

Portfólio Técnico: Modelagem e Otimização em Redes Complexas

Autor: Jandson Fahel Oliveira de Freitas

Doutorando em Física Estatística: Sistemas Complexos

MSc. em Física Teórica e Computacional

Bacharel em Física

Contents

1	Características Gerais de Grafos e Medidas	2
1.1	Redes e Grafos	2
1.2	Grau, Grau Médio e Distribuição do Grau	3
1.2.1	Grau	3
1.2.2	Grau Médio	3
1.2.3	Distribuição do Grau	3
1.3	Matriz de Adjacências	4
1.4	Redes reais são esparsas	5
1.5	Caminhos e Distâncias	5
1.6	Coefficiente de Agrupamento	6
1.7	Análise com Redes Reais	6
2	Algoritmo para Otimização Logística em Redes	6
2.1	Descrição do Problema	7
2.2	Metodologia	8
2.3	Resultados	8
3	Algoritmo Baseado em Mapas Reais	10
3.1	Metodologia	11
3.2	Resultados	11
4	Otimização de Busca e Espalhamento em Redes Complexas	14
4.1	Descrição	14
4.2	Modelo e Simulação	14
4.3	Resultados	15

Resumo Técnico

Neste portfólio, apresento projetos que unem modelagem matemática e análise de dados para resolver problemas de otimização em redes reais. As técnicas utilizadas incluem análise de grafos, simulações de redes complexas e algoritmos de otimização, aplicadas em contextos logísticos e de transporte.

1 Características Gerais de Grafos e Medidas

1.1 Redes e Grafos

De maneira simples, podemos dizer que uma rede, ou grafo, como também pode ser chamado, é uma estrutura formada, basicamente, por nós e arestas. Podemos fazer uma analogia com as redes sociais, como o Instagram ou Facebook, por exemplo, onde os nós representam pessoas que estão cadastradas na rede e as arestas caracterizam as conexões entre os indivíduos. Na Figura 1 podemos observar um exemplo de grafo. Na imagem observamos uma rede formada por $N = 6$ nós (representados pelos círculos na cor laranja), as linhas que formam as conexões entre os nós são chamadas de arestas. Neste exemplo temos $L = 6$ arestas. Também podemos considerar as arestas como links entre os nós, estes links podem ser direcionados ou não-direcionados. No exemplo apresentado na Figura 1 temos uma rede não-direcionada.

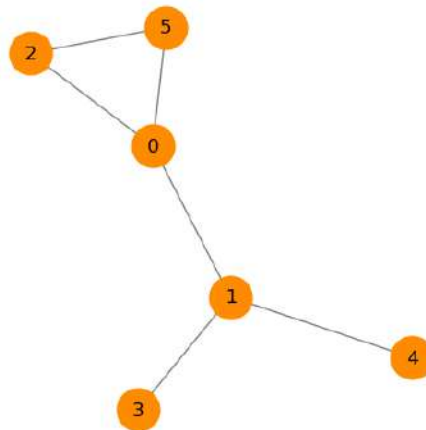


Figure 1: Imagem com objetivo de apresentar as componentes básicas de um grafo, formada por $N = 6$ nós e $L = 6$ arestas. Os nós numerados representam os indivíduos da rede e as arestas as conexões entre eles.

1.2 Grau, Grau Médio e Distribuição do Grau

1.2.1 Grau

Vamos definir o conceito de grau em uma rede. Denotamos como k_i o grau do i -ésimo nó da rede, isto é, k_i representa o número de conexões ou links que um determinado nó possui. Por exemplo, no grafo apresentado na Figura 1 temos $k_0 = k_1 = 3$, $k_3 = k_4 = 1$ e $k_2 = k_5 = 2$.

Em uma rede não-direcionada, o número total de links, L , é dado por:

$$L = \frac{1}{2} \sum_{i=1}^N k_i, \quad (1)$$

onde N é o número total de nós da rede. O fator $\frac{1}{2}$ torna-se necessário devido ao fato de contarmos o mesmo link duas vezes.

1.2.2 Grau Médio

Em uma rede não-direcionada, o grau médio da rede, $\langle k \rangle$ é dado por:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2L}{N}. \quad (2)$$

Em redes direcionadas, temos a seguinte definição, k_i^{in} que representa os links que chegam no i -ésimo nó, e, temos k_i^{out} representando os links que saem do i -ésimo nó para outros nós. Neste caso, temos que

$$k_i = k_i^{in} + k_i^{out}, \quad (3)$$

e,

$$L = \sum_{i=1}^N k_i^{in} = \sum_{i=1}^N k_i^{out}, \quad (4)$$

logo,

$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in} = \langle k^{out} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{out} = \frac{L}{N}. \quad (5)$$

1.2.3 Distribuição do Grau

Esta distribuição, representada por p_k , nos dar a probabilidade de encontrar um nó com grau k . Temos a condição de normalização, isto é,

$$\sum_{k=1}^{\infty} p_k = 1. \quad (6)$$

Para uma rede com N nós,

$$p_k = \frac{N_k}{N}, \quad (7)$$

onde N_k representa o grau do k -ésimo nó. Também podemos escrever

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k. \quad (8)$$

Na Figura 2 podemos observar um histograma do grau de cada nó da rede apresentada na Figura 1.

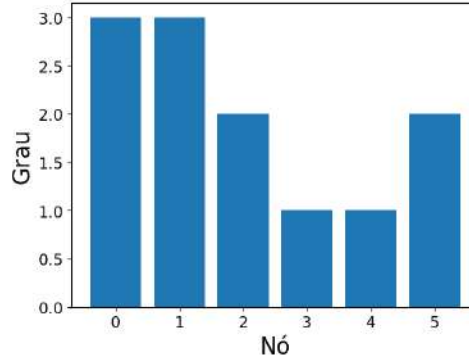


Figure 2: Distribuição do grau da rede mostrada na Figura 1.

1.3 Matriz de Adjacências

Uma descrição completa de uma rede exige que acompanhemos todas as conexões. A maneira mais simples de conseguir isto é através de uma lista completa com todas as conexões da rede. Para a rede mostrada na Figura 1 temos a seguinte lista:

$$\{[0, 1], [0, 2], [0, 5], [1, 3], [1, 4], [2, 5]\}. \quad (9)$$

No entanto, para facilitar cálculos analíticos, torna-se mais conveniente trabalhar com a matriz de adjacências A , de tamanho $N \times N$. As entradas da matriz A são definidas da forma: $A_{ij} = 1$ se existir um link entre i e j , ou, $A_{ij} = 0$ caso não exista um link entre os nós i e j . Os termos em que $i = j$, ou seja, A_{ii} são todos nulos, pois não consideramos uma ligação de um nó com ele próprio.

Para redes não-direcionadas, a matriz A possui duas entradas para cada link. O link $[1, 3]$ implica que $A_{13} = A_{31} = 1$, ou seja, a matriz de adjacências é simétrica. Podemos obter o grau de um determinado nó através de A . Para redes não-direcionadas,

$$k_i = \sum_{j=1}^N A_{ji} = \sum_{i=1}^N A_{ji}, \quad (10)$$

enquanto que, para redes direcionadas,

$$k_i^{in} = \sum_{j=1}^N A_{ij}, \quad k_i^{out} = \sum_{j=1}^N A_{ji}. \quad (11)$$

1.4 Redes reais são esparsas

Em uma rede formada por N nós, o número máximo de arestas varia entre 0 e L_{max} , onde

$$L_{max} = \frac{N(N-1)}{2}. \quad (12)$$

Entretanto, encontramos que $L \ll L_{max}$, portanto, isto implica que a maioria das entradas da matriz A sejam nulas. Logo, é mais vantajoso, computacionalmente, utilizar apenas listas com as conexões para estudar redes grandes pois utilizar a matriz de adjacências irá utilizar muita memória.

1.5 Caminhos e Distâncias

Suponha que queremos saber a distância de um caminho que leva do nó i para o nó j . Definimos a distância entre o nó i e o nó j como sendo a quantidade de arestas que compõem o caminho escolhido. Note que podemos ter diversos caminhos que podem levar do nó i para o nó j , entretanto, vamos focar apenas no caminho mais curto possível entre esses dois nós da rede.

O caminho mais curto entre os nós i e j é denotado por d_{ij} ou d . O caminho mais curto nunca contém loops e também não cruza com ele mesmo. Para redes não-direcionadas, $d_{ij} = d_{ji}$ e em redes direcionadas $d_{ij} \neq d_{ji}$. A menor distância média da rede é dada por:

$$d = \frac{1}{N(N-1)} \sum_{i,j=1, i \neq j, N} d_{ij} \quad (13)$$

Para a rede mostrada na Figura 1, temos $d_{01} = 1$, $d_{02} = 1$, $d_{03} = 2$, $d_{04} = 2$, $d_{05} = 1$, $d_{12} = 1$, $d_{13} = 1$, $d_{14} = 1$, $d_{15} = 2$, $d_{23} = 3$, $d_{24} = 3$, $d_{25} = 1$, $d_{34} = 2$, $d_{35} = 3$ e $d_{45} = 3$. Portanto, $\langle d \rangle = 28/15 \approx 1.87$.

Conhecendo a menor distância entre dois sítios, podemos escrever a matriz de distâncias entre pares $D_{ij} = d_{ij}$. Como D é simétrica, $D_{ji} = d_{ij}$.

1.6 Coeficiente de Agrupamento

Se um sítio A está conectado com o sítio B , e B está conectado com o sítio C , podemos estudar a possibilidade de que A esteja conectado com C . O coeficiente de agrupamento mede o vínculo entre os vizinhos de um determinado nó (ou sítio). Em resumo, o coeficiente de agrupamento C_i mede a densidade de link local em uma rede. Podemos definir o coeficiente de agrupamento de um nó i na forma:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}, \quad (14)$$

onde L_i representa o número de arestas entre os vizinhos k_i do nó i . Por exemplo, para o nó $i = 5$, temos que $C_5 = 1$, pois todos os seus vizinhos estão conectados entre si. Enquanto que para o nó $i = 1$, temos $C_1 = 0$, pois não há conexões entre os seus vizinhos. Para o nó $i = 0$, temos apenas uma única conexão entre os seus primeiros vizinhos, portanto, $C_0 = 1/3$. Temos então que C_i assume apenas valores entre 0 e 1. A média do coeficiente de agrupamento em uma rede é dada por:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i. \quad (15)$$

Essas métricas (grau, distância e agrupamento) são fundamentais para identificar padrões estruturais em redes sociais, biológicas e de transporte, servindo como base para modelos de otimização apresentados nos capítulos seguintes.

1.7 Análise com Redes Reais

Nesta seção iremos utilizar redes reais para fazer uma análise estatística das grandezas apresentadas nas seções anteriores. O caso que iremos estudar é uma rede não-direcionada. Esta rede relaciona páginas da Wikipédia que possuem algum tipo de conexões entre elas, isto é, dois artigos estão conectados se um contém um hiperlink para o outro. Podemos observar a rede formada por $N = 677$ nós na Figura 3. Ao fazer uma análise da distribuição dos graus de cada nó da rede, obtemos a Figura 4. Utilizando a biblioteca Networkx do python, encontramos que a menor distância média no grafo é 3.43. Também medimos o coeficiente de agrupamento para cada nó da rede, como podemos observar na distribuição apresentada na Figura 5. Também encontramos que o diâmetro da rede, isto é, o menor caminho máximo possível, é igual a 9.

2 Algoritmo para Otimização Logística em Redes

Nesta parte, desenvolvo um algoritmo que identifica o nó mais eficiente para servir como centro de distribuição. Toda a simulação foi modelada utilizando apenas a Numpy,

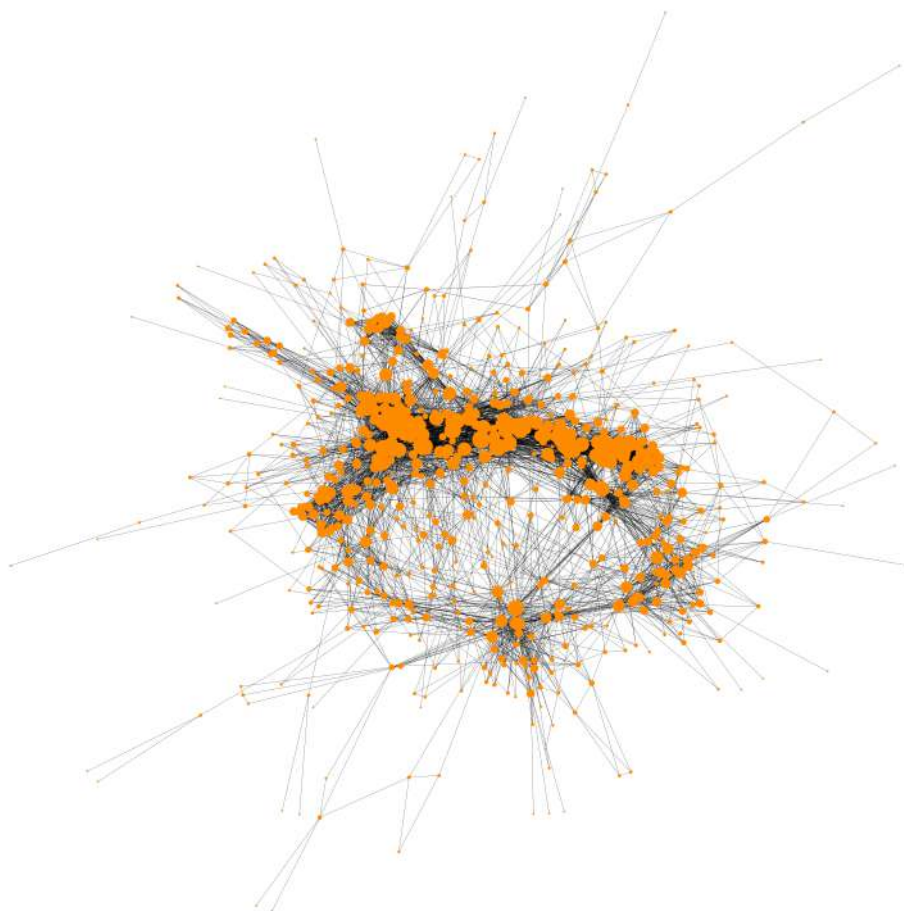


Figure 3: Rede que relaciona páginas da Wikipédia que possuem algum tipo de conexões entre elas. Representação em grafo.

garantindo todo conhecimento do algoritmo. Desde funções mais simples até as funções mais complexas.

2.1 Descrição do Problema

Neste projeto, simulei uma rede Barabási–Albert e observei o comportamento dos nós em trajetos aleatórios. O objetivo é entender qual nó seria o melhor centro de distribuição logística, ou seja, aquele que minimiza a distância total de entrega. Isto é, digamos que você sai do centro de distribuição e precisa entregar em um determinado endereço. Você sai do centro de distribuição, realiza a entrega e retorna para o ponto de partida, recarrega e parte para a próxima entrega. O objetivo do problema é realizar a entrega em todos os sítios da rede percorrendo a menor distância possível. Ou seja, estamos em busca de reduzir recursos, como combustível por exemplo.

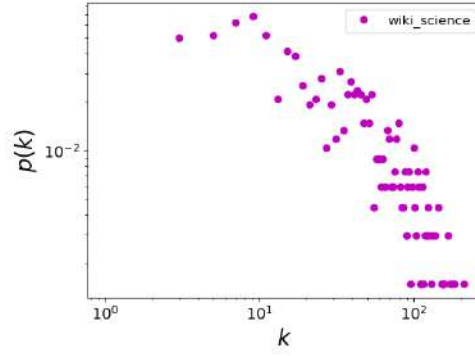


Figure 4: Distribuição do grau da rede que relaciona páginas da Wikipédia.

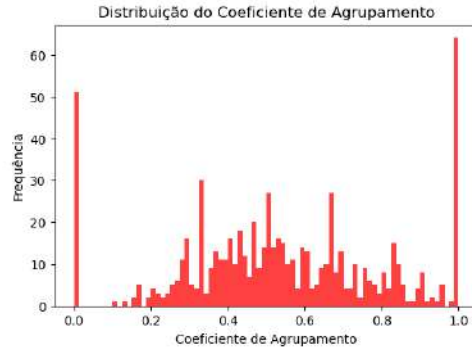


Figure 5: Distribuição do coeficiente de agrupamento da rede que relaciona páginas da Wikipédia.

2.2 Metodologia

A solução é obtida através da análise das distâncias médias entre pares de nós. O nó ótimo i^* é definido como:

$$i^* = \min_i \sum_j 2 \times d(i, j), \quad (16)$$

onde $d(i, j)$ é a menor distância entre os nós i e j .

Realizamos a simulação numa rede do tipo Barabási-Albert. A rede Barabási-Albert é conhecida por gerar hubs altamente conectados, que atuam como pontos de centralidade e influência. Esses hubs aparecem naturalmente devido ao mecanismo de ligação preferencial. Os nós mais conectados (como o nó 2, em amarelo na Figura 6) têm um papel central na estrutura, pois eles são hubs naturais da rede.

2.3 Resultados

A estrutura heterogênea da rede estudada fica clara quando analisamos a distribuição de grau. Podemos observar poucos nós com grau alto (os hubs), enquanto a maioria tem poucos vizinhos, padrão típico de redes scale-free. No histograma de grau (com média e mediana marcadas) mostrado na Figura 7, a média está deslocada para a direita devido

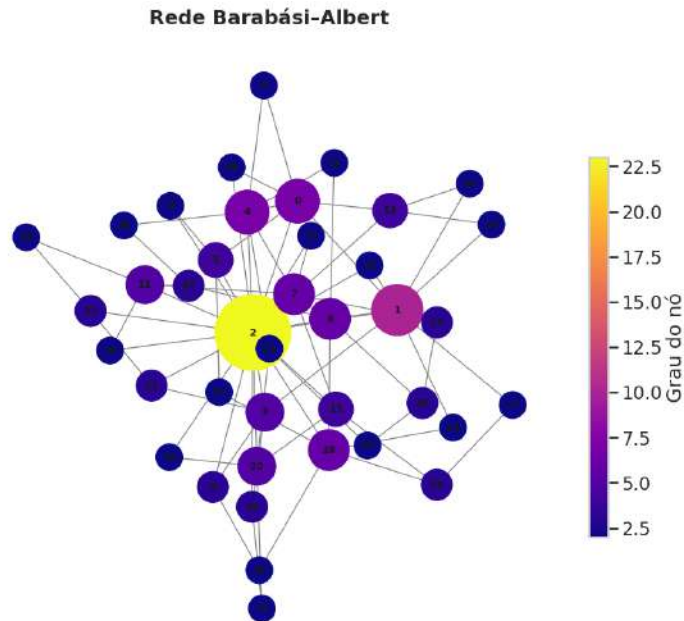


Figure 6: Rede Barabási-Albert. O tamanho do sítio está proporcional ao grau. Também temos um mapa de cores, quanto mais o nó tende ao azul menor será seu grau. Quando a cor tende ao amarelo, maior será o grau. Perceba que o fato de ser uma rede pequena, temos poucos hubs.

à presença dos hubs , evidência clássica da lei de potência nas redes complexas.

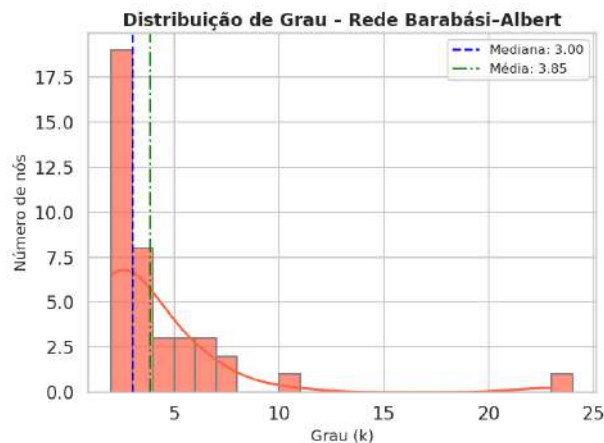


Figure 7: Distribuição de grau da rede 6.

A Matriz de distâncias mínimas entre nós (heatmap colorido) mostrado na Figura 8 mostra que ao calcular a distância mínima entre todos os pares de nós, observamos que a rede possui pequenas distâncias médias, mesmo com poucos hubs. Isso demonstra o efeito pequeno mundo (small-world effect).

Quando analisamos a distância total de entrega (ida e volta), mostrado na Figura 9, para cada nó, observamos que o nó 2 novamente se destaca. Ele apresenta a menor distância média total, o que o torna o melhor centro logístico da rede. Isto acontece

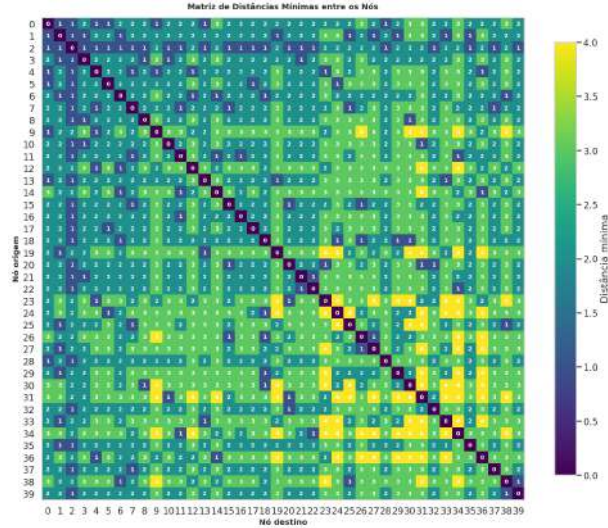


Figure 8: Matriz de distâncias mínimas entre nós da rede 6.

devido ao fato do nó 2 ser um hub. Neste caso, o hub se comporta como um atalho, minimizando a distância de entrega. Observe na Figura 8 que a distância mínima do nó 2 para os outros nós da rede é bem pequena.

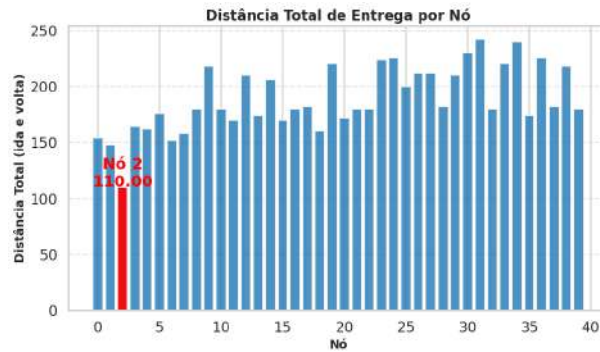


Figure 9: Distância total para realizar todas as entregas versus nó. Observe que o nó 2 é o melhor local para colocar o centro de distribuição, pois ele minimiza a distância de entrega.

Esses conceitos têm aplicações diretas em logística, transporte, comunicações e ciência de dados. Obviamente é um modelo simplificado, mas podemos tirar boas conclusões a partir dele, além de poder criar um modelo mais realista. Simulações feitas em Python com NetworkX, Matplotlib, Numpy, Seaborn e ImageIO.

3 Algoritmo Baseado em Mapas Reais

Nesta seção, demonstro o uso das bibliotecas `osmnx` (OpenStreetMap) e `networkx` para converter mapas reais em grafos e calcular rotas otimizadas. Faço também o uso da biblioteca `folium` para gerar um mapa interativo. Esta seção tem os mesmo objetivos da Seção 2, no entanto, aplicamos em uma malha viária real.

3.1 Metodologia

Este algoritmo foi criado para demonstrar como dados de mapas reais podem ser integrados a modelos de otimização. Primeiramente, com o uso da biblioteca OpenStreetMap obtemos, em forma de grafo, a malha viária do centro de Belo Horizonte. A Figura 10 mostra como as ruas podem ser representadas por arestas (em vermelho) e os cruzamentos podem ser representados por nós (em azul).

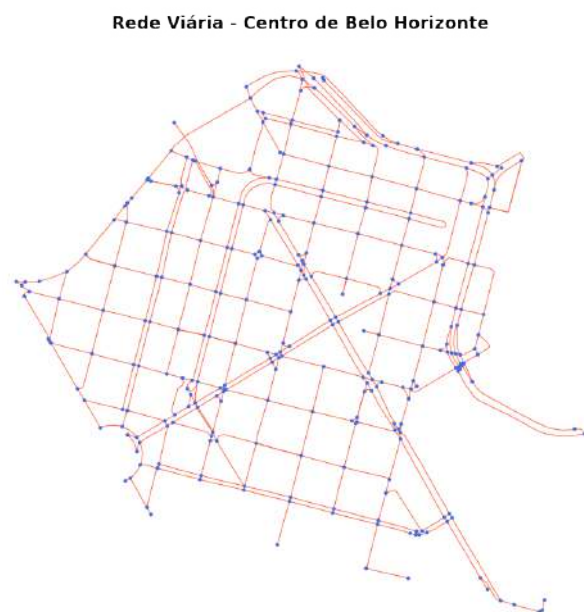


Figure 10: Malha viária do centro de Belo Horizonte representada por um grafo.

Agora com o apoio da biblioteca `folium` conseguimos criar um mapa interativo onde você pode navegar por ele e aproximar em uma região de interesse. Em casos de redes grandes, este recurso é bem viável. Na Figura 11 podemos ter uma noção, conseguimos ver o mapa real juntamente com o grafo obtido na Figura 10.

No arquivo `mapa.html` podemos fazer uma exploração da área estudada. Neste trabalho nos limitamos apenas ao centro de Belo Horizonte, mas podemos definir qualquer região, como cidades, estados ou até mesmo países. No entanto, quanto maior o tamanho da região, maior será o custo computacional. Como este projeto tem objetivo apenas de mostrar a ferramenta, no limitamos em uma pequena região.

3.2 Resultados

Como foi mostrado na Seção 2, é muito importante o cálculo da menor distância entre dois pontos da rede. Isto está diretamente relacionado em otimização de rotas para reduzir os custos com logística. Com o uso da `NetworkX` conseguimos definir o trajeto mais curto entre dois pontos do mapa levando em conta a distância física real. Além do caminho mais curto, também temos acesso às ruas que formam o trajeto. Esta ferramenta tem

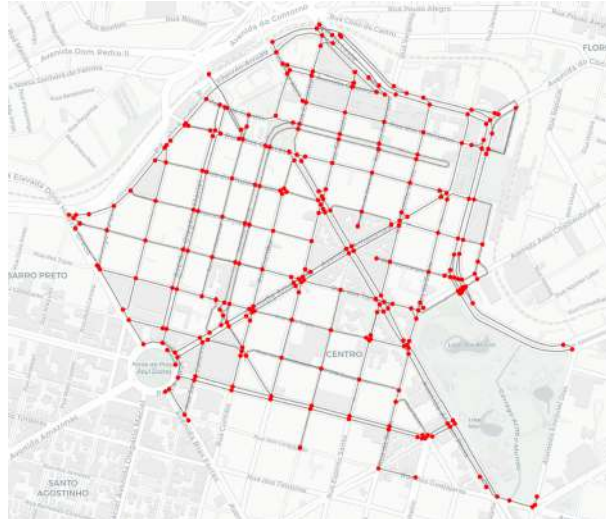


Figure 11: Amostra do mapa interativo do centro de Belo Horizonte.

aplicação direta em otimização de logística.



Figure 12: Trajeto que minimiza a distância entre dois locais do centro de Belo Horizonte.

A Figura 12 mostra a rota realizada para minimizar a distância entre dois locais da cidade. O local de partida está representado por um ponto verde enquanto que o destino final está sendo representado por um ponto vermelho. Sobre as ruas que formaram o trajeto temos a seguinte informação:

A rua inicial: Avenida Álvares Cabral. A rua final: Rua dos Tupis. A rota completa: ['Avenida Álvares Cabral', 'Avenida Afonso Pena', 'Avenida Álvares Cabral', 'Avenida Álvares Cabral', 'Avenida Álvares Cabral', 'Avenida Augusto de Lima', 'Avenida Augusto de Lima', 'Avenida Augusto de Lima', 'Rua Padre Belchior', 'Rua Padre Belchior', 'Rua Padre Belchior', 'Avenida Paraná', 'Avenida Paraná', 'Rua

dos Tupis, *'Rua dos Tupis'*].

Para regiões onde o nome da rua não está cadastrado é necessário introduzir manualmente ou usar apenas as informações de coordenadas. Zonas rurais tendem a apresentar este problema, pois é raro dar-se nome para estas estradas. Felizmente para centros urbanos, este problema ocorre com menos frequência.

Agora vamos transformar o grafo em uma rede complexa não-direcional para que possamos implementar ferramentas que foram utilizadas em rede complexa 6. Perceba que estamos introduzindo uma aproximação, pois a partir deste momento consideraremos que as distâncias entre primeiros vizinhos são iguais. Isto é, estamos dizendo que a distância de um cruzamento para outro cruzamento mais próximo é fixa para toda a rede. Podemos perder um pouco de precisão, mas será muito interessante para aplicação do modelo que encontra o local que otimiza a distribuição logística.

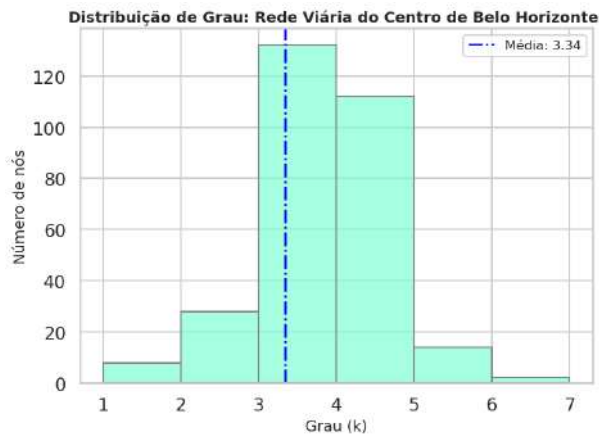


Figure 13: Distribuição de graus da rede que representa o centro de Belo Horizonte.

Com a nova rede já obtida, podemos fazer a distribuição de graus da rede, como mostra a Figura 13. O histograma nos mostra que a rede possui poucos nós com grau 1 e alguns com grau 6. A maior parte dos sítios possuem um grau bem próximo da média. Isto era esperado, visto que boa parte dos cruzamentos possuem quatro direções para seguir. O estudo de uma rede maior torna-se interessante para verificar se esta distribuição se aproxima de uma distribuição normal. Após verificar qual local do centro de Belo Horizonte é a melhor opção para alocar o centro de distribuição, descobrimos que ele está localizado na Rua dos Carijós, como mostra a Figura 14.



Figure 14: Local ideal para ser ponto de distribuição no centro de Belo Horizonte.

4 Otimização de Busca e Espalhamento em Redes Complexas

4.1 Descrição

Este projeto tem como principal objetivo mostrar minha capacidade de implementar um método a partir de um artigo científico. Nesta seção irei reproduzir resultados de textos acadêmicos publicados em periódicos internacionais. O que irei fazer agora será reproduzir um dos resultados do artigo encontrado em: [Riascos and Mateos 2012](#).

4.2 Modelo e Simulação

Neste caso estamos fazendo a parte computacional, visto que no artigo, os resultados são analíticos. O método proposto por Riascos e Mateos consiste em uma generalização das caminhadas de lévy para redes complexas. Ele cria um modelo de navegação de longo alcance onde a probabilidade de um nó ser visitado depende uma distribuição power-law.

Basicamente, ele sorteia a probabilidade de transição entre dois nós da rede, onde esta probabilidade de transição é proporcional a $d_{ij}^{-\alpha}$. Aqui, d_{ij} é o comprimento do caminho mais curto entre o nó i e o nó j . Ele sorteia um nó a ser visitado e é transferido diretamente sem mapear os nós que formam o caminho mais curto. O caminhante realiza uma transição não-local. No modelo do Riascos a probabilidade de transição é dada por:

$$w_{i \rightarrow j} = \frac{d_{ij}^{-\alpha}}{\sum_{l \neq i} d_{il}^{-\alpha}}. \quad (17)$$

Neste problema, temos um alvo numa posição aleatória da rede que deve ser encon-

trado. O objetivo é encontrar a melhor estratégia de modo que a busca torne-se mais eficiente possível.

4.3 Resultados

Após implementar o modelo de buscas aleatórias proposto por Riascos e Mateos em 2012, recuperamos um dos resultados apresentados no artigo. Calculamos o tempo médio de primeira passagem na busca por um alvo em uma rede complexa. Basicamente descobrimos a estratégia que minimiza o tempo de busca por um alvo aleatório em uma rede.

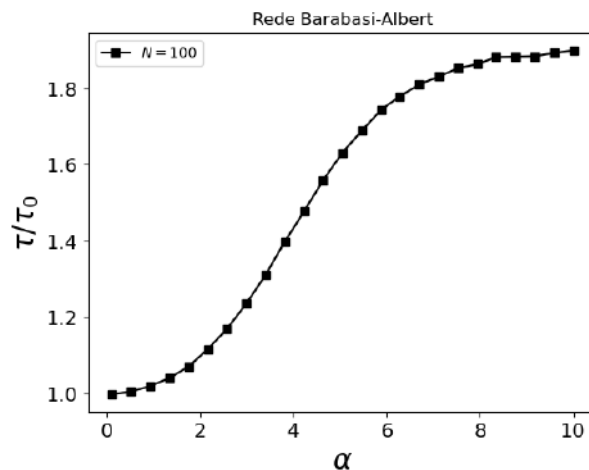


Figure 15: Tempo global τ em função de α . Reprodução de um dos resultados do artigo do Riascos e Mateos.

Digamos que vc tem um problema na rede, mas você não conhece a localização exata, este trabalho tem como objetivo encontrar a melhor estratégia para encontrá-lo o mais rápido possível. A Figura 15 mostra o resultado do tempo médio global τ em função do parâmetro da distribuição power-law α . Como podemos observar, o tempo de busca é minimizado justamente em $\alpha \rightarrow 0$.

Conclusão Geral

O conjunto de projetos apresentados neste portfólio mostra a aplicação prática de técnicas de física estatística, otimização e ciência de redes para resolver problemas reais de logística e análise de dados. As metodologias desenvolvidas podem ser facilmente adaptadas para contextos corporativos envolvendo Pesquisa Operacional e Otimização.

Contatos:

jfahelof@outlook.com

jandson.freitas@ufv.br