

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

ROSANA ROGISKI

**DESEMPENHO DO ALGORITMO GULOSO NA
COLORAÇÃO DE VÉRTICES EM GRAFOS DE SISTEMAS
COMPLEXOS**

TRABALHO DE CONCLUSÃO DE CURSO

CURITIBA

2016

ROSANA ROGISKI

**DESEMPENHO DO ALGORITMO GULOSO NA
COLORAÇÃO DE VÉRTICES EM GRAFOS DE SISTEMAS
COMPLEXOS**

Trabalho de Conclusão de Curso de Graduação, apresentado à disciplina de Trabalho de Conclusão de Curso 2, do Curso de Bacharelado em Sistemas de Informação do Departamento Acadêmico de Informática - DAINF - da Universidade Tecnológica Federal do Paraná - UTFPR, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Murilo Vicente Gonçalves da Silva

Co-orientador: Prof. Dr. Gustavo Alberto Gimenez Lugo

CURITIBA

2016

TERMO DE APROVAÇÃO

Rosana Rogiski

DESEMPENHO DO ALGORITMO GULOSO NA COLORAÇÃO DE VÉRTICES EM GRAFOS DE SISTEMAS COMPLEXOS

Este Trabalho de Conclusão de Curso foi apresentado às do dia de junho de 2016 como requisito parcial para a obtenção do título de BACHAREL EM SISTEMAS DE INFORMAÇÃO, pela Universidade Tecnológica Federal do Paraná. Após deliberação, a Banca Examinadora considerou o trabalho APROVADO.

Prof. Dr. Rodrigo Minetto
UTFPR

Prof. Dr. Ricardo Dutra da Silva
UTFPR

Prof. Dr. Murilo V. Gonçalves da Silva
UTFPR
Orientador

Responsável pelo TCC: Profa. Dra. Anelise Munaretto Fonseca

Coordenador do Curso de Bacharelado em Sistemas de Informação:

Profa. Dra. Marília Abrahão Amaral

A meus pais, com toda gratidão e amor.

AGRADECIMENTOS

Primeiramente agradeço à minha família pelo apoio e por não terem medido esforços para que eu conseguisse concluir mais esta etapa da minha vida.

À Lukas Ferreira Machado, que mesmo estando longe me deu suporte e incentivo, pela amizade, lealdade, companheirismo e amor.

Agradeço aos meus orientadores Prof. Dr. Murilo Vicente Gonçalves da Silva e Prof. Dr. Gustavo Alberto Gimenez Lugo pelas valiosas contribuições feitas no decorrer dos dois anos gastos na elaboração deste trabalho.

Extendo minha gratidão aos Professores Ricardo Dutra da Silva e Rodrigo Minetto, componentes da banca de avaliação, pelas importantes considerações feitas ao final das duas etapas deste trabalho de conclusão de curso.

Aos professores com quem tive contato durante a graduação por contribuírem para a minha formação e me incentivarem a buscar os meus sonhos.

Meu muito obrigado aos amigos que fiz na UTFPR e aqueles que conquistei durante o intercâmbio, pelo carinho e por terem sido meu ponto de apoio durante todos esses anos.

Finalmente, agradeço a todos que direta ou indiretamente contribuíram para que esses últimos anos fossem de muito crescimento pessoal e profissional.

”Não conheço nenhuma fórmula infalível para obter o sucesso, mas conheço uma forma infalível de fracassar: tentar agradar a todos.”

(John F. Kennedy)

RESUMO

ROGISKI, Rosana. DESEMPENHO DO ALGORITMO GULOSO NA COLORAÇÃO DE VÉRTICES EM GRAFOS DE SISTEMAS COMPLEXOS. 99 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

Neste trabalho buscamos estudar o desempenho da técnica gulosa no problema de coloração de vértices em grafos oriundos de Sistemas Complexos. Para tanto, foram selecionados grafos de vinte e cinco Sistemas Complexos e obtidos seus grafos subjacentes sobre os quais foram executados os testes. Foram analisadas duas heurísticas gulosas para a escolha do próximo vértice a colorir: a primeira utilizando a heurística do maior grau para seleção do próximo vértice a colorir e a segunda utilizando a heurística da maior saturação combinada com maior grau. Os resultados foram comparados com a solução ótima e alguns limitantes do problema de coloração de vértices.

Palavras-chave: Coloração de Vértices, Teoria dos Grafos, Sistemas Complexos, Algoritmo Guloso

ABSTRACT

ROGISKI, Rosana. PERFORMANCE OF GREEDY ALGORITHM FOR VERTEX COLORING IN GRAPHS FROM COMPLEX SYSTEMS. 99 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2016.

In this work we aim to study the performance of greedy technique in the vertex coloring problem for graphs from Complex Systems. For doing that, twenty-five Complex Systems were selected, and from them we obtained the underlying graph on which the tests were performed. Two greedy heuristics for choosing the next vertex to be colored were analyzed: the first one uses the greatest degree heuristic for selecting the next vertex to be colored, and the second one applies the greatest saturation heuristic combined with the greatest degree heuristic. The results were compared with the optimal solution and some of the boundaries to the vertex coloring problem.

Keywords: Vertex Coloring, Graph Theory, Complex Systems, Greedy Algorithm

LISTA DE FIGURAS

FIGURA 1	– Problema das Sete Pontes de Königsberg.	13
FIGURA 2	– Grafo Não Direcionado.	19
FIGURA 3	– Grafo direcionado.	19
FIGURA 4	– Exemplos de subgrafos.	21
FIGURA 5	– Grafo com cinco vértices e seis arestas.	22
FIGURA 6	– Exemplo de Multigrafo.	23
FIGURA 7	– Tipos de Grafos.	23
FIGURA 8	– Exemplos de grafo conexo e desconexo.	24
FIGURA 9	– Exemplo de grafo 2-conexo.	24
FIGURA 10	– Exemplos de Grafo bipartido e bipartido completo.	26
FIGURA 11	– Os grafos (a) e (b) são k -regulares enquanto o grafo (c) é não k -regular.	26
FIGURA 12	– Operação de União e Intersecção de Grafos.	27
FIGURA 13	– Representação da remoção de uma aresta do grafo G	28
FIGURA 14	– Representações da contração de uma aresta do grafo G	29
FIGURA 15	– Representação do grafo G por meio de matriz de adjacências.	30
FIGURA 16	– Representação do grafo G por meio de uma lista de adjacências. ..	31
FIGURA 17	– Duas colorações gulosas possíveis para o grafo G	35
FIGURA 18	– Exemplificação do Teorema 3.3.	41
FIGURA 19	– Exemplo da aplicação do Teorema 3.3.	42
FIGURA 20	– Gadget para redução de 3-CNF-SAT para 3-coloração - Para a cláusula $(x_1 \vee \neg x_2 \vee x_3)$	45
FIGURA 21	– Gadget colorido utilizado na redução da cláusula $(x_1 \vee \neg x_2 \vee x_3)$ para instância de 3-coloração.	47
FIGURA 22	– Exemplos de Coeficientes de Clusterização para o nó central (escuro).	56
FIGURA 23	– Rede de amizade de uma escola americana.	58
FIGURA 24	– Exemplos de redes geradas pelo modelo de Erdős e Rényi.	61
FIGURA 25	– Comparação da Distribuição de Graus da Wikipedia e do obtido pelo Modelo de Erdős e Rényi.	63
FIGURA 26	– Redes obtidas pelo Modelo de Watts e Strogatz.	65
FIGURA 27	– Relação entre distâncias médias e coeficiente de clusterização médio para valores intermediários de p	66
FIGURA 28	– Exemplo de rede obtida utilizando o modelo de Barabási para $n = 200$	69
FIGURA 29	– Exemplo de seleção do próximo vértice a colorir utilizando a heurística de maior grau.	74
FIGURA 30	– Exemplo de seleção do próximo vértice a colorir utilizando a heurística de maior saturação - Passo 1.	75
FIGURA 31	– Exemplo de seleção do próximo vértice a colorir utilizando a heurística de maior saturação - Passo 2.	76
FIGURA 32	– Fluxograma do algoritmo para cálculo do número cromático ótimo.	77

LISTA DE TABELAS

TABELA 1	– Graus dos Nós.	20
TABELA 2	– Detalhes das redes complexas selecionadas.	81
TABELA 3	– Características das redes complexas selecionadas.	82
TABELA 4	– Resultados da coloração gulosa.	84
TABELA 5	– Resultados da coloração gulosa.	86
TABELA 6	– Aproximação dada pela coloração gulosa em relação ao número cromático.	88
TABELA 7	– Limitantes inferiores e superiores do número cromático.	90
TABELA 8	– Aproximação garantida pelo algoritmo guloso em relação ao número clique.	92

LISTA DE SÍMBOLOS

n	Número de vértices de um grafo
m	Número de arestas de um grafo
$d(v)$	Grau do vértice v
$\delta(G)$	Grau mínimo de um grafo
$\Delta(G)$	Grau máximo de um grafo
$\omega(G)$	Número de clique de um grafo
$\alpha(G)$	Número de independência de um grafo
$\chi(G)$	Número cromático de um grafo
$\langle k \rangle$	Grau de conexão médio de uma rede
ρ	Densidade de uma rede
s_i	Força do vértice i
$\langle s \rangle$	Força média de uma rede
$P(k)$	Distribuição de graus de uma rede
f_k	Fração de vértices com grau k
$\langle l \rangle$	Distância média de uma rede
l	Diâmetro de uma rede
c_i	Coefficiente de clusterização do vértice i
$\langle c \rangle$	Coefficiente de clusterização da rede
B_v	Grau de intermediação do vértice v
$\langle B \rangle$	Grau de intermediação médio de uma rede
$C(v)$	Grau de proximidade do vértice v
$E(m)$	Probabilidade de que existam m arestas em uma rede
$G(n,p)$	Modelo de criação para redes aleatórias
$ S $	Cardinalidade do conjunto de redes que podem ser geradas pelo Modelo de Rede Aleatória
$P[G(n,p) \rightarrow E]$	Probabilidade de que o modelo de Rede Aleatória gere uma rede com o conjunto E de arestas
$P[G(n,p) \rightarrow E = m]$	Número de redes com m arestas gerados pelo modelo de Rede Aleatória
$P[D = k]$	Probabilidade de um vértice qualquer possuir grau k no modelo de Rede Aleatória
$\langle l \rangle$	Comprimento médio de caminho
$\ell(p)$	Distância média da rede gerada no modelo de rede Mundo Pequeno
$c(p)$	Coefficiente de clusterização de uma rede gerada no modelo de rede Mundo Pequeno
$\Pi(k_i)$	Probabilidade de um nó i receber conexão com um novo vértice
γ	Coefficiente de uma função do tipo Lei de Potência
$P(k,t)$	Distribuição de graus para uma rede gerada no modelo de Rede Livre de Escala

SUMÁRIO

1 INTRODUÇÃO	11
1.1 JUSTIFICATIVA	15
1.2 OBJETIVOS	16
1.2.1 Objetivo Geral	16
1.2.2 Objetivos Específicos	16
1.3 ORGANIZAÇÃO DO TRABALHO	16
2 CONCEITOS INTRODUTÓRIOS	18
2.1 TÓPICOS EM TEORIA DE GRAFOS	18
2.2 TIPOS DE GRAFOS	22
2.3 OPERAÇÕES EM GRAFOS	27
2.4 REPRESENTAÇÃO COMPUTACIONAL	29
3 COLORAÇÃO DE VÉRTICES	32
3.1 LIMITAÇÕES DO NÚMERO CROMÁTICO	33
3.2 POLINÔMIO CROMÁTICO	38
3.3 CARÁTER NP-DIFÍCIL	43
4 REDES COMPLEXAS	49
4.1 CARACTERIZAÇÃO TOPOLÓGICA	52
4.2 MODELOS DE CRESCIMENTO	59
4.2.1 Rede aleatória - Modelo de Erdős e Rényi	59
4.2.2 Rede mundo pequeno - Modelo de Watts e Strogatz	64
4.2.3 Rede Livre de Escala - Modelo de Barabási e Albert	67
5 METODOLOGIA	72
6 RECURSOS DE HARDWARE E SOFTWARE	79
6.1 RECURSOS DE HARDWARE	79
6.2 RECURSOS DE SOFTWARE	79
7 RESULTADOS	81
8 CONCLUSÕES E TRABALHOS FUTUROS	95
REFERÊNCIAS	97

1 INTRODUÇÃO

Redes estão em todo lugar, desde a internet com a estrutura da *World Wide Web*, até economia, transmissão de doenças e eventos terroristas, a ideia de formação de redes permeia a cultura moderna (NEWMAN, 2003) uma vez que os sistemas reais onde componentes interagem entre si podem ser bem representados por uma estrutura de redes. Isso ocorre porque este tipo de estrutura propicia uma representação simples mas muito útil na investigação dos sistemas complexos presentes na natureza ou criados pelo homem (NEWMAN, 2008).

Estruturas complexas são capazes de descrever uma grande variedade de sistemas de grande importância para a humanidade. Por exemplo, uma célula é melhor descrita como uma rede complexa de produtos químicos interligados por meio de reações químicas. A internet pode ser melhor compreendida se vista como uma complexa rede de roteadores e computadores interligados de alguma forma por fios ou outro tipo de tecnologia (ALBERT; BARABÁSI, 2002). O mesmo ocorre nas relações sociais, onde fenômenos como a propagação de ideias pode ser melhor compreendida quando se percebe que a sociedade é uma rede de pessoas conectadas entre si por meio de vínculos de amizade, relações familiares ou ligações profissionais, por exemplo (BARABÁSI; BONABEAU, 2003).

Essas estruturas são exemplos de sistemas que levaram um grupo diverso de cientistas a investigar a topologia das redes complexas, num esforço conjunto que atrai físicos, matemáticos, sociólogos, biólogos e profissionais da Computação (NEWMAN, 2003). Isso porque a estrutura de redes é encontrada nas mais diferentes áreas do conhecimento - informática, ciências sociais, economia, ecologia, biologia e engenharia por exemplo (ALBERT; BARABÁSI, 2002). Em todas essas áreas do conhecimento existem sistemas reais que formam redes complexas cujos vértices representam elementos do sistema e cujas arestas representam as interações entre eles (AMARAL; OTTINO, 2004).

Qualquer sistema natural ou artificial formado por muitas partes que interagem entre si pode ser representado por redes (RODRIGUES, 2007). No entanto, os chamados sistemas complexos são mais interessantes do que outros tipos de sistemas, devido às suas

características específicas que os diferenciam dos sistemas simples ou complicados, especialmente pela auto-organização que apresentam. Nos sistemas complexos, os elementos são capazes de se organizar e interagir de tal forma que não é possível compreender o sistema apenas considerando as características de cada elemento, já que a interação entre os elementos e destes com o ambiente promovem mudanças no comportamento geral do sistema, uma vez que os elementos são capazes de se adaptar e reorganizar de acordo com a situação observada (AMARAL; OTTINO, 2004).

Esse é o caso, por exemplo, de bandos de pássaros ou do trânsito de uma cidade: nesses casos, não é possível compreender o todo apenas considerando os elementos isoladamente, uma vez que a interação entre os elementos é tão importante quanto o seu comportamento individual, o que os diferencia de sistemas complicados, como uma nave espacial, por exemplo. Ainda que esta última represente um sistema extremamente intrincado, ela é composta por partes muito bem determinadas, com funções específicas e comportamentos conhecidos, de tal forma que é possível prever o comportamento do sistema caso um dos seus elementos apresente falha. Nos sistemas complexos, isso não é verdade, uma vez que os efeitos da adaptabilidade e organização dos elementos não podem ser totalmente previstos.

Sistemas complexos podem ser bem representados por uma estrutura de redes, comumente chamada de rede complexa. Tradicionalmente, o estudo de redes complexas está apoiado na teoria de grafos, já que o primeiro se utiliza da segunda para representação e investigação de sua estrutura e propriedades fundamentais.

Historicamente, o estudo de redes possui uma longa história na matemática e ciência. Pode-se afirmar que sua origem remonta ao ano de 1736 quando o matemático Leonard Euler se interessou em solucionar o chamado problema das sete pontes de Königsberg. Na cidade de Königsberg (atualmente Kaliningrado, na Rússia), havia uma ilha chamada Kneiphoff formada por dois braços do rio Pregel. Havia sete pontes no entorno dessa ilha e o desafio girava em torno da possibilidade de criar um caminho que atravessasse todas as pontes, passando apenas uma vez em cada uma delas (NEWMAN, 2003).

A Figura 1 apresenta uma ilustração desse problema. Em (a) é ilustrada a cidade de Königsberg. (b) contém uma representação esquemática da área das sete pontes e por fim, em (c) está ilustrada a representação de Euler para o problema.

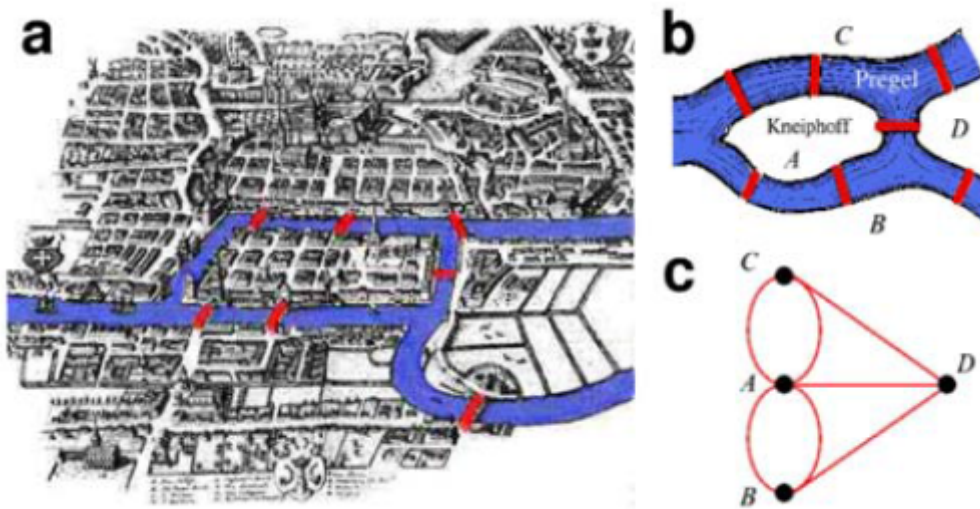


Figura 1: Problema das Sete Pontes de Königsberg.

Fonte: Amaral e Otino (2004).

Para solucionar esse problema e provar que não havia como passar por todas as pontes sem repetí-las, Euler utilizou um grafo, objeto matemático consistindo de pontos chamados de vértices ou nós e linhas chamadas de arestas ou *links*, capaz de abstrair os detalhes do problema, deixando o foco sobre o problema real da conectividade. Mais tarde esse problema foi inserido na linguagem matemática como o problema do caminho euleriano. Na Figura 1 (c), está representado o grafo utilizado por Euler, que abstraiu o problema criando quatro regiões, A, B, C e D. Cada uma dessas regiões deu origem a um nó no grafo e as pontes são representadas pelas linhas (arestas) contidas no grafo (NEWMAN, 2003).

Desde então, grafos têm sido utilizados para representar sistemas reais porque ao abstrair os detalhes do problema, é possível descrever as características topológicas importantes com uma clareza que seria impossível se todos os detalhes do problema fossem mantidos.

Desde 1930, pesquisadores vêm estudando empiricamente as redes sociais com base nos fundamentos da sociologia (NEWMAN, 2008). Mas é a partir de 1950, como resposta ao crescente interesse da comunidade científica, que a teoria dos grafos passou a ser utilizada para auxiliar cientistas na compreensão de dados de estudos etnográficos. Conceitos clássicos da teoria dos grafos foram adaptados para a análise de redes sociais, de tal forma que aspectos como centralidade, comprimento de caminho, cliques e componentes conexas têm sido utilizadas para discutir questões como status, influência, coesão,

regras sociais e identidade em redes sociais (NEWMAN, 2003).

Por mais de 40 anos, a ciência tratou as redes complexas como completamente randômicas, apoiadas no paradigma proposto em 1959 por Paul Erdős e Alfréd Rényi, que propunham que os sistemas poderiam ser modelados pela conexão dos seus vértices através de links aleatórios, o que resultava num sistema democrático, onde a grande maioria dos nós possuía aproximadamente o mesmo número de links, revelando uma distribuição de número de links no formato de Poisson com cauda curta, já que era raro encontrar um nó cujo número de links diferisse muito da média da rede (BARABÁSI; BONABEAU, 2003).

No final da década de 1990, dois outros trabalhos buscaram apresentar modelos que fossem compatíveis com o crescimento observado em redes reais. O primeiro deles foi proposto por Watts e Strogatz (1998) e o segundo por Albert e Barabási (2002). Ambos eram capazes de recriar importantes características percebidas em redes reais, embora nenhum deles fosse capaz de reproduzir todas as propriedades encontradas na exploração de diferentes tipos de redes estudadas ao longo das últimas décadas. Embora a teoria de redes complexas tenha obtido grande sucesso, há um vasto campo de pesquisa que pode ser explorado, devido às limitações em termos de medidas e métodos de análise, caracterização e classificação de redes reais (RODRIGUES, 2007).

Enquanto a teoria da redes complexas era desenvolvida, a teoria dos grafos também seguia seu curso, com a proposição e prova de novos conceitos, teoremas e aplicações.

Em 1852, enquanto tentava colorir o mapa de municípios da Inglaterra, Francis Guthrie percebeu que apenas quatro cores seriam suficientes para colorir o mapa de tal forma que municípios fronteiriços recebessem cores distintas. Isso deu origem à conjectura das quatro cores que afirma que as regiões de qualquer mapa podem ser coloridas mediante o uso de quatro cores distintas, sem que regiões vizinhas recebam cores iguais (CHARTRAND; ZHANG, 2009).

Esse problema despertou o interesse de muitos cientistas, mas foi só em 1976, já na era da computação, que Appel e Haken apresentaram uma prova para tal conjectura. Para tanto, Appel e Haken dividiram o problema em várias subclasses de mapas e escreveram um algoritmo computacional para verificar se um mapa poderia ser colorido com quatro cores em todos os possíveis casos (FORMANOWICZ; TANÁŠ, 2012; CHARTRAND; ZHANG, 2009).

Desde então, muito se tem estudado sobre coloração em grafos, não somente visando a coloração de mapas mas pensando na coloração de arestas, vértices ou grupos de vértices. Tal interesse se deve ao fato de que coloração de grafos pode ser utilizada

para modelar diversos problemas reais, tais como alocação de recursos (NAVARRO; COELHO, 2009), agendamento, horários, otimização de registradores, plataformas de trem, atribuição de tarefas, redes de comunicação (REGO; SANTOS, 2011) e armazenamento de produtos químicos (BONDY; MURTY, 2008).

Devido às inúmeras aplicações e relevância prática e teórica, a coloração de vértices em grafos é um dos problemas mais estudados em teoria dos grafos. Esse problema faz parte da classe dos problemas NP-Difícil (*NP-Hard*), ou seja, os algoritmos exatos conhecidos demandam tempo exponencial de resolução e é pouco provável que ele possa ser resolvido de maneira exata em tempo polinomial (BONDY; MURTY, 2008; LEWIS; PAPANIMITRIOU, 1998).

1.1 JUSTIFICATIVA

Tendo em vista que coloração de vértices em grafos é um dos temas mais relevantes em teoria de grafos e considerando que muitos problemas reais podem ser modelados por redes complexas, este trabalho se propõe a verificar o desempenho dos algoritmo guloso para coloração de vértices, quando aplicado a grafos originários de redes complexas que possuem uma distribuição de graus na forma de lei de potência (*Power-Law*).

A aplicação de algoritmos gulosos representa uma alternativa à utilização de algoritmos ótimos, cuja utilização em grafos com grande quantidade de vértices se torna inviável, visto que exige tempo exponencial de processamento. Sendo assim, com os algoritmos ótimos que conhecemos hoje, seria improvável encontrar solução ótima, pois em alguns casos, dada a tecnologia que temos na atualidade, não há máquina com poder de processamento e espaço em disco ou memória suficiente para encontrar a solução ótima em tempo plausível.

Com essa análise, busca-se compreender como as características das redes complexas podem ser utilizadas pela computação para melhorar algoritmos e aplicações em grafos de redes reais, uma vez que, se muitos grafos de sistemas reais são do tipo lei de potência, ou seja, possuem poucos vértices com alto grau e muitos vértices com baixo grau, suas características, em tese, poderiam ser melhor exploradas pela computação na busca de algoritmos mais eficientes.

1.2 OBJETIVOS

Nesta seção são apresentados os objetivos traçados para a execução deste projeto de pesquisa.

1.2.1 OBJETIVO GERAL

Analisar o desempenho do algoritmo guloso para coloração de vértices em grafos de redes complexas com distribuição de graus na forma de lei de potência.

1.2.2 OBJETIVOS ESPECÍFICOS

Este trabalho possui como objetivos específicos:

- Avaliar o desempenho de algoritmos de heurística gulosa para coloração de vértices em relação ao algoritmo ótimo.
- Comparar o desempenho de algoritmos de heurística gulosa para coloração de vértices em relação a seus limitantes inferiores.
- Compreender como as características das redes complexas podem ser utilizadas para melhorar os algoritmos utilizados na computação.

1.3 ORGANIZAÇÃO DO TRABALHO

Para uma melhor organização, o trabalho foi subdividido em duas grandes etapas que juntas colaboram para que se atinja o objetivo final. A primeira se refere à teoria de coloração de grafos e redes complexas e a segunda que se dedica à execução dos algoritmos a serem detalhados posteriormente.

A primeira etapa está subdividida em dois grandes blocos: estudo da teoria de grafos, especialmente no que concerne à coloração de vértices em grafos. Esse estudo contempla os conceitos mais importantes numa revisão da área. Um segundo bloco se dedica a analisar definições, métricas e características mais relevantes de redes complexas, especialmente no que concerne a redes complexas do tipo lei de potência.

Na segunda grande etapa do trabalho, os algoritmos para coloração de grafos são aplicados ao conjunto de grafos selecionado e o desempenho dos algoritmos é analisado e discutido.

Este trabalho está organizado da seguinte maneira: no capítulo 2 são abordados conceitos relativos à teoria dos grafos. São apresentadas definições importantes a serem utilizadas para a definição de conceitos relativos à coloração de vértice em grafos que são abordados no capítulo 3.

No capítulo 4 são tratadas as redes complexas: principais características, métricas, estrutura e os modelos de crescimento de redes mais importantes encontradas na literatura especializada. Já o capítulo 5 apresenta em detalhes a metodologia adotada na execução deste projeto.

O capítulo 6 discute e analisa os recursos de hardware e software necessários para a execução das tarefas relativas a este projeto.

Por fim o capítulo 7 apresenta e analisa os resultados obtidos nos experimentos realizados, restando ao capítulo 8 a enumeração das conclusões e trabalhos futuros que podem ser conduzidos na mesma linha de pesquisa que o presente trabalho de conclusão de curso.

2 CONCEITOS INTRODUTÓRIOS

Este capítulo é dedicado a apresentar alguns conceitos importantes acerca da teoria de grafos que serão utilizados ao longo desta monografia. Os conceitos apresentados neste capítulo são de domínio da área de teoria de grafos e podem ser encontradas nas seguintes bibliografias: Bondy e Murty(1979), West (2002), Diestel (2005), Bondy e Murty(2008) e Feofiloff (2011).

2.1 TÓPICOS EM TEORIA DE GRAFOS

Muitas situações reais podem ser convenientemente descritas através de diagramas chamados grafos, que consistem em um conjunto de pontos e linhas interligando pares desses pontos (BONDY; MURTY, 1979, 2008). Esses pontos, por exemplo, podem representar pessoas e as linhas pares de amigos; ou os pontos podem representar aeroportos e as linhas a existência de um vôo direto entre os dois aeroportos.

A abstração matemática desse tipo de situação dá origem ao conceito de grafo (BONDY; MURTY, 1979) assim chamado porque eles podem ser representados graficamente para facilitar a compreensão de suas propriedades (BONDY; MURTY, 2008). Convencionou-se denominar os pontos como vértices e as linhas são denominadas arestas.

Definição 2.1. *Um grafo G é denotado por um par ordenado $(V(G), E(G))$, onde $V(G)$ é um conjunto de vértices e $E(G)$ é um conjunto de conjuntos de dois vértices, chamados de arestas do grafo. São denotadas por $n = |V(G)|$ e $m = |E(G)|$ as cardinalidades dos conjuntos $V(G)$ e $E(G)$, respectivamente.*

Considere a Figura 2. Nela é apresentado o grafo G onde $V(G) = \{a, b, c, d\}$ e $E(G) = \{\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}\}$. Usualmente, denotamos uma aresta $\{x, y\} \in E(G)$ por xy .

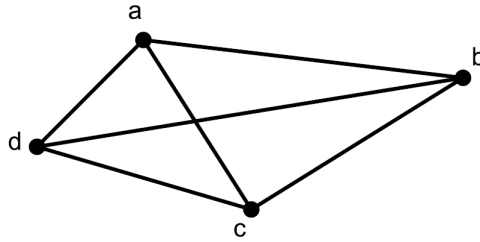


Figura 2: Grafo Não Direcionado.

Fonte: Autoria própria.

A Figura 2 ilustra um grafo não direcionado. Nesse tipo de grafo, as relações são simétricas: a aresta bc , por exemplo, equivale à aresta cb . Esse tipo de grafo é adequado para representação de relações simétricas, como co-autoria de artigos, vizinhança entre cidades ou atividades incompatíveis.

Porém, há outras situações onde as relações entre os vértices é unidirecional, ou seja a direção em que ocorre a relação é importante. Isso acontece, por exemplo, em relações de confiança, hiperlinks em páginas *web* ou citações de artigos. Grafos que representam esse tipo de situação são ditos direcionados.

Definição 2.2. Um grafo direcionado G é um par ordenado $(V(G), E(G))$, consistindo de um conjunto $V(G)$ de vértices e um conjunto $E(G)$ de pares ordenados de vértices de $V(G)$, chamados de arestas (ou arestas direcionadas). Se existe uma aresta $a = (u, v)$ então a direção da aresta é saindo de u e chegando em v .

A Figura 3 ilustra o grafo direcionado G , onde o conjunto de vértices é dado por $V(G) = \{a, b, c, d\}$ e conjunto de arestas é dado por $E(G) = \{ab, ad, bc, ca, db\}$.

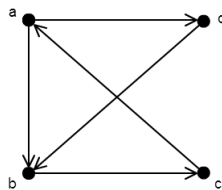


Figura 3: Grafo direcionado.

Fonte: Autoria própria.

Na Figura 3 a aresta ab indica que há uma relação de a para b , mas o inverso não é verdade. Ou seja, a relação representada pela aresta ab é unidirecional. Assim, o significado das arestas da Figura 3 é diferente do representado na Figura 2, onde as arestas representam relações bidirecionais.

Definição 2.3. Dizemos que dois vértices v e w são adjacentes (ou vizinhos) se há uma aresta $a = (v, w)$ ou $a = (w, v)$ em $E(G)$. A aresta a é dita incidente aos vértices v e w .

Ou seja, dois vértices são ditos adjacentes se são extremidades de uma mesma aresta. Assim, se $e = (u, v) \in E(G)$, os vértices u e v são adjacentes entre si (outra forma de nomear a relação de u e v é utilizando o termo vizinhos).

A quantidade de arestas incidentes a um vértice dá origem ao conceito de grau de um vértice. Além de ser fundamental para diversos resultados expressos no Capítulo 3, ele define o tamanho do conjunto de vizinhos de um vértice.

Definição 2.4. Em um grafo $G = (V(G), E(G))$, o grau de um vértice v , denotado por $d(v)$, é o número total de arestas de G , incidentes a v . O grau mínimo de um grafo G é o número $\delta(G) := \min\{d(v) : v \in V(G)\}$. Ou seja, $\delta(G)$ é igual ao menor grau dentre o conjunto de vértices de G . Já o grau máximo do grafo G é dado por $\Delta(G) := \max\{d(v) : v \in V(G)\}$. Em outras palavras, $\Delta(G)$ é dado pelo grau do vértice que possui o maior número de arestas incidentes.

No caso dos grafos direcionados, é possível identificar dois tipos de graus de vértices: grau de saída que se refere à quantidade de arestas que partem do vértice e grau de chegada que se refere à quantidade de arestas que chegam num determinado vértice.

Assim, a Tabela 1 apresenta os graus dos nós do grafo ilustrado na Figura 3.

Tabela 1: Graus dos Nós.

Aresta	Grau de saída	Grau de chegada
a	2	1
b	1	2
c	1	1
d	1	1

Outro assunto de interesse ao longo desta monografia diz respeito ao conceito de subgrafo. Bondy e Murty (1979) apresentam a seguinte definição para subgrafo:

Definição 2.5. Um grafo H é um subgrafo de G , escrito como $H \subseteq G$ se $V(H) \subseteq V(G)$ e $E(H) \subseteq E(G)$. Isso implica que um possível subgrafo de G é o próprio grafo G . Se $H \subseteq G$, mas $H \neq G$, dizemos que H é um subgrafo próprio de G e denota-se essa relação por $H \subset G$.

A Figura 4 apresenta dois exemplos de subgrafos para o grafo da Figura 2.

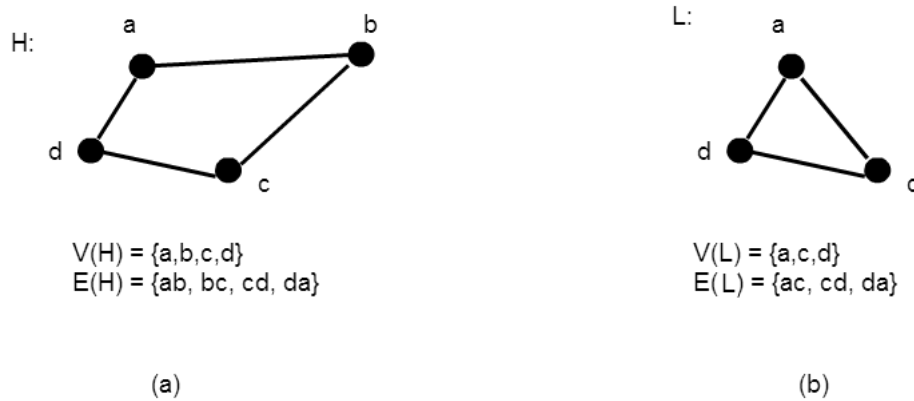


Figura 4: Exemplos de subgrafos.

Fonte: Autoria própria.

Tendo em vista que os grafos são compostos por arestas e vértices, é possível traçar caminhos entre pares de vértices de um grafo. Assim, a seguir são definidos passeio, caminho e ciclo em grafos.

Definição 2.6. *Seja $G = (V(G), E(G))$ um grafo. Um passeio $P = (v_1, \dots, v_k)$ em G é uma sequência de vértices de G tal que, para todo par consecutivo de vértices (v_{i-1}, v_i) , $2 \leq i \leq k$, temos que $v_{i-1}v_i$ é uma aresta de G . Se não houver vértices repetidos em P , dizemos que P é um caminho. Já um ciclo no grafo G é uma sequência de vértices $C = (v_1, v_2, \dots, v_k)$ se $P = (v_1, \dots, v_k)$ é um caminho em G e $v_1v_k \in E(G)$.*

Há ainda outro tipo de caminho que interessa para algumas definições utilizadas no decorrer do próximo capítulo. Trata-se do caminho Hamiltoniano.

Definição 2.7. *Um caminho que contém todos os vértices de um grafo G é dito um caminho hamiltoniano de G . Da mesma forma, um ciclo que contém todos os vértices de G é chamado de ciclo hamiltoniano. Se o grafo G contém um ciclo hamiltoniano, então ele é definido como grafo hamiltoniano.*

O conceito de caminho em grafo permite definir a distância entre vértices:

Definição 2.8. *A distância $d(u, v)$ entre dois vértices $u, v \in V$ é igual ao tamanho do menor caminho em G que começa em u e termina em v .*

Em teoria dos grafos, o conceito de conjunto independente é relevante e útil para algumas definições feitas ao longo dos próximos capítulos.

Definição 2.9. *Um subconjunto S de $V(G)$ é chamado de conjunto independente ou conjunto estável de G se $S \subseteq V(G)$ e não existem arestas de $E(G)$ entre qualquer par de elementos de S . Um conjunto independente $I(G)$ é dito máximo se todos os outros conjuntos*

independentes de G possuem cardinalidade menor ou igual à cardinalidade de $I(G)$. Por outro lado, ele é dito maximal se mais nenhum vértice puder ser adicionado ao conjunto independente sem que apareça uma aresta entre dois vértices do conjunto independente.

Seja G o grafo apresentado na Figura 5:

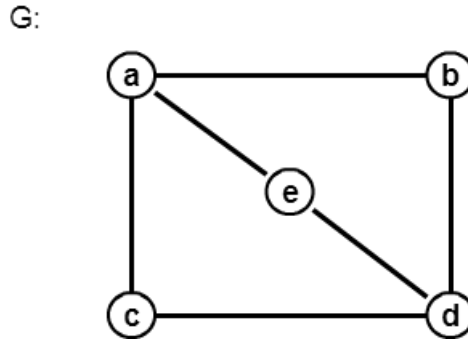


Figura 5: Grafo com cinco vértices e seis arestas.

Fonte: Autoria própria.

Os conjuntos independentes $I(G) = \{a, d\}$ e $I'(G) = \{b, c, e\}$ são maximais, pois se for incluído qualquer outro vértice de G num desses conjuntos, aparecerá uma aresta entre os vértices de $I(G)$ ou $I'(G)$. O conjunto independente $I'(G)$ é máximo, pois nenhum outro conjunto independente de G possui cardinalidade maior que $I'(G)$.

2.2 TIPOS DE GRAFOS

Nesta seção são definidos alguns tipos de grafos utilizados em diversas definições e provas de teoremas. Inicialmente apresentamos os conceitos de grafo simples, nulo e completo.

Definição 2.10. Um grafo G que não possui arestas, isto é, $E(G) = \emptyset$, é dito nulo e seus vértices são ditos isolados.

Definição 2.11. Um grafo G é dito completo, ou clique, se, para quaisquer $a, b \in V(G)$, distintos, tem-se que $ab \in E(G)$, ou seja, o grafo possui arestas entre todos os pares de vértices do grafo.

Definição 2.12. Um grafo G é denominado multigrafo se G é um par ordenado $(V(G), E(G))$ onde $V(G)$ é o conjunto de vértices e $E(G)$ é uma função $f(e_i)$ tal que $f(e_i) = \{\{u, v\} | u, v \in V(G)\}$. Se $f(a) = \{\{u, u\} | u \in V(G)\}$, dizemos que a aresta a é um laço. Por outro lado, se $f(e_1) = \{\{u, v\} | u, v \in V(G)\}$ e $f(e_2) = \{\{u, v\} | u, v \in V(G)\}$, tal que $e_1 \neq e_2$, dizemos que as arestas e_1 e e_2 são arestas múltiplas ou paralelas.

A Figura 6 apresenta um exemplo de multigrafo.

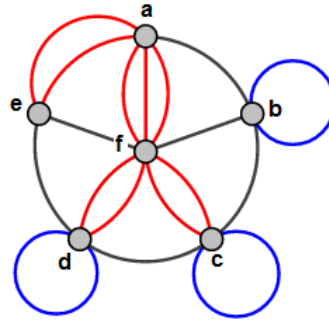


Figura 6: Exemplo de Multigrafo.

Fonte: Autoria própria.

As arestas coloridas em azul são chamadas de laços e as arestas vermelhas são ditas arestas múltiplas ou paralelas. Um grafo G que não possua arestas múltiplas ou laços é, então, chamado de grafo simples.

Exemplos das estruturas definidas acima podem ser vistas na Figura 7.

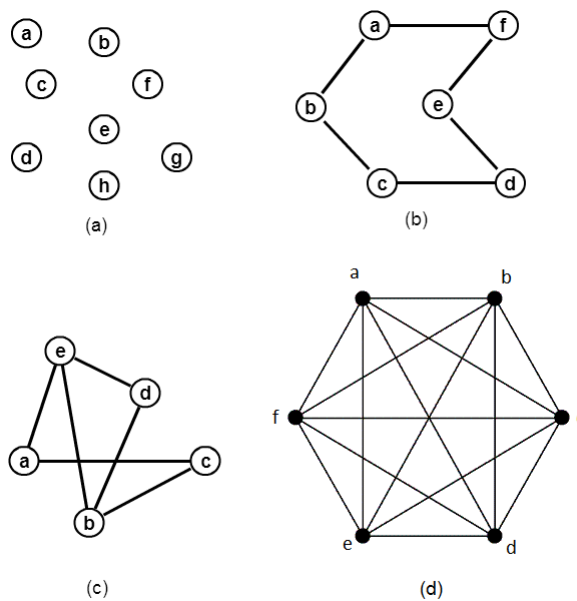


Figura 7: Tipos de Grafos.

Fonte: Autoria própria.

O grafo representado em (a) exemplifica um grafo nulo; os grafos (b) e (c) ilustram grafos simples; em (d) é apresentado um grafo completo.

Pode-se, ainda definir o conceito de grafo conexo:

Definição 2.13. Um grafo G é dito conexo se para qualquer par de vértices $v, w \in V(G)$, existe um caminho com extremos v e w . Caso isso não aconteça, o grafo é dito desconexo.

Considere o grafo G apresentado na Figura 8.

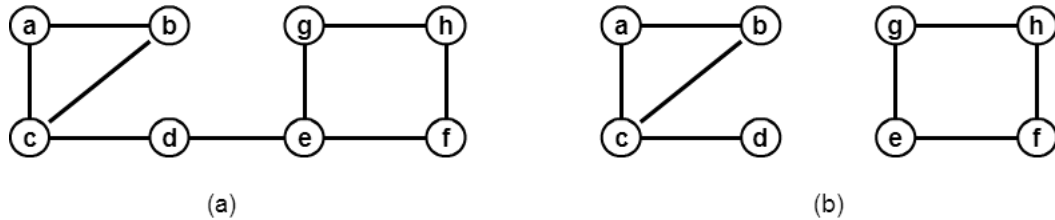


Figura 8: Exemplos de grafo conexo e desconexo.

Adaptado de Lozano (2007).

Em (a) é apresentado um grafo conexo, uma vez que há um caminho entre qualquer par de vértices. Já em (b) é ilustrado um grafo desconexo. Um grafo desconexo é formado por um conjunto de componentes conexas. Em (b), há duas componentes conexas: a primeira formada pelos vértices $\{a, b, c, d\}$ e suas respectivas arestas e a segunda formada pelos vértices $\{f, g, h, i\}$ e suas respectivas arestas.

Utilizando o conceito de caminho e tendo em vista a definição de grafo conexo, pode-se caracterizar os grafos 2-conexos. Para tanto, é necessário inicialmente definir o conceito de caminhos disjuntos. De acordo com West (2002) ,

Definição 2.14. Seja o grafo G e sejam u e v dois vértices de G . Dois caminhos de u para v são internamente disjuntos se eles não possuem vértices internos comuns aos dois caminhos. Ou seja, os únicos vértices que aparecem simultaneamente nos dois caminhos são u e v .

Seja o grafo G apresentado na Figura 9. Os caminhos $X = (1, 2, 4)$ e $Y = (1, 3, 4)$ são disjuntos. Já os caminhos $W = (1, 2, 3, 4)$ e $Z = (1, 3, 4)$ não são disjuntos, pois o vértice 3 aparece nos dois caminhos, W e Z .

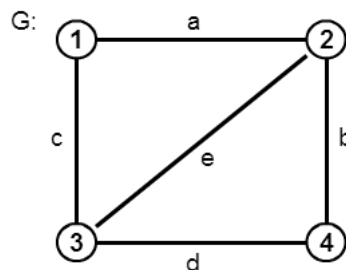


Figura 9: Exemplo de grafo 2-conexo.

Fonte: Autoria própria.

Sendo assim, um grafo 2-conexo, cujo exemplo é ilustrado na Figura 9, é, então definido como:

Definição 2.15. *Um grafo G é k -conexo, com $k \geq 1$, se existem no mínimo k caminhos disjuntos entre todos os pares de vértices de G . Assim, um grafo é dito 2-conexo se existem no mínimo 2 caminhos disjuntos para quaisquer dois vértices distintos u e v de G .*

Quando se analisam os grafos em busca de diferentes subgrafos, é possível encontrar subgrafos completos que são chamados de cliques. Isso dá origem a dois conceitos importantes definidos a seguir:

Definição 2.16. *Uma clique em um grafo G é um subconjunto G' onde $V(G') \subseteq V(G)$ e $E(G') \subseteq E(G)$ tal que para todo par $v_i, v_j \in V(G') | v_i \neq v_j$, existe uma aresta $e = \{v_i, v_j\}, \in E(G')$. Ou seja, para todo par de vértices, existe uma aresta ligando o par de vértices.*

Definição 2.17. *O número de clique de um grafo G , denotado por $\omega(G)$ é o tamanho máximo de um conjunto de pares adjacentes de vértices em G . Em outras palavras, o número de clique é o número de vértices da maior clique encontrada em G .*

De forma semelhante, o número de independência de um grafo, denotado por $\alpha(G)$, é o tamanho do maior conjunto independente de G .

Nos grafos da Figura 8, é possível identificar uma clique, formada pelos vértices $\{a, b, c\}$ e suas arestas. Portanto, o número de clique dos grafos (a) e (b) é três. Ao mesmo tempo, um conjunto independente máximo dos grafos (a) e (b) é dado pelos vértices $\{a, d, g, f\}$ e portanto, $\alpha(G) = 4$.

Abaixo é definido um tipo especial de vértice, chamado vértice de corte, cuja importância reside no fato de que se ele for removido do grafo, torna tal grafo desconexo.

Definição 2.18. *Um vértice $v \in V(G)$ é dito vértice de corte se ao ser retirado do conjunto $V(G)$ juntamente com as arestas incidentes a ele, torna o grafo G desconexo.*

No grafo G ilustrado na Figura 8 (a), o vértice d é um vértice de corte, pois se ele for removido do grafo juntamente com as suas arestas incidentes, provoca o surgimento de duas componentes conexas e torna o grafo desconexo.

Quando se abordam os conjuntos independentes de um grafo, está implícita a tentativa de subdividi-lo em conjuntos distintos de vértices. Alguns grafos podem ser divididos em dois conjuntos independentes disjuntos. Esse tipo de grafo é, então, dito bipartite. Formalmente tem-se que:

Definição 2.19. Um grafo G é dito bipartido se pode ser dividido em dois conjuntos independentes disjuntos não vazios φ e γ de tal forma que toda aresta de G possui um dos vértices em φ e o outro em γ . φ e γ , então, são chamados de conjunto partição. Se todos os vértices de φ são adjacentes a todos os vértices de γ , então G é dito bipartido completo. Nesse caso, sendo $|\varphi| = a$ e $|\gamma| = b$, denota-se o grafo G por $K_{a,b}$.

Exemplos desses dois tipos de grafos são apresentados na Figura 10. O grafo G ilustra um grafo bipartido com conjuntos partições dados por $\varphi = \{a, b, c, d\}$ e $\gamma = \{e, f, g\}$. Em (b) é apresentado o grafo H , bipartido completo. Esse grafo é denotado por $K_{3,3}$.

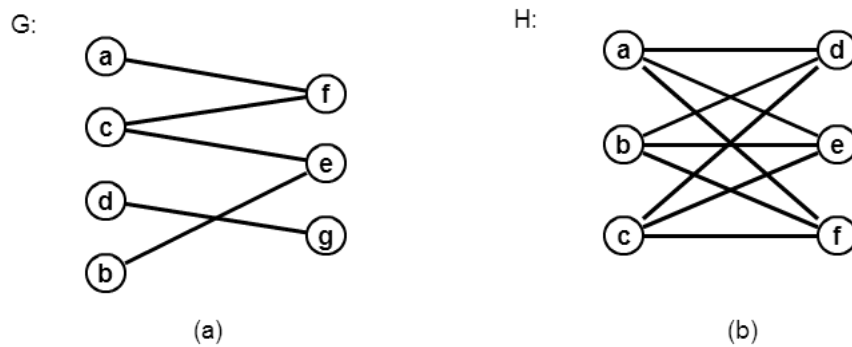


Figura 10: Exemplos de Grafo bipartido e bipartido completo.

Fonte: Autoria própria.

Definição 2.20. Um grafo G é denominado k -regular se o grau de todos os seus vértices é k , onde k é um número inteiro positivo. Se isso não ocorre, então G é dito não k -regular.

Na Figura 11 são apresentados exemplos de grafos desse tipo. (a) e (b) ilustram grafos k -regulares. O grafo G é dito 2-regular e o grafo H apresentado em (b) é 3-regular, pois todos os seus vértices possuem grau 3. Sendo assim, pode-se afirmar que todo grafo completo é também k -regular. Por fim, em (c) é ilustrado um grafo não k -regular.

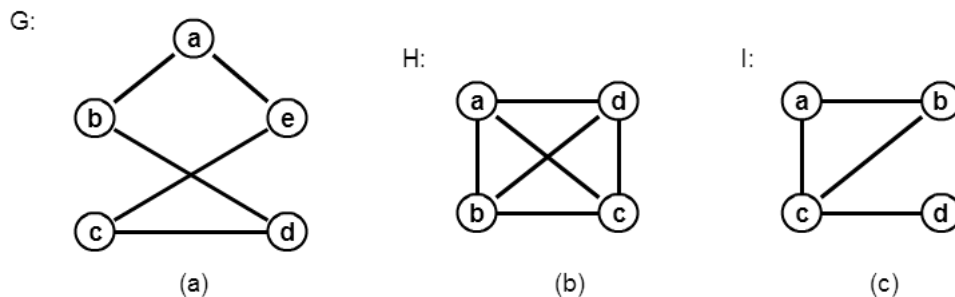


Figura 11: Os grafos (a) e (b) são k -regulares enquanto o grafo (c) é não k -regular.

Fonte: Autoria própria.

2.3 OPERAÇÕES EM GRAFOS

Grafos são formados por um conjunto de vértices e arestas. Então é possível definir algumas operações com os vértices e as arestas dos grafos. Essas operações serão utilizadas no Capítulo 3 no estudo da coloração de vértices.

Inicialmente é definida a operação de união de grafos:

Definição 2.21. A união dos grafos G_1, \dots, G_k denotada por $G_1 \cup \dots \cup G_k$, resulta num grafo H onde o conjunto de vértices é dado por

$$V(H) = \bigcup_{i=1}^k V(G_i) \quad (2.1)$$

e o conjunto de arestas é formado por

$$E(H) = \bigcup_{i=1}^k E(G_i) \quad (2.2)$$

Bondy e Murty (2008), por sua vez, definem a operação de intersecção de grafos nos seguintes termos:

Definição 2.22. A intersecção dos grafos G_1, \dots, G_k , representada por $G_1 \cap \dots \cap G_k$, resulta no grafo H onde o conjunto de vértices é dado por

$$V(H) = \bigcap_{i=1}^k V(G_i) \quad (2.3)$$

e o conjunto de arestas é formado por

$$E(H) = \bigcap_{i=1}^k E(G_i) \quad (2.4)$$

Um exemplo da aplicação das operações de união e intersecção de dois grafos é apresentado na Figura 12.

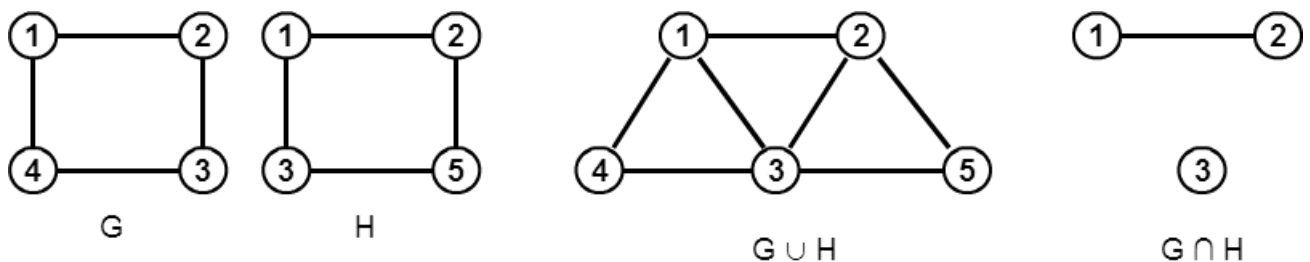


Figura 12: Operação de União e Intersecção de Grafos.
Adaptado de Bondy e Murty (2008).

É importante assinalar que o resultado da intersecção apresentará conjunto de vértices diferente de \emptyset (vazio) se houver pelo menos um vértice de mesmo nome em todos os grafos que participam da operação de intersecção. Se os nomes dos vértices forem diferentes em todos os grafos, $G_1 \cap \dots \cap G_k = H$, então $V(H) = \emptyset$ e $E(H) = \emptyset$.

Localmente, é possível realizar a remoção e contração de arestas. Essas duas operações são importantes para alguns teoremas de coloração de vértices e por isso são aqui definidas.

West (2002) define remoção de aresta da seguinte forma:

Definição 2.23. *Seja o grafo G onde o conjunto de arestas é dado por $E(G)$ e a aresta $e = \{v, w\} \in E(G)$, onde v e w são vértices de G . O resultado da remoção da aresta e é um grafo G' tal que $E(G') = E(G) - \{v, w\}$. A operação de remoção de arestas é denotada por $G - e$.*

Dessa forma, ao ser removida uma aresta de um grafo G qualquer, o grafo resultante terá o mesmo conjunto de vértices que G . Já seu conjunto de arestas será o mesmo $E(G)$ exceto pela aresta removida que será excluída do conjunto de arestas resultante, conforme pode ser verificado no exemplo apresentado na Figura 13.

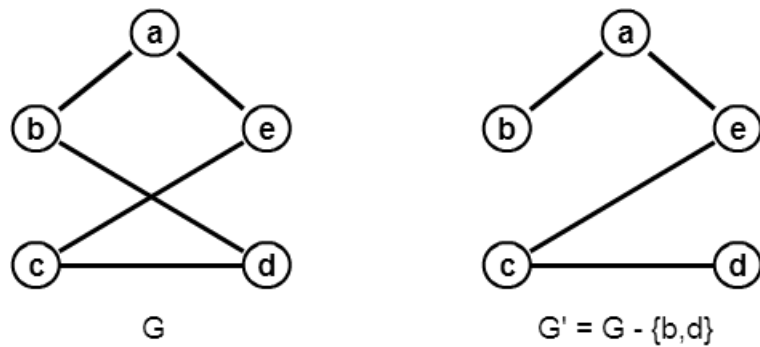


Figura 13: Representação da remoção de uma aresta do grafo G .

Fonte: Autoria própria.

Por fim, a última operação de interesse a ser definida é a contração de arestas, apresentada na próxima definição enunciada.

Definição 2.24. *Em um grafo G , contrair a aresta $e = \{u, v\}$, onde u e v são vértices de G , significa substituir u e v por um único vértice w , tal que as arestas incidentes a u e v , exceto a aresta u, v , passarão a incidir em w e as arestas paralelas e laços são removidos. A operação de contração de aresta é denotada por G/e .*

A Figura 14 (b) ilustra a contração da aresta $\{a,b\}$ do grafo G . Em (c) é apresentado o resultado da operação $G/\{b,d\}$.

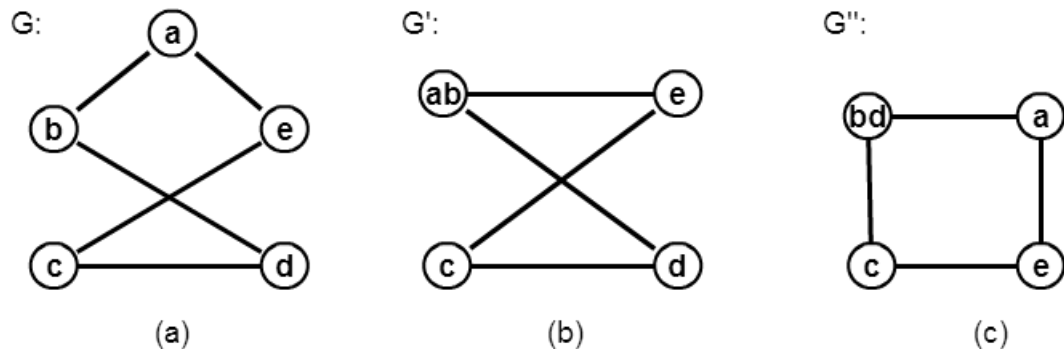


Figura 14: Representações da contração de uma aresta do grafo G .

Fonte: Autoria própria.

2.4 REPRESENTAÇÃO COMPUTACIONAL

A representação dos grafos utilizada nas seções anteriores é muito interessante visualmente, mas pouco prática do ponto de vista computacional, pois computacionalmente falando, é necessária uma forma de representação que não admita ambiguidades e possa ser facilmente manipulada por algoritmos. Duas formas importantes de representar a estrutura de um grafo de tal forma que possa ser armazenada e manipulada computacionalmente são a matriz de adjacências e a lista de adjacências. Dependendo das características do grafo e dos algoritmos a serem utilizados, uma ou outra estrutura será mais indicada (CORMEN et al., 2009).

Uma das formas padrão de representação de um grafo é pela utilização de uma matriz que codifica todas as arestas contidas no grafo que se pretende representar. Tal matriz é quadrada e cada elemento $A(i, j)$ representa o par de vértices (i, j) . Se houver uma aresta entre os vértices i e j , então o elemento $A(i, j)$ será igual a 1. Caso contrário, $A(i, j) = 0$.

Formalmente, a matriz de adjacências de um grafo $G = (V(G), E(G))$ consiste numa matriz $n \times n$, $M_{adj} = [a_{ij}]$ onde:

$$a_{ij} = \begin{cases} 1 & \text{se } (i, j) \in E(G) \\ 0 & \text{caso contrário} \end{cases}$$

Em grafos não direcionados, esta matriz é uma matriz simétrica. Se o grafo não possui laços, os elementos da diagonal principal da matriz de adjacências são iguais a 0.

Esse tipo de estrutura requer $\Theta(|V(G)|^2)$ memória, independente do número de arestas existentes em G . Portanto, ela é indicada para grafos densos (quando $|E(G)|$ é semelhante a $|V(G)|^2$ ou quando é necessário saber rapidamente se existe uma aresta entre dois vértices quaisquer (CORMEN et al., 2009). A Figura 15 apresenta um exemplo de representação do grafo G por meio de uma matriz de adjacências.

Matriz de Adjacências

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

Grafo G

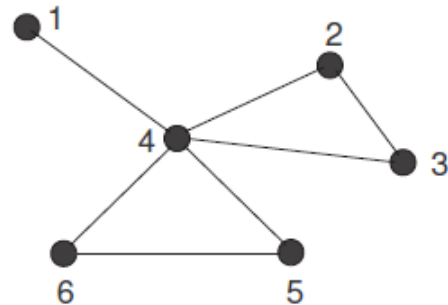


Figura 15: Representação do grafo G por meio de matriz de adjacências.

Fonte: Lozano (2007).

Já a representação do grafo $G = (V(G), E(G))$ por meio da lista de adjacências consiste num vetor com $|V(G)|$ listas, uma para cada vértice de $V(G)$. Para cada $u \in V(G)$, a lista de adjacências $Adj[u]$ contém todos os vértices v tal que $(u, v) \in E(G)$. Ou seja, $Adj[u]$ contém todos os vértices adjacentes a u (CORMEN et al., 2009). Sendo assim, a lista de adjacências referente ao grafo G da Figura 15 está representada na Figura 16 abaixo:

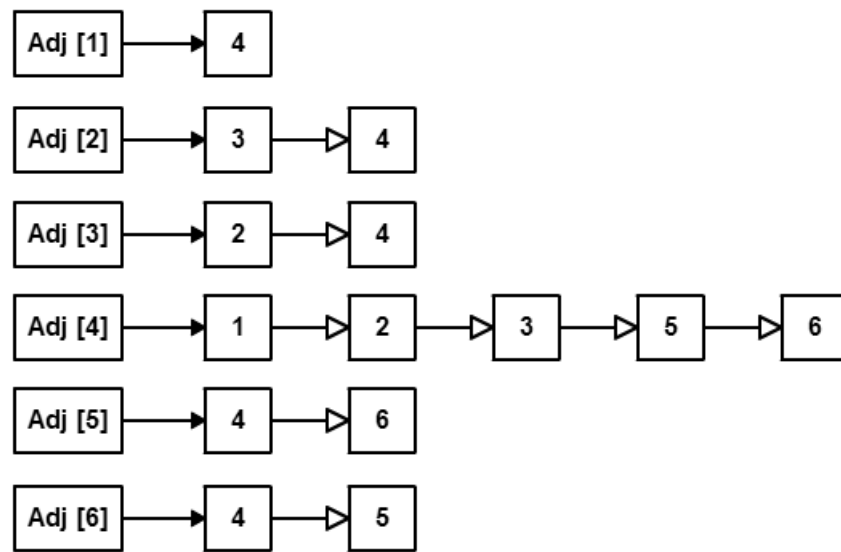


Figura 16: Representação do grafo G por meio de uma lista de adjacências.

Fonte: Autoria própria.

Esse tipo de estrutura é mais indicado para a representação de grafos esparsos (quando $|E(G)|$ é muito menor que $|V(G)|^2$), pois requer $\Theta(|V(G)| + |E(G)|)$ espaço em disco. Uma desvantagem potencial reside no fato de que para saber se uma aresta existe, é necessário fazer uma busca na lista de adjacências do vértice em questão, o que é mais lento do que o acesso direto a uma posição, como ocorre na representação por meio de matrizes.

3 COLORAÇÃO DE VÉRTICES

Este capítulo é dedicado a apresentar os conceitos relativos à coloração de vértices em grafos. Os tópicos abordados são baseados no disposto nos seguintes materiais: Costa et al. (2008), Feofiloff et al. (2011), Bondy e Murty (2008), West (2002), Lozano (2007) e Chartrand e Zhang (2009).

Uma coloração de vértices de um grafo é uma atribuição de cores aos vértices de G . Formalmente temos que,

Definição 3.1. *Uma k -coloração de um grafo G é uma partição $\{X_1, X_2, \dots, X_k\}$ de V , onde X_i denota um conjunto (possivelmente vazio) de vértices de G que receberam a cor i . Os conjuntos X_i , são, então, chamados de classes de cor.*

Dessa forma, uma coloração possível para um grafo G qualquer é a atribuição de uma cor distinta para cada vértice. Além disso, essa definição permite que um grafo qualquer seja colorido com apenas uma cor, já que a definição não faz nenhuma ponderação a respeito da forma como se dará a coloração.

Definição 3.2. *Uma k -coloração própria de vértices, ou simplesmente uma k -coloração apropriada de um grafo G é a atribuição de k cores, $\{1, 2, \dots, k\}$ para os vértices de G de tal forma que dois vértices adjacentes distintos não possuam a mesma cor. Se $\{X_1, \dots, X_k\}$ é uma coloração dos vértices de G , X_i é uma classe de cor e k é o número de cores utilizadas para colorir o grafo. Neste caso, G é dito k -colorível.*

Alternativamente, uma k -coloração apropriada pode ser vista como uma partição $\{X_1, X_2, \dots, X_k\}$ de $V(G)$ onde X_i é o conjunto (possivelmente vazio) de vértices com a cor i tal que os vértices de cada partição não são adjacentes entre si. Cada conjunto X_i é, então, chamado de classe de cor.

Na continuação deste documento, o termo k -coloração será utilizado como sinônimo de k -coloração própria apenas para simplificação.

Encontrar uma k -coloração para um grafo é uma tarefa relativamente simples quando não há restrições para o valor de k , já que nesse caso podem ser usadas tantas cores quando necessárias e, portanto, basta atribuir cores diferentes para cada um dos vértices de G . Porém, colorir o grafo G com o menor número possível de cores pode ser uma tarefa extremamente difícil. O número mínimo de cores necessárias para colorir um grafo é chamado de número cromático.

Formalmente, têm-se que

Definição 3.3. *O número cromático de um grafo G , denotado por $\chi(G)$, é o menor inteiro positivo k tal que G seja k -colorível. Quando isso ocorre, diz-se que a coloração é mínima e que o grafo G é $\chi(G)$ -cromático.*

Assim, um grafo G é 1-colorível se é um grafo nulo e é 2-colorível se e somente se G é um grafo bipartite. Para um grafo nulo, é intuitiva a coloração com apenas uma cor, uma vez que se não existem arestas, não há vértices adjacentes e todos os vértices podem, então, ser coloridos com apenas uma cor.

Com relação aos grafos bipartites, observa-se que seus vértices podem ser divididos em dois conjuntos partição onde os elementos de cada conjunto não são adjacentes entre si e, portanto, podem ser coloridos com a mesma cor. Como existem dois conjuntos partição, segue que duas cores são suficientes para colorir o conjunto de vértices do grafo.

Já um grafo completo, usualmente denotado por K_n , onde n é o número de vértices do grafo, não pode ser colorido com menos de n cores, uma vez que num grafo completo, cada vértice é adjacente a todos os outros vértices do grafo. Neste caso, $\chi(K_n) = n$.

3.1 LIMITAÇÕES DO NÚMERO CROMÁTICO

Para grafos bem conhecidos ou que possuem um número pequeno de vértices, é fácil determinar o número cromático. Porém, determinar o número cromático de um grafo qualquer geralmente é uma tarefa difícil. Assim, nesta seção serão apresentados alguns limitantes para o número cromático.

Um grafo G com n vértices é n -colorível e por consequência, $\chi(G) \leq n$. Ao mesmo tempo, $\chi(K_n) = n$. De modo semelhante, se G é um grafo e possui um subgrafo K_r , então $\chi(G) \geq r$. Isso é intuitivo já que serão utilizadas ao menos r cores para colorir K_r .

Essas duas afirmações limitam o número cromático de um grafo com base na quantidade de vértices que ele contém. A maioria dos limitantes superiores para o número

cromático se originam nos algoritmos utilizados para produzir a coloração. Um algoritmo muito simples utilizado na coloração de vértices possui a abordagem dita gulosa. Essa abordagem será utilizada posteriormente nos experimentos propostos. Bondy e Murty (2008) apresentam a heurística gulosa da seguinte forma:

HEURÍSTICA PARA COLORAÇÃO GULOSA DE VÉRTICES

INPUT: um grafo G com n vértices

OUTPUT: uma coloração de G

1. Ordene os vértices de G numa ordem linear: v_1, v_2, \dots, v_n e as possíveis cores na ordem $1, 2, \dots, n$.
 2. Pinte os vértices, um a um, seguindo a ordem, atribuindo a v_i o menor inteiro positivo ainda não atribuído a um dos seus vizinhos já coloridos.
-

Dessa forma, inicia-se a coloração atribuindo a cor 1 ao vértice v_1 e passa-se a analisar o vértice v_2 . Se v_1 e v_2 são adjacentes, atribui-se a v_2 a cor 2. Caso contrário a cor 1 será atribuída ao vértice v_2 .

Isso indica que a coloração obtida é dependente da ordem estabelecida para os vértices. Assim, se os vértices forem indexados de formas distintas, é possível obter colorações diferentes. Nesse caso, a quantidade de cores utilizadas na coloração também pode variar. De fato, sempre existe uma ordem dos vértices do grafo G que gera uma coloração gulosa ótima, mas em geral é difícil saber qual das $n!$ ordenações possíveis deve ser utilizada.

A Figura 17 ilustra o grafo G e duas colorações possíveis, em função da ordenação de vértices escolhida.

Proposição 3.1. *Para todo grafo G tem-se que $\chi(G) \leq \frac{1}{2}\sqrt{2|E(G)| + \frac{1}{4}}$ (COSTA, 2011).*

Para ver que a proposição acima é verdadeira, considere o seguinte: suponha $1, \dots, k$ uma coloração mínima para o grafo G . Então, para todo $i \neq j$, existe uma aresta com um dos extremos em X_i e o outro em X_j . Dessa forma, $|E(G)| \geq \binom{k}{2}$. Mas $\binom{k}{2} = \frac{k^2 - k}{2}$, o que dá origem à função $f(k) = k^2 - k - 2|E(G)|$. Resolvendo essa função e analisando-se seu sinal, obtém-se que $k \leq \frac{1 + \sqrt{8|E(G)| + 1}}{2}$, o que equivale ao resultado descrito na Proposição 3.1.

Outro resultado interessante, obtido a partir da abordagem gulosa, é apresentado no Teorema 3.1.

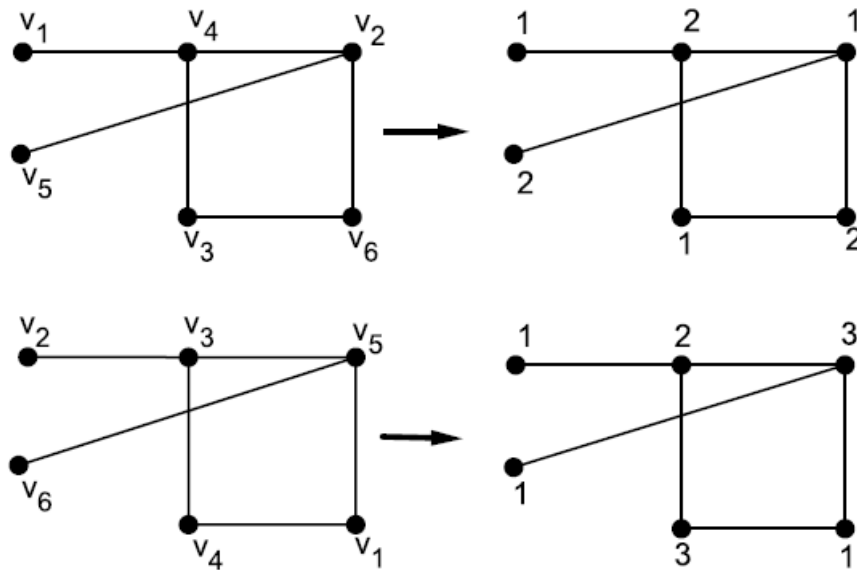


Figura 17: Duas colorações gulosas possíveis para o grafo G .
Fonte: Costa (2011).

Teorema 3.1. *Qualquer grafo G é $(\Delta + 1)$ -colorível, onde Δ é o grau máximo de G .*

Demonstração: Ordene os vértices de G de modo a dispô-los na ordem decrescente de grau. Conecte cada vértice a seus vizinhos. Dessa forma, cada vértice terá no máximo $\Delta(G)$ vizinhos anteriores na ordenação. Portanto, ao colorir tal grafo, uma coloração gulosa não necessitará mais do que $\Delta(G) + 1$ cores. Como resultado, $\chi(G) \leq \Delta(G) + 1$. \square

Esse teorema indica que, independente da ordem em que os vértices forem ordenados na abordagem gulosa, a coloração obtida nunca será maior que $\Delta + 1$. Ou seja, este é um limitante superior para o algoritmo com abordagem gulosa, pois no pior caso possível a coloração com heurística gulosa utilizará $(\Delta + 1)$ cores. Esse teorema pode ser provado de diferentes formas, mas aqui foi utilizada a prova apresentada por West (2012).

Esse limitante, $(\Delta(G) + 1)$, é apertado para cliques e grafos que são ciclos ímpares, mas pode não ser bom para outros tipos de grafos, como por exemplo, no caso do grafo $M = M_{1,20}$. Esse grafo, que é bipartite (com uma das classes de cor de cardinalidade 1 e a outra com cardinalidade 20), possui grau máximo $\Delta(M) = 20$. Pelo teorema 3.1, M é 21-colorível, o que é um resultado ruim, tendo em vista que um grafo bipartite é 2-colorível, como discutido na Seção 3.

Em 1967, Welch-Powell propôs um algoritmo para coloração de vértices com heurística gulosa, onde os vértices são coloridos na ordem decrescente dos seus graus. Essa abordagem parte do princípio de que quanto maior for o grau de um vértice, mais difícil será colorí-lo, já que por ter mais vértices adjacentes que os demais vértices, ele fica

mais restrito quanto à seleção de uma cor e, portanto, deveria ser colorido o mais cedo possível. O algoritmo em questão é descrito a seguir (LIPSCHUTZ; LIPSON, 2004):

ALGORITMO DO MAIOR PRIMEIRO PARA COLORAÇÃO DE VÉRTICES

INPUT: um grafo G com n vértices

OUTPUT: uma coloração de G

1. Ordene os vértices de G em ordem decrescente de graus: v_1, v_2, \dots, v_n tal que $d(v_1) \geq d(v_2) \geq \dots \geq d(v_n)$.
 2. Pinte os vértices, um a um, seguindo esta ordem, atribuindo a v_i o menor inteiro positivo ainda não atribuído a um dos seus vizinhos já coloridos até que todos os vértices de G estejam coloridos.
 3. Retorne a coloração obtida.
-

Esse algoritmo dá origem à Proposição 3.2:

Proposição 3.2. *Se um grafo G possui sequência de grau dos vértices dada por $d_1 \geq \dots \geq d_n$, onde d_i é o grau do vértice i , então $\chi(G) \leq 1 + \max_i \min\{d_i, i - 1\}$.*

Essa proposição trata justamente da aplicação do algoritmo guloso para uma ordenação decrescente da ordem dos vértices. Quando o i -ésimo vértice v_i é colorido, ele tem no máximo $\min\{d_i, i - 1\}$ vizinhos anteriores, de modo que no máximo $\min\{d_i, i - 1\}$ cores aparecem nos seus vizinhos. Então, quando uma cor é atribuída a v_i , no máximo $1 + \min\{d_i, i - 1\}$ cores são atribuídas à coloração. Isso vale para cada um dos vértices, então ao maximizar i obtém-se o limitante superior para a coloração dos vértices do grafo.

A Proposição 3.2 é tão boa quanto o Teorema 3.1, uma vez que no pior caso a Proposição 3.2 fornece uma coloração que é igual a $1 + \Delta(G)$.

O Teorema 3.2, conhecido como teorema Brooks, enunciado a seguir, apresenta mais um importante limitante para o problema da coloração de vértices.

Teorema 3.2. *Se G é um grafo conexo que não é um ciclo ímpar e nem um grafo completo, então $\chi(G) \leq \Delta(G)$.*

Demonstração: Seja G um grafo conexo (não ciclo ímpar ou grafo completo) e seja $k = \Delta(G)$. Assume-se que $k \geq 3$, pois se $k \leq 1$, G é o grafo K_1 ou um grafo desconexo. Se $k = 2$, como G não pode ser um ciclo ímpar, G é um ciclo par e portanto, $\Delta(G) = \chi(G) = 2$. A prova será dividida em dois casos e baseada na demonstração sugerida por Bondy e Murty (2008).

Caso 1: G não é um grafo regular.

Seja x um vértice de G com grau δ e T uma árvore de busca em profundidade de G com raiz em x . De acordo com a heurística gulosa, G será colorido com as cores $1, 2, \dots, \Delta$, sendo que a cada passo será colorida uma das folhas da subárvore de T induzida pelos vértices não coloridos de T , atribuindo a menor cor disponível e terminando com a coloração do vértice x . Quando um vértice v diferente de x é colorido, ele é adjacente a pelo menos um vértice ainda não colorido de T e é adjacente em G a no máximo $d(v) - 1 \leq \Delta - 1$ vértices já coloridos. Então é atribuído a ele uma das cores $1, 2, \dots, \Delta$. Quando x é colorido, também é possível atribuir uma das cores $1, 2, \dots, \Delta$ porque $d(x) = \delta \leq \Delta - 1$ e existe uma cor disponível que pode ser atribuída a x . Dessa forma, a heurística gulosa produz uma Δ -coloração para G .

Caso 2: G é um grafo regular. Se G possui um vértice de corte x , então $G = G_1 \cup G_2$ onde G_1 e G_2 são conectados e $G_1 \cap G_2 = \{x\}$. Como o grau de x em G_i é menor que $\Delta(G)$, nenhum subgrafo G_i é regular e então, $\chi(G_i) \leq \Delta(G_i)$, com $i = 1, 2$ e $\chi(G) = \max\{\chi(G_1), \chi(G_2)\} \leq \Delta(G)$.

Se cada árvore de busca em profundidade de G é um caminho hamiltoniano com raiz em uma das suas pontas, então G é um ciclo, um grafo completo ou um grafo bipartido completo. Como, por hipótese, G não é um ciclo ímpar e nem um grafo completo, então $\chi(G) = 2 \leq \Delta(G)$.

Se G não tem vértice de corte (G é 2-conexo). Suponha que T é uma árvore de busca em profundidade de G que não é um caminho. Seja x um vértice de T com pelo menos dois filhos y e z . Como G é 2-conexo, $G - y$ e $G - z$ são conexos. Então, y e z são ambos folhas de T ou possuem descendentes próprios que são adjacentes a ancestrais de x . Então, $G' = G - \{y, z\}$ é conexo.

Seja T' uma árvore de busca em profundidade em G' com raiz x . Atribui-se a cor 1 aos vértices y e z (y e z não são adjacentes pois são filhos de um mesmo vértice). Colore-se T' com a heurística gulosa. Por ter sido atribuída a cor 1 à y e z (vizinhos de x), quando o vértice x for colorido haverá uma cor disponível para atribuir à x e, portanto, será obtida uma Δ -coloração para G . \square

O teorema de Brooks não é limitante interessante para todos os casos. Suponha o grafo bipartite $K_{1,15}$. Pelo teorema de Brooks, este grafo é 15-colorível, o que é um péssimo limitante, pois por se tratar de um grafo bipartite, ele pode ser colorido com apenas duas cores.

Finalmente, é possível limitar inferiormente a coloração pelo disposto na seguinte

proposição:

Proposição 3.3. *Para todo grafo G , tem-se que $\chi(G) \geq \omega(G)$ e $\chi(G) \geq \frac{n(G)}{\alpha(G)}$*

Relembrando que:

$\chi(G)$ é o número cromático de G ;

$n(G)$ é a cardinalidade do conjunto $V(G)$;

$\omega(G)$ é o número de clique de G ;

$\alpha(G)$ é o número de independência de G .

Essa proposição traz duas afirmações. A primeira, $\chi(G) \geq \omega(G)$, é verdadeira porque a coloração de um clique exige a utilização de cores distintas para todos os seus vértices, uma vez que há arestas interligando todos os vértices pertencentes à clique.

Já a segunda afirmação, $\chi(G) \geq \frac{n(G)}{\alpha(G)}$, é justificada porque cada classe de cor é um conjunto independente de G . Se existem $n(G)$ vértices em G e a maior classe de cor de G possui $\alpha(G)$ cores, então, todas as classes de cores possuem no máximo $\alpha(G)$ elementos e, portanto, para colorir todo o grafo são necessárias ao menos $\frac{n(G)}{\alpha(G)}$ cores.

3.2 POLINÔMIO CROMÁTICO

Quando se estuda a coloração de grafos, é importante saber não só seu número cromático, mas também a quantidade de diferentes colorações possíveis com determinadas cores. Essa informação é proporcionada pelo polinômio cromático, cujo conceito foi introduzido em 1912 por Birkhoff (1912) para atacar a conjectura das quatro cores (LOZANO, 2007; BONDY; MURTY, 2008), já citada na introdução deste documento.

A ideia central do polinômio cromático é determinar o número cromático de um grafo qualquer e o número de colorações associadas. Formalmente, podemos defini-lo como:

Definição 3.4. *O polinômio cromático, $P(G,k)$, é uma função que conta o número de colorações do grafo G com k cores.*

Essa definição dá origem à seguinte proposição:

Proposição 3.4. *O menor k que satisfaz $P(G,k) > 0$ é o número cromático de G .*

Esse resultado é válido em função do significado da Definição 3.3. Dela deriva que o número cromático de um grafo é o menor número de cores necessárias para se obter

uma coloração própria para G . $P(G, k)$ conta o número de colorações distintas do grafo G com a utilização de k cores. Logo, se $k < \chi(G)$, então $P(G, k) = 0$. Assim, $P(G, k) > 0$ se $k \geq \chi(G)$. Se k é minimizado para $P(G, k) > 0$, obtém-se o número cromático de G .

Para alguns grafos é fácil determinar o polinômio cromático apenas com a utilização dos princípios básicos de contagem. A seguir são apresentados alguns exemplos de obtenção do polinômio cromático. Para tanto, considere um conjunto de cores W com k cores distintas que podem ser utilizadas na coloração.

- Grafo nulo: Como não existem arestas no grafo nulo, seus vértices podem ser coloridos com a utilização de qualquer combinação entre as k cores do conjunto W . Assim, escolhendo-se um vértice ele pode ser colorido com qualquer uma das k cores. Isso vale para todos os vértices de G . Dessa forma, o polinômio cromático de um grafo nulo é dado por:

$$P(G, k) = k^n \quad (3.1)$$

onde n é o número de vértices de G .

- Grafo completo: Se G é um grafo completo, existem k opções para colorir o primeiro vértice, $k - 1$ escolhas para o segundo vértice, $k - 2$ cores disponíveis para o terceiro vértice e assim por diante. Assim, o polinômio cromático de um grafo completo, de acordo com Bondy e Murty (2008), é dado por:

$$P(G, k) = k(k - 1)(k - 2)(k - 3) \dots (k - n + 1) \quad (3.2)$$

onde n é o número de vértices de G .

- Árvore: Ao colorir o primeiro vértice, v_i de uma árvore há k opções de escolha. Os vértices adjacentes a v_i podem ser coloridos com qualquer uma das $k - 1$ cores de W . Além disso, restam $k - 1$ cores possíveis para colorir os vértices que não são adjacentes a v_i , mas que são adjacentes aos demais vértices já coloridos, pois a cor utilizada para colorir v_i pode ser usada novamente. Dessa forma, o polinômio cromático de uma árvore é dado por:

$$P(G, k) = k(k - 1)^{n-1} \quad (3.3)$$

onde n é o número de vértices de G .

No entanto, se G é um grafo qualquer, não é fácil determinar a função $P(G, k)$. O teorema da remoção-contracção, enunciado a seguir, ilustra uma forma de obter recursivamente o polinômio cromático $P(G, k)$ para um grafo G :

Teorema 3.3. *Se G é um grafo simples e $e = (u, v)$ é uma aresta qualquer de G , então $P(G, k)$ satisfaz a seguinte fórmula recursiva:*

$$P(G, k) = P(G - e, k) - P(G/e, k) \quad (3.4)$$

Para demonstração deste teorema serão analisados dois casos, de acordo com o proposto por Lozano (2007) e Bondy e Murty (2008):

Caso 1: Em todas as colorações de $G - e$, os vértices u e v possuem cores diferentes.

Neste caso, é possível adicionar a aresta e ao grafo $G - e$ sem que haja alteração no número de colorações distintas com k cores para G . Então, G e $G - e$ têm o mesmo número de k -colorações.

Caso 2: Em todas as colorações de $G - e$, os vértices u e v possuem a mesma cor.

Se u e v possuem a mesma cor, é possível contrair esses vértices no grafo $G - e$ e o número de colorações com k cores não se altera neste grafo. Logo, o número de k -colorações de $G - e$ e G/e é igual.

Como os casos 1 e 2 são disjuntos, tem-se que:

$$P(G - e, k) = P(G, k) + P(G/e, k) \quad (3.5)$$

Logo,

$$P(G, k) = P(G - e, k) - P(G/e, k)$$

□ (3.6)

A Figura 18 apresenta uma ilustração da validade do Teorema 3.3. Para tanto, é considerado o grafo G e um conjunto com k cores disponíveis para a coloração.

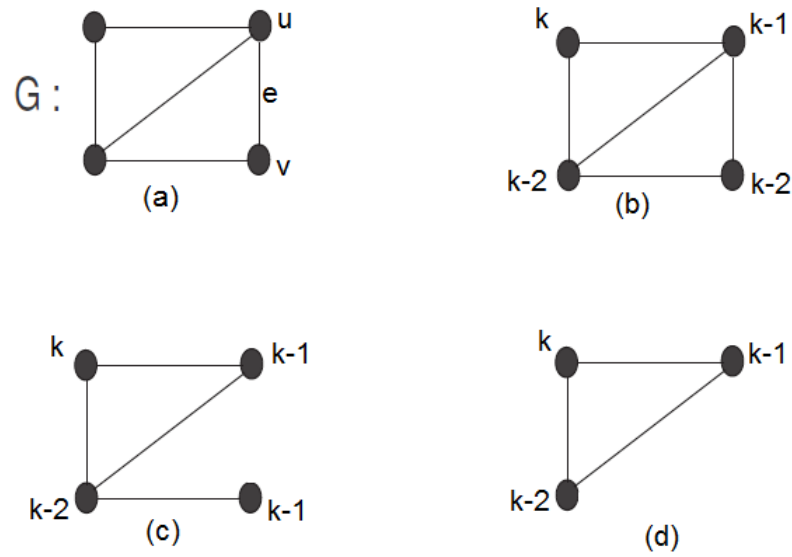


Figura 18: Exemplificação do Teorema 3.3.

Fonte: Autoria própria.

Em (a) está representado o grafo G que será utilizado na representação do Teorema 3.3. Em (b) está assinalada a quantidade de cores disponíveis para a coloração de cada vértice de G . O polinômio cromático associado ao grafo G é dado por:

$$P(G, k) = k(k-1)(k-2)^2. \quad (3.7)$$

Na figura (c) está representada a quantidade de cores para a coloração após a remoção da aresta e . O polinômio cromático associado a esse grafo é então:

$$P(G-e, k) = k(k-1)^2(k-2). \quad (3.8)$$

Por fim, a figura (d) representa o número de cores disponíveis para a coloração do grafo G/e . Nesse grafo foi efetuada a contração da aresta e . O polinômio cromático do grafo da figura (d) é dado por:

$$P(G/e, k) = k(k-1)(k-2). \quad (3.9)$$

Substituindo as Equações 3.8 e 3.9 no enunciado pelo teorema 3.3 temos que:

$$\begin{aligned} P(G, k) &= k(k-1)^2(k-2) - k(k-1)(k-2) \\ P(G, k) &= (k-1)[k(k-1)(k-2)] - k(k-1)(k-2) \\ P(G, k) &= (k-2)[k(k-1)(k-2)] \\ P(G, k) &= k(k-1)(k-2)^2. \end{aligned} \quad (3.10)$$

Que é igual ao resultado obtido na Equação 3.7, mostrando assim a validade do teorema para o grafo G . Claramente, $P(G, k)$ é um polinômio de grau 4. O Teorema 3.4 apresenta um resultado mais geral para a caracterização de $P(G, k)$ para grafos simples.

Teorema 3.4. $P(G, k)$ para um grafo simples G é um polinômio .

Para demonstrar a validade deste teorema, considere G um grafo simples com n vértices e m arestas. Com base no Teorema 3.3 é possível obter dois grafos, $G - e$ e G/e , cada um com no máximo $m - 1$ arestas, pois no processo de contração, arestas paralelas e laços são removidos. Repetindo o mesmo processo para os dois grafos obtidos, serão gerados 4 grafos com no máximo $m - 2$ arestas. Continuando este processo será obtido um grafo nulo. Um grafo nulo possui $P(G, k)$ polinomial e portanto, um grafo qualquer também possuirá $P(G, k)$ polinomial, já que o processo de contração e remoção de arestas pode ser repetido até que reste apenas um grafo nulo. Neste caso, o polinômio cromático do grafo G será a soma dos polinômios cromáticos dos grafos nulos.

Dessa forma, pode-se perceber que se G é um grafo simples com n vértices, então $P(G, k)$ tem grau n . Um exemplo da utilização deste teorema é apresentado na Figura 19. Nela é mostrado passo-a-passo a aplicação do Teorema 3.3. No último passo da remoção/contração, há somente grafos nulos, o que demonstra a validade do último teorema.

$$\begin{aligned}
 P(G, k) &= \begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} = \begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} - \begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} = \left(\begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} - \begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} \right) - \left(\begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} - \begin{array}{c} \circ \\ | \\ \circ - \circ \end{array} \right) \\
 &= \left(\begin{array}{c} \circ \\ \circ \\ \circ \end{array} \right) - 3 \left(\begin{array}{c} \circ \\ \circ \\ \circ \end{array} \right) + 3 \left(\begin{array}{c} \circ \\ \circ \end{array} \right) - \left(\begin{array}{c} \circ \end{array} \right) = k^4 - 3k^3 + 3k^2 - k = k(k-1)^3
 \end{aligned}$$

Figura 19: Exemplo da aplicação do Teorema 3.3.

Fonte: Bondy e Murty (2008)

A seguir são enunciados outros resultados que garantem que $P(G, k)$ é um polinômio.

Teorema 3.5. O polinômio $P(G, k)$ de qualquer grafo com n vértices é um polinômio com coeficientes inteiros, sendo o maior termo k^n , termo constante igual a zero e os demais termos alternando de sinal.

Este teorema pode ser provado por indução na cardinalidade do conjunto $E(G)$, denominado m . A prova é baseada no proposto por Bondy e Murty (2008). Para tanto, considere G um grafo com n vértices e m arestas.

i) CASO BASE: Se $m = 0$, então o grafo é nulo e $P(G, k) = k^n$ e esse polinômio satisfaz trivialmente as condições expressas no teorema.

i) HIPÓTESE DE INDUÇÃO: Suponha que o teorema é válido para todo o grafo com $m - 1$ arestas, onde $m \geq 1$. Será provado que o teorema é, também, válido para todo grafo com m arestas.

Pelo teorema 3.3 temos que $P(G, k) = P(G - e, k) - P(G/e, k)$. $G - e$ e G/e possuem menos que m arestas, então pela hipótese de indução, existem inteiros não negativos a_i e b_i , onde $i = 1, \dots, n - 1$, tais que:

$$P(G - e, k) = \sum_{i=1}^{n-1} (-1)^{n-1} a_i k^i + k^n \quad (3.11)$$

e

$$P(G/e, k) = \sum_{i=1}^{n-1} (-1)^{n-1-i} b_i k^i + k^{n-1} \quad (3.12)$$

Como $P(G, k) = P(G - e, k) - P(G/e, k)$, temos que:

$$P(G, k) = \sum_{i=1}^{n-1} (-1)^{n-1} a_i k^i + k^n - \sum_{i=1}^{n-1} (-1)^{n-1-i} b_i k^i + k^{n-1} \quad (3.13)$$

$$P(G, k) = \sum_{i=1}^{n-1} (-1)^{n-i} (a_i + b_i) k^i + k^n \quad (3.14)$$

Então, $P(G, k)$ satisfaz as condições expressas no teorema. \square

Os teoremas apresentados nesta seção mostram que é possível determinar uma função que indica a quantidade de colorações possíveis para um dado grafo, considerando um número k de cores disponíveis. Pelo polinômio cromático é possível determinar o número cromático de um grafo. Porém, determinar essa função, pode ser tão ou mais difícil que encontrar o número cromático do grafo através de uma busca exaustiva.

3.3 CARÁTER NP-DIFÍCIL

O problema de coloração de vértices foi provado NP-Difícil por Karp (1972) em 1972. Os resultados da prova apresentada por Karp implicam que colorir um grafo 3-colorível com três cores é NP-Difícil e o mesmo pode ser estendido para grafos k -coloríveis (KHANNA et al., 1993).

Ocorre que para provarmos que um problema pertence à classe dos problemas NP-Difícil, precisamos primeiramente provar que sua versão de decisão pertence à classe de problemas NP-Completo. Assim, esta seção dedica-se a apresentar a prova de NP-Completude do problema de coloração de vértices. A prova aqui apresentada é baseada no descrito por Cormen et al. (2009), onde detalhes adicionais podem ser encontrados.

Um dos problemas de coloração de vértices mais simples diz respeito a verificar se um dado grafo é 2-colorível, ou seja, decidir se ele pode ser colorido com apenas duas cores. Esse problema pode ser resolvido em tempo polinomial e por isso pertence à classe dos problemas P.

Para checar se um grafo possui coloração própria com duas cores basta fazer uma busca em largura no grafo. Inicialmente atribui-se a cor a a um dos nós (nó raiz) e então executa-se a busca em largura, colorindo os nós vizinhos ao nó raiz com a cor b . Na sequência, visita-se os nós vizinhos aos coloridos no nível anterior, e tenta-se colorir os nós com a cor a , seguindo o processo até que não restem mais nós a serem coloridos ou até que um dos nós não possa ser colorido nem com a cor a nem com a cor b . Se todos os nós puderem ser coloridos com as cores a ou b , então o grafo em questão é 2-colorível. Caso contrário, duas cores não são suficientes para colorir o grafo e ele é não 2-colorível.

A versão de decisão para o problema de coloração de vértices pode ser enunciado da seguinte forma: Dado um grafo $G = (V, E)$, existe uma coloração de vértices que utiliza no máximo k cores? Claramente esse é um problema de decisão cuja resposta será sim ou não. Responder a esse problema de decisão claramente exige tempo computacional exponencial, uma vez que potencialmente todas as possíveis colorações com até k cores precisam ser testadas para que se possa obter uma resposta ao problema enunciado. Por outro lado, esse problema pode ser resolvido em tempo polinomial se e somente se a sua versão de otimização puder ser resolvida em tempo polinomial, uma vez que a versão de decisão é um caso particular da versão de otimização.

Sendo assim, para provar que o problema de otimização de coloração de vértices pertence à classe dos problemas NP-Difícil, é necessário provar que uma versão do problema de decisão pertence à classe dos problemas NP-Completo. Para tanto, considerar-se-á o problema da 3-coloração de grafos 3-coloríveis.

Para provar que o problema de 3-coloração é NP-Completo, é necessário reduzir uma instância de um problema provadamente NP-Completo para uma instância do problema de 3-coloração de vértices. Sendo assim, será feita a redução do problema 3-CNF-SAT para o problema de 3-coloração.

Dada uma fórmula ϕ , instância do problema 3-CNF-SAT, com m cláusulas definidas em n variáveis x_1, x_2, \dots, x_n constrói-se um grafo $G = (V, E)$ da seguinte forma:

1. Crie um vértice para cada cada variável presente na fórmula ϕ
2. Crie um vértice para a negação de cada variável presente em ϕ
3. Crie três vértices especiais: *BLUE*, *TRUE*, *FALSE*. Conecte-os em um triângulo. Conecte cada vértice representando as variáveis de ϕ ou as suas negações ao vértice especial *BLUE*
4. Conecte cada vértice que representa uma das variáveis de ϕ ao vértice que representa a negação de tal variável
5. Para cada cláusula da fórmula ϕ , construa o gadget da Figura 20

Suponha que a cláusula para quem se busca criar o gadget seja $(x_1 \vee \neg x_2 \vee x_3)$. O gadget terá o seguinte formato:

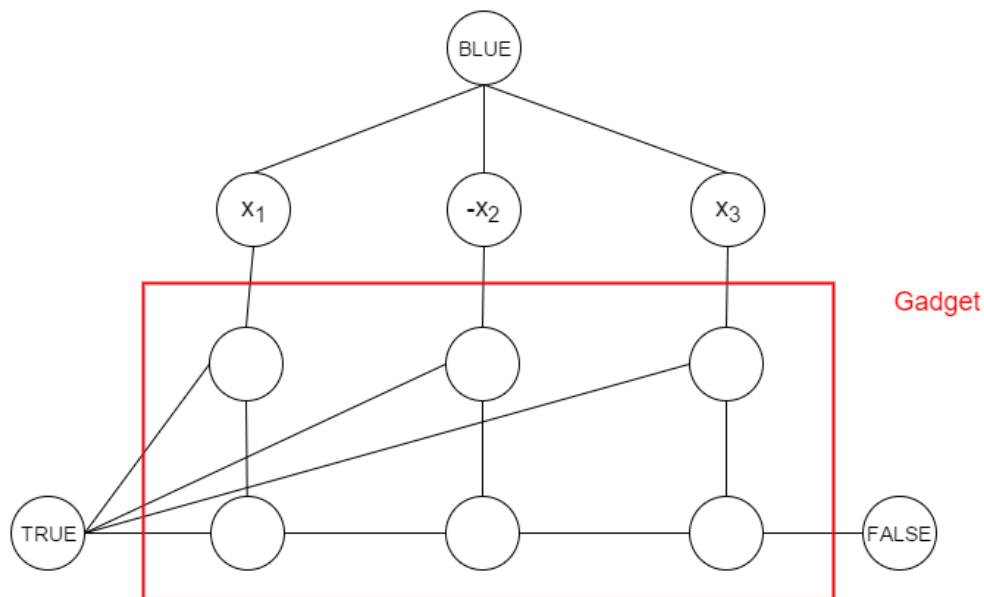


Figura 20: Gadget para redução de 3-CNF-SAT para 3-coloração - Para a cláusula $(x_1 \vee \neg x_2 \vee x_3)$.

Fonte: Autoria própria.

O triângulo formado pelo nós *TRUE*, *FALSE*, *BLUE* e a conexão dos vértices correspondentes aos literais presentes em ϕ e suas respectivas negações garante que cada literal é verdadeiro ou falso, não pode ser ambos ao mesmo tempo. A ligação entre os vértices literais presentes em ϕ com suas negações, por outro lado, garante que os valores

que ambos irão assumir serão opostos. Por fim, o gadget relativo a cada cláusula de ϕ garante que pelo menos um literal em cada cláusula será verdadeiro.

Na sequência, atribui-se cores aos nós da seguinte forma:

1. Pinte o vértice especial *BLUE* com a cor azul.
2. Pinte o vértice especial *TRUE* com a cor verde.
3. Pinte o vértice especial *FALSE* com a cor vermelha.
4. Pinte os vértices representando literais verdadeiros (x_n) com a cor verde e os vértices referentes à literais falsos ($\neg x_n$) com a cor vermelha.
5. Escolha um dos vértices que representa literal verdadeiro. Pinte o vértice abaixo do vértice escolhido com a cor vermelha e o mais abaixo com a cor azul.
6. Em cada um dos gadgets criados atribua a cor vermelha aos vértices abaixo de um vértice colorido com verde. Ao vértice abaixo deste atribua a cor azul.
7. Atribua aos vértices do meio do gadget a cor azul.
8. Pinte os demais vértices conforme de acordo com o induzido pelas cores atribuídas aos vértices vizinhos.

Se tal grafo puder ser colorido utilizando essas regras, então o grafo gerado com a redução é 3-colorível e conseqüentemente a fórmula ϕ é satisfazível. Caso contrário, o grafo gerado não é 3-colorível e a instância de 3-CNF-SAT que o originou é não satisfazível.

Suponha que a cláusula para quem se busca criar o gadget seja dada por $(x_1 \vee \neg x_2 \vee x_3)$. A coloração resultante para o gadget dessa cláusula é apresentada na Figura 21.

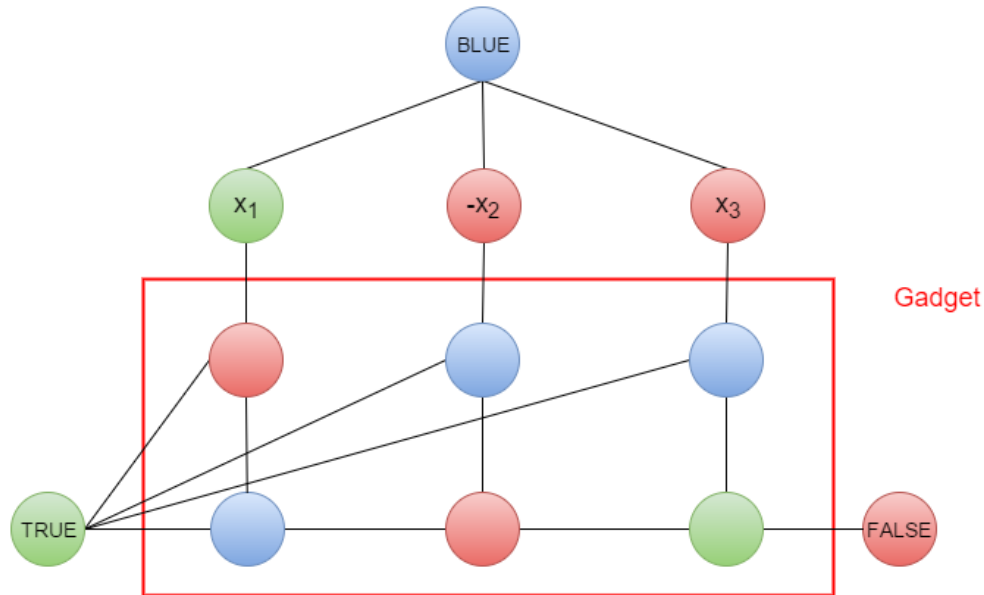


Figura 21: Gadget colorido utilizado na redução da cláusula $(x_1 \vee \neg x_2 \vee x_3)$ para instância de 3-coloração.

Fonte: Autoria própria.

A redução de uma instância de 3-CNF-SAT para o problema da coloração de vértices é feita claramente em tempo polinomial e como existe uma nítida dependência entre a solução da instância do problema 3-CNF-SAT e a solução do problema de coloração pode-se afirmar que 3-coloração pertence à classe dos problemas NP-Completo.

O problema da 3-coloração é um caso específico do problema da k -coloração. Como ao menos um caso de k -coloração (para $k=3$) foi provado NP-Completo, k -coloração pertence, por consequência, ao conjunto dos problemas NP-Completo. Finalmente, como a versão de decisão do problema de k -coloração pertence à classe dos problemas NP-Completo, sua versão de otimização pertence à classe dos problemas NP-Difícil.

Tendo em vista que o problema da coloração de vértices é NP-Difícil, várias tentativas tem sido feitas nas últimas décadas na busca de algoritmos polinomiais que garantam algum tipo de aproximação para o problema. Em 1974 Jhonson (1974) propôs um algoritmo polinomial com garantia de aproximação da ordem de $O(\frac{n}{\log n})$ onde n é o número de vértices do grafo. Pode-se perceber que essa aproximação depende do tamanho da instância, piorando sua garantia de aproximação à medida que aumenta o tamanho da instância a ser analisada.

Depois dessa proposta, demorou oito anos para que houvesse algum tipo de aperfeiçoamento dessa aproximação. Isso só ocorreu em 1983, quando Wigderson (1983) publicou artigo apresentando um algoritmo de custo polinomial com garantia de aproximação

igual a $O(\frac{n(\log \log n)^2}{(\log n)^2})$ para o problema da coloração de vértices.

Em 1990, Berger e Rompel (1990) apresentaram um algoritmo com custo computacional dado por $O(\frac{n(\log \log n)^3}{(\log n)^3})$, uma melhoria de $\log \log n$ em relação ao algoritmo proposto por Wigderson. Por fim, em 1993, Hallórsson (1993) propôs um novo algoritmo eficiente com garantia de aproximação igual a $O(\frac{n(\log \log n)^2}{\log^3 n})$. Este é o mais recente resultado para k -coloração, sendo possível encontrar outras aproximações para problemas específicos como a 3-coloração de vértices.

4 REDES COMPLEXAS

Neste capítulo é apresentada uma introdução à teoria de redes complexas, abordando alguns problemas reais que podem ser modelados e estudados sob esse enfoque, bem como características estruturais importantes, métricas utilizadas no seu estudo e modelos de crescimento que buscam criar redes que se assemelhem a redes reais. Todavia, não se pretende esgotar o assunto, porque isso fugiria ao escopo deste trabalho, mas criar uma referência rápida sobre diversos conceitos que serão úteis e importantes nas discussões que seguem.

Suponha uma fonte qualquer despejando continuamente areia sobre uma superfície plana. Inicialmente, observa-se que a pilha de areia é baixa e ocorrem pequenos deslizamentos à medida que seu tamanho aumenta. Conforme a pilha aumenta de tamanho deslizamentos mais intensos vão ocorrendo, atingindo toda a pilha, até que ela atinja uma certa organização e se torne estacionária. Esse é um exemplo típico de um sistema complexo, pois embora se conheçam bem as características estruturais dos grãos de areia (massa, formato), não é possível modelar o comportamento global ou prever quando e onde os deslizamentos vão ocorrer e nem a sua intensidade, apenas pela decomposição e análise das suas partes, já que os deslizamentos são decorrentes da interação dos grãos de areia com seus vizinhos (MELLO, 2010). O mesmo ocorre ao se observar o comportamento de pessoas de uma grande metrópole num horário de grande movimento. Embora se conheça bem a massa e o formato de um ser humano, não é possível prever o comportamento de todo o grupo apenas por conhecer dados estruturais dos seus componentes.

Essa dificuldade em inferir o comportamento do sistema como um todo a partir da decomposição e compreensão das partes componentes é uma das características das redes complexas e mostra que a conexão e interação das partes influencia fundamentalmente o comportamento individual e global do sistema. Disso surge a necessidade de se compreender como as partes se conectam e a importância dessas conexões para o problema que se quer estudar (FIGUEIREDO, 2011).

De acordo com Mello (2010), é difícil estabelecer uma definição suficientemente

ampla sobre o que seja um sistema complexo, mas em linhas gerais, pode-se afirmar que um sistema complexo é um sistema com um grande número de elementos capazes de interagir entre si e com o ambiente, de tal forma que pode ser identificado algum tipo de evolução nele.

Os sistemas complexos normalmente possuem um grande número de elementos que agem de acordo com regras que podem não ser muito bem compreendidas, ou que podem mudar ao longo do tempo. Suponha um bando de gansos em processo migratório e um *Boeing* 747-400. Num primeiro momento, pode-se acreditar que o *Boeing*, por possuir 3×10^6 peças, é um sistema mais complexo que o bando de gansos. Porém, o avião é composto por peças que precisam trabalhar em conjunto, mas com papéis bem definidos e regras bem determinadas. Por outro lado, o bando de gansos é um sistema adaptável que interage com o meio ambiente e tem a capacidade de se auto-organizar e por isso, o primeiro sistema é apenas um sistema complicado, enquanto o segundo é um sistema complexo (AMARAL; OTTINO, 2004).

Isso mostra que a quantidade de elementos presentes no sistema não é fundamental para determinar se um sistema é dito complexo ou não, pois o todo é muito mais que a soma das suas partes (AMARAL; OTTINO, 2004; MELLO, 2010).

Dessa forma, a principal característica dos sistemas complexos é a sua capacidade de exibir organização sem que qualquer princípio de organização externa seja aplicado, ou seja, sua auto-organização e emergência de comportamentos (AMARAL; OTTINO, 2004). Além disso, há um caráter de universalidade, uma vez que sistemas pertencentes à uma mesma classe em geral possuem propriedades semelhantes (RODRIGUES, 2007).

Atualmente, as ferramentas utilizadas para o estudo de sistemas complexos são suportadas por três áreas principais: dinâmica não-linear, física estatística e teoria de redes. A dinâmica não-linear, que tem como caso particular a teoria do caos, busca explicar como sistemas simples são capazes de produzir resultados complexos e seu mais famoso exemplo é a equação logística ¹. Essa técnica possui aplicações em várias áreas do conhecimento, como geofísica, fisiologia, neurofisiologia, engenharia e física (AMARAL; OTTINO, 2004). Já a física estatística, ao lado da relatividade e mecânica quântica, foi responsável por grandes avanços conceituais e técnicos no estudo da física no início do século XX, por trazer uma nova concepção para o desenvolvimento de previsões e introduzir o conceito de modelos discretos (AMARAL; OTTINO, 2004).

Finalmente, a teoria de redes complexas, a mais recente dentre as citadas, é de

¹Maiores informações sobre a equação logística podem ser encontrados em Mello (2010).

suma importância para a compreensão de sistemas complexos. Um dos grandes desafios dessa área é o estudo da natureza da topologia dos sistemas complexos, uma vez que as interações frequentemente ocorrem de forma não-linear e são sujeitas a influências externas, de tal forma que a rede de interações pode ser desconhecida na maior parte do tempo (MELLO, 2010).

A modelagem de sistemas complexos por meio de redes complexas tem ganho relevância devido à capacidade de processamento dos computadores modernos, o que permitiu a percepção de características não triviais comuns aos mais diferentes tipos de redes. Ao mesmo tempo, a disponibilização de novos bancos de dados a respeito da estrutura de redes reais em diversas áreas do conhecimento fornece uma base empírica nunca antes disponível tanto em termos de precisão quanto em quantidade de dados, o que favorece o aumento no interesse em problemas envolvendo redes complexas (FERNANDES, 2012; MELLO, 2010).

Uma rede é uma abstração que permite codificar algum tipo de relacionamento entre pares de objetos (FIGUEIREDO, 2011). Assim, redes podem ser formadas por qualquer tipo de elemento, como por exemplo moléculas, neurônios, indivíduos, agentes financeiros, páginas *Web*, computadores ou cidades. Entre cada par de elementos, pode ou não existir algum tipo de relacionamento. Dessa forma, podem ser definidos diferentes tipos de relação, mesmo sobre um mesmo conjunto de elementos. Por exemplo, sobre um conjunto de indivíduos podem ser definidas relações como contato sexual, confiança, amizade ou parentesco.

Por esta razão, redes complexas possuem um caráter multidisciplinar, uma vez que podem ser estudadas sob a ótica de diversas áreas do conhecimento. Ela se utiliza da computação como ferramenta de modelagem, tratamento e análise de dados (ALMEIDA, 2009).

Em geral, redes complexas são representadas por meio de grafos, onde os nós representam os elementos do problema que se está estudando, que de alguma forma precisam ser individualizados. Já os links (ou arestas) entre os nós (ou vértices) representam algum tipo de interação entre os elementos, intrínseca ao problema que se está estudando (FERNANDES, 2012).

Nesse tipo de sistema, em geral, a distância física é irrelevante, porém a complexidade e ambiguidade das relações entre os elementos torna difícil a modelagem e compreensão da dinâmica e estrutura da rede (MOURA, 2012). Assim, descobrir, caracterizar e modelar a estrutura das redes é aspecto central no estudo das redes complexas,

isso porque a estrutura das redes tem influência nos fenômenos que operam sobre as redes. Caracterizar a estrutura de uma rede quer dizer enumerar aspectos capazes de capturar e resumir a estrutura da rede, de forma a codificar toda a informação presente na mesma. Porém, codificar toda informação que pode ser extraída de um sistema complexo não necessariamente ajuda a compreender as suas características, já que pode haver informação demasiada que confunde mais do que esclarece (FIGUEIREDO, 2011). Em função disso, extrair e definir que informações são determinantes no comportamento de um sistema complexo torna-se uma tarefa fundamental.

Diversas propriedades estruturais podem ser utilizadas para caracterizar as redes, isso porque é possível que a presença de uma determinada propriedade implique em alguma característica que independa do restante da estrutura da rede. Isso é semelhante à uma situação onde uma pessoa quer descrever para um amigo alguém que conheceu recentemente. Essa descrição provavelmente é baseada em características como cor da pele, dos olhos, cabelos, estatura, sexo e silhueta. Uma descrição completa só seria possível pelo DNA, que embora contenha todas as informações possíveis, contribui muito pouco para construir uma imagem da pessoa, ao passo que informações como o sexo, trazem consigo uma série de características inerentes aos indivíduos daquele sexo e contribuem fortemente para a construção de uma imagem do indivíduo (FIGUEIREDO, 2011).

Dessa forma, a próxima seção é dedicada a apresentar propriedades estruturais importantes na caracterização de redes complexas.

4.1 CARACTERIZAÇÃO TOPOLÓGICA

As redes complexas podem apresentar topologias diversas e características intrínsecas ao domínio estudado. Utilizando-se algumas métricas, as redes podem ser modeladas e caracterizadas.

A mais simples propriedade de uma rede é o seu tamanho, ou seja o número de elementos que a compõem (n) e a quantidade de relacionamentos existentes entre esses elementos ($|E(G)|$). Com essas duas informações é possível calcular o grau de conexão médio da rede, $\langle k \rangle$, que é dado pela média aritmética dos graus de todos os nós da rede, denotado como:

$$\langle k \rangle = \frac{2 |E(G)|}{n} \quad (4.1)$$

onde n é o número de nós (vértices) da rede e $|E(G)|$ é a quantidade de *links* (arestas)

presentes na rede.

Tendo em vista que o grau de conexão dos nós de uma rede pode assumir valores distintos, a função de densidade de probabilidade dos seus graus pode ajudar a inferir o tipo específico da rede, ou seja, se ela obedece à alguma lei de formação ou tem caráter aleatório (LOPES, 2011; ALMEIDA, 2009).

Conhecendo o grau de conexão médio da rede é possível calcular a densidade (ρ) da rede. Essa medida representa a fração de arestas que a rede possui levando em consideração a quantidade máxima possível de arestas em um sistema com aquela quantidade de vértices. Para tanto, é necessário determinar o número máximo de arestas que a rede poderia ter. O maior número possível de arestas ocorre quando cada nó possui $(n - 1)$ arestas. Sendo assim, o total de arestas é dado por $\frac{n(n-1)}{2}$ pois há n vértices. A divisão por 2 é justificada porque em uma rede não direcionada, cada aresta está sendo contada duas vezes (FIGUEIREDO, 2011). Assim, a densidade de uma rede é dada por:

$$\rho = \frac{|E(G)|}{n(n-1)/2} = \frac{\langle k \rangle}{n-1} \quad (4.2)$$

onde $|E(G)|$ é a quantidade de arestas presentes na rede, n é o número de vértices da rede e $\langle k \rangle$ é o grau de conexão médio da rede.

Algumas redes podem possuir pesos em suas arestas, ou seja, as ligações entre pares de vértices podem possuir pesos diferentes. Nestes casos, é possível definir a medida de **força** (*strength*) do vértice que é dada por s_i (ALMEIDA, 2009):

$$s_i = \sum_{j=1}^n w_{ij} \quad (4.3)$$

onde w_{ij} representa o peso da ligação entre os nós i e j e n é a quantidade de vértices da rede.

Já a força média de $\langle s \rangle$ uma rede é dada por:

$$\langle s \rangle = \frac{1}{n} \sum_{i=1}^n s_i \quad (4.4)$$

onde s_i é a força do vértice i e n é a quantidade de vértices da rede.

A medida de força da rede é similar à medida de grau de conexão médio da rede, sendo que a primeira é aplicada a redes que possuem pesos nas arestas.

Outra propriedade estrutural importante das redes diz respeito à sua distribuição

de graus $P(k)$, pois ela fornece uma informação completa em redes onde os graus dos nós conectados não são correlacionados, ou seja, a probabilidade de um nó possuir grau k independe do grau dos seus nós vizinhos (MELLO, 2010). De acordo com Figueiredo (2011), a fração de vértices com grau k , f_k , é dado pela seguinte equação:

$$f_k = \frac{n_k}{n} \quad (4.5)$$

onde n_k é o número de vértices com grau k e n é o número total de vértices da rede.

Em redes complexas é comum o interesse na fração de vértices que possuem grau maior que um inteiro qualquer k . A partir da Equação 4.5, é possível definir essa métrica, chamada de distribuição complementar cumulativa do grau (CCDF):

$$F_k = 1 - \sum_{i=0}^{k-1} f_i \quad (4.6)$$

onde k é o grau que se busca analisar e f_i é a fração de vértices com grau i .

Tão importante quanto as métricas já detalhadas é o estudos dos caminhos presentes numa rede, valor obtido pela soma da quantidade de arestas presentes no caminho entre dois vértices i e j . Em uma rede é comum haver diversos caminhos entre dois nós quaisquer, cada qual com custo diferente. Assim, a distância entre dois vértices é dada pelo custo do menor caminho entre esses dois vértices. Se os vértices estão em componentes conexas diferentes, o custo do caminho é infinito (FIGUEIREDO, 2011).

Fenômenos como comunicação e transporte em uma rede são fortemente impactados pelas distâncias mínimas entre os nós, daí a sua importância. A distância média $\langle l \rangle$ da rede é dada por:

$$\langle l \rangle = \frac{2}{n(n-1)} \sum_{i,j \in V, i \neq j} l_{ij} \quad (4.7)$$

onde n representa a quantidade de nós da rede, i e j são vértices da rede e l é a distância do menor caminho entre i e j .

Dessa forma, para uma rede totalmente conectada, tem-se que $\langle l \rangle = 1$. Para redes aleatórias é possível fazer uma aproximação da distância média. Sendo k a quantidade de vizinhos de um nó, cerca de k^l vértices encontram-se a uma distância de no máximo l arestas do nó. Então, $n \approx k^l$ e $l \approx \frac{\log(n)}{\log(k)}$. Já redes com graus que seguem uma lei de potência, possuem distâncias ainda menores, sendo no máximo proporcional a $\frac{\log(n)}{\log(\log(n))}$ (MELLO, 2010).

Isso indica que existe um caminho relativamente curto entre os pares de vértices da rede. Esse efeito chamado de mundo pequeno é devido ao fato de que a distância de separação entre os vértices cresce de forma mais lenta que o tamanho da rede. O conceito de mundo pequeno não é novo, pois foi proposto por Stanley Milgran em 1967 (NEWMAN et al., 2006) como resultado de seu experimento. Em seu experimento Milgran enviou centenas de cartas a pessoas escolhidas aleatoriamente, residentes nas cidades americanas de Kansas e Nebraska, pedindo que as pessoas que as recebessem entregassem a carta ao destinatário ou a um conhecido que ele imaginasse que pudesse conhecer o destinatário. Das 160 cartas enviadas, 42 chegaram ao destino e, na média, passaram por 5,5 pessoas até chegar ao destino, o que foi arredondado para 6 e deu origem à ideia de que há uma pequena distância entre os indivíduos (RODRIGUES, 2007).

Em redes reais, é comum encontrar essa propriedade, de tal forma que atualmente sabe-se que a maioria das redes complexas apresenta essa característica (RODRIGUES, 2007) que tem implicações importantes sobre os processos que ocorrem nas redes, já que possíveis perturbações podem se propagar rapidamente, como na velocidade de propagação de rumores ou disseminação de epidemias (MELLO, 2010).

Relacionada à questão das distâncias há, ainda, o conceito de diâmetro da rede, l , definido como o maior caminho mínimo entre dois vértices quaisquer da rede. Isso é dado por:

$$l = \max_{i,j \in V} l_{ij} \quad (4.8)$$

onde i e j são vértices da rede.

Tão importante quanto conhecer o grau de um vértice, é compreender como os vizinhos de um vértice se relacionam. Uma propriedade que captura essa relação é o coeficiente de clusterização, também chamado de coeficiente de agrupamento. Ele expressa a presença de ciclos mínimos (triângulos) em uma rede (ALMEIDA, 2009). Em outras palavras, o coeficiente de clusterização expressa a probabilidade de dois vizinhos do vértice $v \in V$ serem vizinhos também. Numa rede real que expresse, por exemplo, uma relação de amizade, essa métrica indica a probabilidade de que dois amigos de alguém sejam amigos entre si (MELLO, 2010).

O coeficiente de agrupamento ou coeficiente de clusterização de um vértice representa uma medida de quão conectados estão os vizinhos de um determinado vértice, sendo dado pela razão entre a quantidade de arestas observadas entre os vizinhos de um vértice e a quantidade máxima possível entre esses vértices. Seja v um vértice da rede

que se quer estudar e considere o grau de v como $k - 1$. A quantidade máxima de arestas possíveis para o subgrafo formado por v e seus vizinhos é dado por $\frac{k(k-1)}{2}$. Dessa forma, o coeficiente de clusterização, c_i , é dado por:

$$c_i = \frac{2|E_i(G)|}{k(k-1)} \quad (4.9)$$

onde $|E_i(G)|$ é o número de arestas observadas entre o vizinhos de i e k é o grau do vértice i .

O coeficiente de clusterização não está definido para vértices com grau um ou zero. Apenas em vértices com grau maior essa métrica pode ser aplicada (FIGUEIREDO, 2011). Utilizando-se o coeficiente de clusterização dos vértices de uma rede, pode-se calcular o coeficiente de clusterização de uma rede, $\langle c \rangle$, da seguinte forma:

$$\langle c \rangle = \frac{1}{n} \sum_{i \in V} c_i \quad (4.10)$$

onde n é o número de vértices da rede e c_i é o coeficiente de clusterização do vértice i .

Analogamente, no cálculo do Coeficiente de Clusterização médio da rede são considerados apenas os vértices que possuem essa propriedade definida, ou seja, apenas para vértices com grau maior que um. Para uma rede completamente conectada, $\langle c \rangle = 1$ (MELLO, 2010). A Figura 22 mostra exemplos de Coeficientes de Clusterização para algumas redes pequenas.

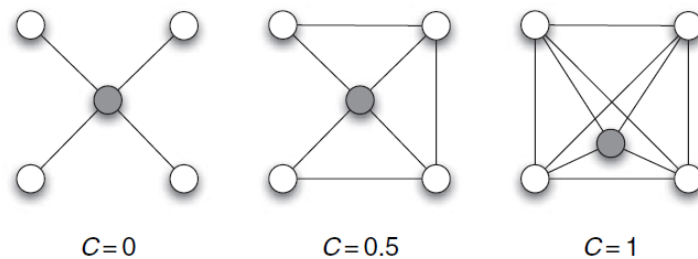


Figura 22: Exemplos de Coeficientes de Clusterização para o nó central (escuro).

Fonte: Fernandes (2012).

É possível, ainda, verificar a importância de um nó devido a sua localização e sua relação com os vizinhos. Essas relações podem ser medidas por meio de duas métricas distintas: grau de intermediação (*betweenness centrality*) e grau de proximidade (*closeness centrality*).

O grau de intermediação de um vértice v pode ser visto como uma medida da

elasticidade da rede e , de acordo com Newman (2003), parece seguir uma lei de potência para muitas redes reais. Ele exprime o efeito causado no caminho mínimo entre vértices da rede se v for removido (NEWMAN, 2003) da rede.

Em outras palavras, o grau de intermediação de um vértice v , B_v , é dado pela quantidade de caminhos mínimos entre dois vértices quaisquer da rede que passam por v (ALMEIDA, 2009). Essa medida é expressa pela equação:

$$B_v = \sum_{ij} \frac{\sigma(i, v, j)}{\sigma(i, j)} \quad (4.11)$$

onde $\sigma(i, v, j)$ é o número de caminhos mínimos entre i e j que passam por v e $\sigma(i, j)$ é a quantidade de caminhos mínimos existentes entre i e j .

Numa rede é importante analisar o comportamento da rede como um todo e não somente o comportamento de cada um de seus componentes, e, por isso, é importante conhecer o grau de intermediação médio da rede, $\langle B \rangle$, que representa o quão homogênea a rede é em termos da importância dos nós para os caminhos mínimos entre os vértices da rede. Essa média é dada por:

$$\langle B \rangle = \frac{1}{n} \sum_{v \in V} B_v \quad (4.12)$$

onde V representa o conjunto de vértices da rede e B_v é o grau de intermediação do vértice v .

Já o grau de proximidade indica o quão próximo um vértice v está do restante dos nós da rede. Assim, vértices centrais, que possuem grau de proximidade mais baixo, possuem distância média menor para os demais nós da rede (NEWMAN, 2008).

O grau de proximidade de vértice v , $C(v)$ é a média dos caminhos mínimos dele para os outros nós da rede (ALMEIDA, 2009). Ele é calculado pela equação:

$$C(v) = \sum_{v \neq u} \frac{1}{d(v, u)} \quad (4.13)$$

onde v e u são vértices da rede e $d(v, u)$ é a distância mínima entre os vértices v e u .

Outra característica bastante comum em redes complexas é a formação de comunidades, que são subgrafos muito conectados internamente, mas com poucas conexões externas. Ou seja, uma comunidade é formada por um conjunto de vértices densamente conectados entre si e com relativamente poucas arestas para fora do grupo (ALMEIDA,

2009; MELLO, 2010).

Em redes reais, existe essa tendência à formação de grupos, pois, por exemplo, as pessoas têm uma tendência a se conectarem a grupos com que se identificam, de tal forma que o surgimento de comunidades é favorecido. Um exemplo dessa tendência é apresentado por Moody (2001) em artigo que apresenta os resultados de um estudo a respeito da rede de amizades entre alunos adolescentes de uma escola americana. A Figura 23 apresenta uma ilustração dos resultados encontrados por Moody (2001):

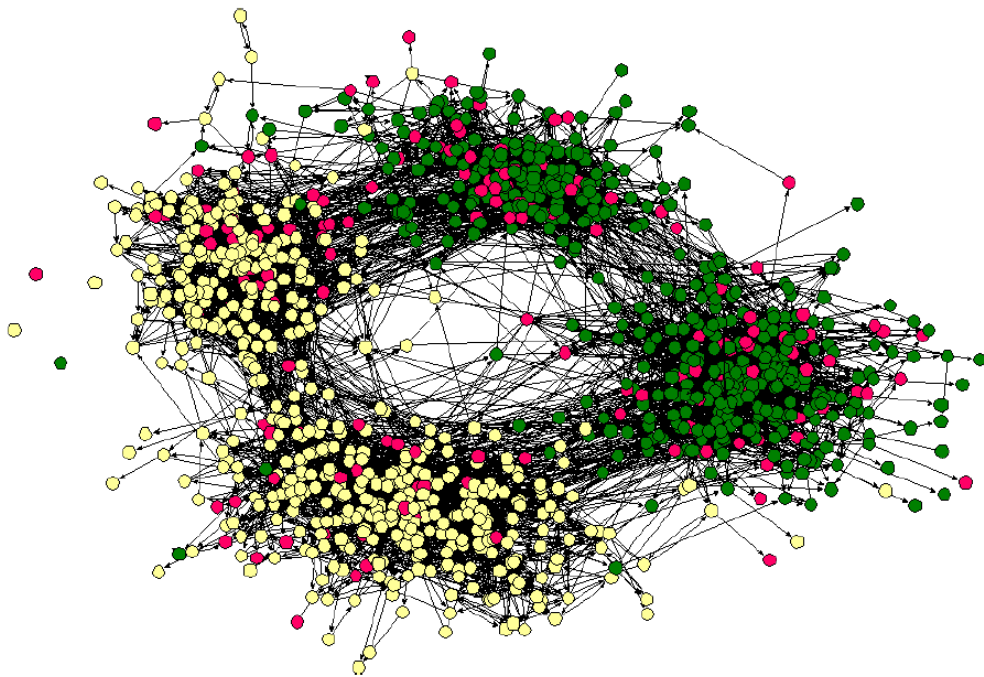


Figura 23: Rede de amizade de uma escola americana.

Fonte: Mello (2010).

Nesta figura é possível observar a divisão dos alunos em quatro comunidades. Os vértices estão coloridos de acordo com a raça dos indivíduos. A diagonal que passa pela porção superior esquerda da imagem, mostra a separação dos indivíduos por raça, enquanto que a outra diagonal mostra a separação entre alunos do Ensino Fundamental e Médio (MELLO, 2010). Da mesma forma que neste exemplo, é de se esperar que a formação desse tipo de estrutura seja verificada em redes de outros tipos, como rede de citações ou agrupamento de páginas *web*.

De modo mais geral, as redes reais podem ser usadas para representar as mais diversas relações entre elementos. O estudo de muitas das redes reais levou à observação de que independentemente da finalidade delas, algumas características são comuns a todas elas, o que de certa forma é surpreendente já que a princípio não há qualquer razão para

que, por exemplo, uma rede da *web* compartilhe características estruturais com a rede de neurônios humana ou a rede rodoviária de uma cidade.

A primeira característica recorrente em muitas das redes reais diz respeito à sua distribuição de graus, já que muitas das redes reais possuem distribuição de graus dos nós com cauda longa, muito bem representadas por uma lei de potência. É o que ocorre, por exemplo, em redes como *web*, *Twitter*, rede de atores, internet e rede metabólica (FIGUEIREDO, 2011).

Além disso, é possível perceber que muitas dessas redes também possuem distância relativamente pequena entre seus vértices, inclusive entre os mais distantes da rede. Em geral, esse valor fica ordens de grandeza menor que o número de nós encontrado na rede. Muitas delas apresentam distância média entre os nós da ordem de $\log(n)$.

Por fim, redes reais costumam ser bastante esparsas, com densidade muito baixa, cuja ordem de grandeza é menor que 1. Apesar disso, essas redes são altamente conectadas com quase todos os seus vértices pertencendo a uma mesma componente conexa. Isso ocorre, por exemplo, com a rede da *web*, internet, rede de neurônios ou uma rede metabólica.

4.2 MODELOS DE CRESCIMENTO

Historicamente, diversas tentativas têm sido feitas para construir um modelo de rede que consiga explicar o crescimento de redes complexas. Esta seção apresenta alguns dos modelos matemáticos mais encontrados na literatura e que tentam capturar ou representar as características encontradas nas redes reais.

4.2.1 REDE ALEATÓRIA - MODELO DE ERDÖS E RÉNYI

Até a metade do século XX, uma das áreas de estudo da teoria de grafos era voltada à proposição de modelos matemáticos para criação de grafos e a catalogação de suas propriedades (ALMEIDA, 2009). Na década de 60, os influentes matemáticos húngaros Paul Erdős e Alfred Rényi propuseram um modelo de crescimento de redes cujas ligações entre os vértices eram aleatórias, e ao estudar as consequências desse modelo descobriram diversas propriedades muito interessantes que serão citadas no decorrer desta seção (FIGUEIREDO, 2011).

Eles propuseram um modelo, conhecido como modelo $G(n, p)$, no qual as ligações entre os vértices da rede são totalmente aleatórias, ou seja, não há critérios para privile-

giar uma ou outra ligação, de tal forma que apenas dois parâmetros são necessários para caracterizar o modelo: n e p , onde n é o número de vértices da rede e p é a probabilidade de que exista uma ligação entre dois vértices quaisquer. Cada aresta é incluída de forma independente, de tal forma que a probabilidade p é aplicada a cada uma das $\frac{n(n-1)}{2}$ possíveis arestas. Assim, a probabilidade de que existam m arestas em uma rede é dado por $E(m) = p^{\frac{n(n-1)}{2}}$ (FIGUEIREDO, 2011; ALMEIDA, 2009; ALBERT; BARABÁSI, 2002).

Por depender de apenas dois parâmetros o modelo $G(n, p)$ também é conhecido como modelo binomial. Ele não determina a estrutura da rede e sim o processo que gera a rede (FIGUEIREDO, 2011). Por isso, há diversas redes possíveis para uma mesma configuração de n e p . A cardinalidade do conjunto de redes, (conjunto amostral S de $G(n, p)$), $|S|$, que pode ser gerado pelo modelo é dado por

$$|S| = 2^{\frac{n(n-1)}{2}} \quad (4.14)$$

onde n é o número de vértices da rede.

Essa expressão indica que o espaço amostral cresce rapidamente quando se aumenta a quantidade de nós presentes na rede, de tal forma que o espaço amostral de uma rede com 25 vértices é maior do que a quantidade de átomos do universo. No entanto, algumas redes são bem mais prováveis de ocorrer do que outras e algumas delas são quase impossíveis de se obter (FIGUEIREDO, 2011).

A probabilidade do modelo $G(n, p)$ obter uma rede representada pelo conjunto de arestas $E = \{e_1, e_2, \dots, e_m\}$ pode ser obtida se for considerado que toda aresta de E deve estar presente na rede e não aparece nenhuma outra aresta que esteja fora de E . A probabilidade de que uma aresta esteja em E é dada por p . Ao mesmo tempo, a probabilidade de uma aresta não aparecer em E é dada por $1 - p$ (FIGUEIREDO, 2011). Assim a probabilidade de que o modelo $G(n, p)$ gere uma rede com o conjunto E de arestas, $P[G(n, p) \rightarrow E]$, de acordo com Albert e Barabási (2002), é dada por

$$P[G(n, p) \rightarrow E] = p^m (1 - p)^{\frac{n(n-1)}{2} - m} \quad (4.15)$$

onde n é o número de vértices da rede, m é o número de arestas em E e p é a probabilidade de que a aresta e_i apareça na rede.

A Equação 4.15 apresenta a probabilidade de obter uma rede com o conjunto de arestas E , embora seu valor só dependa da quantidade de arestas contidas no conjunto E . No entanto, em alguns casos, o interesse gira em torno da obtenção de redes com

características específicas, por exemplo, redes com m arestas, independente de quais sejam essas arestas. Neste caso, o número de redes com m arestas, $P[G(n, p) \rightarrow E = m]$, é dado pelo número de diferentes combinações entre as m dentre as $\frac{n(n-1)}{2}$ arestas possíveis (FIGUEIREDO, 2011). Dessa forma, têm-se que

$$P[G(n, p) \rightarrow E = m] = C_{\frac{n(n-1)}{2}, m} p^m (1-p)^{\frac{n(n-1)}{2} - m} \quad (4.16)$$

onde $C_{\frac{n(n-1)}{2}, m}$ é a combinação de $\frac{n(n-1)}{2}$, m a m arestas, n é o número de vértices da rede, m é o número de arestas em E e p é a probabilidade de que a aresta e_i esteja presente na rede.

Esse modelo é bastante simples, mas se apresenta como uma abordagem interessante para o estudo de redes complexas, pois numa rede assim construída, não se conhecem as razões e mecanismos que propiciam o surgimento de arestas, de tal forma que, no máximo, se pode afirmar que um link entre dois nós pode existir com probabilidade p . Devido à sua simplicidade, os resultados obtidos nesse modelo são importantes como fator de comparação com outros mecanismos de formação de ligações entre nós (FERNANDES, 2012).

Diferentes probabilidades geram redes com densidades diferentes, como é ilustrado na Figura 24.

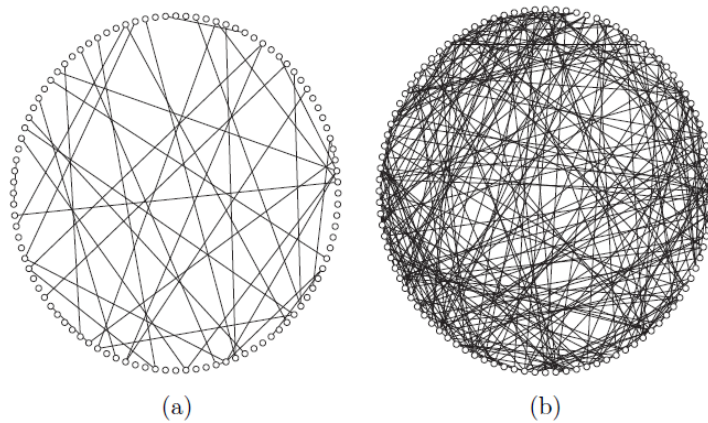


Figura 24: Exemplos de redes geradas pelo modelo de Erdős e Rényi.

Fonte: Fernandes (2012).

As redes foram criadas com 100 vértices. Em (a) utilizou-se probabilidade $p = 0,009$ e em (b) tem-se $p = 0,05$. Como esperado, quanto maior a probabilidade de existência de uma aresta, mais densa é a rede gerada.

Cada link da rede contribui para o grau de dois vértices da rede. Portanto, de

acordo com Fernandes (2012), o grau médio de um grafo construído sob essa ótica é dado por:

$$\langle k \rangle = \frac{2 p n(n-1)}{n} = (n-1)p \cong n p \quad (4.17)$$

onde n é o número de vértices da rede, p é a probabilidade de que a aresta e_i esteja presente na rede e a simplificação $n p$ é válida para redes com n grande.

O grau médio indica a estrutura da conectividade da rede. Se $\langle k \rangle < 1$, a rede não possui ciclos e é constituída por componentes conexas pouco conectadas. Se $\langle k \rangle > 1$, há um componente gigante a quem a maioria dos nós está ligada. Já, se $\langle k \rangle \geq \ln(n)$ raramente se encontrará uma rede com elementos isolados (FERNANDES, 2012)

Uma vez considerada a criação de uma rede de acordo com esse modelo, é possível definir a probabilidade de um vértice ter grau k . Para um vértice qualquer da rede, o seu grau d é uma variável aleatória que varia entre 0 e $n-1$. A probabilidade de incidência de cada uma das arestas é p (FIGUEIREDO, 2011) e para que o grau do vértice seja k , apenas k arestas devem ser incidentes a ele. Dessa forma, a probabilidade de um vértice qualquer possuir grau k , $P[D = k]$, é dado por:

$$P[D = k] = C_{(n-1),k} p^k (1-p)^{n-1-k} \quad (4.18)$$

onde $C_{(n-1),k}$ é a combinação de $(n-1)$ vértices, k a k , n é o número de vértices da rede, k é o grau do vértice e p é a probabilidade de que a aresta e_i pertença à rede.

Embora seja fundamental para o estudo de Redes Complexas, esse modelo não é muito eficiente para representar redes reais, pois produz redes com estrutura muito diferentes de redes reais, como pode ser visto quando se compara a distribuição de graus produzida pelo Modelo de Erdős e Rényi e a rede da Wikipedia. A Figura 25 apresenta a diferença na distribuição de graus entre o produzido pelo modelo e o observado na rede da Wikipedia (FIGUEIREDO, 2011).

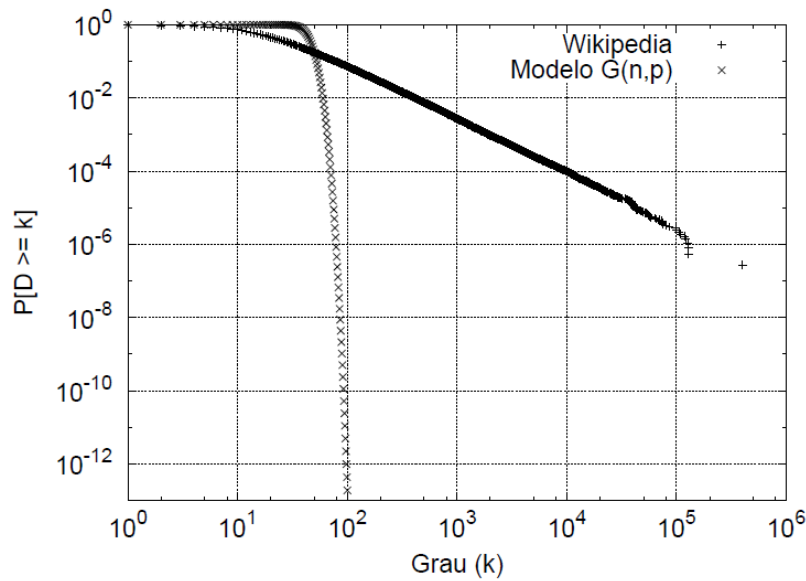


Figura 25: Comparação da Distribuição de Graus da Wikipedia e do obtido pelo Modelo de Erdős e Rényi.

Fonte: Figueiredo (2011).

Pode-se observar que a rede real possui uma distribuição de graus com cauda pesada, o que não é obtido pelo Modelo de Erdős e Rényi. Analogamente, o coeficiente de clusterização das redes produzidas por este modelo costuma ser muito menor do que o observado em redes reais (FIGUEIREDO, 2011). Como essas são propriedades fundamentais para a estrutura das redes, é necessário considerar que o modelo em questão nem sempre apresenta resultados satisfatórios.

Outro ponto de destaque com relação a esse modelo, diz respeito ao comprimento de caminho médio, $\langle l \rangle$ da rede, que de acordo com Fernandes (2012) pode ser relacionado a n e $\langle k \rangle$ da seguinte forma:

$$\langle l \rangle \approx \frac{\ln(n)}{\ln(\langle k \rangle)} \quad (4.19)$$

onde n é o número de vértices da rede e $\langle k \rangle$ é o grau médio da rede.

Como consequência, percebe-se que $\langle l \rangle$ cresce logaritmicamente com o tamanho de n e não como uma lei de potência, expressando, assim, o efeito de mundo pequeno comum em redes reais, a ser abordado na próxima seção.

4.2.2 REDE MUNDO PEQUENO - MODELO DE WATTS E STROGATZ

No final da década de 1990, Duncan Watts e Steven Strogatz perceberam que muitas redes reais possuíam muito mais triângulos formados por três vértices do que o Modelo de Rede Aleatória gerava para valores de vértices e arestas iguais aos de redes reais. Dessa forma, perceberam que seria necessário um modelo mais sofisticado do que o proposto por Erdős e Rényi para modelar redes reais (ALMEIDA, 2009).

Assim, o Modelo de Rede Mundo Pequeno foi motivado pela observação de que muitas redes reais apresentavam propriedades não bem contempladas pelo modelo Rede Aleatória, especialmente com relação à baixa distância entre os vértices da rede, alta clusterização e estrutura de redes esparsa (FIGUEIREDO, 2011; NEWMAN et al., 2006).

O Modelo de Watts e Strogatz é baseado numa configuração de rede regular, esparsa e com alta clusterização, com mudanças aleatórias nas arestas para induzir a formação de caminhos pequenos.

Este modelo inicia com n nós dispersos num formato circular, onde cada um dos nós está conectado com os seus k vizinhos mais próximos, sendo $\frac{k}{2}$ nós para cada lado do nó, gerando assim uma rede esparsamente conectada. Aleatoriamente, as arestas são religadas a outros vértices com probabilidade p , laços e arestas duplicadas não são permitidos. Esse processo introduz $\frac{p \times n \times k}{2}$ arestas de longa distância, ou seja, nós conectados a vértices distantes (que não os $k/2$ vizinhos imediatos) (ALBERT; BARABÁSI, 2002).

A Figura 26 apresenta exemplos de redes geradas pelo algoritmo proposto por Watts e Strogatz para uma rede com 20 nós e probabilidade de reconexão variando entre 0 e 1.

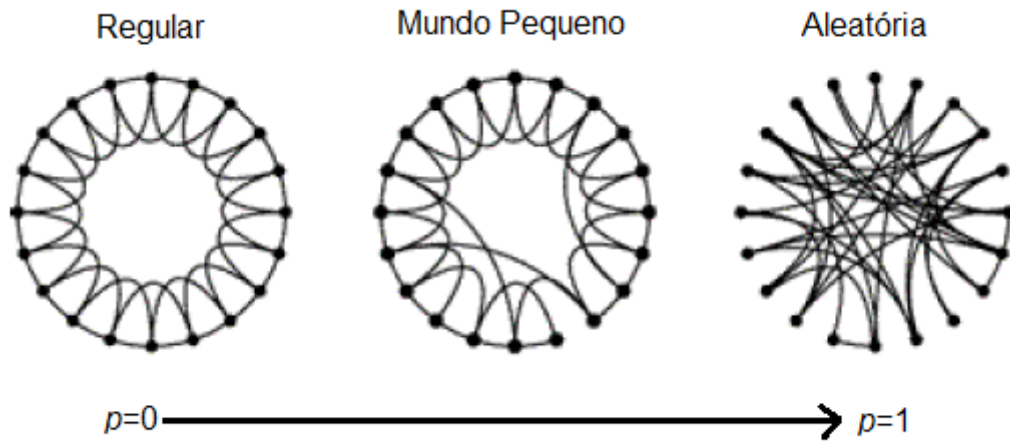


Figura 26: Redes obtidas pelo Modelo de Watts e Strogatz.

Adaptado de: Watts e Strogatz (1998).

Quando a probabilidade de reconexão das arestas é 0, a rede não muda com o passar do tempo, e permanece sempre regular. À medida que a probabilidade p é incrementada, mais arestas são religadas a vértices não vizinhos. Por outro lado, se $p = 1$, são obtidas redes aleatórias. É em algum ponto onde a probabilidade de reconexão é maior que zero e menor que um surgem as redes Mundo Pequeno e propriedades interessantes podem ser encontradas.

Redes Mundo Pequeno refletem características de sistemas sociais reais. Por exemplo, considere uma rede de pessoas. É esperado que os componentes dessa rede sejam amigos de indivíduos de seu círculo de convivência (escola, trabalho, vizinhos de rua), mas possuam alguns poucos amigos que moram longe, seja em outros estados ou países, o que no modelo de Watts e Strogatz é representado pelos vértices em que há a reconexão e criação de arestas de longa distância (ALBERT; BARABÁSI, 2002).

Considere $\ell(p)$ e $c(p)$ como a distância média da rede e o coeficiente de clusterização quando a probabilidade de reconexão da rede é p . Quando $p = 0$, todos os vértices são idênticos pois mantém-se a estrutura original, de tal forma que a média do coeficiente de clusterização e distância média da rede são iguais aos valores de um vértice qualquer, o que é dado por:

$$c(0) = \frac{n}{2k} \quad (4.20)$$

$$\ell(0) = \frac{3(k-1)}{2(2k-1)} \approx \frac{3}{4} \quad (4.21)$$

onde n é o número de vértices da rede e k é o grau médio da rede.

Como era esperado, uma rede com $p = 0$ possui coeficiente de clusterização alto e distância média proporcional à quantidade de nós na rede. No outro extremo, quando $p = 1$, o modelo converge para uma Rede Aleatória e o grau médio da rede é dado por $2k$, obtendo-se os seguintes valores para $c(p)$ e $\ell(p)$:

$$c(1) \approx \frac{2k}{n} \quad (4.22)$$

$$\ell(1) \approx \frac{\lg n}{\lg p} \quad (4.23)$$

onde n é o número de vértices da rede e k é o grau médio da rede.

Como pode ser observado, a rede terá uma clusterização muito baixa, já que o coeficiente de clusterização é inversamente proporcional à quantidade de nós na rede. Ao mesmo tempo, é esperada uma pequena distância entre os vértices, já que a distância média entre os vértices da rede é proporcional a $\lg n$.

Com base na Figura 27, é possível observar o que acontece com $c(p)$ e $\ell(p)$ para valores intermediários de p :

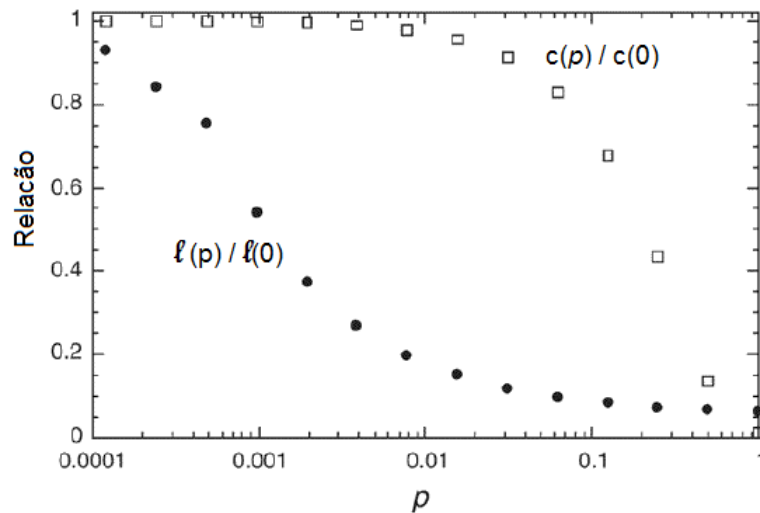


Figura 27: Relação entre distâncias médias e coeficiente de clusterização médio para valores intermediários de p .

Adaptado de: Watts e Strogatz (1998).

A figura mostra que há valores de p , onde o coeficiente de clusterização é relativamente alto e a distância entre os nós é baixa. Dessa forma, pode-se afirmar que o Modelo de Watts e Strogatz é capaz de induzir redes de mundo pequeno, de onde se origina o

nome Mundo Pequeno (ou *small world*) (FIGUEIREDO, 2011).

Intuitivamente, é possível perceber porque isso acontece. Quando p é bem pequeno, poucas arestas são reposicionadas, mas por serem escolhidas de forma aleatória, formam atalhos na rede, reduzindo drasticamente a distância média da rede. Por outro lado, como poucas arestas são reposicionadas, a estrutura inicial da rede é muito pouco afetada, de tal forma que o coeficiente de clusterização se mantém alto. À medida que p aumenta, há mais reposicionamentos de arestas e a distância média da rede tende a ficar ainda menor. Analogamente, o coeficiente de clusterização tende a diminuir já que a estrutura da rede é afetada (FIGUEIREDO, 2011).

Embora o modelo tenha despertado o interesse da comunidade científica há que se tecer algumas críticas a ele, pois ele deixa de capturar alguns aspectos importantes presentes em redes reais, no que tange, por exemplo, à distribuição de grau dos nós que difere de redes reais por não possuir cauda longa (não segue uma distribuição de lei de potência). Além disso, esse modelo é bastante estético e não ajuda a compreender os motivos que levam ao surgimento das características das redes reais, motivo pelo qual outros modelos e variações foram propostos como tentativa de suprir as deficiências do Modelo de Watts e Strogatz. Esses outros modelos não serão abordados neste trabalho pois excederia em muito o escopo proposto, mas podem ser encontrados na literatura da área.

4.2.3 REDE LIVRE DE ESCALA - MODELO DE BARABÁSI E ALBERT

Os modelos de Rede Aleatória e Mundo Pequeno significaram grandes avanços na teoria de redes complexas ao conseguirem reproduzir características importantes de redes reais. No entanto, eles falham na modelagem da distribuição de graus dos vértices, já que não produzem distribuições de graus no formato de Lei de Potência, como ocorre em boa parte das redes reais. Mesmo redes reais que fogem à distribuição em Lei de Potência, se aproximando de uma distribuição exponencial, têm distribuição de graus bem diferente das obtidas pelos Modelos de Erdős e Rényi ou Watts e Strogatz (FERNANDES, 2012) .

No final da década de 1990, Albert-László Barabási e Réka Albert estavam estudando a rede da Web e perceberam que ela possuía uma distribuição de graus muito particular, muito bem representada por uma lei de potência. Em suas pesquisas, descobriram que mais de 80 por cento das páginas *web* possuíam menos de quatro links para outras páginas, enquanto que uma minoria de páginas, cerca de 0,01 por cento das páginas possuía mais de mil links, fenômeno que ficou conhecido como ligação preferencial

(ou *preferencial attachment*) (BARABÁSI; BONABEAU, 2003).

Eles buscaram compreender quais são os mecanismos mais importantes que levam a essa distribuição de graus característica quando da formação das redes reais. Como resultado, propuseram um modelo de rede baseado no crescimento do número de nós na rede e no fato de que novas ligações são criadas preferencialmente com os nós que já possuem um grau alto.

Esses dois princípios parecem bastante razoáveis quando são consideradas redes de páginas *web* ou redes de citações científicas, por exemplo. No primeiro caso, constantemente novas páginas são criadas e é comum que páginas importantes sejam mais citadas do que páginas mais inexpressivas. Já no caso de redes de citações, à medida que mais artigos são publicados, mais autores são adicionados à rede e é esperado que os artigos importantes de cada área sejam citados pelos novos artigos. Dessa forma, há uma certa ligação preferencial de novos nós com aqueles que já possuem maior importância (maior número de citações).

Esses dois ingredientes, crescimento e ligação preferencial, inspiraram a criação do Modelo Livre de Escala (ou *Scale-Free*) que origina uma distribuição de graus no formato de lei de potência.

O algoritmo proposto por Barabási e Albert, inicia com um número pequeno de nós (n_0) e a cada passo é adicionado um novo nó com m arestas, onde $m \leq n_0$, de tal forma que o novo nó recebe arestas com m diferentes nós presentes no sistema. A probabilidade de um nó i receber conexão com o novo nó, $\Pi(k_i)$, depende do grau de i (k_i):

$$\Pi(k_i) = \frac{k_i}{\sum_{j=0}^n k_j} \quad (4.24)$$

onde n é o número de vértices da rede (excluído o novo nó) e k_i é o grau do nó i .

Ao final de t passos do algoritmo surgirá uma rede com $n = t + n_0$ nós e $m \times t$ arestas. Utilizando simulações numéricas é possível verificar que a rede evolui para um estado de escala invariante, onde a probabilidade de um nó possuir k arestas segue uma lei de potência com coeficiente $\gamma = 2.9 \pm 0.1$, o que independe do valor de m , único parâmetro do modelo (ALBERT; BARABÁSI, 2002)². Analogamente, este expoente independe do tempo t transcorrido, desde que esse tempo seja grande (BARABÁSI et al., 2000).

²Uma distribuição de potência é uma função de probabilidade no formato $p(x) = Cx^{-\gamma}$, onde C é a constante de normalização (uma vez que as probabilidades para todos x devem somar 1) e γ é o expoente (ou coeficiente) da lei de potência

Como consequência do processo de ligação preferencial, na rede resultante aparecem nós ditos *hubs*, que são os nós com grau alto. A figura 28 apresenta um exemplo de uma rede gerada de acordo com o modelo proposto por Barabási e Albert, onde os nós *hubs* estão posicionados mais ao centro da rede.

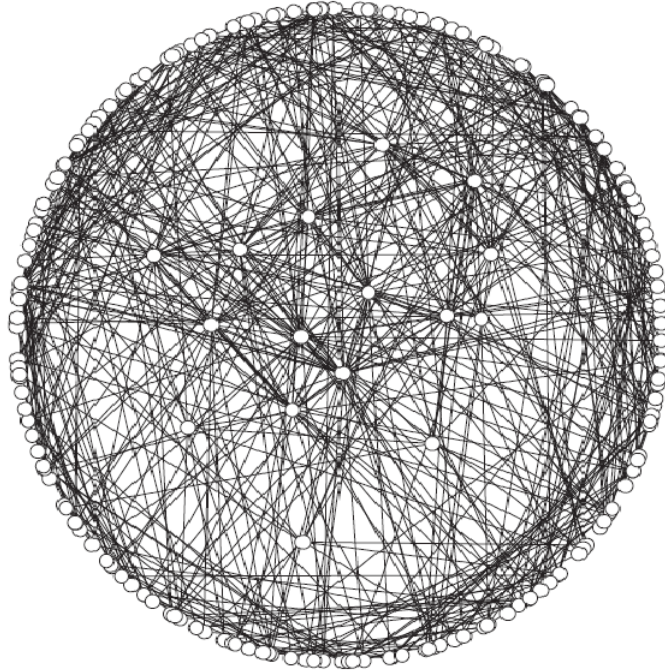


Figura 28: Exemplo de rede obtida utilizando o modelo de Barabási para $n = 200$.

Fonte: Fernandes (2012).

Apesar de terem sido descobertas num estudo que tentava mapear a *web*, as redes Livre de Escala aparecem em outras áreas do conhecimento, como é o caso de redes de negócios, rede de atores de *Hollywood*, metabolismo celular de diferentes organismos, interação de proteínas. Observando essas redes, notou-se que a distribuição de graus segue uma lei de potência com coeficiente (ou expoente) variando entre 2 e 3 (BARABÁSI; BONABEAU, 2003).

É importante notar que a influência da rede inicial utilizada no modelo de Barabási e Albert diminui à medida em que a rede cresce, se anulando para um tempo t suficientemente longo. É possível mostrar (ver Albert e Barabási (2002)) que, assumindo uma distribuição de valores contínuos para k , é gerada uma rede que no passo t possui distribuição de graus dada por:

$$P(k,t) = 2m^2 \frac{(m_0/2m) \langle k \rangle_0 k^{-3}}{t + m_0} \quad (4.25)$$

onde m é o número de arestas da rede, m_0 é o número de arestas presentes na rede inicial

e $\langle k \rangle_0$ é a média de graus da rede no instante t_0 .

Para t muito grande, quando $t \rightarrow \infty$, essa equação pode ser reduzida a:

$$P(k) = 2mk^{-3} \quad (4.26)$$

onde m é o número de arestas da rede.

Demonstrando, assim, a independência do expoente de k com relação ao tamanho da rede.

É possível, ainda, obter a forma como se comporta o comprimento de caminho mínimo médio de uma rede criada de acordo com o modelo de Barabási e Albert:

$$\langle l \rangle \approx \frac{\ln(n)}{\ln(\ln(n))} \quad (4.27)$$

onde n é o número de nós da rede.

Esse valor indica que as redes criadas por este modelo apresentam comprimento de caminho mínimo médio bem menor do que o gerado pela topologia proposta por Erdős e Rényi. Ao mesmo tempo, o coeficiente de agrupamento pode ser obtido analiticamente, (ver Albert e Barabási (2002) para maiores detalhes), e é dado por:

$$\langle c \rangle = \frac{m}{8n} (\ln(n))^2 \quad (4.28)$$

onde n é o número de nós da rede e m é o número de arestas da rede.

Apesar de se obter um coeficiente de clusterização maior do que o obtido a partir de redes aleatórias, esse resultado é bastante diferente do obtido para redes Mundo Pequeno onde o coeficiente de clusterização independe do tamanho da rede.

Devido às suas características estruturais, as redes geradas pelo modelo de Barabási e Albert são mais robustas do que aquelas geradas pelo modelo de Erdős e Rényi.

Suponha uma rede real, onde alguns nós podem falhar, é o caso, por exemplo, da internet que possui milhões de roteadores que de tempos em tempos apresentam falha. Neste caso, se poucos nós falharem, o impacto das falhas é desprezível na rede, já que por possuir muitos nós com grau baixo, a tendência é que as falhas ocorram principalmente em nós com poucas conexões, de tal forma que não haverá um grande dano à rede como um todo. Por outro lado, se muitos nós falharem, cresce a probabilidade de que nós *hubs* falhem. Se nós *hubs* falham, o efeito sobre a rede pode ser devastador (FIGUEIREDO,

2011). De fato, para falhas de até 5% dos nós, a comunicação entre os nós da rede não é afetada (ALBERT et al., 2000).

Já em redes puramente aleatórias, se poucos nós falharem o impacto será desprezível, mas à medida em que aumenta o número de falhas, a rede começa a se fragmentar. Se o número de nós com falha for muito grande, a rede será desmembrada em muitas componentes conexas, todas muito pequenas.

Por outro lado, a sobrevivência à falhas determinísticas (ataques direcionados aos *hubs*), onde a falha ocorre de acordo com a ordem decrescente dos graus dos nós, tem impacto maior em redes Livres de Escala. Nessas, ataques direcionados têm um efeito devastador sobre a rede, pois os nós *hubs* são os primeiros a falhar, fazendo com que rapidamente decaia a habilidade de comunicação entre os nós restantes, já que os nós *hubs* são importantes na ligação entre os diversos nós da rede (ALBERT et al., 2000). Como redes de Livre Escala seguem uma distribuição de grau dos nós em forma de lei de potência, a falha de cerca de 3% dos nós já é suficiente para causar a desconexão de toda a rede (NEWMAN, 2008). Já no caso das redes aleatórias, todos os nós possuem praticamente o mesmo grau, então remover os nós com maior grau tem o mesmo efeito que a remoção aleatória de nós (FIGUEIREDO, 2011).

Dessa forma, pode-se concluir que o modelo proposto por Barabási e Albert é altamente robusto a falhas aleatórias, mas extremamente frágil a falhas determinísticas, direcionadas aos vértices com maior grau. O efeito de ataques direcionados pode ser preocupante quando se considera redes de comunicação por exemplo, mas pode ser muito positiva em outras situações, como é o caso de uma campanha de vacinação. Vacinar apenas pessoas consideradas *hubs* em teoria seria suficiente para evitar que doenças se alastrem e uma epidemia seja instalada (o difícil neste caso seria identificar quem são as pessoas hubs) (NEWMAN, 2008).

5 METODOLOGIA

Diante da importância crescente do estudo das redes complexas, sentiu-se a necessidade de explorar essas redes sob o ponto de vista algorítmico. Redes são muito bem representadas por grafos. Para muitos autores, essas áreas têm, inclusive, a mesma origem. A teoria dos grafos apresenta diversos problemas interessantes e cuja solução têm grande utilidade para o ser humano. Um desses problemas é a coloração de vértices, não só pelas aplicações práticas que possui, como também pelo fato de ser este um dos problemas clássicos da área e um dos mais estudados devido à sua importância conceitual, conforme destacado na introdução deste trabalho.

O problema da coloração de vértices é NP-Difícil, o que significa que os algoritmos conhecidos que garantem a solução ótima têm custo computacional exponencial no tamanho da instância. Sendo assim, neste trabalho objetiva-se investigar se algoritmos com heurística gulosa apresentam resultados satisfatórios para alguns problemas de teoria dos grafos.

Este trabalho é parte de um projeto maior desenvolvido por professores da Universidade Tecnológica Federal do Paraná, que busca verificar o desempenho de algoritmos gulosos em redes complexas.

O fato de já existir um trabalho publicado que utiliza a estratégia gulosa para obter a cobertura por vértices em redes complexas, com resultados interessantes (para maiores detalhes ver SILVA et al. (2013)), apoia a suspeita de que o mesmo pode ocorrer para o problema da coloração de vértices.

Dessa forma, faz-se necessário primeiramente definir o conjunto de redes reais aos quais será aplicado o algoritmo guloso para coloração de grafos. Foram selecionadas 25 redes distintas disponíveis em bases de dados reconhecidas pela comunidade científica. Foram utilizadas as bases de redes complexas *Gephi* (GEPHI Datasets) , *Pajek* (PAJEK Datasets) e *Stanford* (Stanford Large Network Data Collection) para a seleção das redes a serem estudadas, de tal forma que fossem contemplados os diferentes tipos de redes

complexas: rede tecnológica, social, informação, biológica, de citação, de comunicação e colaboração.

As redes estudadas são as mesmas sobre as quais aplicou-se o problema da cobertura de vértices por algoritmo guloso em SILVA et al. (2013), uma vez que se objetiva caracterizar o comportamento de algoritmos gulosos em redes complexas. Algumas das redes selecionadas são não-direcionadas e outras são direcionadas, com quantidade de vértices variando no intervalo entre 235 e 75.879 elementos. Para os casos onde as redes selecionadas são direcionadas, é usado o grafo subjacente que representa a rede em questão.

As redes selecionadas foram convertidas para um formato único, uma vez que, por serem originárias de bases de dados distintas, apresentam organização distinta. Com essa etapa buscou-se padronizar as entradas do algoritmo para facilitar os processos que se seguem a este passo.

Na sequência, foram aplicados alguns algoritmos sobre o conjunto de grafos selecionado. O primeiro algoritmo aplicado é o algoritmo com heurística gulosa para coloração de vértices. Duas heurísticas serão utilizadas: a primeira delas utilizando a heurística de maior grau para a seleção do próximo vértice a ser colorido. A segunda heurística utiliza o critério de maior saturação combinado com a heurística de maior grau para definir qual vértice será o próximo a receber cor.

O primeiro algoritmo seleciona para colorir o vértice que apresentar maior grau. Suponha que se busque determinar o próximo vértice a colorir no caso apresentado no exemplo dado pela Figura 29. Neste caso, os vértices c , d e g já foram coloridos porque eram aqueles que nas interações anteriores possuíam os maiores graus.

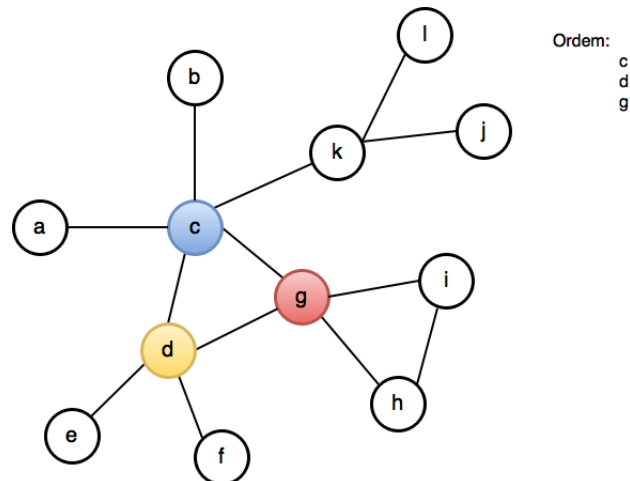


Figura 29: Exemplo de seleção do próximo vértice a colorir utilizando a heurística de maior grau.

Fonte: Autoria própria.

Nesse ponto, restam nove vértices para coloração. Os vértices a , b , e , f , l e j possuem grau 1. Já os vértices i e h possuem grau 2 e o vértice k possui grau 3. Por essa razão, o próximo nó a ser colorido será o vértice k .

O algoritmo que ilustra a aplicação dessa heurística é, então, descrito a seguir.

ALGORITMO GULOSO COM HEURÍSTICA DE MAIOR GRAU

INPUT: um grafo G com n vértices

OUTPUT: número de cores utilizadas para colorir G

1. Selecione o vértice v de maior grau dentre aqueles que ainda não receberam cores.
 2. Atribua a primeira cor disponível para v .
 3. Retorne ao passo 1 até que todos os vértices estejam coloridos.
-

O segundo algoritmo de coloração gulosa de grafos, algoritmo guloso com heurística de maior saturação, utiliza outra heurística para seleção do próximo vértice a ser colorido. Ele utiliza uma combinação entre o grau de saturação do vértice e o número de conexões do mesmo para definir o vértice a ser colorido a cada iteração do algoritmo de coloração gulosa. Entende-se o grau de saturação de um vértice como sendo o número de cores já utilizados para colorir os vizinhos de tal vértice. Assim, o vértice que tiver maior grau de saturação será selecionado para coloração. Caso dois ou mais vértices empatem no maior grau de saturação, aquele de maior grau será escolhido para a coloração. Este algoritmo foi proposto por (ALOMARI; SABRI, 2006) que ao executar testes de coloração de vértices em redes aleatórias obteve resultados melhores ao utilizar a combinação de

grau de saturação e grau médio como heurística para seleção de próximo vértice a colorir, se comparado aos resultados obtidos utilizando-se apenas o grau médio dos vértices como critério de seleção. Embora grafos de redes complexas tenham características bem diferentes das encontradas em grafos de redes aleatórias, objetivou-se verificar o desempenho deste segundo algoritmo, já que parece fazer sentido utilizar o grau de saturação como critério de escolha para o próximo vértice a ser colorido, uma vez que quanto maior o número de cores utilizado na coloração dos vizinhos, mais difícil encontrar uma cor para colorir o vértice em questão.

Suponha dois passos desse algoritmo para o grafo ilustrado na Figura 30. Até o ponto ilustrado nessa figura, o vértice c , primeiro a ser colorido recebeu cor porque era o que possuía o maior grau entre todos. No passo seguinte, d recebeu cor porque dentre os vértices sem colorir era o que apresentava o maior grau entre aqueles com saturação 1.

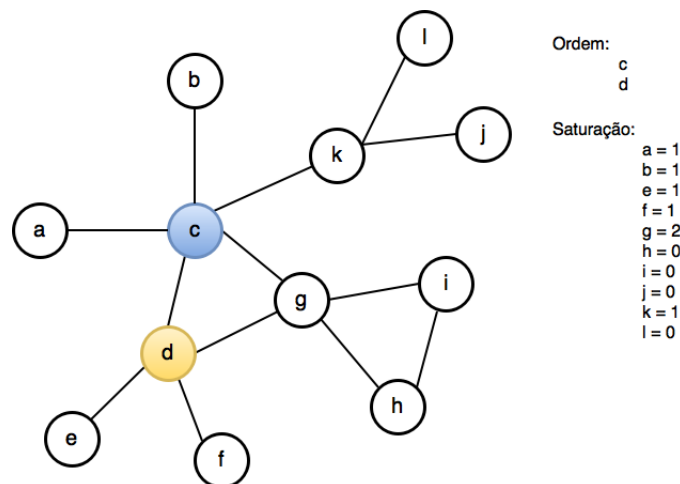


Figura 30: Exemplo de seleção do próximo vértice a colorir utilizando a heurística de maior saturação - Passo 1.

Fonte: Autoria própria.

Nesse ponto da coloração, os vértices a , b , e , f , e k possuem saturação igual a 1, pois apenas uma cor foi utilizada para colorir os vizinhos desses vértices. Os vértices h , i , j e l possuem saturação 0, já que nenhum vizinho foi colorido. Por fim, o vértice g possui grau de saturação 2 porque os seus vizinhos c e d receberam cores diferentes. Sendo assim, g será o próximo vértice a ser colorido.

Um segundo passo desse algoritmo é apresentado na Figura 31.

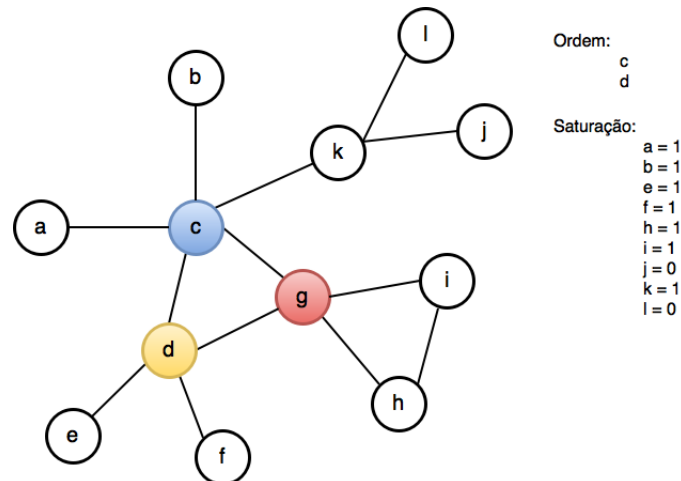


Figura 31: Exemplo de seleção do próximo vértice a colorir utilizando a heurística de maior saturação - Passo 2.

Fonte: Autoria própria.

Os vértices c , d e g já estão coloridos e busca-se o vértice de maior saturação para ser o próximo a receber cor. Observando-se a saturação dos vértices ainda sem cor, têm-se que os vértices j e l possuem saturação igual a 0 e os vértices a , b , e , f , h , i e k possuem grau de saturação igual a 1. Como critério de desempate, então será utilizado o maior grau dos vértices. Analisando-se os vértices a , b , e , f , h , i e k percebe-se que dentre esses, k possui o maior grau e por isso ele será o próximo a ser colorido.

A aplicação dessa heurística é descrito no algoritmo descrito a seguir.

ALGORITMO GULOSO COM HEURÍSTICA DE MAIOR SATURAÇÃO

INPUT: um grafo G com n vértices

OUTPUT: número de cores utilizadas para colorir G

1. Selecione o vértice v de maior saturação dentre aqueles que ainda não receberam cores.
 2. Em caso de empate, ou seja, se mais de um vértice possui o maior grau de saturação, selecione dentre esses aquele que possui o maior grau. Chame-o de v .
 3. Atribua a primeira cor disponível para v .
 4. Retorne ao passo 1 até que todos os vértices estejam coloridos.
-

Os resultados obtidos por esse algoritmo foram comparados com o resultado ótimo calculado com o auxílio do software *SAGE Library* (SAGE Math) que possui um módulo dedicado a algoritmos em grafos. Evidentemente, essa comparação foi feita sempre que possível, pois por se tratar de um problema NP-Difícil, é esperado que o software citado não consiga encontrar o resultado ótimo para os grafos com alto número de nós.

O software *SAGE Library* calcula o número cromático dos grafos utilizando um algoritmo de programação linear cujo fluxograma abreviado é apresentado na Figura 32.

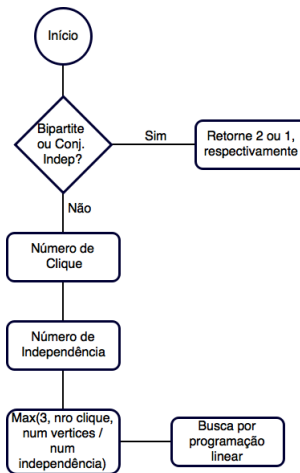


Figura 32: Fluxograma do algoritmo para cálculo do número cromático ótimo.

Fonte: Autoria própria.

Pode-se observar que o algoritmo ótimo inicializa-se verificando se a instância em questão é um grafo bipartite ou um conjunto independente. Caso isso seja verdade, o algoritmo retorna o número cromático 2 ou 1 respectivamente. Se esse critério não for satisfeito, então o algoritmo busca os limitantes do número cromático. Primeiramente, o algoritmo busca o número de clique do grafo em questão. Na sequência, o algoritmo busca o número de independência da instância. Em seguida, o algoritmo inicia a busca do número cromático com um valor k dado pelo máximo valor entre 3, número de clique e divisão do número de nós pelo número de independência. Essa busca, é então realizada utilizando-se o paradigma de programação linear.

Um dos limitantes do número cromático é o tamanho da maior clique presente no grafo. Tendo isso em mente, foi utilizado o algoritmo ótimo para a descoberta do número de clique das redes. Com isso, têm-se um limitante inferior para o problema. Analogamente, o grau máximo da rede é outro limitante inferior da coloração de vértices. Esse valor será encontrado para todas as redes, já que é um problema de ordem de complexidade linear e não haverá problemas em obtê-lo. Outro limitante verificado diz respeito à divisão do número de vértices pelo número de independência do grafo. Dessa forma, o algoritmo *SAGE Library* foi executado nos grafos selecionados para a determinação do número de independência dos grafos.

Na codificação do algoritmo com heurística gulosa para coloração de vértices de grafos foi utilizado o paradigma de programação estruturado, mais especificamente

a linguagem de programação *C*. A escolha dessa linguagem de programação se deve ao fato de que ela possui uma sintaxe simples e flexível, permitindo a criação de códigos enxutos e mais velozes do que outras linguagens. Além disso, para esta linguagem de programação existem diversos compiladores e bibliotecas intensivamente usadas, o que diminui a quantidade de falhas referentes à erros de biblioteca ou compilação.

Uma vez executados os algoritmos propostos, seus resultados foram comparados com relação ao desempenho do algoritmo guloso em relação à coloração mínima indicada pelo algoritmo ótimo. Além disso, foi verificado o quanto esses resultados se distanciam dos limitantes inferiores da coloração de vértices.

Depois de analisados os resultados, serão tecidas considerações qualitativas sobre os motivos e as estruturas das redes que favorecem a obtenção de tais resultados. Buscar-se-á, também, tecer comparações qualitativas com os resultados obtidos nos algoritmos aplicados no trabalho desenvolvido por SILVA et al. (2013) no que for plausível.

6 RECURSOS DE HARDWARE E SOFTWARE

Neste capítulo são descritos os recursos de *hardware* e *software* necessários para a realização dos experimentos que envolvem este trabalho.

6.1 RECURSOS DE HARDWARE

Para execução do projeto há duas demandas diferentes em termos de recursos de *hardware*. A codificação dos algoritmos foi feita em uma máquina com capacidade de processamento de 8GB de memória RAM, processador 2.7 GHz Intel Core i5 e sistema Operacional OSX.

Já a execução dos algoritmos exigirá uma máquina com poder de processamento e memória muito maior. É tomada como base de cálculo quanto à necessidade de *hardware* a configuração utilizada por Santos, Gimenez-Lugo e Silva (2013) em seus experimentos: R710 DELL Server executando com Intel Xenon Quad Core 5500 e 5600 e 60GB de Memória Swap + 12GB de Memória Física. Assim, o ideal é a execução dos algoritmos em máquina com configuração parecida ou superior à descrita.

Tomando-se como base a configuração utilizada por Santos, Gimenez-Lugo e Silva (2013), utilizou-se um Supercomputador classificado entre os Top500 do mundo, com 121 TeraFLOPS, 2 CPU de 16 cores cada e 64GB de memória RAM em cada nó. Para os experimentos foram alocados 64 GB de memória para processamento de cada algoritmo, seja utilizando o software SAGE *Library* ou o código C desenvolvido.

6.2 RECURSOS DE SOFTWARE

O desenvolvimento deste projeto envolveu dois recursos de *software* distintos. Para a codificação do algoritmo com heurística gulosa para coloração de vértices utilizou-se o software *XCode Beta*. Este *software* é disponibilizado para desenvolvedores *Apple*, de tal forma que não há custos quanto a esta ferramenta de desenvolvimento.

Para a execução dos algoritmos que calculam os resultados ótimos para coloração de vértices, clique máximo e número de independência foi utilizado o conhecido *software* SAGE *Library*, *software* matemático gratuito e *open-source*, desenvolvido sob a licença GPL por uma comunidade de programadores e matemáticos, que busca ser uma alternativa para os principais sistemas proprietários de *software* matemático. Ele foi construído a partir de cerca de 100 pacotes de código aberto, sendo direcionado ao estudo de matemática elementar e avançada, pura e aplicada, o que inclui conceitos matemáticos de álgebra básica, cálculo fundamental para a teoria avançada, criptografia, computação numérica, álgebra comutativa, teoria de grupo, análise combinatória, teoria dos grafos e álgebra linear exata, sendo por isso bem adequado para o ensino e a pesquisa.

Esse *software* é suportado nas plataformas *Linux*, *Windows*, *OSX* e *Unix*. Trata-se de um *software* livre, com código fonte e executável disponível em <http://git.sagemath.org/>. A escolha deste recurso se deve ao fato de ser este um dos poucos *softwares* matemáticos gratuitos e também pela facilidade de uso do mesmo. Por ser uma ferramenta *open-source*, não haverá custos para a sua utilização, de tal forma que não houveram custos de *hardware* ou *software* para o desenvolvimento deste trabalho de conclusão de curso.

7 RESULTADOS

Inicialmente, para a execução dos testes, foram selecionados vinte e cinco grafos representativos de redes complexas reais, obtidos das bases de dados Gephi, Pajek e Stanford. Buscou-se selecionar redes que contivessem número de vértices e arestas distintos, para que fosse possível obter resultados não viciados em relação ao tamanho do grafo sobre os quais seriam executados os testes. Dessa forma, as redes selecionadas representam diferentes tipos de redes e possuem quantidades de vértices que variam de 235 a 75879. A tabela 2 apresenta a descrição das redes nas quais foram executados os testes propostos.

Tabela 2: Detalhes das redes complexas selecionadas.

ID	Rede	Vértices	Arestas	Tipo	Classificação	Origem
1	Airlines	255	1297	Não direcionada	Tecnológica	Gephi
2	USAir	332	2126	Não direcionada	Tecnológica	Pajek
3	Codeminer	724	1017	Direcionada	Social	Gephi
4	CPanAuthors	839	2222	Não direcionada	Social	Gephi
5	EuroSis	1285	6462	Não direcionada	Informação	Gephi
6	Oclinks	1899	13838	Não direcionada	Informação	Gephi
7	YeastS	2361	7182	Não direcionada	Biológica	Pajek
8	CA-QrQc	5242	14496	Não direcionada	Citação	Stanford
9	p2p-Gnutella08	6301	20777	Direcionada	Comunicação	Stanford
10	Wiki-Vote1	7115	8347	Direcionada	Social	Stanford
11	p2p-Gnutella09	8114	26013	Direcionada	Informação	Stanford
12	p2p-Gnutella06	8717	31525	Direcionada	Informação	Stanford
13	p2p-Gnutella05	8846	31839	Direcionada	Informação	Stanford
14	CA-HepTh	9877	25998	Não direcionada	Citação	Stanford
15	p2p-Gnutella04	10876	39994	Direcionada	Colaboração	Stanford
16	CA-AstroPh	18772	198110	Não direcionada	Colaboração	Stanford

(continua)

Tabela 2: Detalhes das redes complexas selecionadas.

(conclusão)

ID	Rede	Vértices	Arestas	Tipo	Classificação	Origem
17	p2p-Gnutella25	22687	54705	Direcionada	Informação	Stanford
18	CA-CondMat	23133	93497	Não direcionada	Colaboração	Stanford
19	p2p-Gnutella24	26518	65369	Direcionada	Informação	Stanford
20	Cit-HepTh	27770	352324	Direcionada	Citação	Stanford
21	p2p-Gnutella30	36682	88328	Direcionada	Informação	Stanford
22	Email-Enron	36692	183831	Não direcionada	Comunicação	Stanford
23	Brightkite_edges	58228	214078	Não direcionada	Social	Stanford
24	p2p-Gnutella31	62586	147892	Direcionada	Informação	Stanford
25	soc-Epinions1	75879	405740	Direcionada	Social	Stanford

A quantidade de arestas indicada na tabela 2 representa a rede resultante após eliminados loops e multiarestas. Essa tabela indica a quantidade de vértices e arestas sob os quais foram executados os testes propostos nesse documento. Nos casos de redes direcionadas, os testes foram executados sobre o grafo subjacente obtido a partir do grafo da rede original.

As principais características dos grafos representativos das redes complexas selecionadas estão descritas na tabela 3.

Tabela 3: Características das redes complexas selecionadas.

ID	Grau Médio	Largura Média de caminho	Diâmetro	Densidade	Total de Comunidades	# Componentes Conexas	Coefficiente de Clustering	Total de Triângulos
1	11,038	2,31	4	0,047	5	1	0,652	3688
2	12,807	2,73	6	0,039	8	1	0,749	12181

(continua)

Tabela 3: Características das redes complexas selecionadas.

(conclusão)

ID	Grau Médio	Largura Média de caminho	Diâmetro	Densidade	Total de Comunidades	# Componentes Conexas	Coefficiente de Clustering	Total de Triângulos
3	2,809	8,13	19	0,004	29	15	0,288	167
4	5,297	2,79	6	0,006	9	1	0,531	2032
5	10,058	3,89	10	0,008	17	6	0,381	12117
6	14,574	3,055	8	0,008	13	4	0,138	14319
7	6,084	4,376	11	0,003	124	59	0,271	4659
8	5,531	6,048	17	0,001	390	354	0,687	48282
9	6,595	4,642	9	0,001	24	2	0,015	2383
10	2,346	18,71	39	0	64	13	0,001	2
11	6,412	4,766	10	0,001	31	6	0,014	2354
12	7,233	4,571	10	0,001	28	1	0,009	1142
13	7,199	4,596	9	0,001	29	3	0,009	1112
14	5,264	5,945	18	0,001	472	428	0,6	28404
15	7,355	4,635	10	0,001	29	1	0,008	934
16	21,107	4,193	14	0,001	326	290	0,677	1352167
17	4,823	5,544	11	0	67	13	0,009	806
18	8,083	5,352	15	0	613	567	0,706	173623
19	4,93	5,417	11	0	68	11	0,009	986
20	25,374	4,278	15	0,001	165	143	0,33	1479195
21	4,816	5,749	11	0	68	12	0,011	1590
22	10,02	4,025	13	0	1263	1065	0,716	727044
23	7,353	7,371	18	0	786	547	0,271	494728
24	4,726	10,817	11	0	77	12	0,01	2024
25	10,694	10,694	11,549	0	870	2	0,261	1624481

Pode-se observar que a densidade de todas as redes analisadas é muito baixa,

variando de 0 a 0,047. Isso indica que embora as redes sejam compostas por muitos vértices, a quantidade de arestas existente é muito menor do que a quantidade máxima possível de arestas. Ao mesmo tempo, o grau médio dos vértices varia entre 2,346 e 25,374, o que ilustra que há uma certa diferença entre as redes selecionadas, pois embora todas tenham densidade média muito baixa, a média de conexões dos vértices da rede varia significativamente. Vale destacar, ainda, que essas redes apresentam o fenômeno do caminho estreito, ou seja, a largura média do caminho entre dois vértices quaisquer da rede é relativamente baixo, variando entre 2,31 e 18,71. Finalmente, cabe destacar a variação no coeficiente de clusterização das redes selecionadas. O mesmo varia entre 0,001 e 0,749, demonstrando que há uma certa variedade quanto à tendência à formação de grupos entre os vizinhos de um vértice qualquer da rede.

Após analisadas as características das redes selecionadas foram então executados os algoritmos de coloração de vértices sobre as redes processadas (loops e arestas múltiplas foram removidas). Foram aplicados dois algoritmos de coloração gulosa, cuja diferença reside na forma de escolha do próximo vértice a ser colorido. No primeiro deles, algoritmo guloso com heurística de maior grau, a escolha do vértice a ser colorido leva em conta apenas o grau dos vértices. Assim, o vértice ainda não colorido que possui maior grau tem prioridade para coloração, já que acredita-se que quanto maior for o grau de um vértice mais difícil será atribuir uma cor previamente utilizada na coloração de outros vértices.

Já no segundo algoritmo de coloração gulosa de grafos, algoritmo guloso com heurística de maior saturação, utiliza a heurística de maior saturação para definir o próximo vértice a ser colorido, como detalhado na metodologia deste trabalho. O maior grau é, então, utilizado como critério de desempate caso mais de um vértice detenha a saturação máxima do grafo.

Sendo assim, a tabela 4 apresenta o número de cores utilizadas para a coloração dos grafos ao se executar os algoritmos de coloração com heurística gulosa descritos acima.

Tabela 4: Resultados da coloração gulosa.

ID da Rede	Heurística do Maior Grau	Heurística da Maior Saturação
1	12	11
2	24	23
3	5	4
4	9	8
(continua)		

Tabela 4: Resultados da coloração gulosa.

(conclusão)

ID da Rede	Heurística do Maior Grau	Heurística da Maior Saturação
5	13	13
6	11	10
7	9	9
8	44	44
9	7	7
10	4	3
11	7	6
12	7	6
13	7	5
14	32	32
15	7	6
16	57	57
17	7	5
18	26	26
19	7	6
20	28	26
21	7	5
22	29	25
23	40	40
24	7	6
25	42	31

Pode-se perceber que os resultados obtidos com o algoritmo guloso utilizando a heurística da maior saturação foram melhores do que aqueles resultantes da heurística de maior grau. O algoritmo de maior saturação resulta em resultados iguais ou melhores que o algoritmo de maior grau, uma vez que a heurística de maior saturação utiliza menos cores para colorir os grafos.

Para que seja possível analisar o desempenho do algoritmo guloso é preciso compará-lo com a coloração ótima ou os limitantes inferiores e superiores do número de coloração. Dessa forma, utilizou-se o software *Sage* para calcular a coloração ótima e

os limitantes inferiores dados pelo número clique e da relação com o número de independência.

Para cálculo da coloração ótima, o software *Sage* utiliza programação linear inteira. Inicialmente, a função que calcula o número cromático verifica se o grafo em questão é bipartite ou um conjunto independente de vértices. Se esse for o caso, o algoritmo retorna 2 ou 1 respectivamente. Caso o grafo não se encaixe em nenhuma dessas classificações, o algoritmo busca o tamanho da maior clique presente no grafo e o tamanho do maior conjunto independente. Em seguida, inicia-se a busca de uma coloração com k cores onde o menor valor de k é dado pelo máximo entre 3, $\omega(G)$ (número clique) e $\frac{|g|}{\alpha(G)}$ (total de vértices dividido pelo número de independência do grafo), incrementando-se o valor de k toda vez que há falha para um determinado k . Esse algoritmo retorna o número cromático do grafo.

O número clique, por outro lado, é calculado por meio da interface do *Sage* com o software *Cliquer* (CLIQUER,), desenvolvido por Sampo Niskanen com base no algoritmo exato branch-and-bound criado por Patric Östergård. O *Cliquer* é um conjunto de rotinas desenvolvidas na linguagem C que almeja encontrar a maior clique presente num grafo. O algoritmo é exato e seus detalhes podem ser encontrados em Östergård (2002).

Por fim, o tamanho do maior conjunto independente pode ser calculado utilizando a interface com o software *Cliquer* ou Programação Linear. Nos experimentos realizados, optou-se por utilizar a interface com o *Cliquer* devido a tempo de execução.

A tabela 5 apresenta os resultados obtidos para a coloração ótima dos grafos das redes selecionadas. Cada uma das instâncias foram computadas com limite de tempo igual a 7 dias (ou 168 horas), tempo limite máximo permitido pelo computador onde foram executados os testes.

Tabela 5: Resultados da coloração gulosa.

ID da Rede	Número Cromático	Maior Grau	Maior Saturação
1	11	12	11
2	timeout	24	23
3	4	5	4
4	8	9	8
5	timeout	13	13

(continua)

Tabela 5: Resultados da coloração gulosa.

(conclusão)

ID da Rede	Número Cromático	Maior Grau	Maior Saturação
6	timeout	11	10
7	9	9	9
8	44	44	44
9	timeout	7	7
10	3	4	3
11	timeout	7	6
12	timeout	7	6
13	timeout	7	5
14	32	32	32
15	timeout	7	6
16	timeout	57	57
17	timeout	7	5
18	timeout	26	26
19	timeout	7	6
20	timeout	28	26
21	timeout	7	5
22	timeout	29	25
23	timeout	40	40
24	timeout	7	6
25	timeout	42	31

Pode-se perceber que na maioria das instâncias o tempo limite de execução foi atingido sem que a coloração ótima fosse encontrada pelo algoritmo. Ao mesmo tempo, pode-se verificar que o fator decisivo para o estouro do tempo limite de execução não está diretamente relacionado ao tamanho da instância de entrada, uma vez que por exemplo a rede de ID 2 possui menos vértices e arestas que o grafo de ID 3. E isso se repete para outras instâncias.

A princípio pode parecer que mesmo para os menores grafos selecionados, a quantidade de combinações é muito grande para que um algoritmo possa chegar ao seu final com sucesso, por se tratar este de um problema NP-Difícil. No entanto, o caráter NP-

Difícil de um problema indica o pior caso possível, o que não implica que toda e qualquer instância será de difícil solução. O fato de coloração de vértices ser um problema NP-Difícil indica que para todo algoritmo polinomial é possível encontrar uma instância tal que a solução da mesma se torna difícil (ou seja, exige-se tempo exponencial para que a solução ótima seja encontrada a menos que $P = NP$).

O fato de que o tempo limite de execução é atingido pode se dever a algumas possibilidades. Uma razão possível é que no caso da coloração de vértices as estruturas que pioram o desempenho do algoritmo de programação linear utilizado para solução do problema são encontradas com alguma frequência. Por outro lado, o algoritmo de coloração de vértices começa a buscar a solução ótima para k cores, onde k é dado pelo máximo entre 3, $\omega(G)$ (número clique) e $\frac{|G|}{\alpha(G)}$ (número de nós dividido pelo número de independência do grafo). Se os valores do número clique e número de independência do grafo foram distantes do número cromático do grafo, o algoritmo pode precisar executar muitos passos até que a coloração ótima seja encontrada, ocorrendo o estouro do tempo limite de execução. Finalmente, para algumas instâncias o algoritmo de coloração ótima pode ter ficado preso na função que busca o tamanho do maior conjunto independente, já que ao se executar apenas a busca do número independente do grafo houve estouro de tempo limite na busca de tal valor, como pode ser visto na sequência desta seção.

Para os casos em que a coloração ótima foi encontrada, é possível calcular a aproximação garantida pelo algoritmo guloso. Esses dados são apresentados na tabela 6.

Tabela 6: Aproximação dada pela coloração gulosa em relação ao número cromático.

ID da Rede	Maior Grau	Maior Saturação
1	1,0909	1,0
2	-	-
3	1,25	1,0
4	1,125	1,0
5	-	-
6	-	-
7	1,0	1,0
8	1,0	1,0
9	-	-
10	1,3333	1,0
(continua)		

Tabela 6: Aproximação dada pela coloração gulosa em relação ao número cromático.

(conclusão)

ID da Rede	Maior Grau	Maior Saturação
11	-	-
12	-	-
13	-	-
14	1,0	1,0
15	-	-
16	-	-
17	-	-
18	-	-
19	-	-
20	-	-
21	-	-
22	-	-
23	-	-
24	-	-
25	-	-
Média	1,1142	1,0

Analisando-se os dados da tabela acima pode-se perceber que para o algoritmo guloso que utiliza a heurística da maior saturação, a aproximação é de 1,0. Ou seja, o algoritmo guloso em questão encontrou o número cromático do grafo. O algoritmo guloso com heurística maior grau, por outro lado, encontrou resultados que são 1,1142 do ótimo, em média. No entanto, esses resultados podem estar viciados já que esses valores são calculados apenas sobre os casos onde foi possível encontrar a coloração ótima. As instâncias com '-' nas células referentes à aproximação indicam as instâncias onde não foi possível encontrar o número cromático.

Uma vez que os resultados acima apresentados podem estar viciados pelo fato de representarem uma pequena fatia das instâncias selecionadas, torna-se necessário analisar o desempenho do algoritmo guloso em relação aos limitantes inferiores e superiores da coloração de vértices.

O limitante inferior $\frac{|G|}{\alpha(G)}$ se deve ao fato de que o conjunto independente de um

grafo divide o mesmo em grupos de vértices cujos elementos podem ser coloridos com a mesma cor. Existe a possibilidade de todos os conjuntos independentes de um grafo G possuírem o mesmo tamanho, de onde surge o limitante inferior dado por $\frac{|G|}{\alpha(G)}$. Ou seja, G pode ser dividido em no mínimo $\frac{|G|}{\alpha(G)}$ grupos, permitindo o surgimento de tal limitante inferior.

Além disso, é importante analisar o valor do grau do vértice de maior grau do grafo. O grau máximo do grafo é parte do limitante superior da coloração de vértices, uma vez que para colorir um grafo são necessárias no máximo $\Delta(G) + 1$ cores, onde $\Delta(G)$ é o grau máximo do grafo. Isso se explica pelo fato de que se todos os vizinhos do vértice de maior grau v_i formam uma clique entre si, então cada um dos vizinhos de v_i receberá uma cor diferente na coloração, de tal forma que para colorir v_i será necessária a utilização de uma cor distinta das aplicadas aos vizinhos. Se pegarmos o vértice de maior grau, então seu grau + 1 será um limitante superior para a coloração de vértices.

Por fim, tão importante quanto os demais limitantes, é o número de clique do grafo. Nenhum grafo pode ser colorido com menos de k cores onde k é o número de clique do grafo. Isso ocorre porque numa clique todos os vértices estão conectados entre si, e, portanto, uma coloração própria exige uma cor para cada vértice. Sendo assim, o número de clique é um dos limitantes inferiores do problema de coloração de vértices, motivo pelo qual deve ser analisado.

A tabela 7 apresenta os resultados encontrados para os limitantes inferiores e superiores para o problema da coloração de vértices. São apresentados o número clique e o limitante inferior dado por $\frac{|G|}{\alpha(G)}$, onde $\alpha(G)$ é o número de independência do grafo. Além disso, é apresentado o grau máximo de cada grafo.

Tabela 7: Limitantes inferiores e superiores do número cromático.

ID da Rede	$\omega(G)$	$\alpha(G)$	$\frac{ G }{\alpha(G)}$	$\Delta(G)$	$\Delta(G) + 1$
1	11	139	1,69	130	131
2	22	183	1,89	139	140
3	4	533	1,35	55	56
4	8	723	1,16	329	330
5	13	688	1,86	100	101
6	7	1150	1,65	255	256
7	9	1598	1,47	66	67

(continua)

Tabela 7: Limitantes inferiores e superiores do número cromático.

(conclusão)

ID da Rede	$\omega(G)$	$\alpha(G)$	$\frac{ G }{\alpha(G)}$	$\Delta(G)$	$\Delta(G) + 1$
8	44	2459	2,13	81	82
9	5	4247	1,48	97	98
10	3	5749	1,23	63	64
11	5	5540	1,46	102	103
12	4	5312	1,64	115	116
13	4	5418	1,63	88	89
14	32	4896	2,01	65	66
15	4	6528	1,66	103	104
16	57	-	-	504	505
17	4	16670	1,36	66	67
18	26	-	-	281	282
19	4	19309	1,37	355	356
20	23	-	-	2468	2469
21	4	27414	1,33	55	56
22	20	-	-	1383	1384
23	37	36361	1,60	1134	1135
24	4	46893	1,33	95	96
25	23	53599	1,41	3044	3055

Observando-se a tabela que descreve os limitantes do número cromático pode-se perceber que o número de independência do grafo não parece representar um bom limitante inferior para fins de comparação com o número cromático, uma vez que em torno da metade dos vértices de cada grafo está contido no seu maior conjunto independente. Por esta razão, o número de independência dos grafos é em torno de 2, muito distante dos valores encontrados para o número cromático, nos casos em que ele pode ser encontrado. Então, embora o número de independência mostre que para todos os grafos analisados o número cromático aproximado ou exato está correto, esse limitante não parece adequado para fins de determinação de razão de aproximação. O mesmo se aplica para o limitante superior dado pelo grau máximo dos grafos. Este limitante superior corrobora os resultados obtidos pelos algoritmos de heurística gulosa, mas é pouco útil para fins de definição

de razão de aproximação garantida pelo algoritmo de coloração com estratégia gulosa.

O número clique, por sua vez, parece ser um bom indicador de comparação, já que seu valor se altera entre as diferentes instâncias de redes selecionadas. Sendo assim, dentre todos os limitantes citados, o número clique parece ser aquele que fornece o melhor limitante para fins de comparação.

Assim sendo, a tabela 8 apresenta a razão de aproximação garantida pelo algoritmo guloso para o problema da coloração de vértices em relação ao número clique nas instâncias analisadas.

Tabela 8: Aproximação garantida pelo algoritmo guloso em relação ao número clique.

ID da Rede	$\omega(G)$	Maior Grau	Maior Saturação
1	11	1,09	1,0
2	22	1,09	1,04
3	4	1,25	1,0
4	8	1,125	1,0
5	13	1,0	1,0
6	7	1,57	1,42
7	9	1,0	1,0
8	44	1,0	1,0
9	5	1,4	1,4
10	3	1,33	1,0
11	5	1,4	1,2
12	4	1,75	1,5
13	4	1,75	1,25
14	32	1,0	1,0
15	4	1,75	1,5
16	57	1,0	1,0
17	4	1,75	1,25
18	26	1,0	1,0
19	4	1,75	1,5
20	23	1,21	1,13
21	4	1,75	1,25
22	20	1,45	1,25

(continua)

Tabela 8: Aproximação garantida pelo algoritmo guloso em relação ao número clique.

(conclusão)

ID da Rede	$\omega(G)$	Maior Grau	Maior Saturação
23	37	1,08	1,08
24	4	1,75	1,5
25	23	1,82	1,34
Média		1,36	1,16

Podemos perceber que a aproximação média garantida para o algoritmo guloso é muito boa, e como esperado, é melhor para o algoritmo que utiliza heurística de maior saturação na seleção do próximo vértice a colorir. Cabe ressaltar que a aproximação dada pelo algoritmo guloso pode ser maior do que a indicada na tabela. Isso ocorre porque o número cromático dos grafos pode ser maior que o número de clique encontrado para os mesmos, já que o número de clique é um limitante inferior para o problema da coloração de vértices.

Contudo, o que se pode afirmar é que no pior caso, o algoritmo guloso com heurística de maior grau na média conseguiu, em média, uma aproximação de 1,36 da solução ótima e o algoritmo guloso com heurística de maior saturação garantiu uma solução que na média é 1,16 da ótima. Isso mostra que o algoritmo guloso tem boa performance em redes de sistemas complexos, corroborando os resultados encontrados por Silva et al (2013) com relação à utilização de algoritmos gulosos em grafos que representam sistemas complexos.

Um fato interessante que merece ser destacado diz respeito ao fato de que aparentemente o número cromático dos grafos analisados está relacionado ao número de clique do mesmo. A impressão que se tem é que o número cromático, é de fato, determinado pelo número de clique, já que nas instâncias em que foi possível obter o número cromático exato esses valores são iguais aos do número de clique. O mesmo se aplica a outras instâncias quando comparados o número cromático aproximado pelo algoritmo guloso e o número de clique encontrados.

Considerando que a literatura apresenta um algoritmo polinomial que garante aproximação na ordem de $O\left(\frac{(n(\log \log n)^2)}{\log^3 n}\right)$, pode-se afirmar que os algoritmos gulosos testados neste experimento possuem eficiência maior do que o melhor algoritmo encontrado

na literatura. Isso indica a validade da utilização de algoritmos com heurística gulosa para grafos reais. Como grafos reais possuem características bem específicas, não se pode afirmar que o bom desempenho do algoritmo guloso nas instâncias testadas pode ser estendido para outros tipos de problema. Porém, para grafos de redes reais, a utilização de algoritmos com heurística gulosa é totalmente válida no trato de problemas reais que possam se apresentar.

Considerando que este trabalho é parte de um projeto maior, algo que se destaca refere-se ao fato de que deve existir algo na estrutura dos grafos de sistemas complexos que faz com que o algoritmo de coloração por programação linear utilizado no cálculo da coloração ótima fique preso em algum ponto e não consiga obter o resultado mesmo para instâncias não tão grandes.

Esse resultado difere consideravelmente dos resultados obtidos por SILVA et al. (2013) e mesmo dos resultados obtidos neste trabalho para problemas NP-Difíceis em grafos (número clique e número de independência). No caso do problema da coloração de vértices, encontrar o número cromático não está diretamente relacionado ao número de vértices e arestas da instância a ser analisada, o que parece ser verdade para o problema da cobertura por vértices investigada por SILVA et al. (2013). Quanto ao número de clique, não é possível tecer considerações, já que foi possível encontrar a solução para todas as instâncias, embora tenha sido perceptível a diferença de tempo necessária para a obtenção das soluções para as maiores instâncias analisadas. Já no problema do número de independência, foi possível encontrar a solução ótima para a grande maioria das instâncias, sendo que novamente o tamanho da instância não parece estar fortemente ligada a encontrar ou não a solução ótima.

Sendo assim, estender esse trabalho para entender que estruturas dos grafos de sistemas complexos favorecem alguns dos algoritmos citados e quais estruturas pioram o desempenho dos algoritmos utilizados se apresenta como uma interessante linha de pesquisa. Com isso, acredita-se, será possível construir algoritmos que evitem as estruturas prejudiciais e favoreçam bons resultados mais positivos para problemas em grafos.

8 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho buscamos estudar o desempenho do algoritmo guloso na coloração de vértices oriundos de sistemas complexos. Dentre os grafos dos vinte e cinco sistemas complexos analisados, em apenas sete casos foi possível encontrar a solução ótima, nos demais casos ocorreu estouro do tempo limite de execução do algoritmo de busca da solução ótima. Há várias possíveis explicações para este fato, dentre elas: aproximação por número de clique e número de independência muito diferente do número cromático dos grafos, poucas combinações com k cores geram uma coloração válida, algoritmo de busca da solução ótima fica preso na busca pelo número de independência do grafo onde se executa o algoritmo de coloração ótima, conforme detalhado no capítulo 7.

Outro ponto que merece destaque é que no caso dos grafos analisados, a dificuldade em encontrar o número cromático não está diretamente relacionada ao tamanho da instância. Foi possível encontrar o número cromático para instâncias de tamanho maior do que algumas onde ocorreu *timeout* na execução. Isso demonstra que há outras razões fundamentais para que ocorra o estouro do tempo limite de execução. Possivelmente, o que ocorre é que as estruturas que fazem com que o algoritmo de Programação Linear utilizado na busca pela solução ótima aparecem nos grafos selecionados e fazem com que a performance do algoritmo seja ruim. Talvez se outra abordagem para a busca da solução ótima fosse utilizada isso não ocorresse. Porém, coloração de vértices é um problema de otimização combinatorial e, portanto, é esperado que em muitas instâncias o algoritmo exija muito tempo para solução, uma vez que este tipo de problema é um dos mais difíceis que se conhece.

Aparentemente, o número cromático dos grafos selecionados é bem próximo (ou igual) ao número de clique desses grafos, como pode ser observado nos resultados obtidos. Isso indica que em grafos de redes reais, se for mais simples o cálculo do número clique (o que parece ser verdade, considerando o encontrado nos experimentos realizados), este dará uma boa aproximação do número cromático desejado, podendo ser um bom substituto ao cálculo do número cromático em muitas situações.

A aproximação garantida pelo algoritmo guloso com heurística combinada de maior saturação e maior grau é bem melhor do que aquela garantida pelo melhor algoritmo polinomial apontado na literatura que indica uma solução $O(\frac{n(\log \log n)^2}{\log^3 n})$ da ótima. A aproximação garantida pelo algoritmo guloso, em média 1,16 do ótimo, é bem melhor do que a melhor garantida pela literatura. Obviamente isso não é válido para todo tipo de grafo, mas para grafos reais, que encontramos nos problemas reais do dia-a-dia, a utilização do algoritmo guloso certamente não pode ser desprezada ou ignorada.

Os resultados obtidos nesse trabalho corroboram o que foi encontrado por SILVA et al. (2013). Tanto para o problema da cobertura por vértices apresentado por SILVA et al. (2013) quanto no problema da coloração de vértices, a aproximação encontrada pelo algoritmo guloso é considerável, mostrando que é válido estudar como o algoritmo guloso se comporta para outros problemas em grafos. Talvez a abordagem gulosa garanta uma aproximação boa o suficiente para aplicações em grafos de problemas reais, sendo, portanto, útil estudar como heurísticas gulosas se comportam nos problemas de mundo real.

Os resultados encontrados neste trabalho sugerem outros possíveis trabalhos futuros. O primeiro deles envolve estudar o comportamento do algoritmo guloso para outros problemas em grafos, por exemplo, para o problema do número de independência de grafos e problema de coloração de arestas. Isso poderá indicar a importância da heurística gulosa para problemas em grafos reais.

Outro trabalho futuro derivado deste refere-se a analisar as estruturas e características dos grafos de sistemas complexos que favorecem ou prejudicam o desempenho do algoritmo guloso para coloração de vértices. Entender porque o algoritmo guloso consegue bons resultados pode permitir o surgimento de novos algoritmos ou novas abordagens eficientes para solução de problemas em grafos. Esse tipo de trabalho exige abordagens e metodologias totalmente diferentes das aplicadas neste trabalho, provavelmente exigindo mais tempo do que o disponível para a realização de um trabalho de conclusão de curso.

Por outro lado, entender quais estruturas dos grafos de sistemas complexos prejudicam a performance do algoritmo de Programação Linear na busca pela solução ótima pode contribuir para a criação de novos algoritmos capazes de minimizar os problemas encontrados pelo algoritmo utilizado e, portanto, permitir soluções melhores para o problema da coloração de vértices, possivelmente com custo computacional inferior aos atualmente utilizados.

REFERÊNCIAS

- ALOMARI, H.; SABRI, K. E. New graph coloring algorithms. **American Journal of Mathematics and Statistics**, v. 2, n. 4, p. 739–741, 2006.
- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, v. 74, n. 1, p. 47–97, Jan 2002.
- ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. **Error and attack tolerance of complex networks**. 2000. Nature 406.
- ALMEIDA, L. J. **Detecção de comunidades em redes complexas utilizando estratégia multinível**. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação - ICMP - USP, 2009.
- AMARAL, L.; OTTINO, J. Complex networks. **The European Physical Journal B - Condensed Matter and Complex Systems**, Springer-Verlag, v. 38, n. 2, p. 147–162, 2004. ISSN 1434-6028. Disponível em: <<http://dx.doi.org/10.1140/epjb/e2004-00110-5>>.
- BARABÁSI, A. L.; ALBERT, R.; JEONG, H. Scale-free characteristics of random networks: the topology of the world-wide web. **Physics**, v. 281, p. 69–77, 2000.
- BARABÁSI, A. L.; BONABEAU, E. Scale-free networks. **Scientific American**, v. 288, p. 50–59, 2003.
- BERGER, B.; ROMPEL, J. A better performance guarantee for approximate graph coloring. **Algorithmica**, v. 5, n. 1, p. 459–466, 1990. ISSN 1432-0541. Disponível em: <<http://dx.doi.org/10.1007/BF01840398>>.
- BIRKHOFF, G. D. A determinant formula for the number of ways of coloring a map. **Annals of Mathematics**, Annals of Mathematics, v. 14, n. 1/4, p. pp. 42–46, 1912. ISSN 0003486X. Disponível em: <<http://www.jstor.org/stable/1967597>>.
- BONDY, J. A.; MURTY, U. S. **Graph Theory With Applications**. [S.l.]: Elsevier Science Publishing So. Inc., 1979.
- BONDY, J. A.; MURTY, U. S. R. **Graph Theory**. [S.l.]: Springer, 2008.
- CHARTRAND, G.; ZHANG, P. **Chromatic Graph Theory**. [S.l.]: United States of America, 2009.
- CLIQUER. Disponível em: <<https://users.aalto.fi/~pat/cliquer.html>>.
- CORMEN, T. H. et al. **Introduction to Algorithms**. [S.l.]: Cambridge, Massachusetts, 2009.
- COSTA, L. F. et al. Characterization of complex networks: A survey of measurements. In: **Advances in Physics**. [S.l.: s.n.], 2008.

- COSTA, P. P. da. **Teoria de Grafos e suas Aplicações**. Dissertação (Mestrado) — Universidade Estadual Paulista “Júlio de Mesquita Filho”, 2011.
- DIESTEL, R. **Graph Theory**. [S.l.]: New York, 2005.
- FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**. [S.l.: s.n.], 2011.
- FERNANDES, L. D. **Padrões de Turing e processos dinâmicos em Redes Complexas**. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2012.
- FIGUEIREDO, D. R. Introdução a redes complexas. In: _____. [S.l.]: Atualizações em Informática 2011, 2011. cap. 7, p. 303–358.
- FORMANOWICZ, P.; TANÁŠ, K. A survey of graph coloring - its types, methods and applications. **Foundations of Computing and Decision Sciences**, v. 37, p. 223–238, 2012.
- GEPHI Datasets. Disponível em: <<https://github.com/gephi/gephi/wiki/Datasets>>.
- HALLÓRSSON, M. M. A still better performance guarantee for approximate graph coloring. **Inf. Process. Lett.**, Elsevier North-Holland, Inc., Amsterdam, The Netherlands, The Netherlands, v. 45, n. 1, p. 19–23, jan. 1993. ISSN 0020-0190. Disponível em: <[http://dx.doi.org/10.1016/0020-0190\(93\)90246-6](http://dx.doi.org/10.1016/0020-0190(93)90246-6)>.
- JHONSON, D. S. Worst case behaviour of graph coloring algorithms. **Proc. 5th Southeastern Conf. on Combinatorics, Graph Theory, and Computing.**, Congressus Numerantium X, p. 513–527, 1974.
- KARP, R. M. **Reductibility Among Combinatorial Problems**. [S.l.]: Plenum Press, 1972.
- KHANNA, S.; LINIAL, N.; SAFRA, S. On the hardness of approximating the chromatic number. **Proceedings of the 2nd Israel Symposium on Theory and Computing Systems**, p. 250–260, 1993.
- LEWIS, H. R.; PAPADIMITRIOU, C. H. **Elements of Theory of Computation**. [S.l.]: New Jersey - United States of America, 1998.
- LIPSCHUTZ, S.; LIPSON, M. L. **Teoria e Problemas de Matemática Discreta**. 3. ed. [S.l.]: Bookman, 2004.
- LOPES, G. A. W. **Um Modelo de Rede Complexa para Análise de Informações Textuais**. Dissertação (Mestrado) — Centro Universitário da FEI, 2011.
- LOZANO, D. **Modelagem Matemática e Aplicações do Problema de Coloração em Grafos**. Dissertação (Mestrado) — Universidade Estadual Paulista, 2007.
- MELLO, A. G. S. **Aplicação de Redes Complexas para a Definição de Vizinhança na Otimização por Enxame de Partículas**. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2010.
- MOODY, J. Race, school integration, and friendship segregation in america. **American Journal of Sociology**, v. 107, p. 679–716, 2001.

- MOURA, L. M. N. de P. **Identificação de comunidades e exploração visual em redes sociais: rede do comércio internacional europeu**. Dissertação (Mestrado) — Faculdade de Engenharia da Universidade do Porto, 2012.
- NAVARRO, F.; COELHO, F. **Estudo de Coloração Aplicado ao Problema de Alocação de Horário de Professores**. 2009. Disponível em: <<http://www.unipac.br/bb/tcc/tcc-5486411b6e467ad999a2a853d986d947.pdf>>.
- NEWMAN, M. The physics of networks. **Physics Today**, v. 61, Issue 11, p. 33–38, 2008.
- NEWMAN, M.; BARABÁSI, A. L.; WATTS, D. J. **The Structure and Dynamic of Networks**. [S.l.]: United States of America, 2006.
- NEWMAN, M. E. J. The structure and function of complex networks. **SIAM REVIEW**, v. 45, p. 167–256, 2003.
- ÖSTERGARD, P. R. J. A fast algorithm orthe maximum clique problem. **Discrete Applied Mathematics**, v. 120, p. 197–207, 2002.
- PAJEK Datasets. Disponível em: <<http://vlado.fmf.uni-lj.si/pub/networks/data/>>.
- REGO, M. F.; SANTOS, H. G. **Algoritmos para o Problema de Coloração de Grafos**. 2011. Disponível em: <<http://www.decom.ufop.br/menotti/paa111/files/PCC104-111-ars-11.1-MarceloFerreiraRego.pdf>>.
- RODRIGUES, F. A. **Caracterização, classificação e análise de redes complexas**. Tese (Doutorado) — Universidade de São Paulo, 2007.
- SAGE Math. Disponível em: <<http://www.sagemath.org>>.
- SILVA, M. O. D.; GIMENEZ-LUGO, G. A.; SILVA, M. V. G. D. Vertex cover in complex networks. **International Journal of Modern Physics C**, v. 24, n. 11, p. 1350078, 2013. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S0129183113500782>>.
- STANFORD Large Network Data Collection. Disponível em: <<https://snap.stanford.edu/data/>>.
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of /'small-world' networks. **Nature**, n. 393, p. 440–442, June 1998.
- WEST, D. B. **Introduction to Graph Theory**. [S.l.: s.n.], 2002.
- WEST, D. B. **Introduction to Graph Theory**. Second. [S.l.: s.n.], 2012.
- WIGDERSON, A. Improving the performance guarantee for approximate graph coloring. **J. ACM**, ACM, New York, NY, USA, v. 30, n. 4, p. 729–735, out. 1983. ISSN 0004-5411. Disponível em: <<http://doi.acm.org/10.1145/2157.2158>>.