

Ciência das Redes - Desafio parte 1

Maria Fernanda Azolin, Mariana Bittencourt Junghans

April 14, 2024

1 Caracterização do problema

Neste desafio, é disponibilizada uma rede de mobilidade, com nós representando locais e arestas conectando esses locais se a mesma pessoa visitou os dois locais. O peso da aresta representa quantas vezes as pessoas visitaram os dois locais. No entanto, algumas arestas do grafo original foram removidas, e o desafio é descobrir quais foram essas arestas.

2 Análise do grafo modificado

2.1 Análise geral

A rede possui 18991 arestas e 4575 nós. É uma rede não-direcionada, com distribuição de grau seguindo a lei de potência, ou seja, muitos nós com baixo grau e poucos nós com grau elevado. O coeficiente de clusterização médio é baixo, cerca de 0,045. O grau médio é cerca de 8,3.

2.2 Análise de comunidades

A análise de comunidades em um grafo nos permite identificar agrupamentos naturais de nós que estão densamente interconectados, refletindo padrões de relações e interações entre os elementos da rede. Esses padrões podem nos guiar e ajudar na predição de links do grafo, pois as comunidades tendem a exibir uma forte coesão interna, e com isso é viável indicar possíveis conexões ausentes.

Para fortificar nossa análise, usamos dois algoritmos distintos, Louvain e Leiden. Como seus resultados foram muito similares, a análise se mostrou robusta e consistente.

2.2.1 Análise com algoritmo de Louvain

A partir do algoritmo de Louvain, identificamos **813 comunidades** e uma **modularidade de partição de 0,42819**. Apesar de termos 813 comunidades, apenas as 4 maiores possuíam um número significativo de nós, e a partir da quinta maior, esse número já era bem menos expressivo, como podemos identificar no ranking:

1. Comunidade id 4: **916 nós**
2. Comunidade id 0: **868 nós**
3. Comunidade id 5: **743 nós**

4. Comunidade id 2: **692 nós**
5. Comunidade id 31: 121 nós
6. Comunidade id 19: 49 nós
7. Comunidade id 84: 26 nós
8. ...

Assim, decidimos gerar um subgrafo com apenas as quatro maiores comunidades. Nesse novo grafo, tínhamos apenas 3219 nós, representando 70% do original.

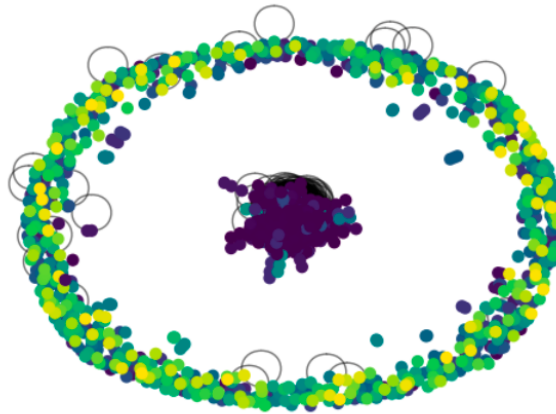


Figure 1: Comunidades a partir do algoritmo de Louvain

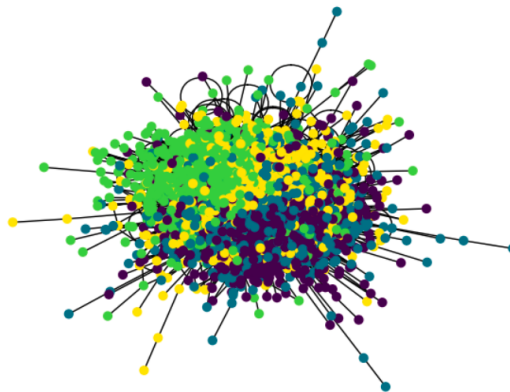


Figure 2: 4 maiores comunidades a partir do algoritmo de Louvain

2.2.2 Análise com algoritmo de Leiden

A partir do algoritmo de Leiden, identificamos **814 comunidades** e uma **modularidade de partição de 0,4215385**. Apesar de termos 814 comunidades, apenas as 4 maiores possuíam um número significativo de nós, e a partir da quinta maior, esse número já era bem menos expressivo, como podemos identificar no ranking:

1. Comunidade id 4: **911 nós**

2. Comunidade id 0: **900 nós**
3. Comunidade id 5: **750 nós**
4. Comunidade id 2: **574 nós**
5. Comunidade id 31: 113 nós
6. Comunidade id 19: 76 nós
7. Comunidade id 84: 59 nós
8. ...

Assim, decidimos gerar um subgrafo com apenas as quatro maiores comunidades. Nesse novo grafo, tínhamos apenas 3135 nós, representando 69% do original.

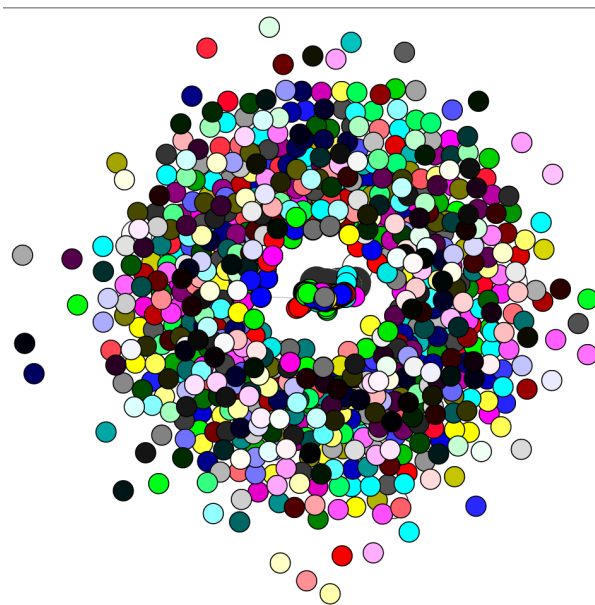


Figure 3: Comunidades a partir do algoritmo de Leiden

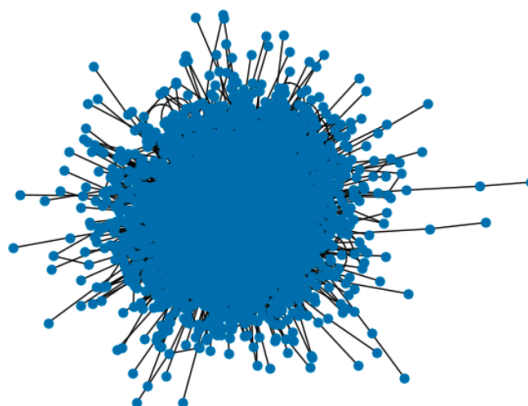


Figure 4: 4 maiores comunidades a partir do algoritmo de Leiden

3 Análise dos resultados

A distribuição de graus da rede, o seu baixo coeficiente de clusterização médio e baixo grau médio trazem a conclusão de que essa é uma rede “Scale-free” ou rede de “ligações preferenciais”, ou seja, nessa rede os nós mais importantes tendem a receber mais ligações que os outros nós menos relevantes.

No entanto, isso pode ser causado também pelo fato de que nessa rede várias arestas foram removidas para possibilitar a conclusão do desafio proposto.

Os locais com maior centralidade de grau são em sua maioria restaurantes e bares, indicando que esse tipo de estabelecimento é muito popular e quem visita esses locais geralmente visita diversos outros de categoria parecida.

Por outro lado, os locais com maior coeficiente de clusterização são de categorias diversas, indicando que podem ser próximos a restaurantes e bares (pois seus vizinhos estão mais conectados).

Já com relação às comunidades, a robustez dos resultados pode ser percebida pela consistência dos mesmos, já que ambos os algoritmos trouxeram um número de comunidades (813 e 814) e modularidade (0,42819 e 0,42154) muito próximos. Adicionalmente, esse valor de modularidade representa uma divisão satisfatória das comunidades.

Nessa análise, observou-se que as maiores comunidades detectadas são compostas por um número significativo de nós. Isso sugere que existem agrupamentos bem definidos na rede, onde os estabelecimentos dentro de uma mesma comunidade tendem a ter avaliações semelhantes entre si.

Além disso, esses agrupamentos identificados podem ser úteis para segmentar os estabelecimentos com base em padrões de avaliação semelhantes. Isso pode ajudar na análise de tendências de preferência dos consumidores ou na identificação de grupos de estabelecimentos que possam colaborar em estratégias de marketing, por exemplo.