,03	0,51	-0,06
3	-0,03	-0,12	-0,09	0,29
4	0,51	-0,09	-1,11	0,12
5	-0,06	0,29	0,12	-0,56

Figura 218: Exemplo Assortividade.

	2	3	4	5
2	0	2	5	1
3	2	2	3	2
4	5	3	2	2
5	1	2	2	0
SOMA:	8	9	12	5

MÉDIA =
2*8+3*9+4*12+5*5 / 34
= 3,41

Figura 219: Exemplo Assortividade.

	2	3	4	5
2	0	2	5	1
3	2	2	3	2
4	5	3	2	2
5	1	2	2	0
SOMA:	8	9	12	5

VARIÂNCIA =
$8*(2-3{,}41)^2+9*(3-3{,}41)^2+12*(4-3{,}41)^2+5*(5-3{,}41)^2$ / 34=
34,23 / 34 =
1,0067

DESVIO-PADRÃO =
√1,0067 = 1,0034

Figura 220: Exemplo Assortividade.

	2	3	4	5	SOMA:
2	0	2	5	1	8
3	2	2	3	2	9
4	5	3	2	2	12
5	1	2	2	0	5

VARIÂNCIA =
$8*(2-3,41)^2+9*(3-3,41)^2+12*(4-3,41)^2+5*(5-3,41)^2$ / 34=
34,23 / 34 =
1,0067

DESVIO-PADRÃO =
$\sqrt{1,0067}$ = 1,0034

Figura 221: Exemplo Assortividade.

Assortatividade em Redes Reais

Network	n	r	
Physics coauthorship (a)	52 909	0.363	SOCIAIS
Biology coauthorship (a)	1 520 251	0.127	
Mathematics coauthorship (b)	253 339	0.120	
Film actor collaborations (c)	449 913	0.208	
Company directors (d)	7 673	0.276	
Internet (e)	10 697	−0.189	Biológicas e Tecnológicas
World-Wide Web (f)	269 504	−0.065	
Protein interactions (g)	2 115	−0.156	
Neural network (h)	307	−0.163	
Marine food web (i)	134	−0.247	
Freshwater food web (j)	92	−0.276	

Figura 222: Tabela com exemplos de Assortividade em Redes Sociais, Biológicas e Tecnológicas.

- N redes aleatórias que apresentavam $r > 0$
- N redes aleatórias que apresentavam $r < 0$
- N redes aleatórias que não apresentavam $r = 0$

Adotou o seguinte procedimento:

- com probabilidade p (variando entre $0,1$ e $0,9$) marcava cada uma das arestas como escolhida (ou não);
- as arestas escolhidas representam arestas que vão difundir certa informação ou serão removidas por falha ou ataque;
- Verificou-se o ponto crítico de cada rede:
  - probabilidade em que a remoção das arestas selecionadas implica em quase desconexão da rede, ou
  - probabilidade em que a passagem de informação pelas arestas selecionadas implica que quase todos os nós receberam a informação

A relação entre a assortatividade e o ponto critico se mostra na Figura 223). As caraterísticas das redes assortativas são:

- Tem um ponto crítico mais baixo, é desconectada mais facilmente .
- Não precisa de muito esforço para transmitir e espalhar informações.
- É desconectada rapidamente em ataques aleatórios.

As caraterísticas das redes disassortativas são:

- Tem um ponto crítico mais alto, demora mais para desconectar.
- Tem um custo maior para a transmissão de informação.
- Resistentes a ataques aleatórios.

Verificou-se também a velocidade do espalhamento da informação em função da assortatividade (Figura 224). Nas Redes assortativas A velocidade de transmissão é menor, portanto temos mais tempo para medidas preventivas. Nas Redes disassortativas A velocidade de transmissão é maior, em caso de espalhamento de informação será difícil conter a mesma.

As Redes Sociais são, em geral, assortativas!!! Logo elas são capazes de transmitir as doenças mais facilmente, são mais difíceis de imunizar (vacinar poucos nós para desconectar a transmissão), porém existe um tempo maior para o processo de imunização (Figura 225).

As Redes Tecnológicas são, em geral, disassortativas!!! Logo falhas no sistema dificilmente atinge boa parte da rede, são mais fáceis de identificar os pontos mais importantes para proteger mas, se um ataque direcionado ocorrer, desconecta a rede muito mais rapidamente (Figura 226).

Para saber mais, pode-se consultar os textos:

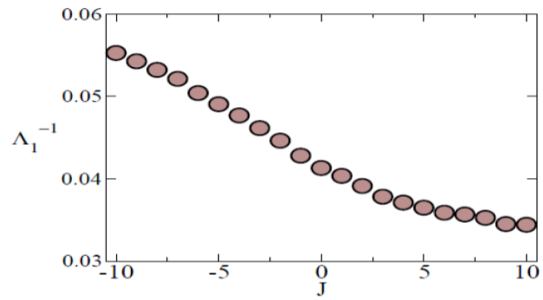


Figura 223: Assortatividade e Ponto Critico.

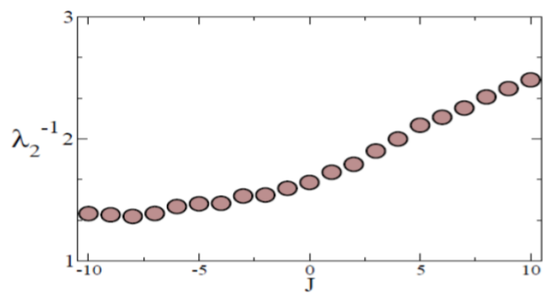


Figura 224: Velocidade do espalhamento da informação em função da assortatividade.

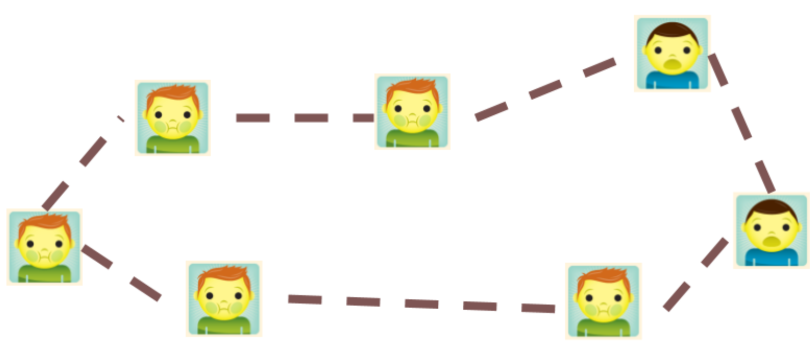


Figura 225: Assortatividade nas doenças.

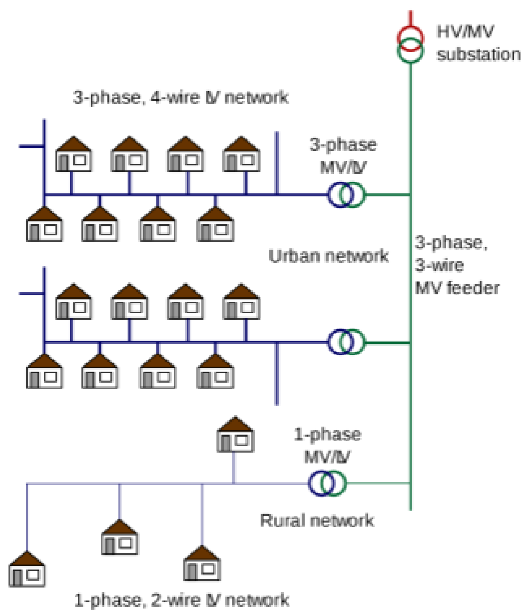


Figura 226: Assortatividade nas Tecnologia.

- Noh, J.D. Percolation transition in networks with degree-degree correlation. Physical Review E, v.76, n.2. 2007.
- Newman, M.E.J. Assortative mixing in networks. Physical Review Letters, v. 89, n. 20. 2002.
- G. D'Agostino, A. Scala, V. Zlatic, G. Caldarelli. Robustness and Assortativity for Diffusion-like Processes in Scale-free Networks, 2012

Apresentamos como é a Assortatividade na rede de Seguidores de Twitter: Assortatividade de grau = -0.21 , Assortatividade de local = 0.009 , Assortatividade de idioma = 0.18 (Figura 226).

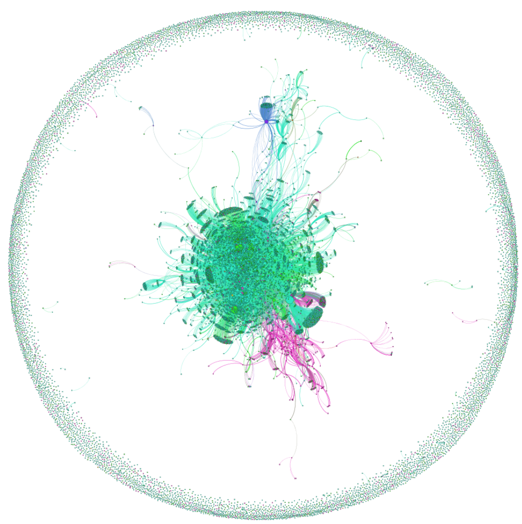


Figura 227: Assortatividade nas Redes de Seguidores em Twitter.

8.2 Comunidades

Como vimos anteriormente, em redes sociais é possível verificar grupos fechados de pessoas (pessoas que se conhecem mutuamente) e, em alguns nós, pessoas que interligam esses grupos. Isoladamente cada grupo não pode contribuir com muita informação nova, pois a informação dentro daquele grupo já é de conhecimento de todos que estão ali. Porém, ao interagirem com outros grupos, novos conhecimentos surgem e levam a uma interação complexa (Figura 226).

8.2.1 Particionamento de Redes

Estudar como funciona o processo de comunicação e difusão de informação na rede. Particionamento é o procedimento de divisão da rede em várias sub-redes

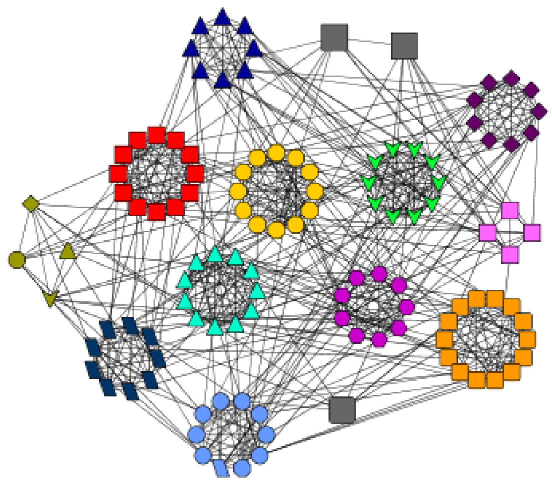


Figura 228: Grupos fechados nas redes conformam comunidades.

de tal forma a identificar as comunidades existentes. Exemplos de utilidades para o particionamento:

- Redes sociais para descobrir comunidades com interesses semelhantes;
- Redes biológicas para descobrir funcionalidades de genes para combater doenças.

Estudar como funciona o processo de comunicação e difusão de informação na rede. É importante conhecer como ocorre o processo de comunicação em uma rede pois podemos:

- Melhorar a forma como a comunicação é feita;
- Controlar a informação e alterá-la em nosso favor;
- Proteger ou preservar o processo de comunicação.

Uma serralheria americana estava com problemas para convencer os funcionários a adotar novas práticas de trabalho. Em particular, os hispânicos rejeitaram em massa o novo método. Ao analisar o nó mais central seguindo critérios de betweenness e identificar as partições da rede, percebeu-se que o problema se resolveria apenas convencendo o Juan. (Figura 229).

Em algumas redes reais as partições de nós são facilmente percebidas visualmente (Figura 230). Mas muitas vezes não é possível ver onde começa e onde termina uma determinada partição (Figura 231). Por isso é necessário utilizar algum método para determinar tais partições e indicar qual partição é melhor.

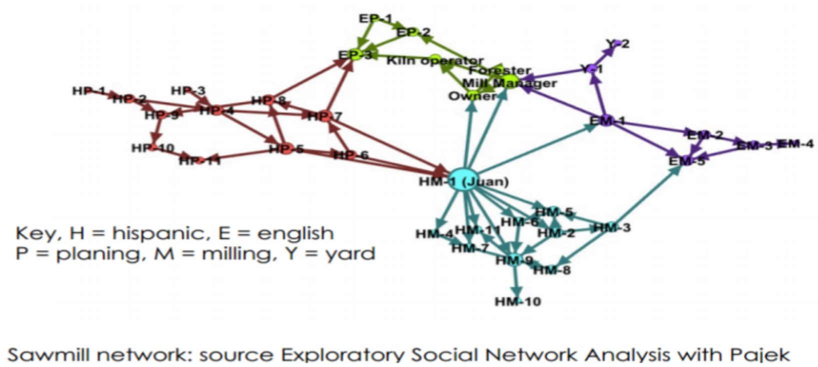


Figura 229: Rede de uma serraria, que tem o Juan como nó central.

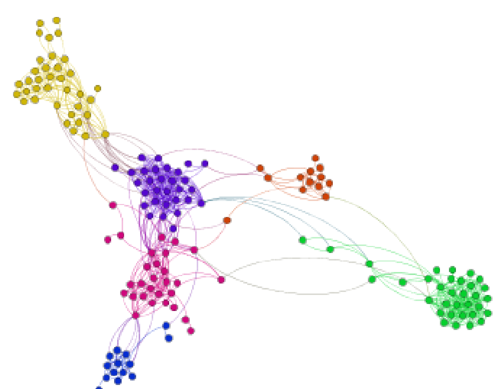


Figura 230: Rede com poucos nós e com partições facilmente identificáveis.

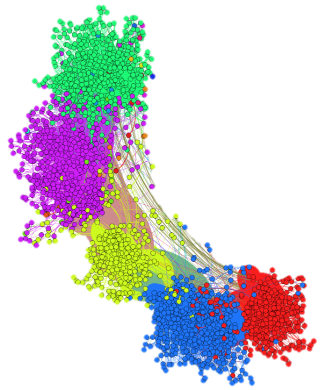


Figura 231: Rede com muitos nós e difícil identificação de particionamento.

8.2.2 Métodos de Particionamento

Antes de criar métodos para particionar a rede precisamos definir o que define uma partição:

- Ligações mútuas: todos conhecem todos em um grupo.
- Ligações frequentes: todos conhecem a maioria do grupo.
- Proximidade: todos estão separados por no máximo n nós.

Um particionamento ideal é aquele em que todas as partições formem uma REDE COMPLETA, ou seja, todos conhecem todos. O objetivo é cortar um número mínimo de arestas de tal forma que obtemos n COMPONENTES CONEXOS que formem REDES COMPLETAS (Figura 232).

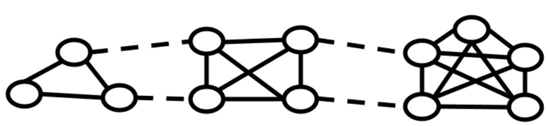


Figura 232: Uma Rede dividida em componentes conexos.

O CLIQUE de um grafo é um grupo de nós que formam uma rede completa, ou seja, com todos os nós interligados entre si (Figura 233).

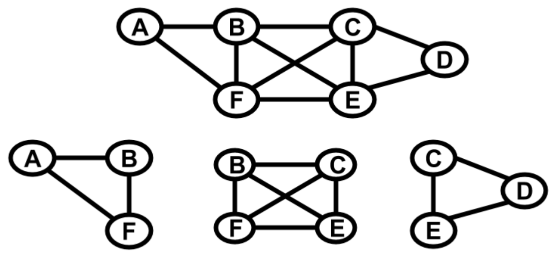


Figura 233: Clique de um Grafo.

O CLIQUE MÁXIMO de uma rede é um clique que não é um subconjunto de nenhum outro clique. Em outras palavras é um clique em que não existe nenhum outro nó não pertencente a ele que, se adicionado, forme também um clique (Figura 234).

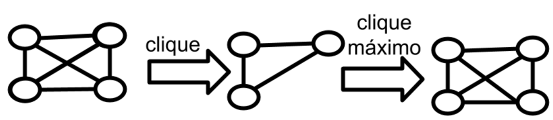


Figura 234: Clique Máximo.

Achar os cliques de um grafo pode ser pensado como enumerar todos os subgrafos e todas as possibilidades de arestas nesse subgrafo.

$$\frac{n(n-1)\dots(n-k+1)}{1\cdot 2\dots\cdot k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

$$\sum_{1\leq k\leq n}\binom{n}{k} = 2^n - 1$$

Se temos n nós, temos n subgrafos de tamanho 1, $n(n-1)/2$ subgrafos de tamanho 2, $n(n-1)(n-2)/6$ subgrafos de tamanho 3,... e assim por diante.

- n = 1, 1 possibilidade diferente;
- n = 2, 3 possibilidades;
- n = 10, 1.023 possibilidades;
- n = 100, 1.267.650.600.228.229.401.496.703.205.375 possibilidades diferentes.

Na prática é impossível determinar grupos tão bem definidos que formam cliques. Logo, o objetivo do particionamento torna-se encontrar os conjuntos de nós que mais se aproximam de cliques . Por exemplo, grupos de nós que tenham ligações com pelo menos k nós do mesmo grupo. Procedimento conhecido como K-CORE. Qual o valor de k para as partições da figura acima? (Figura 235).

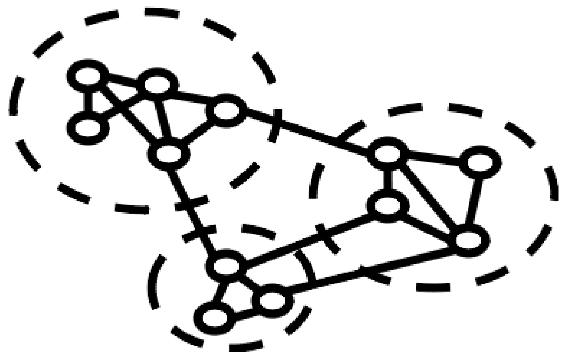


Figura 235: Particionamento por Cliques.

8.2.3 Girvan-Newman

Em algumas redes as estruturas formadas por cliques não indicam a sua funcionalidade. E se as comunidades não são formadas por cliques, mas por alguma particularidade da estrutura? Vimos anteriormente a centralidade betweenness de um nó, que mede quantos caminhos mais curtos passam por aquele nó. Essa medida indica nós que servem como pontes entre dois conjuntos distintos de nós. Se modificarmos essa medida para calcularmos a centralidade das ARESTAS, podemos encontrar as arestas que dividem dois grupos de nós (Figura 236).

O método de Girvan-Newman é baseado na centralidade betweenness DE ARESTA. Esse método é chamado DIVISIVO pois, a cada passo divide a rede em duas formando uma hierarquia de partições (Figura 237, Figura 238) Mais informação pode se encontrar no link `http://www1.maths.leeds.ac.uk/statistics/workshop/lasr2006/proceedings/pinney-talk.pdf`.

8.3 Particionamento Espectral

Algumas propriedades interessantes de uma rede aparecem ao calcularmos os autovalores e autovetores de uma matriz de adjacência modificada. Vamos utilizar a matriz Laplaciana que é formada pela matriz de graus dos nós do grafo subtraída da matriz de adjacência:

$$L = G - A$$

Agora vamos fazer alguns experimentos em uma rede simples (Figura 241). Reparem que o número de AUTOVALORES iguais a ZERO representa o número de COMPONENTES CONEXOS da rede (Figura 242). E como se comportam

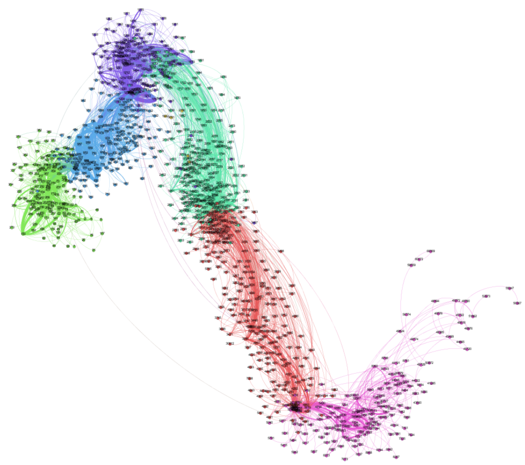


Figura 236: Rede particionada por cliques, mas se as comunidades não fossem formadas por cliques?

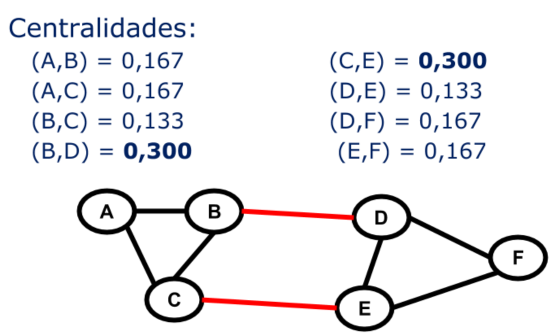


Figura 237: Exemplo Girvan Newman

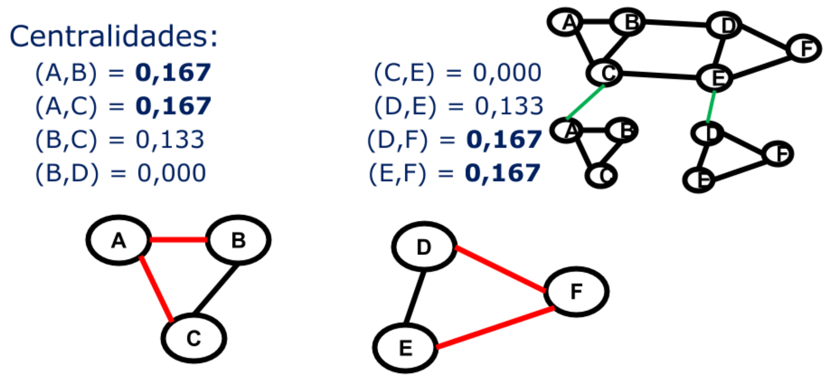


Figura 238: Exemplo Girvan Newman

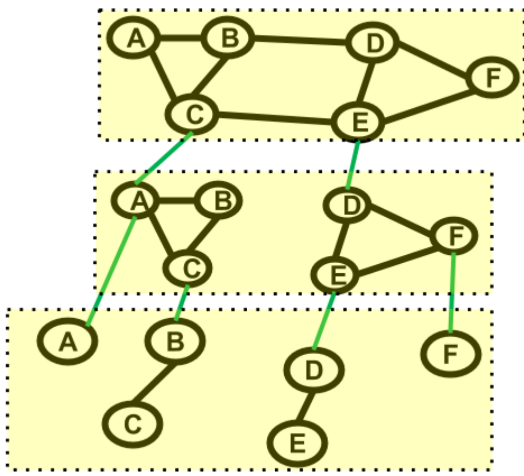
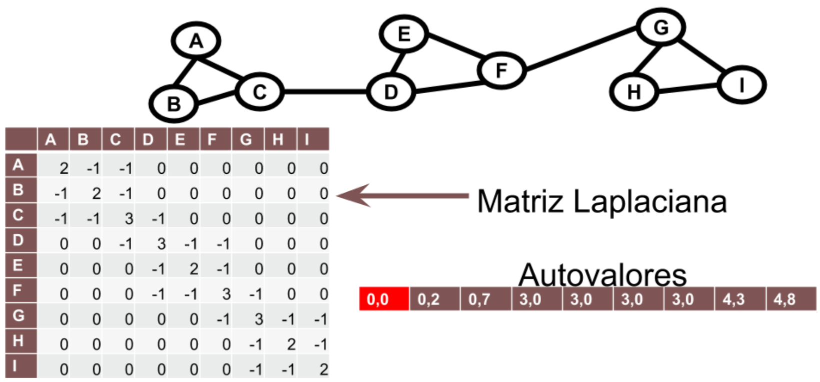


Figura 239: Exemplo Girvan Newman

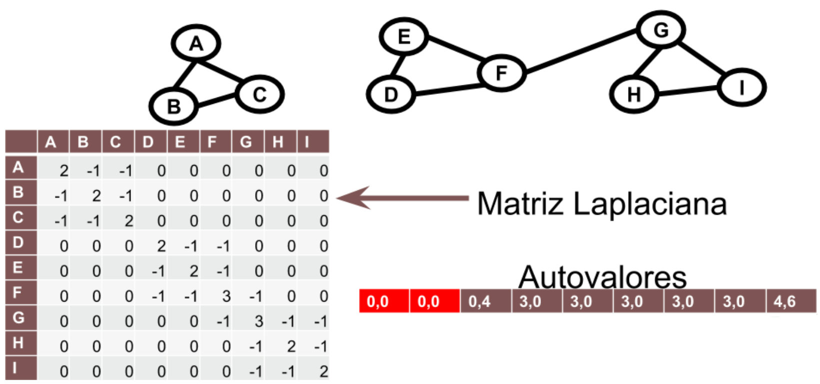
Figura 240: Redes que tem uma pessoa, `http://www.touchgraph.net/`.

os autovetores correspondentes? (Figura 243). Dá para perceber claramente que os autovetores identificam os nós de cada componente conexo! E no primeiro caso, em que temos uma rede totalmente conectada, como se comportam o segundo e terceiro autovetor? (Figura 244). O segundo autovetor mostra claramente a separação entre dois grupos distintos.No terceiro autovetor podemos perceber a formação de 3 grupos, mas isso porque os nós estão ordenados corretamente! (Figura 245).



	A	B	C	D	E	F	G	H	I
A	2	-1	-1	0	0	0	0	0	0
B	-1	2	-1	0	0	0	0	0	0
C	-1	-1	3	-1	0	0	0	0	0
D	0	0	-1	3	-1	-1	0	0	0
E	0	0	0	-1	2	-1	0	0	0
F	0	0	0	-1	-1	3	-1	0	0
G	0	0	0	0	0	-1	3	-1	-1
H	0	0	0	0	0	0	-1	2	-1
I	0	0	0	0	0	0	-1	-1	2

Figura 241: Exemplo Particionamento Espectral.



	A	B	C	D	E	F	G	H	I
A	2	-1	-1	0	0	0	0	0	0
B	-1	2	-1	0	0	0	0	0	0
C	-1	-1	2	0	0	0	0	0	0
D	0	0	0	2	-1	-1	0	0	0
E	0	0	0	-1	2	-1	0	0	0
F	0	0	0	-1	-1	3	-1	0	0
G	0	0	0	0	0	-1	3	-1	-1
H	0	0	0	0	0	0	-1	2	-1
I	0	0	0	0	0	0	-1	-1	2

Figura 242: Exemplo Particionamento Espectral.

O segundo autovetor, conhecido como vetor Fiedler, tem as seguintes propriedades:

- Tem autovalor correspondente maior que ZERO se e somente se a rede for conectada e;

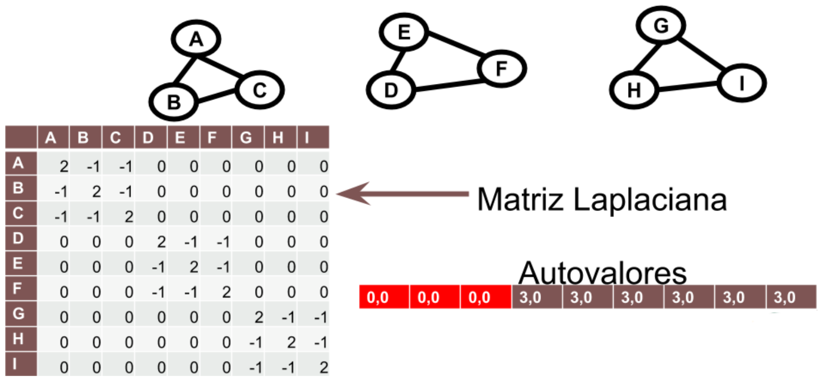


Figura 243: Exemplo Particionamento Espectral.

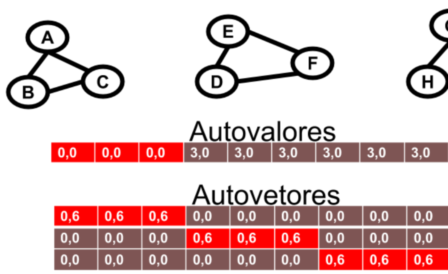


Figura 244: Exemplo Particionamento Espectral.

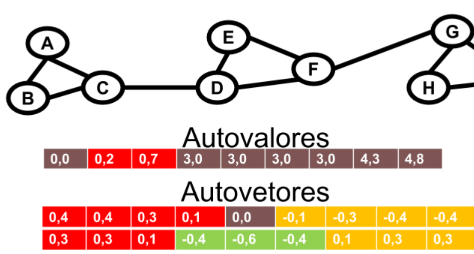


Figura 245: Exemplo Particionamento Espectral.

- Particiona a rede em duas redes distintas repartindo em torno de um nó que divide a rede de forma simétrica.

O autovalor correspondente ao vetor Fiedler é conhecido como conectividade algébrica e indica o quão conectada é a rede. Esse valor varia entre:

$$\frac{4}{nD} \leq \sigma \leq 1$$

onde D é o diâmetro da rede e n é o número de nós. As partições são definidas como:

- Nós em que os valores correspondentes do autovetor são maiores que 0 representam o GRUPO 1;
- Nós em que os valores correspondentes do autovetor são menores que 0 representam o GRUPO 2.

Os valores dos elementos do autovetor indica a conectividade de cada nó de acordo com sua vizinhança (Figura 246). O segundo autovetor indica a conectividade se a rede fosse particionada em dois, o terceiro se ela fosse particionada em três, e assim por diante (Figura 247). Eles tem relação com a frequência de onda em um fio em vibração (Figura 248).

G

2					
	3				
		3			
			3		
				3	
					2

A

0	1	1	0	0	0
1	0	1	1	0	0
1	1	0	0	1	0
0	1	0	0	1	1
0	0	1	1	0	1
0	0	0	1	1	0

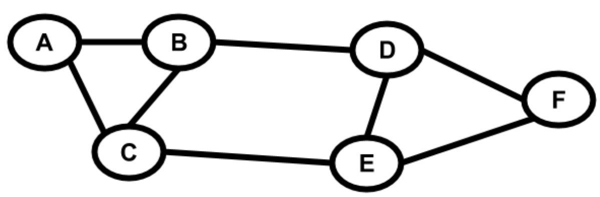


Figura 246: Exemplo Particionamento Espectral

Fontes recomendadas:

- A. Pothen, H. Simon, K.-P. Liou, "Partitioning sparse matrices with eigenvectors of graphs", SIAM J. Mat. Anal. Appl. 11:430-452 (1990)
- M. Fiedler, "Algebraic Connectivity of Graphs", Czech. Math. J., 23:298-305 (1973)
- M. Fiedler, Czech. Math. J., 25:619-637 (1975)

L

2	-1	-1	0	0	0
-1	3	-1	-1	0	0
-1	-1	3	0	-1	0
0	-1	0	3	-1	-1
0	0	-1	-1	3	-1
0	0	0	-1	-1	2

Autovetor Fiedler

0,5773	Grupo 1
0,2887	Grupo 1
0,2887	Grupo 1
-0,2887	Grupo 2
-0,2887	Grupo 2
-0,5773	Grupo 2

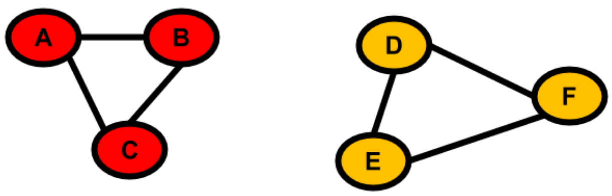


Figura 247: Exemplo Particionamento Espectral

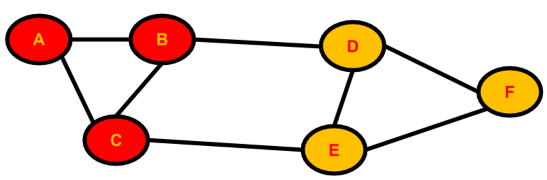


Figura 248: Exemplo Particionamento Espectral

8.4 Modularidade

Muitos problemas podem ser resolvidos matematicamente se existir uma função-objetivo que mede a qualidade de determinada solução. Para medir a qualidade de particionamento de redes uma das medidas mais populares é a MODULARIDADE (Figura 249).

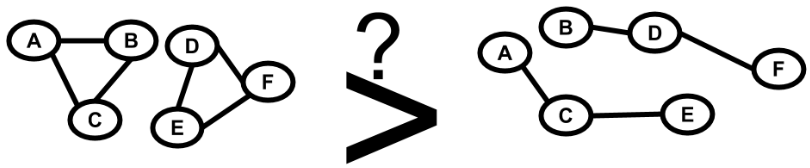


Figura 249: Exemplo Qualidade no particionamento.

Dada uma divisão do grafo em g grupos e uma matriz E de tamanho $g * g$, onde $e_{i,j}$ representa uma fração de arestas na rede que liga o grupo i ao grupo j, definimos a modularidade como:

$$Q = \sum_i \left(e_{i,i} - \left(\sum_j e_{i,j} \right)^2 \right)$$

Vamos gerar 2 grupos aleatórios. Ex.: A, C, E e B, D, F e calcular a modularidade. As arestas vermelhas são as arestas que seriam removidas para formar tal partição (Figura 250). Existem 4 arestas ao todo ligando cada grupo com ele mesmo em um total de 16 arestas, contando ida e volta (Figura 251). Existem 4 arestas ao todo ligando cada grupo com o outro mesmo em um total de 16 arestas, contando ida e volta (Figura 252). Fazendo os cálculos temos (Figura 253). Se trocarmos o nó E pelo nó B, ficaremos com os grupos A, B, C e D, E, F (Figura 254). Existem 6 arestas ao todo ligando cada grupo com ele mesmo em um total de 16 arestas, contando ida e volta (Figura 255). Existem 2 arestas ao todo ligando cada grupo com o outro em um total de 16 arestas, contando ida e volta (Figura 256). Fazendo os cálculos temos (Figura 257). Ao final calculamos as Modularidades (Figura 258).

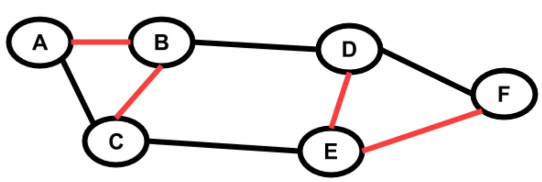


Figura 250: Exemplo de Calculo da Modularidade.

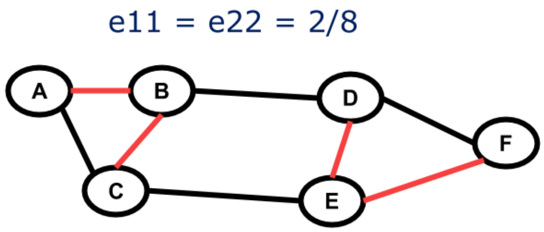


Figura 251: Exemplo de Calculo da Modularidade.

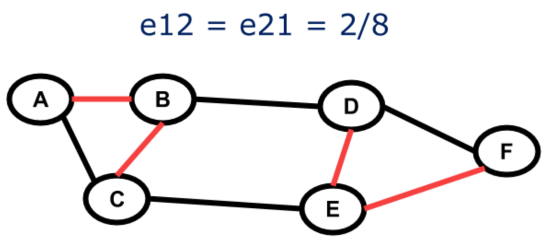


Figura 252: Exemplo de Calculo da Modularidade.

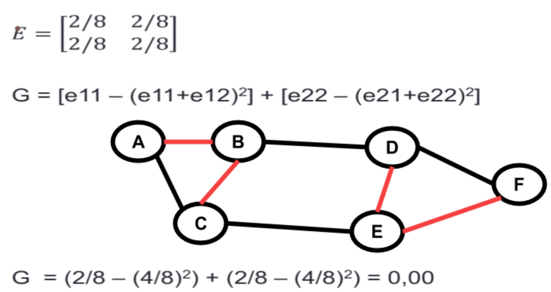


Figura 253: Exemplo de Calculo da Modularidade.

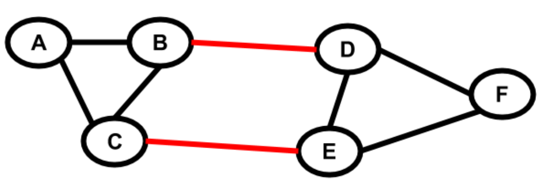


Figura 254: Exemplo de Calculo da Modularidade.

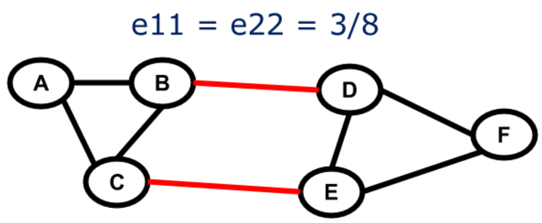


Figura 255: Exemplo de Calculo da Modularidade.

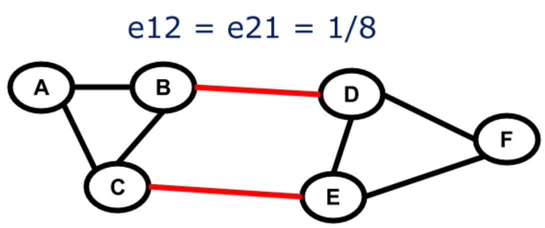


Figura 256: Exemplo de Calculo da Modularidade.

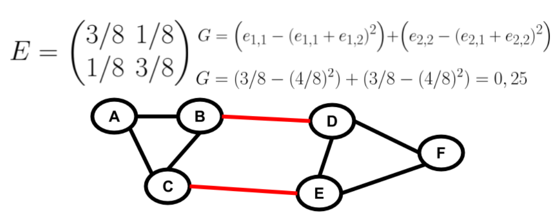


Figura 257: Exemplo de Calculo da Modularidade.

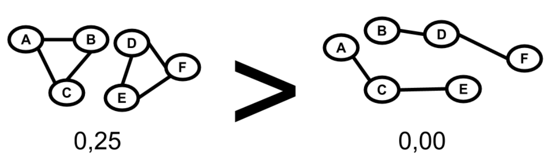


Figura 258: Exemplo de Calculo da Modularidade.

Um algoritmo simples para encontrar uma partição com alta modularidade:

- Inicie com uma partição inicial e calcule G;
- Calcule G para todas as possíveis trocas de nós entre grupos (inclusive simplesmente mover um nó de um grupo ao outro);
- Se alguma troca for favorável, efetue a troca;
- Repita até não existir nenhuma outra troca favorável

A modularidade permite uso de algoritmos mais simples conhecidos como métodos heurísticos. Eles são capazes de lidar com grandes problemas utilizando de forma eficiente o processamento computacional MAS não garante encontrar a melhor solução possível. Por exemplo Amazon: "quem comprou X também comprou Y".

9 Redes de Computadores

As redes de computadores surgiram da necessidade de distribuir o processamento de tal forma a superar as limitações dos computadores (Figura 259).

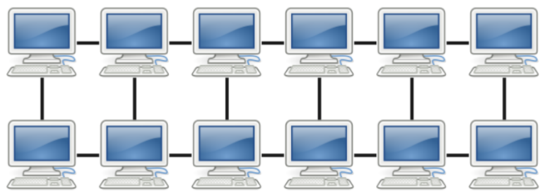


Figura 259: Rede de Computadores `https://commons.wikimedia.org/wiki/File:Gnome-computer.svg`

A primeira rede que se tem conhecimento é conhecida como SAGE (Semi-Automatic Ground Environment) e era utilizada para processar dados de diversos radares de forma a produzir uma imagem única correspondente a uma grande área. Foi utilizada pelo exército americano desde os anos 50 até meados dos anos 80 (Figura 260).

Figura 260: Rede SAGE `https://en.wikipedia.org/wiki/Semi-Automatic_Ground_Environment`.

O projeto ARPANET foi criado em 1966 com o objetivo de permitir o controle remoto de computadores a longa distância. Inicialmente o projeto foi financiado pela ARPA (agora conhecida como DARPA, agência financiadora de pesquisa do exército americano) e posteriormente foi financiado pela National Science Foundation. Esse projeto pode ser considerado o precursor do que conhecemos hoje como internet (Figura 261).

Figura 261: ARPANET `https://en.wikipedia.org/wiki/ARPANET`.

Com esse projeto foram criados protocolos de comunicação para permitir envio e recebimento de dados para compartilhamento de arquivos, troca de mensagens, e-mails, etc. Um protocolo é um conjunto de regras que permite a comunicação entre duas ou mais entidades. Essas regras costumam envolver sintaxe, semântica, sincronização e recuperação de erros (como o CRC visto em Natureza da Informação) (Figura 262).

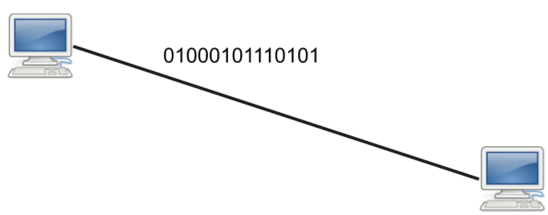


Figura 262: Comunicação entre dois computadores.

Nessa rede os nós são os componentes como computadores, roteadores e qualquer aparelho que serve para interconectar os computadores ou os aparelhos que utilizam essa rede, as arestas representam os componentes físicos por onde as informações são transmitidas, a conexão de uma rede pode ser ponto a ponto em que um computador está ligado diretamente a outro, e portanto a informação vai de um ponto A a um ponto B, ou de múltiplo acesso, em que a informação é transmitida por um canal observável por todos os computadores da rede. Nesse método são utilizadas medidas de proteção para que apenas o computador de destino possa compreender a mensagem transmitida (Figura 263).

A organização da rede é feita por diversas camadas de abstração! Como exemplo, considere uma mensagem enviada do computador A para o computador B (Figura 264). A camada de aplicação define um protocolo que torna

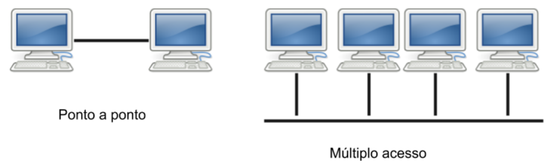


Figura 263: Na direita : Redes Ponto a ponto, na esquerda : Rede de Múltiplo acesso.

transparente o caminho que os dados deve passar para chegar ao destino. Ele simula o caso em que A e B estão conectados diretamente. Na camada de transporte se encarrega de dividir a informação sendo transmitida em pequenos pedaços, negociar a forma de transmissão e verificação e correção de erros. A camada de internet verifica qual o próximo nó que os pacotes devem ser enviados para que se aproxime do computador de destino, ele tem uma visão local do caminho mais curto através de tabelas de roteamento. Finalmente a camada de ligação se encarrega de efetivamente transportar os dados entre dois nós.

Imagine que o computador A quer receber um arquivo de uma página hospedada no computador B. Porém eles estão em países diferentes! (Figura 265). Geralmente esses computadores estão conectados a um roteador que define uma rede local (Figura 266). A primeira barreira é encontrar a saída para a internet!

Cada aparelho possui duas identificações: media access control address (MAC address), que é um endereço fixo identificando o aparelho com um número único (supostamente) de 48 bits. Esse número está associado diretamente ao hardware da rede, seja a placa de rede, placa de wifi, bluetooth, roteador. Em tese esse número é um identificador único para cada placa física de comunicação (Figura 267).

O outro endereço, um endereço lógico, é o Internet Protocol address (IP address) que identifica cada computador conectado a uma rede seguindo um padrão. Esse padrão permite identificar cada componente único da rede e sua localização. Ele é designado a sua máquina no momento em que voce conecta a internet e é retornado assim que você para de usar, ou seja, o número de IP utilizado hoje pode não ser o mesmo amanhã. Esse sistema permitia que esses números fossem representados por apenas 32bits porém, com a popularização da internet logo se tornou insuficiente, levando ao IPv6 que é representado por 128bits (Figura 268).

Além disso, alguns componentes da rede possuem um nome, legível para os usuários, que o identificam. Esses nomes são traduzidos pelo Domain Naming Server (DNS) que traduzem para o endereço IP correspondente! (Figura 269).

Em resumo, os passos são: quem tem acesso a internet? O computador identifica o roteador e envia a requisição para o seu endereço físico! (Figura 270).

O roteador confirma e pergunta qual o endereço (Figura 271).

Com a resposta, ele pergunta para o servidor de DNS mais próximo qual o endereço desse nome (Figura 272).

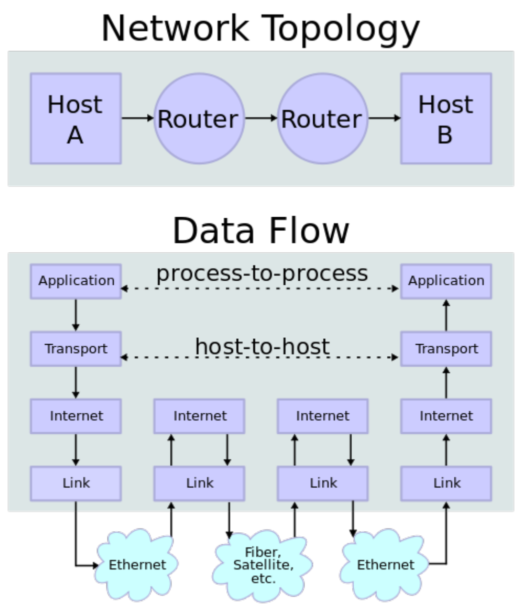


Figura 264: Camadas de Abstração em uma Rede https://en.wikipedia.org/wiki/Internet_protocol_suite#Comparison_of_TCP/IP_and_OSI_layering.

Ele então verifica se algum computador local possui esse endereço....em caso afirmativo envia a requisição diretamente (Figura 273).

em caso negativo, ele consulta a tabela de roteamento para enviar ao computador mais próximo acessível diretamente por ele, esse processo é repetido por cada ponto até que o destino é alcançado (Figura 274).

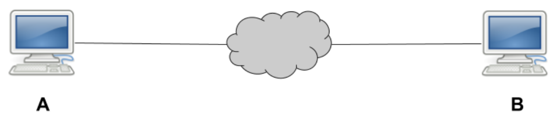


Figura 265: Exemplo de transmissão nas redes.

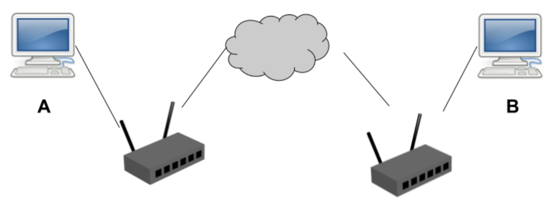


Figura 266: Exemplo de transmissão nas redes.

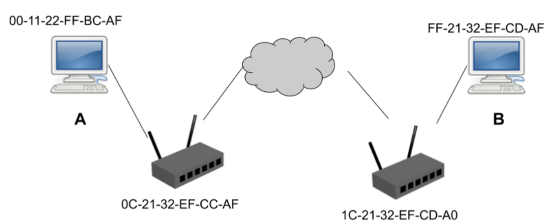


Figura 267: Exemplo de transmissão nas redes.

O endereçamento IP permite a divisão das redes em subredes. Relembrando , o endereço IP é composto por 32 bits divididos em grupos de 8 bits, denominados octetos, e representados por números decimais separados por pontos (Figura 275).Atualmente, três faixas de valores são reservadas para redes privadas (a rede local definida pelo seu roteador):

- 10.0.0.0 - 10.255.255.255 com 16.777.216 endereços possíveis
- 172.16.0.0 – 172.31.255.255 com 1.048.576 endereços possíveis
- 192.168.0.0 – 192.168.255.255 com 65.536 endereços possíveis

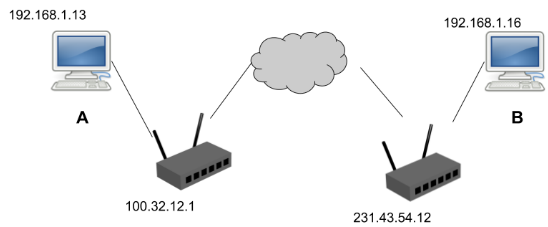


Figura 268: Exemplo de transmissão nas redes.

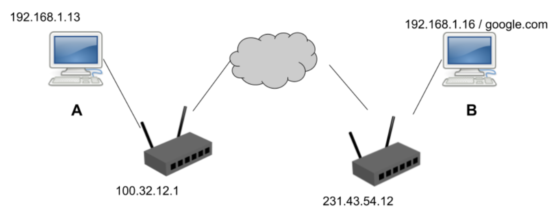


Figura 269: Exemplo de transmissão nas redes.

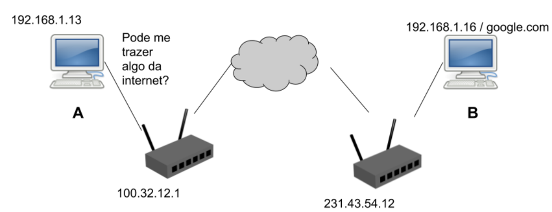


Figura 270: Exemplo de transmissão nas redes.

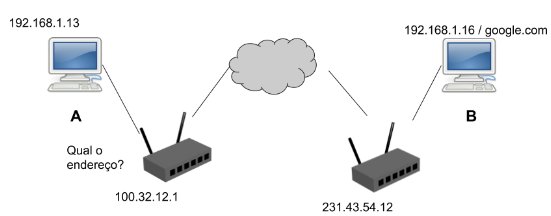


Figura 271: Exemplo de transmissão nas redes.

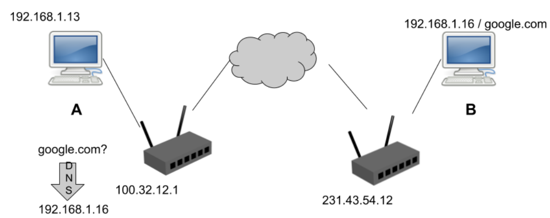


Figura 272: Exemplo de transmissão nas redes.

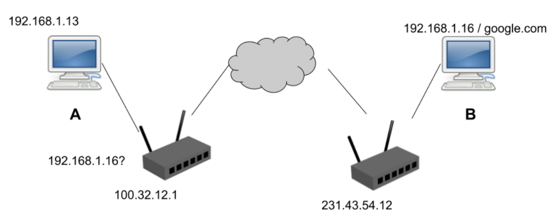


Figura 273: Exemplo de transmissão nas redes.

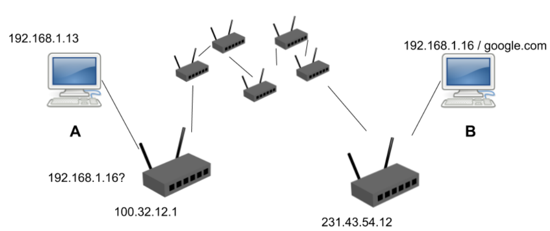


Figura 274: Exemplo de transmissão nas redes.

- 127.0.0.1 - localhost
- 255.255.255.255 - broadcast

Esses endereços não são utilizados como endereços públicos na internet e, portanto, podem ser utilizados para identificação de redes internas. Por exemplo, quando conectados na internet de sua casa os endereços provavelmente começam com 192. Dificilmente você precisará de mais do que 65 mil computadores em sua rede. Empresas maiores eventualmente utilizam os endereços iniciando por 172 ou 10.

Um outro endereço reservado é o 127.0.0.1 que é utilizado como um loopback, ele aponta para sua própria máquina. Ele é utilizado quando queremos executar (ou simular) um serviço de rede sem que ele seja visível para o público externo. Os endereços terminados em 255 são utilizados para transmissão a todos os computadores daquela sub-rede.

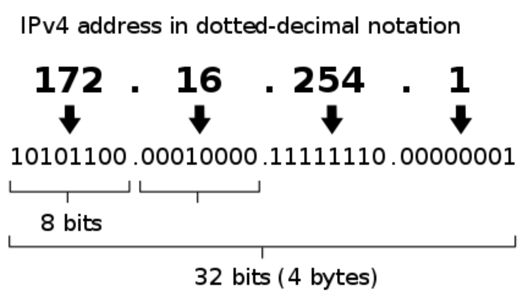


Figura 275: Endereçamento IP.

9.1 Roteamento

9.1.1 Redes Desconhecidas

Uma rede de computadores de larga escala (como a Internet) não tem uma estrutura determinada por algum órgão regulador, ou seja, ela é descentralizada.Nessa rede os nós representam os roteadores de uma rede local e as arestas as conexões físicas entre eles (Figura 276). Existem dois tipos de conexões: inter-redes e intra-redes. As conexões inter-redes são as arestas que interconectam roteadores de uma mesma rede local (Figura 277). As conexões intra-redes são as arestas que interconectam roteadores de redes distintas (Figura 278). Por causa dessa descentralização, as redes não possuem conhecimento de qual a melhor rota para atingir nós que estão localizados em outras redes. Mesmo descobrindo as rotas ótimas de alguma forma elas dificilmente permanecem a mesma pois roteadores podem quebrar e ligações podem congestionar (Figura 279).

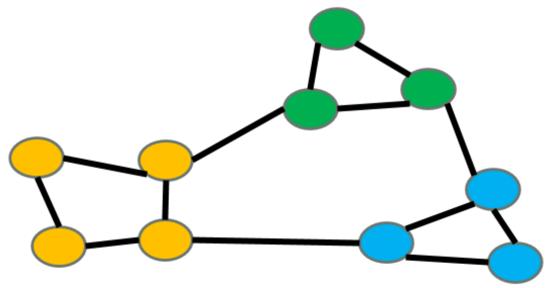


Figura 276: Exemplo Redes desconhecidas.

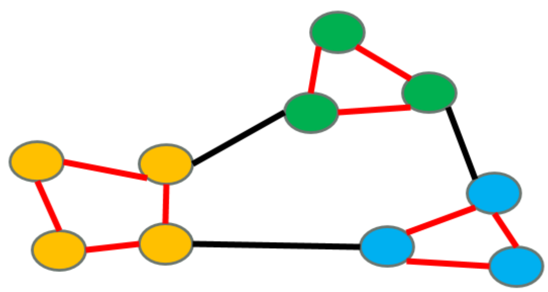


Figura 277: Exemplo Redes desconhecidas.

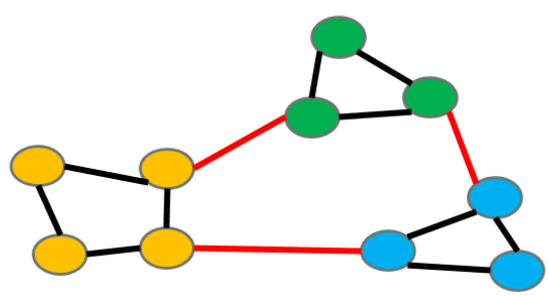


Figura 278: Exemplo Redes desconhecidas.

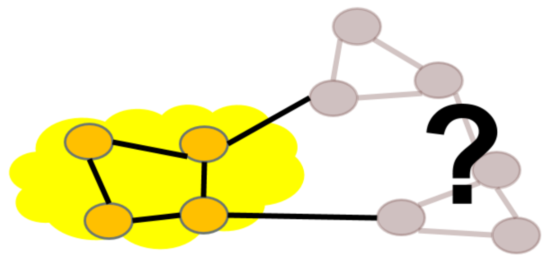


Figura 279: Exemplo Redes desconhecidas.

9.1.2 Situação Hipotética

Em um dia de prova os alunos (nós) podem ser vistos como uma rede formada de acordo com o posicionamento em relação aos seus colegas (arestas existem quando um colega está ao lado, acima ou abaixo) (Figura 280). Os pesos dessas arestas podem ser mensuradas de acordo com o risco do professor ver uma possível trapaça (Figura 281). Um dos alunos resolve passar cola para um colega que se encontra na outra ponta. Ele só consegue ver os colegas imediatamente ao lado, portanto ele não consegue perceber a priori qual o melhor caminho (Figura 282). E, mesmo que ele consiga de alguma forma determinar o melhor caminho, um dos colegas que ajudavam a passar a cola entrega a prova e vai embora. Como determinar o melhor caminho (ou quase melhor) em uma situação como essa e que, ao ocorrer alguma alteração na rede, ser possível determinar um novo caminho o mais rápido possível? (Figura 283)

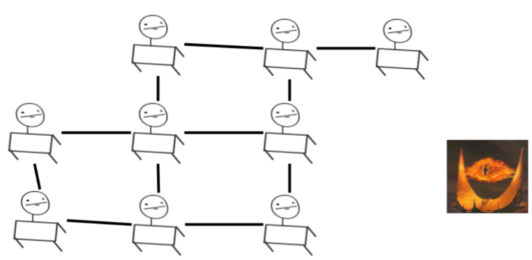


Figura 280: Exemplo situação hipotética.

9.1.3 Algoritmo de Roteamento

Esse problema é resolvido através dos algoritmos de roteamento. Nesses algoritmos existem uma "colaboração" entre cada nó com seus vizinhos para compartilhar a informação conhecida localmente por cada um deles. Dessa

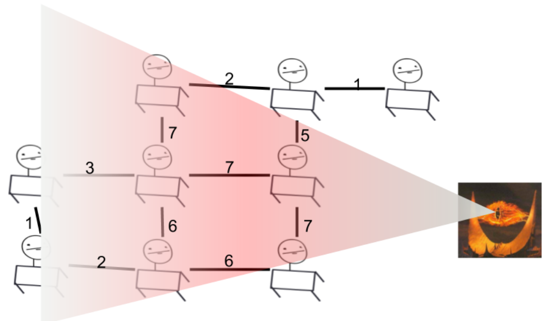


Figura 281: Exemplo situação hipotética.

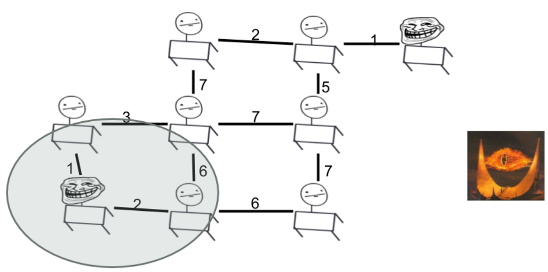


Figura 282: Exemplo situação hipotética.

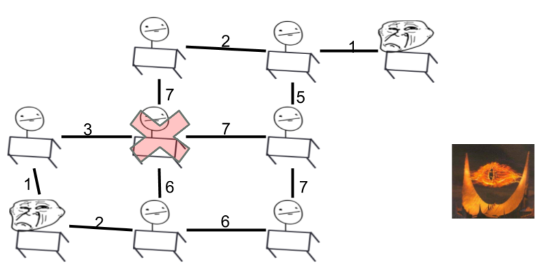


Figura 283: Exemplo situação hipotética.

forma cada roteador é capaz de estimar a melhor rota para enviar a informação (Figura 284).

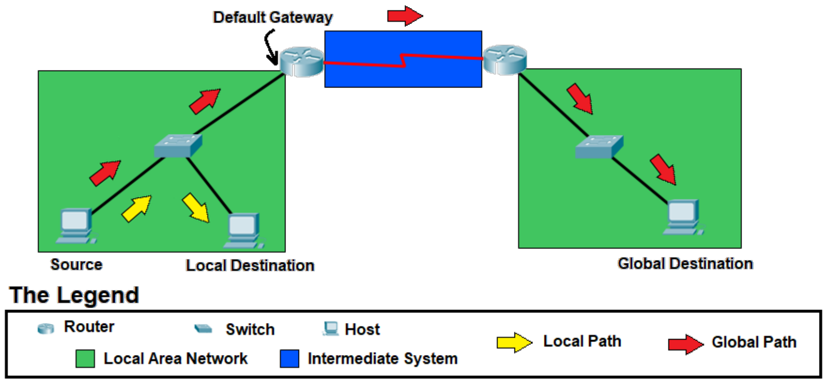
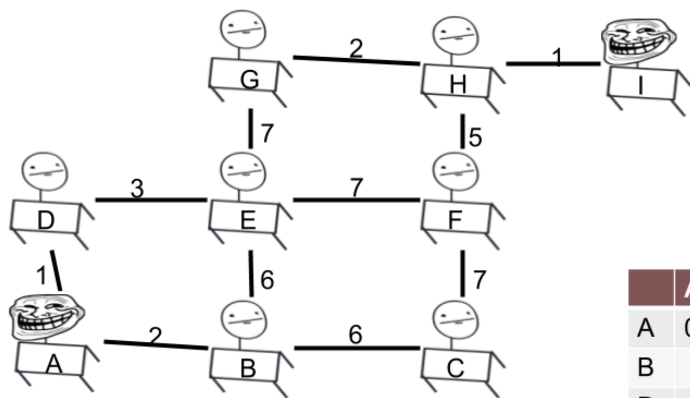


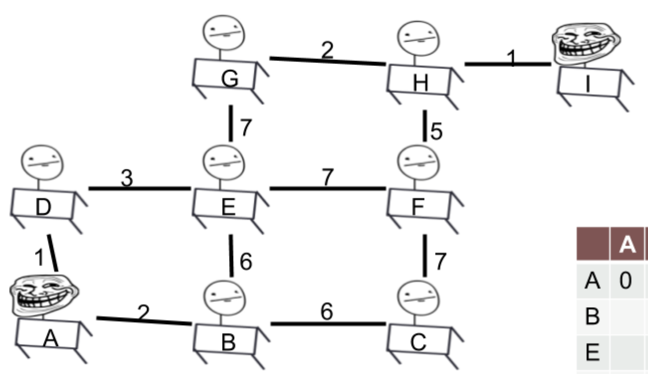
Figura 284: Algoritmo de roteamento `https://commons.wikimedia.org/wiki/File:Unicast_Routing_Terms_and_concepts_-_en.png`

O nó A sabe o custo para enviar informação aos nós B e D. Portanto sua tabela de roteamento inicialmente é (Figura 285). O nó B sabe o custo para enviar informação aos nós A, E e C(Figura 286). O nó D sabe o custo para enviar informação aos nós A e E (Figura 287). Ao conhecer as 3 tabelas, o nó A recalcula as distâncias aumentando sua tabela de caminhos mínimos conhecidos (Figura 288). Os nós A, B e D compartilham suas tabelas de rota entre si e, cada um, terá a seguinte tabela (Figura 289). Quando o nó E envia suas informações locais, a tabela aumenta (Figura 290). As informações obtidas por C pouco alteram a tabela (Figura 291). As informações de F aumentam a tabela (Figura 292). Com G atualizamos a tabela novamente (Figura 293). Com H descobrimos o último nó e completamos a tabela (Figura 294).



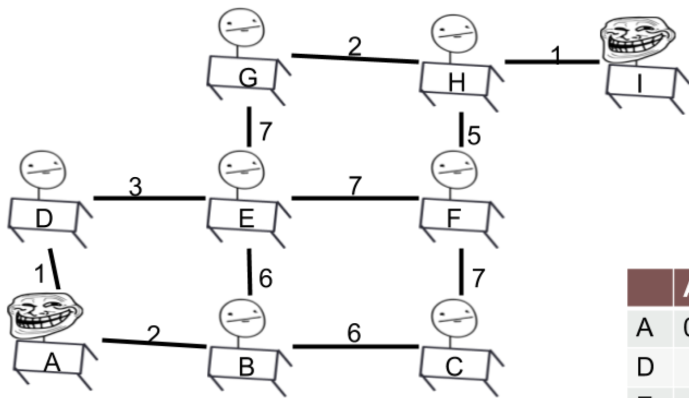
	A	B	D
A	0	2	1
B		0	3
D			0

Figura 285: Exemplo roteamento.



	A	B	E	C
A	0	2	8	8
B		0	6	6
E			0	12
C				0

Figura 286: Exemplo roteamento.



	A	D	E
A	0	1	4
D		0	3
E			0

Figura 287: Exemplo roteamento.

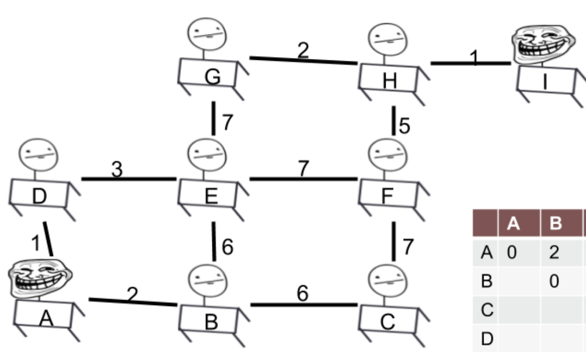
A	A	B	D
A	0	2	1
B		0	3
D			0

B	A	B	E	C
A	0	2	8	8
B		0	6	6
E			0	12
C				0

D	A	D	E
A	0	1	4
D		0	3
E			0

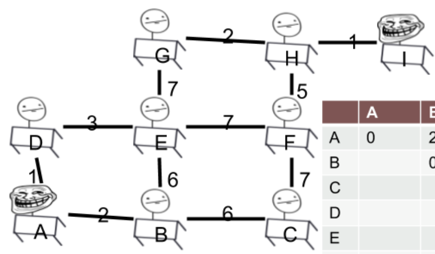
	A	B	C	D	E
A	0	2	8(B)	1	4 (D)
B		0	6	3(A)	6
C			0	9(B)	12
D				0	3
E					0

Figura 288: Exemplo roteamento.



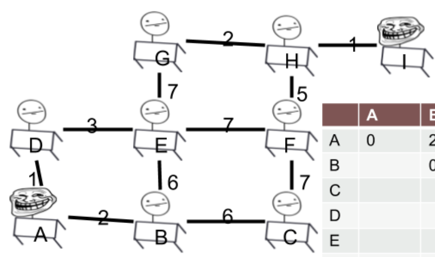
	A	B	C	D	E
A	0	2	8(B)	1	4 (D)
B		0	6	3(A)	6
C			0	9(B)	12
D				0	3
E					0

Figura 289: Exemplo roteamento.



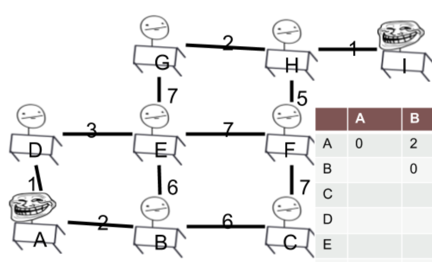
	A	B	C	D	E	F	G
A	0	2	8(B)	1	4(D)	11(D)	11(D)
B		0	6	3(A)	6	13(E)	13(E)
C			0	9(B)	12(B)	19(B)	19(B)
D				0	3	10(E)	10(E)
E					0	7	7
F						0	14(E)
G							0

Figura 290: Exemplo roteamento.



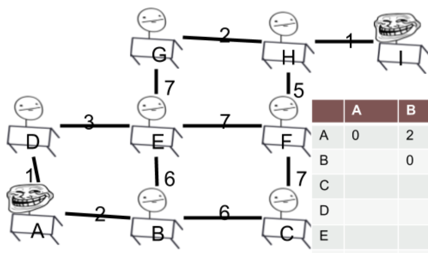
	A	B	C	D	E	F	G
A	0	2	8(B)	1	4(D)	11(D)	11(D)
B		0	6	3(A)	6	13(E)	13(E)
C			0	9(B)	12(B)	7	19(B)
D				0	3	10(E)	10(E)
E					0	7	7
F						0	14(E)
G							0

Figura 291: Exemplo roteamento.



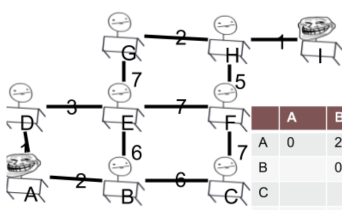
	A	B	C	D	E	F	G	H
A	0	2	8(B)	1	4(D)	11(D)	11(D)	16(D)
B		0	6	3(A)	6	13(E)	13(E)	18(E)
C			0	9(B)	12(B)	7	19(B)	12(F)
D				0	3	10(E)	10(E)	15(E)
E					0	7	7	12(F)
F						0	14(E)	5
G							0	19(E)
H								0

Figura 292: Exemplo roteamento.



	A	B	C	D	E	F	G	H
A	0	2	8(B)	1	4(D)	11(D)	11(D)	13(D)
B		0	6	3(A)	6	13(E)	13(E)	15(E)
C			0	9(B)	12(B)	7	19(B)	12(F)
D				0	3	10(E)	10(E)	12(E)
E					0	7	7	9(G)
F						0	14(E)	5
G							0	2
H								0

Figura 293: Exemplo roteamento.



	A	B	C	D	E	F	G	H	I
A	0	2	8(B)	1	4(D)	11(D)	11(D)	13(D)	14(D)
B		0	6	3(A)	6	13(E)	13(E)	15(E)	16(E)
C			0	9(B)	12(B)	7	19(B)	12(F)	13(F)
D				0	3	10(E)	10(E)	12(E)	13(E)
E					0	7	7	9(G)	10(G)
F						0	14(E)	5	6(H)
G							0	2	3(H)
H								0	1
I									0

Figura 294: Exemplo roteamento.

10 Aplicações

10.1 Análise da rede de interações dos personagens de Breaking Bad

Os dados da rede foram coletados por o Grupo CR 2015 e complementada por Fabrício Olivetti de França e pelo Arthur Feltrin. Membros do Grupo CR 2015:

- Daniel Gomes Cardoso
- Fabiany Cristiny Alves da Silva
- Gustavo Vermelho
- Lucas Saltorelli Sgarbi
- Renato da Silva Freitas
- Wagner Massashi Takahashi

Walter White é um professor de química do segundo grau, descobre um câncer de pulmão. Ele entra em desespero por não dinheiro para deixar para sua família. Um dia seu cunhado Hank leva para uma batida policial de um traficante local. Ele reconhece Jessie Pinkman, seu ex-aluno, e descobre que ele tem feito muito dinheiro produzindo e vendendo metanfetamina. Walter se associa a Pinkman para fabricar uma metanfetamina mais pura e juntar dinheiro para deixar para sua família. Os diversos traficantes locais não gostam da interferência de Walter/Pinkman e tentam fazer com que eles trabalhem para suas redes, trazendo ameaças para a família do Walter. Com o passar dos episódios Walter ganha aliados e adversários (Figura 295).

Figura 295: Banner Breaking Bad

É tido como um dos melhores seriados da televisão pelos críticos. Também foi grande sucesso de público, crescendo a cada temporada (Figura 296). Parte do sucesso é atribuído a trama muito bem elaborada partindo da premissa que "bem e mal são decisões complicadas"e os caminhos escolhidos tem consequências. Uma outra temática recorrente é a devoção a família. Vários personagens fazem tudo em nome dela. Como o criador da série conseguiu manter a coerência e interesse da trama? A rede formada pela interação dos personagens contribuiu para isso? (Figura 297).

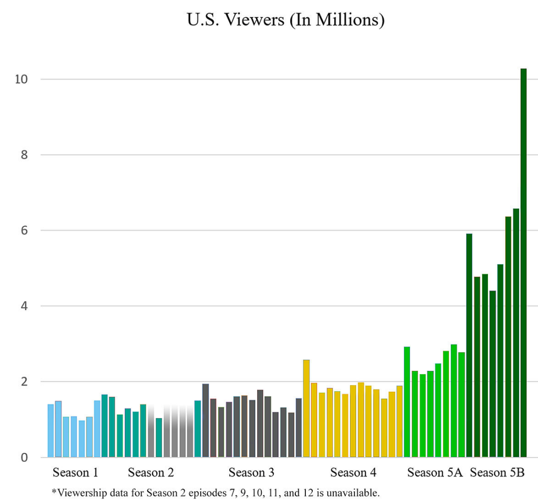


Figura 296: Rating do Breaking Bad ao longo das temporadas.

Figura 297: Interação entre os personagens do Breaking Bad.

A rede é composta por 75 personagens do seriado (nós) e 223 interações (arestas) coletadas ao longo de alguns episódios. Propriedades mensuradas: Grupo, Sexo, Etnia . A rede tem distância média de 2.33, ou seja, espera-se de 2 a 3 passos para um personagem atingir outro. Essa característica, nesse caso, está associada ao universo reduzido introduzido no seriado, onde todos os grupos estão próximos. Talvez uma simbologia de que o perigo está mais próximo do que imaginamos (Figura 298).

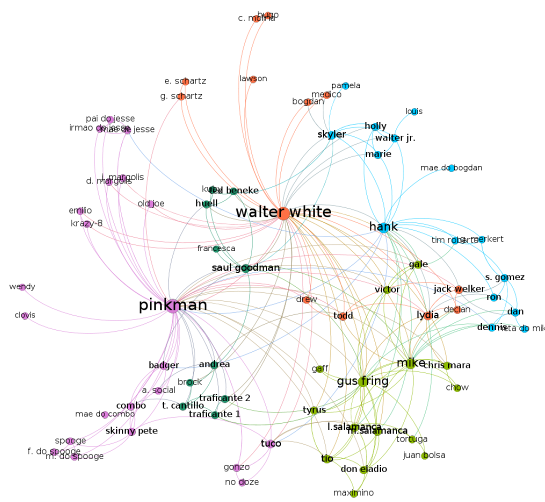


Figura 298: Rede do Trafico no Breaking Bad, com uma distancia media de 2.33, diâmetro de 5 passos e coeficiente de agrupamento de 0.62.

O diâmetro da rede é de 5 passos, ocasionado principalmente por personagens "suporte"que aparecem em uma ou duas cenas do seriado. O Coeficiente de agrupamento médio da rede foi de 0.62, corroborando com o que foi dito anteriormente de um universo fechado, em que os grupos tem muitas ligações entre si. Os nós com centralidade alta geralmente estão associados a personagens suporte que interage com um grupo de personagens reduzidos.

Personagens principais tendem a ter coeficiente de agrupamento baixo, conforme interagem com mais grupos distintos. Vejam que é do interesse de Pinkman e Walter White que os grupos com o qual eles interagem não saibam de suas atividades com outros grupos (Tabela 10.1).

Como é de se esperar com o que foi dito até então, que quanto maior a centralidade de grau, menor o coeficiente de agrupamento. Assim como esses serem os personagens principais da trama, por aparecerem em mais cenas, interagem com um grupo maior (Tabela 10.1).

Em relação a proximidade, embora os personagens principais continuem na mesma posição, o sexto e o sétimo são personagens diferentes (Tabela 10.1). A personagem Lydia teve uma relação de interação estratégica com os "chefões"de grupos que não se conversaram durante o seriado.

A diferença observada nesse rank de Centralidade de Betweeness é a subida de Mike em relação ao Gus Fring. Mike é um personagem ponte e, removendo os principais, era esperado ter essa centralidade alta. Durante o seriado Mike serviu como o intermediador de diversas negociações, quando os "chefões"não

Personagem	Coeficiente de agrupamento
Jesse Pinkman	0.10
Walter White	0.12
Hank	0.19
Mike	0.25
Gus Fring	0.27
Saul Goodman	0.29
Skyler	0.33

Tabela 2: Coeficiente de agrupamento por personagem.

Personagem	Centralidade de Grau
Jesse Pinkman	0.58
Walter White	0.53
Hank	0.28
Gus Fring	0.28
Mike	0.27
Saul Goodman	0.15
Skyler	0.14

Tabela 3: Centralidade de Grau por personagem.

Personagem
Jesse Pinkman
Walter White
Hank
Gus Fring
Mike
Lydia
Saul Goodman

Tabela 4: Centralidade de Proximidade.

queriam aparecer (Tabela 10.1).

Personagem
Jesse Pinkman
Walter White
Hank
Mike
Gus Fring
Saul Goodman
Skyler

Tabela 5: Centralidade de Betweeness

Por causa de como a trama foi montada, a ideia principal era justamente separar os mais próximos dos personagens principais do restante, por isso dois grupos foram formados. O grupo do Saul Goodman é um grupo que tem um envolvimento grande com Walter e Pinkman por fazer parte do processo de lavagem de dinheiro, tendo sua própria rede de contato. Por fim, o grupo dos associados a Gus Fing também é bem característico por ele possuir uma rede própria de produção e distribuição de drogas. Foram detectadas 5 comunidades:

- Familiares e amigos do Walter White
- Familiares e amigos do Pinkman
- Saul Goodman e associados
- Rede do tráfico de Gus Fring
- Personagens de suporte

Esses grupos mostram que a história do seriado procura definir grupos bem fechados que interagem com outros apenas esporadicamente, passando novamente a impressão de "o perigo está mais próximo do que imagina".

Redes disassortativas se caracterizam por terem uma comunicação mais rápida e robustas. A assortatividade de grau da rede é de -0.31. Para esse caso, esse tipo de rede consegue dar um controle maior ao roteirista da série para o destino dos personagens. A morte de um personagem aleatório não tem impacto na rede criada pelos roteiristas. Mesmo personagens principais não destroem a rede por completo. Durante todo o seriado personagens importantes foram mortos para criar expectativa pelos próximos episódios. Porém essas mortes não desconectaram a rede.

A assortatividade do gênero foi de apenas 0.02, não discriminando a relação em função do sexo do personagem. A etnia teve um grau de importância de 0.15 nas interações, indicando que os grupos étnicos estavam agrupados entre si.

A rede segue uma lei de potência, seguindo ligações preferenciais. Como $\alpha < 1$ temos um crescimento sigmoidal, o que representa estabilidade e controle.

É de se esperar que em um filme ou seriado de TV o roteirista não tenha como manter um crescimento explosivo do número de personagens e suas interações. A formação da lei de potência nesse caso ocorre por conta de uma quantidade grande de personagens suporte e apenas alguns poucos principais que interagem com os diversos grupos (Figura 299).

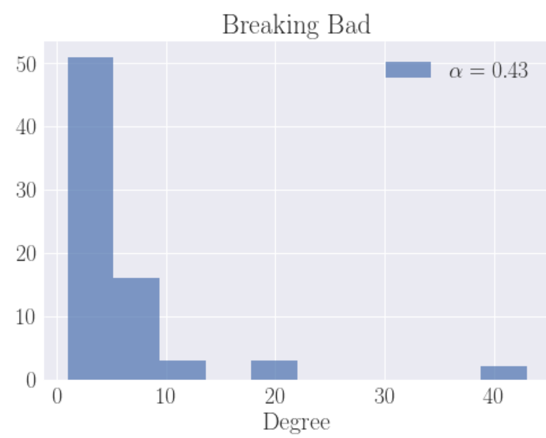


Figura 299: A rede do seriado segue a Lei Potencia.

Algumas conclusões achadas são:

- A rede do seriado Breaking Bad foi construída de tal forma a evidenciar que grupos criminosos podem estar interagindo com pessoas "de bem"sem que elas percebam.
- Ela contém poucos personagens principais bem conectados que são introduzidos inicialmente em seus próprios grupos e acabam interagindo direta ou indiretamente com outros grupos.
- A própria estrutura da rede permite que a interação indireta seja controlada pelos roteiristas e facilmente direcionada para gerar reações a ações individuais.

11 Difusão de Informação

Determinamos até agora:

- Formas de detectar os nós centrais de uma rede, de acordo com o objeto de estudo e;
- Dados transmitidos pela internet;
- Energia transmitida em uma cadeia alimentar;
- Pessoas trafegando em um sistema público de transporte;
- Doenças virais.

Como os nós podem ser agrupados formando comunidades distintas e bem separadas, algumas perguntas surgem:

- Como uma informação se propaga nessa rede?;
- Qual a velocidade da propagação?;
- Ela atinge todos os nós?.

Parte dos estudos de fluxo e difusão de informação se baseiam em estudos epidemiológicos. Esses estudos modelam o espalhamento de doenças em redes sociais (Figura 300).

Existem dois modelos bem conhecidos na difusão da informação:

- SIR (Suscetível-Infectado-Recuperado): corresponde a um modelo em que uma passa pelo estágio de suscetível ao contágio, infectada e recuperada, quando não contrai novamente a doença.
- SIS (Suscetível-Infectado-Suscetível): modelo similar ao anterior, mas ao se curar de uma doença a pessoa se torna novamente suscetível ao contágio.

11.1 Modelo SIR

Esse modelo ocorre em três estágios:

- Suscetível: pode ser contagiada com probabilidade P_{im};
- Infectada: capacidade de contagiar com probabilidade P_c;
- Recuperada: imune à doença.

Inicialmente a pessoa está no estágio "suscetível" em que ela não foi contagiada pela doença mas pode ser contagiada com certa probabilidade P_{im}.Caso ela seja contagiada, essa pessoa passa para o estágio "infectado" e, nesse momento ela tem a "capacidade" de contagiar outras pessoas com probabilidade P_c.Finalmente, após certo tempo, a pessoa infectada se recupera, passando ao estágio "recuperado" e se torna imune à doença não podendo mais transmitir a doença ou ser contagiado por ela. Em um contexto generalizado:

Figura 300: Estudos Epidemiológicos `http://education.mit.edu/tng-community/node/22`.

- Inicialmente a pessoa está no estágio "suscetível" em que ela não recebeu a informação mas pode ser informada com certa probabilidade P_{im};
- Caso ela seja informada, essa pessoa passa para o estágio "infectado" e, nesse momento ela tem a "capacidade" de informar outras pessoas com probabilidade P_c;
- Finalmente, após certo tempo, a pessoa infectada termina de transmitir a informação, passando ao estágio "recuperado" e perde o interesse pela informação não transmitindo ou recebendo ela novamente.

Dado que temos uma população constante de tamanho N:

- $S(t)$ – representa o número de pessoas ainda não infectadas no instante t;
- $I(t)$ – representa o número de pessoas infectadas no instante t;
- $R(t)$ – representa o número de pessoas recuperadas no instante t.

Temos que:

$$S(t) + I(t) + R(t) = N$$

em qualquer instante de tempo.

Agente vai fazer uma pequena pausa no texto para falar um pouco de aquários. Imagine um aquário com água e que a quantidade de água em um instante de tempo é dada pela função $Q(t)$. A taxa de (de)crescimento de água é dada pela derivada dessa função em relação ao tempo. Em condições normais a água se mantém constante durante todo o tempo (Figura 301).

$$\frac{dQ}{dt} = 0$$

Mas se fizermos um furo no aquário, a taxa se torna negativa (e possivelmente constante)(Figura 302).

$$\frac{dQ}{dt} = \lambda$$

E se compensarmos a saída de água jogando água lá dentro temos um equilibro (Figura 303).

$$\frac{dQ}{dt} = \tau - \lambda$$

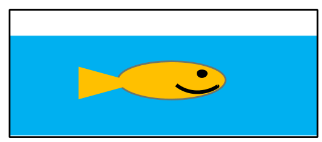


Figura 301: Exemplo aquário com nível constante.

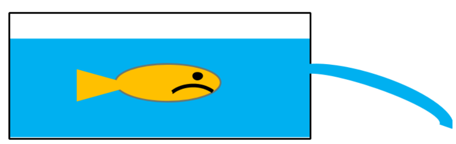


Figura 302: Exemplo aquário com um furo.

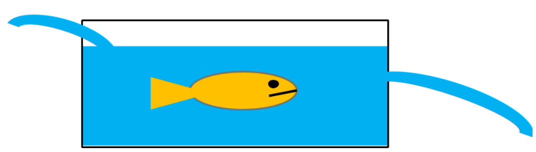


Figura 303: Exemplo aquário com um furo e com entrada de agua.

Retornando ao modelo SIR, para uma pessoa passar de suscetível para infectada ela precisa encontrar uma pessoa infectada.

$$S(t) + I(t) + R(t) = N$$

Dado β como a taxa com que duas pessoas se encontram e I/N a taxa de pessoas infectadas, então a taxa com que pessoas suscetíveis se tornam infectadas é:

$$S \to I = \beta \times S \times \frac{I}{N}$$

Para uma pessoa passar de infectada para recuperada leva um determinado tempo para que o sistema imune combata a doença Dado γ como sendo a taxa com que uma pessoa se recupera da doença e $1/\gamma$ representa a duração média da doença.

$$I \to R = \gamma \times I$$

Então temos que, com essas equações diferenciais podemos encontrar pontos de equilíbrio e modelar o comportamento de uma doença em uma população.

$$\frac{dS}{dt} = \frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Também é possível verificar o quão contagiosa é uma doença. Mas, como isso se aplica em uma rede?

11.2 Modelo SIR em uma rede

Parâmetros:

- p – probabilidade que um nó infectado contagie seu vizinho;
- t_I – tempo de duração do contágio.

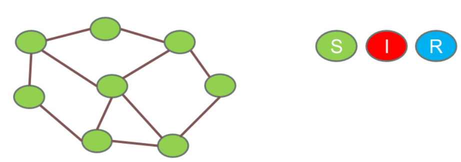


Figura 304: Modelo SIR em uma rede: suscetível(S) - verde, infectada(I) - vermelho, recuperada(R) = azul.

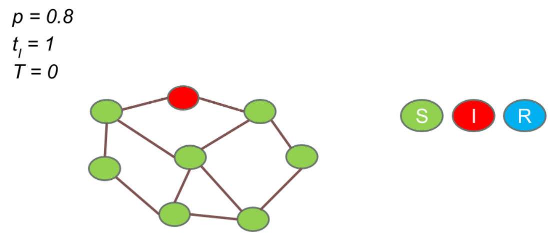


Figura 305: Modelo SIR em uma rede.

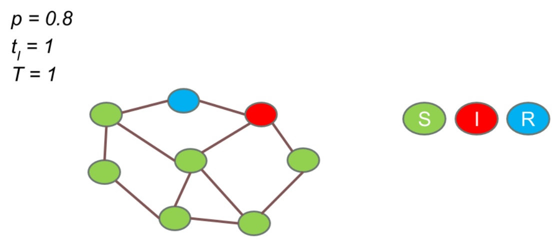


Figura 306: Modelo SIR em uma rede.

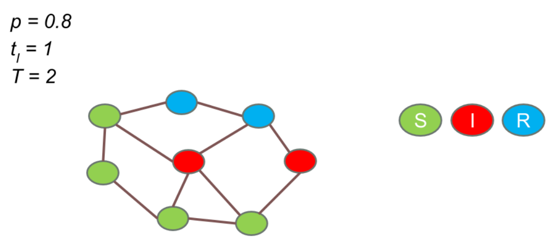


Figura 307: Modelo SIR em uma rede.

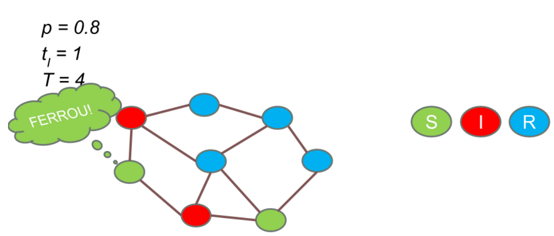


Figura 308: Modelo SIR em uma rede.

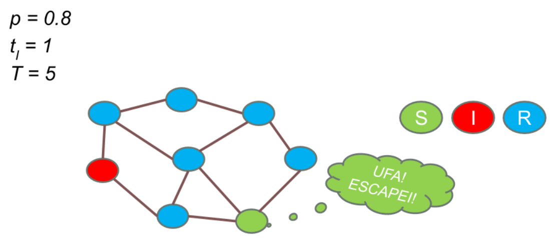


Figura 309: Modelo SIR em uma rede.

11.3 Informação e Contágio

Neste ponto, temos algumas perguntas que precisam ser respondidas:

- Dado um nó i inicialmente infectado, quais nós terão maiores chances de serem infectados?
- Qual nó da rede devo infectar para que tenha a maior chance de contagio na rede inteira?
- Quais nós da rede devo imunizar para que o vírus não se espalhe?

Pudemos reconhecer dois cenários:

- Contágio simples (SIR): nós são infectados com taxa constante.
- Absorção de informação: nós adotam uma informação após serem expostos a ela por uma fração dos nós vizinhos.

Utilizaremos um exemplo para explicar a Absorção de Informação, na rede abaixo, digamos que o nó A e o nó C criaram um produto com a mesma utilidade e querem fazer com que os seus contatos adotem tal produto (Figura 310). Um certo nó adota um novo produto SE a maioria dos nós vizinhos adota tal produto (Figura 311, 312,). O nó A que tem ligações mais fortes conseguiu prevalecer mais rapidamente sua informação, enquanto o nó C conseguiu espalhar para uma rede que A não tem acesso (Figura 313).

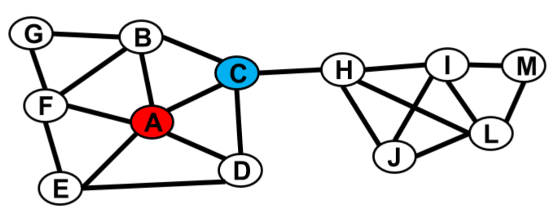


Figura 310: Exemplo Absorção de Informação.

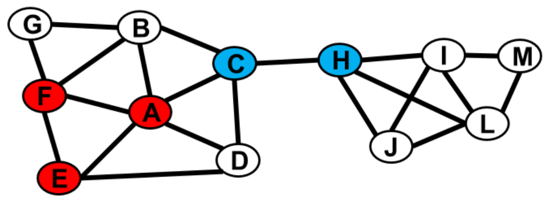


Figura 311: Exemplo Absorção de Informação.

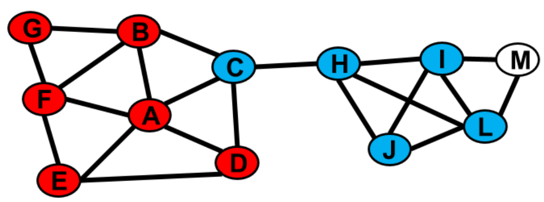


Figura 312: Exemplo Absorção de Informação.

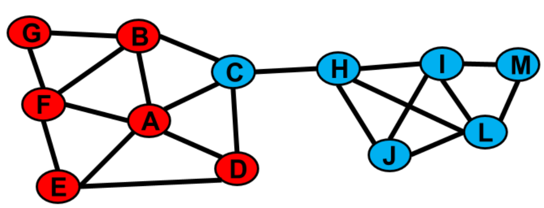


Figura 313: Exemplo Absorção de Informação. http://www.ladamic.com/netlearn/NetLogo4/DiffusionCompetition.html

11.4 Percolação

Outra questão relevante é: a informação consegue atravessar a rede inteira? Essas perguntas podem ser respondida através do fenômeno de PERCOLAÇÃO. O modelo de percolação explica o fenômeno da penetração de fluídos em meios porosos. Imagine uma rede, em forma de grade, que representa a tubulação de água de um prédio. Cada nó é um ponto de interseção e cada aresta um cano ligando dois desses pontos (Figura 314).

Imagine agora que a água, partindo do primeiro nó, passa por um determinado nó com probabilidade p. Essa probabilidade é relacionada com cada nó SER um poro ou não. A pergunta é: a água conseguirá sair do meio poroso? O nó verde representa o nó inicial (Figura 315). Os nós em vermelhos representarão os escolhidos com probabilidade p para intermediar a água (Figura 316).Se p for muito baixo, a água dificilmente conseguirá sair do meio em que se encontra (Figura 317). Esse sistema se torna interessante pois existe um valor pc chamado de ponto crítico que garante a saída da água (Figura 318).

Essa forma de percolação é denominada "percolação de sítios" (site percolation) (Figura 319). Outra forma é a percolação de ligação (bond percolation), as arestas são escolhidas com probabilidade p. Esse processo é muito próximo do modelo de redes aleatórias Erdös-Rényi que veremos na próxima aula.

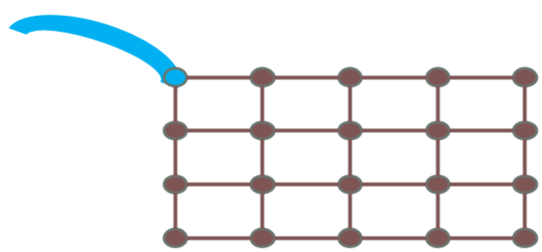


Figura 314: Exemplo Percolação.

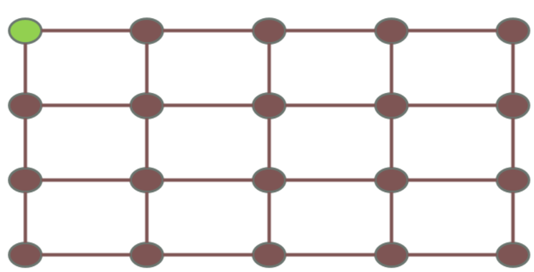


Figura 315: Exemplo Percolação.

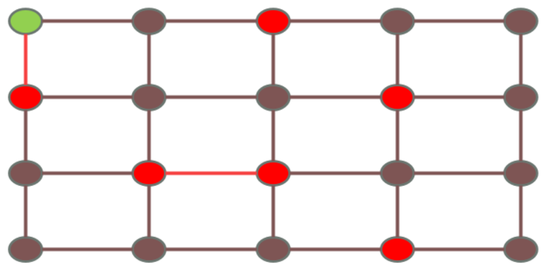


Figura 316: Exemplo Percolação.

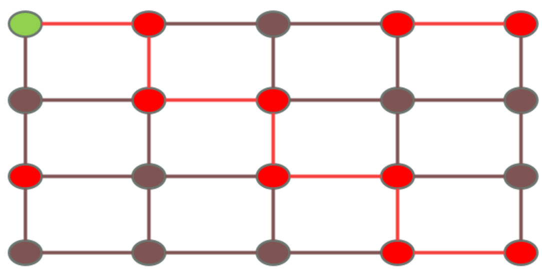


Figura 317: Exemplo Percolação.

11.5 Ponto Crítico de uma Rede

Vamos ver um exemplo de percolação em uma floresta (Figura 319). A partir de qual densidade o fogo sempre se espalha para mais de 60% da floresta? Algo acontece quando mudamos a densidade de 59% para 60%.

Com qual porcentagem de chance de transmissão a doença se espalha para quase 100% da rede em algum momento? Por volta de 3% a informação consegue atingir um maior número de nós. Também reparem que a informação começa a se difundir muito mais rápido a partir desse ponto (Figura 319).

E o modelo SIR? Bom, mas como a percolação se relaciona com o modelo SIR? Se em uma rede social temos:a probabilidade de contágio e; o tempo que uma pessoa pode retransmitir a doença. Utilizamos o modelo da percolação para verificar o quanto da rede está suscetível à doença (Figura 322). Com probabilidade de 80% escolhemos as arestas que servirão de canal de transmissão da doença (Figura 323, Figura 324). Agora, com probabilidade de 20% escolhemos as arestas que servirão de canal de transmissão da doença (Figura 325).

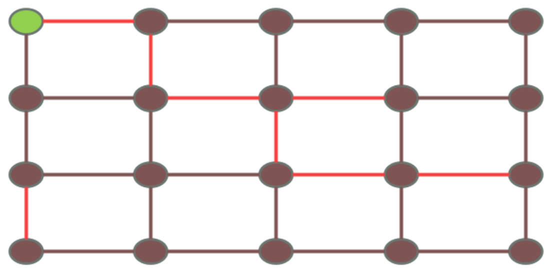


Figura 318: Exemplo Percolação.

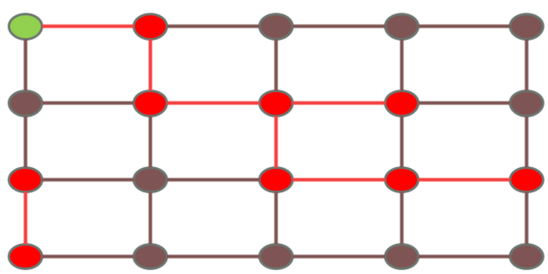


Figura 319: Exemplo Percolação.

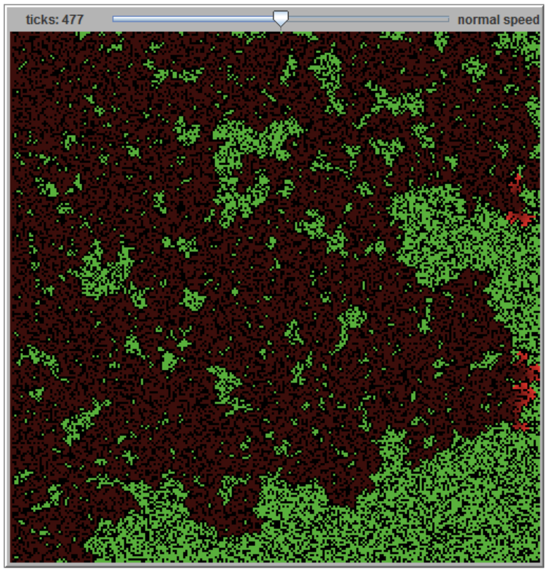


Figura 320: Exemplo de Ponto Critico no Espalhamento de fogo na Floresta. http://netlogoweb.org/launch#http://netlogoweb.org/assets/modelslib/Sample%20Models/Earth%20Science/Fire.nlogo

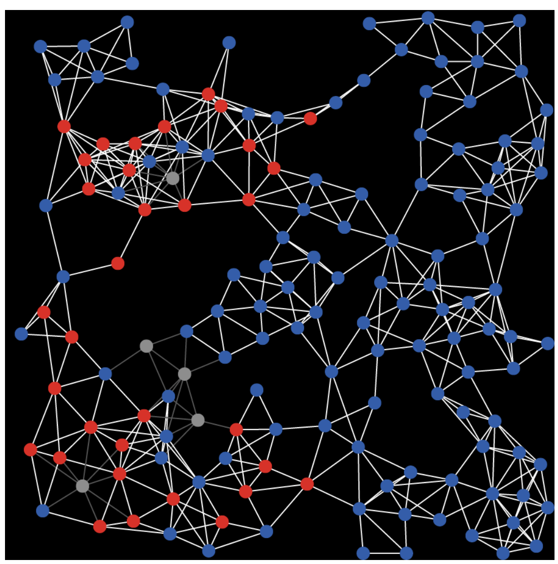


Figura 321: Exemplo de Ponto Critico na transmissão de doenças. http://netlogoweb.org/launch#http://netlogoweb.org/assets/modelslib/Sample%20Models/Networks/Virus%20on%20a%20Network.nlogo

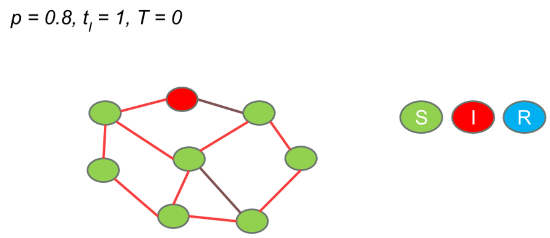


Figura 322: Exemplo SIR Percolação.

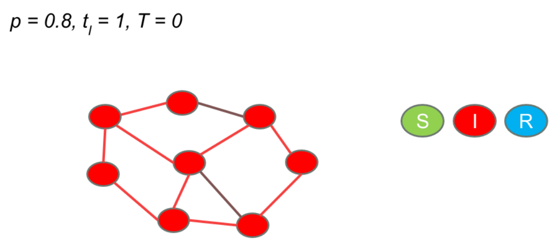


Figura 323: Exemplo SIR Percolação.

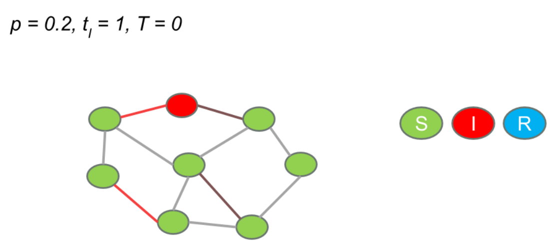


Figura 324: Exemplo SIR Percolação.

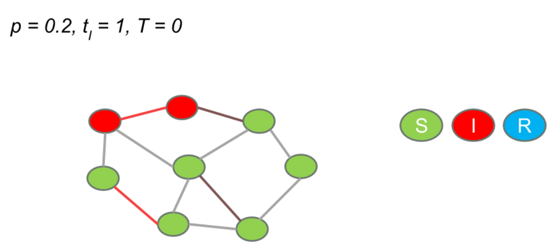


Figura 325: Exemplo SIR Percolação.

11.6 Percolação e Tolerância a Falhas

Outra questão interessante que pode ser respondida com a difusão de informação e a percolação é o quão resistente à falhas é uma rede complexa diante de ataques aleatórios. Se utilizarmos a percolação como a escolha de uma porcentagem de arestas (ou nós) a serem atacados e removidos, qual porcentagem será necessário para cortar a comunicação da rede? Pensando em epidemia, cortar comunicação = imunizar a rede (Figura 326). Fazendo várias simulações a partir do modelo de percolação, podemos verificar com qual valor de probabilidade a rede é percolada. Esse valor é denominado FASE CRÍTICA (Figura 327). Tanto as redes sem escala como as redes aleatórias são resistente a falhas eventuais, porém as redes aleatórias são mais resistentes à ataques direcionados (Figura 328). Redes de interação de proteínas seguem uma lei de potência, ou seja, existem nós nessa rede que servem como "hubs" (Figura 329). Os nós de maior grau estão associados às proteínas essenciais do nosso organismo. Nesse tipo de rede, a ausência de uma proteína pode levar a:

- Morte
- Sem efeito
- Regeneração lenta
- Efeito desconhecido

Um ataque direcionado aos hubs tem grandes chances de ser letal ao organismo através de drogas ou venenos. Além disso, se as proteínas centrais sofrerem mutações também pode ocorrer a morte do ser vivo. Por sorte mutações ocorrem aleatoriamente e em pequenas taxas, e redes sem escala são resistentes a ataques aleatórios.

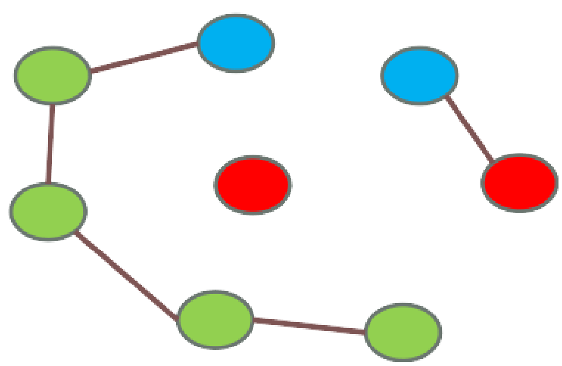


Figura 326: Exemplo SIR Percolação.

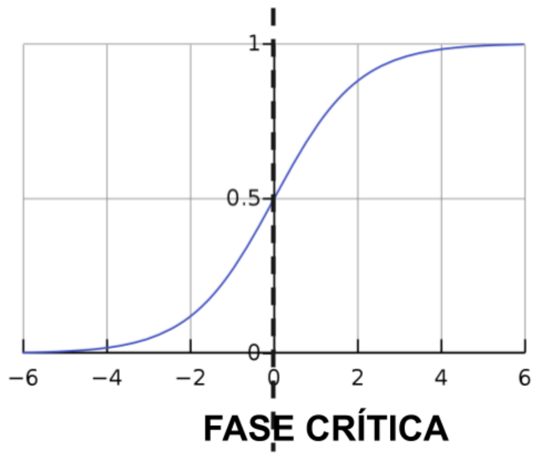


Figura 327: Exemplo SIR Percolação https://commons.wikimedia.org/wiki/File:Logistic-curve.svg.

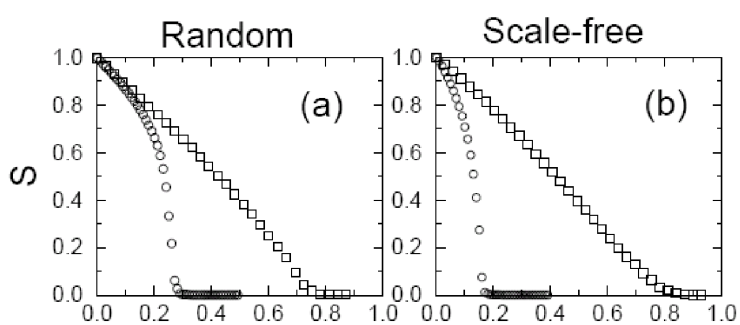


Figura 328: Redes Sem Escala x Redes Aleatórias [1] http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html

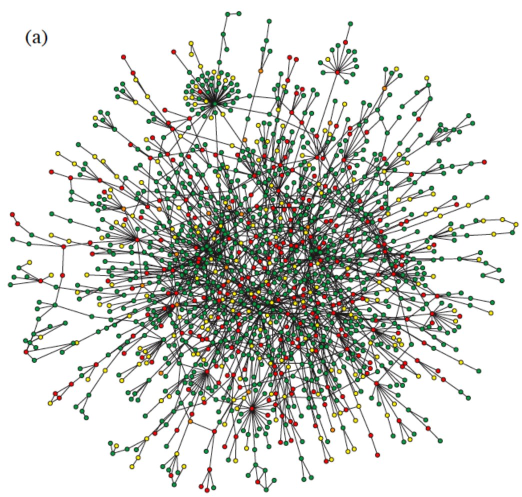


Figura 329: Rede de interação entre proteínas. Se for removido: um nó vermelho é letal, um nó verde ou marrom não letal, um nó amarelo regeneração lenta ou desconhecido. [5].

Referências

[1] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

[2] M. Arita. The metabolic world of escherichia coli is not small. *Proceedings of the National Academy of Sciences*, 101(6):1543–1547, 2004.

[3] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42, 2012.

[4] C. A. Bliss, I. M. Kloumann, K. D. Harris, C. M. Danforth, and P. S. Dodds. Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *Journal of Computational Science*, 3(5):388–397, 2012.

[5] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Physical review letters*, 86(16):3682, 2001.

[6] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[7] P. Hines, S. Blumsack, E. C. Sanchez, and C. Barrows. The topological and electrical structure of power grids. In *2010 43rd Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2010.

[8] J. P. Mena-Chalco, L. A. Digiampietri, F. M. Lopes, and R. M. Cesar. Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, 65(7):1424–1445, 2014.

[9] J. L. Moreno. Who shall survive?: A new approach to the problem of human interrelations. *Nervous and Mental Disease Publishing Company, Washington, DC*, 1934.

[10] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.