

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIA DA COMPUTAÇÃO – BACHARELADO

**GERAÇÃO DE REDES COMPLEXAS
COM COMUNIDADES SOBREPOSTAS E
COMUNIDADES HIERÁRQUICAS**

GUSTAVO HENRIQUE SPIESS

BLUMENAU
2022

GUSTAVO HENRIQUE SPIESS

GERAÇÃO DE REDES COMPLEXAS COM COMUNIDADES SOBREPOSTAS E COMUNIDADES HIERÁRQUICAS

Trabalho de Conclusão de Curso apresentado ao curso de graduação em Ciências da Computação no Centro de de Ciências Exatas e Naturais da Universidade Regional de Blumenau como requisito parcial para a obtenção de grau de Bacharel em Ciências da Computação.

Professor Aurelio Faustino Hoppe, Mestre - Orientador

FOLHA DE ASSINATURAS

Dedico esse trabalho a minha noiva, cuja paciência em me ouvir falar desse trabalho tornou-o possível.

AGRADECIMENTOS

A meu padrinho, Maiko Rafael Spiess, pelo sempre presente incentivo ao estudo.

Ao meu orientador, Aurélio Faustino Hoppe, por acreditar na conclusão desse trabalho.

A minha família, por todos os anos de apoio que foram necessários para chegar até aqui.

Aos amigos que fiz no percurso do bacharelado, pelo apoio recebido.

Aos professores do Departamento de Sistemas e Computação da Universidade Regional de Blumenau por suas contribuições durante os semestres letivos.

“Se eu vi mais longe, foi por estar sobre ombros
de gigantes.”

Isaac Newton

RESUMO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Palavras-chave: Redes complexas. Geração de redes complexas. Comunidades. Comunidades sobrepostas. Comunidades hierárquicas.

ABSTRACT

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Keywords: Complex networks. Complex networks generation. Communities. Overlapping communities. Hierarchical communities

LISTA DE FIGURAS

Figura 1 – Exemplo de grafo	16
Figura 2 – Exemplo	18

LISTA DE QUADROS

Quadro 1 – Função modularidade Q	18
Quadro 2 – Função modularidade estendida EQ	18
Quadro 3 – Exemplo	18

LISTA DE TABELAS

Tabela 1 – Exemplo	19
------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

IE – Internet Explorer

QE – Qternet Explorer

SUMÁRIO

1	Introdução	14
1.1	Objetivos	15
1.2	Estrutura	15
2	Fundamentação teórica	16
2.1	Redes complexas e comunidades	16
3	Considerações finais	20
	Referências	21

1 INTRODUÇÃO

Redes complexas, como definido por [Metz et al. \(2007\)](#), são grafos com uma topologia não trivial. Isso é, são grafos onde parte ou toda a informação de interesse está contida não nos vértices e arestas individualmente, mas em propriedades do conjunto de vértices e arestas.

Como apontado por [Girvan e Newman \(2002\)](#), um dos sistemas do mundo real que se pode modelar em uma rede complexa é o conjunto de relações sociais. Uma modelagem simplista desse sistema apresenta é a representação de cada indivíduo como um vértice, e vértices adjacentes sendo pares de indivíduos que se conhecem. Nesse tipo de sistema um sub grafo completo, denominado clique ([FORTUNATO, 2010](#)), pode ser interpretada como uma propriedade relevante a indicação de que desse conjunto de indivíduos onde todos conhecem todos.

[Girvan e Newman \(2002\)](#) também aponta que outros sistemas, como cadeias alimentares, cadeias de metabolização, redes de transmissão elétrica e redes de computadores podem ser representadas como redes complexas. Muitas vezes propriedades que se observam em redes complexas de um domínio estão presentes também nas redes complexas de outros domínios, mas com interpretações distintas sobre o objeto modelado. O trabalho de [Fortunato \(2010\)](#) indica isso na discussão de múltiplas interpretações do que constitui uma comunidade em uma rede complexa, dividindo-se principalmente em características estruturais, e por semelhança de vértice.

No exemplo do trabalho desenvolvido por [Larger et al. \(2015\)](#), as duas interpretações se encontram presentes, como é característica da literatura a despeito da geração de redes complexas. [Larger et al. \(2015\)](#) descreve o que é chamado na literatura de um modelo de geração algorítmica de redes complexas onde os vértices do grafo estão dispostos em uma nuvem de ponto e a distribuição deles em diferentes comunidades leva em conta sua posição espacial, e as arestas são construídas em função desse pertencimento a uma comunidade. [Akoglu e Faloutsos \(2009\)](#) descreve um modelo mais primitivo, que não realiza a atribuição explícita de comunidades, mas que gera um grafo com essas comunidades ainda assim.

Indica-se, observando o trabalho de [Fortunato \(2010\)](#) de que ha uma vasta literatura a respeito dos processos de detecção dessas comunidades. Oberando-se a literatura da qual os trabalhos de [Larger et al. \(2015\)](#), [Akoglu e Faloutsos \(2009\)](#) e [Slota et al. \(2019\)](#), é indicada a existência dos modelos necessários para a geração de redes complexas com comunidades. No entanto propriedades adjacentes a presença de comunidades para os quais

existe literatura a respeito da detecção, como comunidades hierárquicas e comunidades sobrepostas, parecem estar pouco presentes em modelos de geradores de redes complexas.

1.1 OBJETIVOS

Dado esse contexto, o objetivo do trabalho é a adaptação dos modelos presentes na literatura de geração de redes complexas para a incorporação de comunidades sobrepostas e comunidades hierárquicas.

Os objetivos específicos são:

- a) A construção de um modelo algorítmico de geração de redes complexas que inclua a propriedade de comunidades.
- b) A especificação, dentro desse modelo, de uma *ground truth* de quais vértices pertencem a quais comunidades.
- c) A possibilidade, dentro desse modelo, de comunidades hierárquicas.
- d) A possibilidade, dentro desse modelo, de comunidades sobrepostas.
- e) A representação, dentro desse modelo, dos vértices como uma nuvem de pontos, para a definição de semelhança de vértices por distância.

1.2 ESTRUTURA

Esse trabalho se estrutura em quatro capítulos sendo o primeiro uma introdução aos temas abordados, bem como a apresentação dos objetivos do trabalho.

O segundo capítulo apresenta a fundamentação teórica da pesquisa, descrevendo o estado da arte do objeto de estudo.

O terceiro capítulo discute o desenvolvimento do modelo algorítmico proposto, incluindo ferramentas e técnicas utilizadas. Também são apresentados os blocos de pseudo código do modelo.

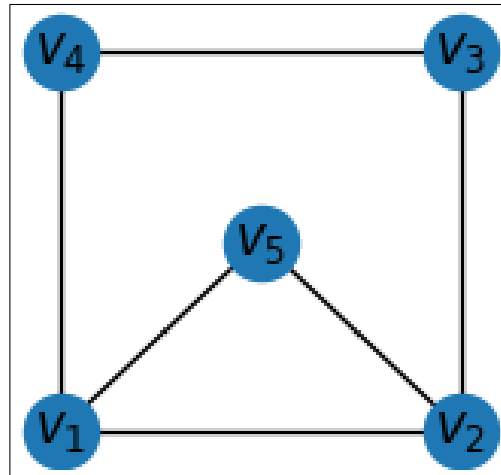
O quarto capítulo compõe os dados obtidos na avaliação dos resultados, bem como quaisquer discussões de implementações futuras ou outras formas de continuação.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 REDES COMPLEXAS E COMUNIDADES

Grafos podem trivialmente ser definidos como $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ onde \mathcal{V} é um conjunto dos vértices de \mathcal{G} e \mathcal{E} é um conjunto de pares não ordenados de vértices adjacentes em \mathcal{G} , i.e. as arestas.

Figura 1 – Exemplo de grafo



Fonte: elaborado pelo autor

No exemplo da [Figura 1](#), pode-se representar o mesmo grafo com $\mathcal{V} = \{v_1, v_2, v_3, v_4, v_5\}$ e $\mathcal{E} = \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_4, v_1\}, \{v_1, v_5\}, \{v_2, v_5\}\}$. Dentro desse grafo, o subgrafo formado pelos vértices v_1, v_2 e v_5 é também um grafo completo. A esse conjunto de vértices que forma um subgrafo completo dá-se o nome de clique ([FORTUNATO, 2010](#)). Essa descrição de clique é utilizada como uma definição inicial do conceito de comunidade, partindo de uma perspectiva local ([FORTUNATO, 2010](#)). Dentro desse conceito de comunidade, nomeia-se as arestas cujos dois vértices estão contidos em uma comunidade como sendo interna à comunidade.

É facilmente observável, no entanto, que essa é uma definição muito limitante de comunidade, é raro que comunidades de pessoas apresentem tanta homogeneidade a ponto de todos os membros conhecerem todos os outros membros. De fato, [Fortunato \(2010\)](#) indica que a definição precisa do que é uma comunidade varia de acordo também com o contexto de estudo, mas que algumas características são universais. Uma comunidade, dentro de qualquer definição, deve ser um sub grafo conexo, por exemplo.

As definições são agrupadas em três classes distintas por [Fortunato \(2010\)](#):

- Definição local

- Definição global
- Definição por similaridade de vértice

Essas definições não são mutuamente exclusivas, mas também não são ortogonais uma a outra. Segundo [Fortunato \(2010\)](#), a definição local parte das características topológicas internas á comunidade. Nominalmente, isso significa a existência de um conjunto considerável de arestas internas a comunidade e um conjunto limitado de arestas para além da comunidade.

A definição global de comunidades é aplicada aos casos onde a presença de clusters é uma característica inerente ao grafo que se está estudando [Fortunato \(2010\)](#). Essa propriedade inerente ao grafo pode ser definida como alguma propriedade dos vértices do objeto em questão e que partindo disso se atribui pertencimento á comunidades, ou ainda por comparação com um exemplo nulo [Fortunato \(2010\)](#). No caso de comparação com um exemplo nulo, define-se uma comunidade pela característica de uma não ser presente dentro de o que é chamado de “grafo aleatório” ([FORTUNATO, 2010](#)). Essa definição de um modelo nulo é crucial para o trabalho de [Girvan e Newman \(2002\)](#), o modelo nulo considerado é um onde o grafo original é alterado de forma aos graus de todos os vértices se manterem, mas a probabilidade de dois vértices estarem ligados é constante independente de quais os vértices.

Por fim, a definição de comunidade por similaridade de vértice se baseia na tendencia de que em muitas aplicações, membros de comunidades são mais similares entre si do que seria esperado de um conjunto do mesmo tamanho escolhido aleatoriamente ([FORTUNATO, 2010](#)). Essa definição se faz visível no trabalhos de [Akoglu e Faloutsos \(2009\)](#) e de [Larger et al. \(2015\)](#). Na observação desses dois trabalhos também é interessante o questionamento de como se define semelhança, [Akoglu e Faloutsos \(2009\)](#) representa os vértices como sequencias de caracteres de tamanhos variáveis em que a probabilidade de dois vértices estarem ligados é maior conforme mais caracteres eles compartilham; e [Larger et al. \(2015\)](#) representa os vértices como pontos em um espaço n -dimensional e define que vértices são mais semelhantes quando a distância euclideana deles é menor.

Também independente de qual definição de comunidade que se esteja utilizando, existem os conceitos de partição e cobertura. Segundo [Fortunato \(2010\)](#), uma partição é uma divisão dos vértices de um grafo tal que cada vértice pertença a um e exatamente um cluster. O caso de um vértice “livre”, não pertencendo a nenhuma comunidade, é trivialmente resolvido incluindo ele á comunidade com a qual ele mais tem jacências.

Mas o caso de vértices que pertençam a mais de uma comunidade, isso é, comunidade que se sobreponham é mais interessante. Fortunato (2010) define uma cobertura como uma divisão dos vértices em clusters onde cada vértice pertence a um ou mais clusters. Fortunato (2010) também descreve o conceito de comunidades hierárquicas, como sendo comunidades cuja estrutura interna também se organiza em clusters de escala menor do que o original.

Por fim, Fortunato (2010) oferece também o conceito de “função de qualidade”, sendo uma função que mapeia uma partição para um espaço de comparação, usualmente em números reais, onde partições que mapeiem para valores maiores são consideradas melhores. Segundo Fortunato (2010) função de qualidade mais comumente utilizada é a modularidade Q de Girvan e Newman (2002).

Quadro 1 – Função modularidade Q

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{K_i K_j}{2m} \right) \delta(C_i, C_j)$$

Fonte: Girvan e Newman (2002)

Essa função no entanto não se aplica adequadamente ao caso de comunidades sobrepostas ou comunidades hierárquicas, para tanto, é necessário utilizar a função de modularidade estendida, conforme desenvolvido por Shen et al. (2009).

Quadro 2 – Função modularidade estendida EQ

$$EQ = \frac{1}{2m} \sum_i \sum_{v \in C_i, w \in C_i} \frac{1}{O_v O_w} \left[A_{vw} - \frac{K_i K_j}{2m} \right]$$

Fonte: Shen et al. (2009)

Figura 2 – Exemplo

Lorem

Quadro 3 – Exemplo

Lorem

Tabela 1 – Exemplo

Lorem

3 CONSIDERAÇÕES FINAIS

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

REFERÊNCIAS

- AKOGLU, L.; FALOUTSOS, C. Rtg: A recursive realistic graph generator using random typing. In: SPRINGER. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. [S.l.], 2009. p. 13–28. Citado 2 vezes nas páginas [14](#) e [17](#).
- FORTUNATO, S. Community detection in graphs. *Physics reports*, Elsevier, v. 486, n. 3-5, p. 75–174, 2010. Citado 4 vezes nas páginas [14](#), [16](#), [17](#) e [18](#).
- GIRVAN, M.; NEWMAN, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 99, n. 12, p. 7821–7826, 2002. Citado 3 vezes nas páginas [14](#), [17](#) e [18](#).
- LARGERON, C. et al. Generating attributed networks with communities. *PloS one*, Public Library of Science, v. 10, n. 4, p. e0122777, 2015. Citado 2 vezes nas páginas [14](#) e [17](#).
- METZ, J. et al. Redes complexas: conceitos e aplicações. São Carlos, SP, Brasil., 2007. Citado na página [14](#).
- SHEN, H. et al. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 388, n. 8, p. 1706–1712, 2009. Citado na página [18](#).
- SLOTA, G. M. et al. Scalable generation of graphs for benchmarking hpc community-detection algorithms. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. [S.l.: s.n.], 2019. p. 1–14. Citado na página [14](#).