

Introdução a grafos e análise de redes sociais

Arthur Alves, Fausto Sá Teles e Flávio Oliveira

LAMFO



Outline

1 Introdução

2 Exemplo LinkedIn

3 Exemplo Facebook

4 Quiz

5 Referências



O que são grafos?

Definição

- São a representação de conexões entre entidades
- Definidos como um par $G = (V, E)$, onde $V = \{v_1, \dots, v_n\}$ é um set de nós e $E = \{\{v_k, v_w\}, \dots, \{v_i, v_j\}\}$ é um set de dois elementos de link, representando a conexão entre dois nós pertencentes à V

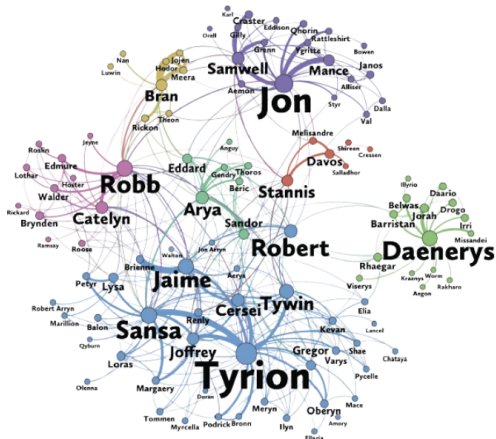
Rede social

Particularidade

- As redes sociais são grafos cujos nós representam pessoas e os links representam as conexões sociais entre eles



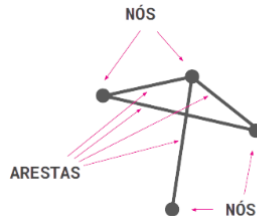
Rede social



Estruturas dos grafos

Principais componentes

- Nós: as entidades da rede, são representados por pontos
- Arestas: as conexões entre as entidades, são representadas por linhas



Propriedades básicas

Propriedades básicas

- Ordem do grafo: quantidade de nós
- Tamanho do grafo: quantidade de links
- Grau do vértice: quantidade de links adjacentes ao nó

Links

Principais conceitos

- Direcionamento: uma aresta é dita direcionada quando a informação que parte de um nó ao outro não pode fazer o caminho de volta por ela mesma.
- Peso



Principais conceitos

- Densidade: razão entre o número de arestas na rede sobre o número total de arestas possíveis entre todos os pares de nós
- Clusters: Subgrupos de nós mais conectados entre si
- Rede assortiva: viés para formação de ligações entre vértices semelhantes
- Rede dissortiva: viés para formação de ligações entre vértices diferentes
- Grau de reciprocidade: Mede a frequência de loops entre nós em redes direcionais

Métricas

Categorias principais

- Métricas de integração
- Métricas de segregação
- Métricas de centralidade
- Métricas de resiliência

Métricas de integração

Principais métricas

- Comprimento médio do caminho mais curto: média de todos os caminhos mais curtos entre todos os pares de nós possíveis
- Eficiência global e local: média do inverso do caminho mais curto

Métricas de segregação

Principais métricas

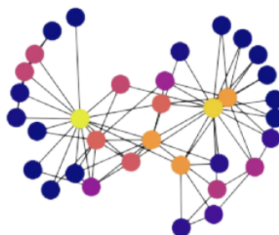
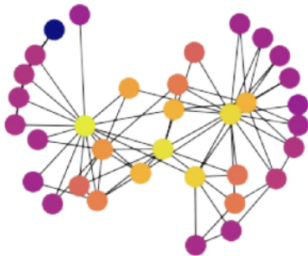
- Coeficiente de cluster: medida do grau de junção entre os nós
- Transitividade: Variação do coeficiente de cluster
- Modularidade: quantifica a divisão de uma rede em conjuntos agregados de nós altamente interconectados (clusters)

Métricas de centralidade

Principais métricas

- Centralidade de grau: número de relações em determinado nó - ou a quantidade de arestas que se ligam a ele
- Proximidade: média da distância entre um nó e todos os outros nós da rede
- Intermediação: quantifica quanto um nó funciona como uma ponte entre outros nós

Métricas de centralidade



Métricas de resiliência

Principais métricas

- Coeficiente de assortividade: quantifica a tendência de nós se conectarem a nós similares. O método mais utilizado para calcular essa correlação é a correlação de Pearson

Outline

1 Introdução

2 Exemplo LinkedIn

3 Exemplo Facebook

4 Quiz

5 Referências



Análise de rede social - LinkedIn do LAMFO

Time Divulgação

- Trabalha pela divulgação da produção científica do LAMFO, com presença no LinkedIn, Instagram, Facebook e YouTube
- LinkedIn é atualmente a rede na qual o LAMFO tem maior penetração, com mais de 500 seguidores em abril/2022

Exemplo prático 1: analisar o perfil dos seguidores do LAMFO

- Por localidade
- Por setor de atuação



Análise de rede social - LinkedIn do LAMFO

Passo a passo da análise

- Baixar dados
- Instalar dependências e importar bibliotecas e dados
- Reduzir dimensões dos dataframes (para melhor visualização)
- Gerar grafos e visualizações

LinkedIn do LAMFO: baixando os dados

LinkedIn interface showing the profile of LAMFO - Machine Learning Laboratory in Finance and Organizations. The page includes a search bar, navigation icons (Início, Minha rede, Vagas, Mensagens, Notificações, Eu), and a dropdown menu for the 'Análise' tab. The main content area displays the company name, logo, and a description: 'LAMFO - Machine Learning Laboratory in Finance and Organizations. Technology for Intelligent Decision Making. Think tanks - Brasília, Distrito Federal - 522 seguidores'.



LinkedIn do LAMFO: baixando os dados

LAMFO - Machine Learning Laboratory in Finance and Organizati... Versão do administrador de conteúdo Visualizar como usuário

Início Conteúdo ▾ **Análise ▾** Atividades Ferramentas administrativas ▾

Análise de seguidores ⓘ Exportar

Todas as datas e horários de análise exibidos em UTC

Exportar relatório de seguidores

Os dados serão exportados em um arquivo .XLS. As datas estão em UTC.

Intervalo de tempo

Últimos 12 meses ▾

Dados diários de 1 de abr. de 2021 a 31 de mar. de 2022

LinkedIn do LAMFO: preparando ambiente e importando dados

▼ Install dependencies, import libraries and data

```
!pip install pyvis --quiet
!pip install --upgrade xlrd

[2] import pandas as pd
    from IPython.core.display import display, HTML
    from pyvis import network as net
    import networkx as nx

[4] df_local = pd.read_excel("/content/sample_data/lamfo_followers.xls", sheet_name="Localidade")

[5] df_local.info()

[6] df_setor = pd.read_excel("/content/sample_data/lamfo_followers.xls", sheet_name="Setor")
```



LinkedIn do LAMFO: dataframes

▼ Lamfo's LinkedIn



```
print(df_local.shape)
df_local_reduced = df_local.loc[df_local['Total de seguidores']>=3]
print(df_local_reduced.shape)
```

```
(66, 2)
```

```
(13, 2)
```



```
[28] print(df_setor.shape)
df_setor_reduced = df_setor.loc[df_setor['Total de seguidores']>=10]
print(df_setor_reduced.shape)
```

```
(77, 2)
```

```
(11, 2)
```

LinkedIn do LAMFO: código - localidades

```
# initialize graph
g = nx.Graph()
g.add_node('lamfo.unb') # initialize Lamfo as central node

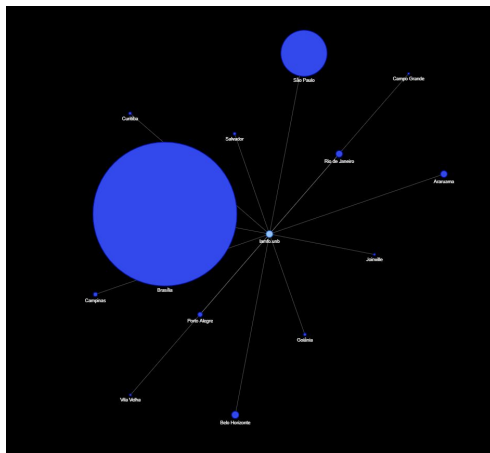
# use iterrows to iterate through the data frame
for _, row in df_local_reduced.iterrows():

    total = f"{row['Total de seguidores']}"
    local= row['Localidade']

    g.add_node(local, size=total, color='#3449eb', title=local)
    g.add_edge('lamfo.unb', local, color='grey')

# generate the graph
nt = net.Network(height='700px', width='700px', bgcolor="black", font_color='white')
nt.from_nx(g)
nt.hrepulsion(node_distance=325)
nt.show('local_graph.html')
display(HTML('local_graph.html'))
```

LinkedIn do LAMFO: visualização - localidades



Link para Datapane - localidades

LinkedIn do LAMFO: código - setores

```
# initialize graph
g = nx.Graph()
g.add_node('lamfo.unb') # initialize yourself as central

# use iterrows to iterate through the data frame
for _, row in df_setor_reduced.iterrows():
    total = f"{row['Total de seguidores']}"
    setor= row['Setor']

    g.add_node(setor, size=total, color='#3449eb', title=setor)
    g.add_edge('lamfo.unb', setor, color='grey')

# generate the graph
nt = net.Network(height='700px', width='700px', bgcolor="black", font_color='white')
nt.from_nx(g)
nt.hrepulsion(node_distance=300)
nt.show('sector_graph.html')
display(HTML('sector_graph.html'))
```

LinkedIn do LAMFO: visualização - setores



Link para Datapane - setores

Outline

1 Introdução

2 Exemplo LinkedIn

3 Exemplo Facebook

4 Quiz

5 Referências



Análise de rede social - Facebook

Redes sociais

- Grande fonte de informação
- Problema complexo
- Exemplo prático 2: análise usando grafos
- Predição de links
- Detecção de comunidades

Análise de rede social - Facebook

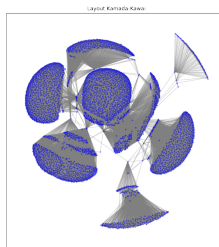
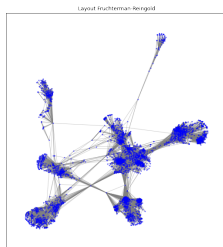
Dataset

- Social circles SNAP Facebook public dataset
- Ego-networks de 10 usuários
- Arestas: existe se dois usuários são amigos
- Features de nós: anonimizadas
- Formação de um grafo único

Análise de rede social - Facebook

Rede agregada

- Biblioteca networkx
- Número de nós: 4039
- Número de arestas: 88234
- Grau médio: 43



Facebook - métricas

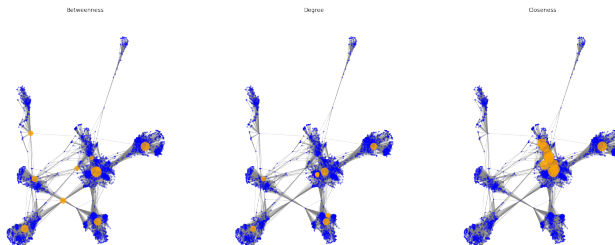
Topologia (visão geral)

- Assortatividade
- Transitividade
- Coeficiente de Clustering

Facebook - métricas

Medidas de centralidade

- Centralidade de Intermediação
- Centralidade de Grau
- Centralidade de Proximidade



Facebook - métricas

Top 10 nós

```

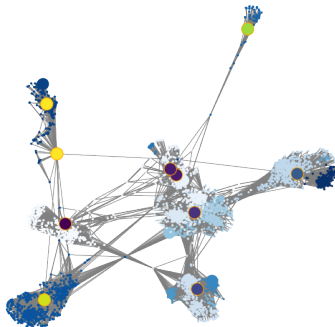
❶ dicio = {'betweenness centrality': bet_cent_list,
           'degree centrality': deg_cent_list,
           'closeness centrality': closeness_cent_list}
metricas = pd.DataFrame(dicio)
metricas

```

	betweenness centrality	degree centrality	closeness centrality
0	107	107	107
1	1684	1684	58
2	3437	1912	428
3	1912	3437	563
4	1085	0	1684
5	0	2543	171
6	698	2347	348
7	567	1888	483
8	58	1800	414
9	428	1663	376

Facebook - detecção de comunidades

Comunidades + ego users



Facebook - machine learning

Predição de links

- Aprendizado supervisionado (train/test sets)
- Biblioteca Stellargraph
- Embeddings (Node2Vec, GraphSAGE)

Facebook - predição de links

Node2Vec

```
[ ] 1 param_grid = {
2     'n_estimators': [10, 25, 50, 100, 200],
3     'max_depth': [5, 10, None]
4 }
5
6 clf = GridSearchCV(estimator=rf, param_grid=param_grid, cv=5)
7 clf.fit(train_embeddings, labels_train)
8 print(clf.best_params_)

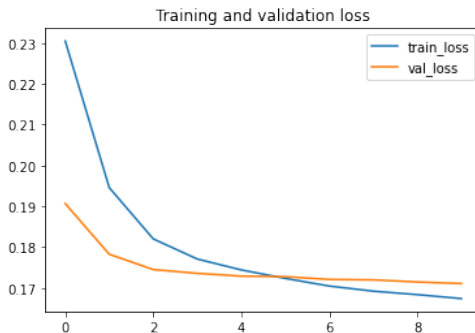
{'max_depth': None, 'n_estimators': 100}

[ ] 1 rfc = RandomForestClassifier(n_estimators=100, max_depth=None, random_state=101)
2 rfc.fit(train_embeddings, labels_train);
3 y_pred = rfc.predict(test_embeddings)
4 print(metrics.classification_report(labels_test, y_pred))
```

	precision	recall	f1-score	support
0	0.98	0.97	0.97	8823
1	0.97	0.98	0.97	8823
accuracy			0.97	17646
macro avg	0.97	0.97	0.97	17646
weighted avg	0.97	0.97	0.97	17646

Facebook - predição de links

GraphSAGE



Facebook - predição de links

GraphSAGE

```
[ ] 1 y_pred = np.round(model.predict(test_flow)).flatten()
    2 print(metrics.classification_report(labels_test, y_pred))
```

	precision	recall	f1-score	support
0	0.99	0.63	0.77	8823
1	0.73	0.99	0.84	8823
accuracy			0.81	17646
macro avg	0.86	0.81	0.80	17646
weighted avg	0.86	0.81	0.80	17646

Outline

1 Introdução

2 Exemplo LinkedIn

3 Exemplo Facebook

4 Quiz

5 Referências



Quiz

Instruções para participar do Quiz no Sli.do

- Use a Câmera do celular para ler o QR Code ou entre no site e utilize o código
- Insira seu nome e entre no Quiz (join)
- Você terá 20 segundos para responder cada questão.
- Lute pela VITÓRIA!



Outline

- 1 Introdução
- 2 Exemplo LinkedIn
- 3 Exemplo Facebook
- 4 Quiz
- 5 Referências**



Referências

Livros

- C. Stamile, A. Marzullo, E. Deusebio. Graph Machine Learning, 2021.
- Alessandro Negro. Graph-Powered Machine Learning, 2021.

Referências

Tutoriais

- Social Network Analysis: From Graph Theory to Applications with Python (Towards Data Science - Medium)
- An Introduction to Graph Theory and Network Analysis (with Python codes) (Analytics Vidhya)
- Game of Thrones: Network Analysis (Kaggle)
- Visualize your LinkedIn Network with Python (Medium)



Referências

Outros exemplos interessantes

- A network of science: 150 years of Nature papers
- GPS Ideológico: Análise do debate político no Twitter (Folha de São Paulo)



Referências

OBRIGADO!

