



Universidade Federal
de São João del-Rei

Instagram: Caracterização topológica de redes sociais

Ana Cláudia Machado

Trabalho desenvolvido na disciplina de Mineração e Análise de
Redes Sociais.

Professor: Vinícius Vieira

São João del-Rei – Maio de 2025

Sumário

1	Introdução	2
2	Coleta e Tratamento de Dados	2
3	Análise de Resultados	2
3.1	Topologia Básica da Rede	3
3.2	Grau dos Vértices	3
3.3	Centralidades na Rede	6
3.3.1	Centralidade por Grau	6
3.3.2	Centralidade por Autovetor	7
4	Conclusão	8

1 Introdução

A topologia de uma rede se refere a como os vértices estão interconectados e como eles se organizam. Nesse sentido, o estudo da topologia da rede é importante para entendermos as relações existentes, de modo a determinar quais vértices são centrais para a rede. Neste trabalho, analisamos as relações em uma das maiores redes sociais da atualidade: o Instagram. A partir de um usuário específico, denominado $user_0$, foi coletado o conjunto de interseção I entre os usuários que ele segue e aqueles que o seguem. Com base nesse conjunto, construímos uma rede direcionada, conectando os usuários que se seguem mutuamente dentro de I . A proposta da análise é comparar a topologia dessa rede com e sem a presença de $user_0$, a fim de compreender como a remoção de $user_0$ afeta a clusterização da rede e as distribuições de grau, observando alterações no número de conexões que cada vértice possui e como isso modifica a estrutura geral da rede. Por fim, identificar quais grupos de indivíduos dependem de sua presença para permanecerem conectados.

2 Coleta e Tratamento de Dados

O *Instagram* não possui uma *API* de extração de dados aberta, de modo que só é possível solicitar alguns dados tendo uma conta de "criador de conteúdo". Diante dessa restrição, os dados necessários para essa análise foram obtidos por meio de *web scraping*, utilizando os scripts disponíveis nos repositórios do GitHub 1 e 2 ¹. Destacamos que esse processo foi extremamente demorado e trabalhoso, uma vez que foi interrompido diversas vezes por bloqueios e limitações impostas pela própria rede social.

Para a coleta, primeiro definimos o conjunto de interseção I entre os usuários seguem $user_0$ e que $user_0$ segue de volta. Com isso, obtivemos 450 usuários, para os quais extraímos quem eles seguiam em comum com $user_0$, de modo a construir a rede. Feita toda a extração, anonimizamos todos os usuários, de modo a manter a privacidade, não expondo no trabalho nenhum nome pessoal ou instituição.

3 Análise de Resultados

Nesta seção, serão detalhados os resultados obtidos a partir de uma série de análises topológicas, buscando compreender como a presença e a ausência do usuário $user_0$ influencia a rede.

¹Repositório 1: https://github.com/redianmarku/instagram-follower-scraper/blob/master/nthn_followers.txt
Repositório 2: <https://github.com/vasilisa-che/instagram-followers-and-bio-scraper/blob/main/bio-scraper.py>

3.1 Topologia Básica da Rede

A rede é direcionada, haja vista que dado dois usuários quaisquer i e j , i pode seguir j , mas não ser seguido de volta. Assim, as conexões que saem e entram em um vértice possuem significados diferentes: uma aresta que entra em um vértice representa um seguidor, ou seja, um usuário interessado naquele perfil; já uma aresta que sai indica que o usuário está seguindo outro, demonstrando interesse no perfil de outro. Em relação à importância das conexões, é mais significativo que um usuário receba arestas, indicando que os outros usuários demonstram interesse em seu perfil.

Com a presença do $user_0$, a rede possui 451 vértices e 28693 arestas, com coeficiente de clusterização médio $C = 0.49$, definido pela Equação 1. Nessa equação, n representa o número total de vértices e C_i o coeficiente de clusterização do vértice i . Esse valor indica que, em média, 49% dos pares de vizinhos de um vértice estão conectados entre si.

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i \quad (1)$$

Ao retirar o $user_0$, a rede passa a ter, logicamente, 450 vértices e 27812 arestas. O número de arestas reduz em 881 unidades, indicando esse é o número total de conexões do $user_0$ com os demais usuários na rede. Já o coeficiente de clusterização médio aumenta para $C = 0,54$.

Esse aumento pode ser explicado pelo fato de que muitos dos vizinhos do $user_0$ não possuíam conexões entre si. Assim, sua presença na rede gera diversos pares de vizinhos desconectados, reduzindo a clusterização média. Ao removê-lo, esses pares deixam de influenciar o cálculo, elevando seu valor em 0.05.

Uma confirmação da análise anterior é o fato de $C_{user_0} = 0.14$, evidenciando que apenas 14% dos seus pares de vizinhos se conectam entre si.

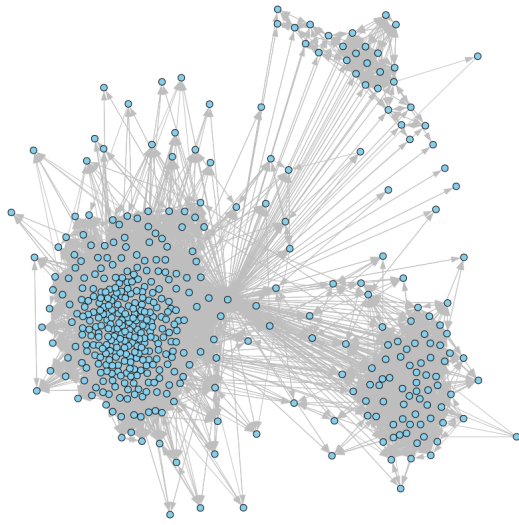
A Figura 1a evidencia a topologia da rede com a presença do $user_0$ enquanto que a Figura 1b mostra as diferenças estruturais após a retirada desse usuário. Percebemos que é possível identificar agrupamentos de usuários, basicamente, sem qualquer ligação entre si, demonstrando a existência de diferentes contextos sociais de interação de $user_0$.

3.2 Grau dos Vértices

Logicamente, $user_0$ possui o maior grau tanto de entrada quanto de saída da rede, uma vez que esta foi construída tendo-o como vértice central. Podemos analisar a distribuição de graus da rede, a fim de identificar

Naturalmente, o $user_0$ possui o maior grau tanto de entrada quanto de saída na rede, uma vez que esta foi construída tendo-o como vértice central. Considerando o grau dos vértices, podemos analisar sua distribuição com o objetivo de entender como as demais conexões ocorrem.

Para isso, temos as distribuições tanto em escala linear quanto logarítmica, o que nos permite analisar graus de magnitudes discrepantes com mais facilidade.

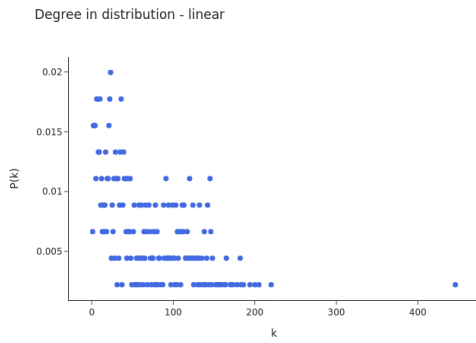


(a) Rede com $user_0$

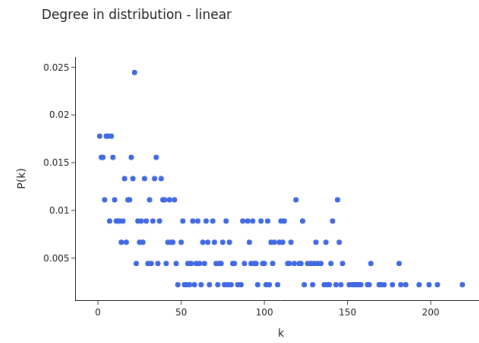


(b) Rede sem $user_0$

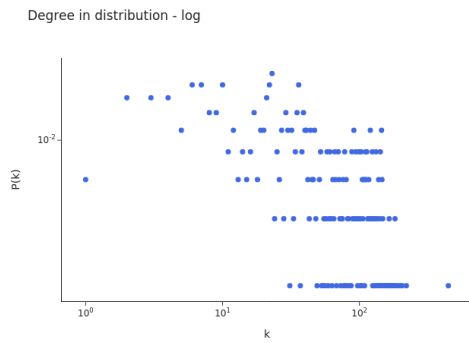
Figura 1: Topologias da rede



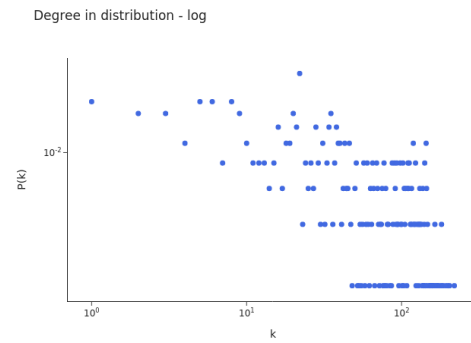
(a) Linear, com $user_0$



(b) Linear, sem $user_0$



(c) Logarítmica, com $user_0$



(d) Logarítmica, sem $user_0$

Figura 2: Distribuições de graus de entrada

Para as distribuições lineares de entrada da Figura ??, não existem mudanças significativas, tendo apenas um ponto com $k > 400$ na Figura 2a, representando o $user_0$. Sem a presença desse vértice, na Figura 2b, os valores de k ficam próximos de 200. É perceptível que há uma maior concentração de graus no intervalo $[50, 150]$, indicando que a maioria dos usuários nessa rede possui o número de seguidores contidos nesse intervalo. Para as distribuições logarítmicas, Figuras 2c e 2d, as conclusões são as mesmas, dado que a magnitude dos graus de entrada não é tão discrepante a ponto de a distribuição logarítmica identificar novos padrões.

Já para as distribuições de graus de saída, Figura 3, as observações feitas para as distribuições de graus de entrada também são válidas. É interessante observar esse comportamento, tendo, como hipótese, o fato de que, numa rede composta apenas por usuários comuns (sem figuras de grande influência, como celebridades), o número de seguidores tende a ser próximo ao número de perfis seguidos.

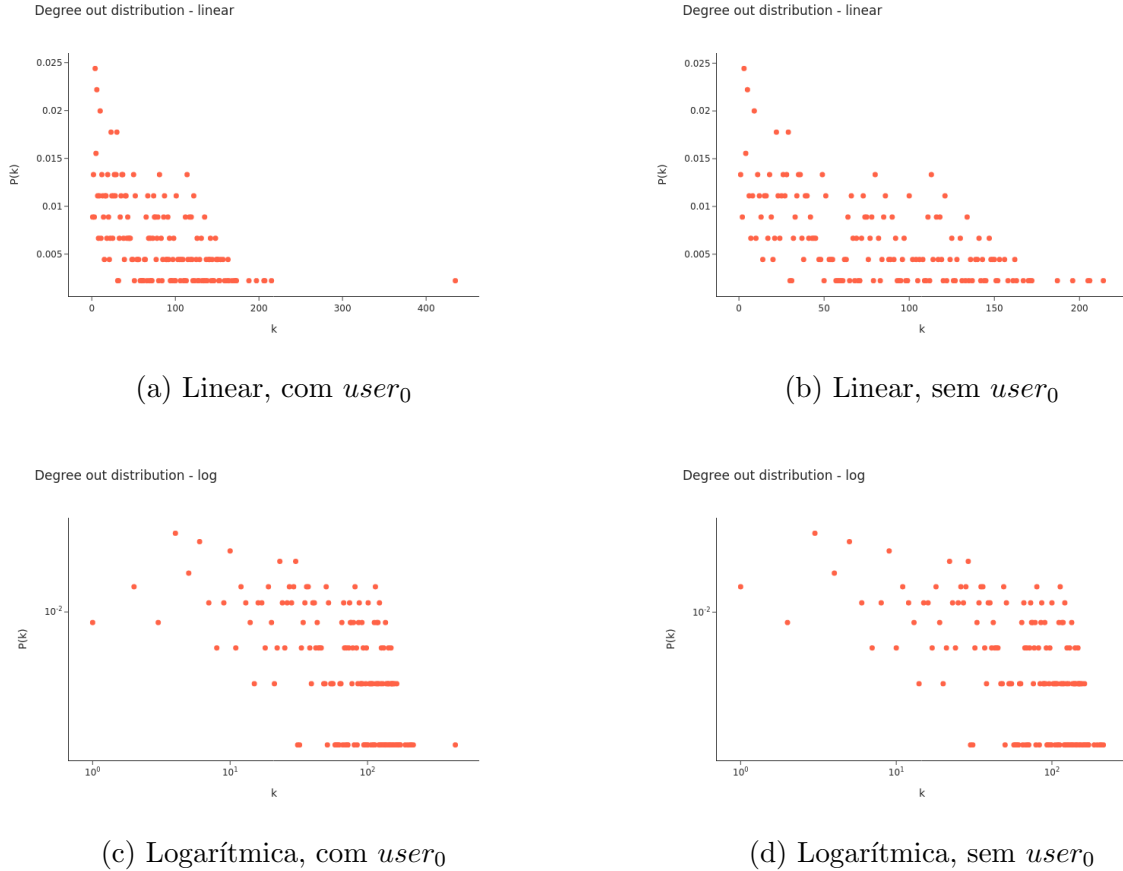


Figura 3: Distribuições de graus de saída

3.3 Centralidades na Rede

Tendo um vértice central pré-definido, a análise das centralidades na rede é uma forma de identificar, além do $user_0$, quais outros usuários são relevantes. Para isso, vamos analisar a rede considerando dois tipos de centralidades diferentes: por grau e por autovetor.

Para melhor visualização, plotamos a rede com os vértices destacados de acordo com seu valor de centralidade, sendo que, quanto maiores os valores, mais grandes e com coloração mais escura o vértice se apresenta.

3.3.1 Centralidade por Grau

A centralidade por grau quantifica a importância de um vértice de acordo com a quantidade de conexões que ele possui. Nesse sentido, para uma análise coerente, consideramos apenas o grau de entrada, pois, como explicado na Seção 3.1, é esse grau de define a importância do vértice na rede.

A tabela 1 evidencia os dez vértices com maior centralidade por grau. Como esperado, o $user_0$ possui a maior centralidade, dado que esse usuário está conectado a todos os outros na rede. Ademais, os seguintes usuários possuem centralidade próximas ao intervalo $[0.4, 0.5]$.

Vértice	Centralidade por Grau
user_0	0.9911
user_354	0.4889
user_259	0.4556
user_56	0.4444
user_203	0.4311
user_323	0.4133
user_367	0.4067
user_19	0.4044
user_191	0.4044
user_167	0.3956

Tabela 1: Centralidade por Grau

Nos gráficos da Figura 4, temos a plotagem dos vértices proporcional à sua centralidade, com o $user_0$ possuindo grande destaque na Figura 4a. Os demais vértices não possuem tanta relevância, ainda mais considerando os agrupamentos menores (localizados na parte superior da imagem). Já o terceiro agrupamento, possui alguns vértices de tamanhos medianos, os quais foram apresentados na tabela 1.

Ao analisar a Figura 4b, temos vértices totalmente isolados e um agrupamento sem conexão com os demais vértices na rede. Fora isso, temos dois agrupamentos que possuem algumas conexões entre si, com os vértices de maior centralidade por grau

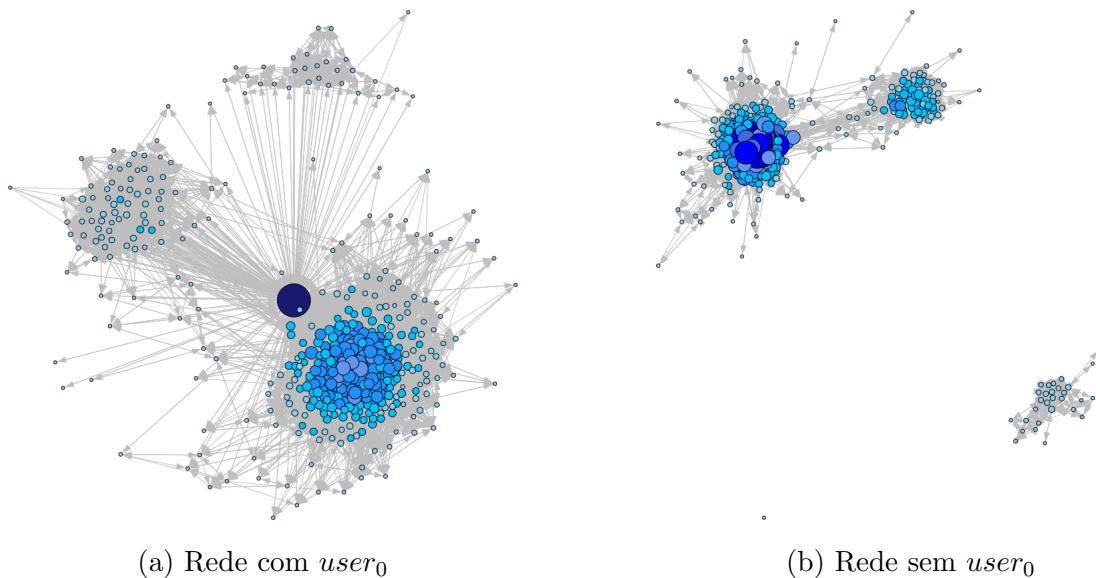


Figura 4: Centralidade por Grau de Entrada

estando no maior desses dois agrupamentos. É interessante notar que, quanto mais isolado do restante da rede um agrupamento estiver e quanto menor o número de vértices que o compõe, menor será a relevância desse vértice.

Para o contexto do *Instagram*, podemos ver que quanto menos seguidores em comum com o $user_0$ um vértice possuir, menor será sua relevância na rede.

3.3.2 Centralidade por Autovetor

Diferentemente da centralidade por grau, que leva em conta apenas o número de conexões de um vértice, a centralidade por autovetor quantifica a qualidade dessas conexões. Nesse contexto, estar ligado a vértices que também são bem conectados e influentes aumenta a importância de um nó. Aplicado ao cenário deste trabalho, ser seguido por um usuário considerado relevante na rede é um indicativo de que o próprio usuário também possui relevância.

Na tabela 2 os dez vértices com maiores centralidades por autovetor são destacados. Em relação a tabela 1, temos que alguns vértices mudam de posição na ordenação. Nesse sentido, para os vértices que adquirem posições mais relevantes, isso indica que, mesmo tendo menos seguidores do que outros, suas conexões são mais interessantes, o que faz essa inversão nas posições ocorrer.

Na plotagem dos gráficos, evidenciados na Figura 5, fica evidente na Figura 5a que os dois agrupamentos que, no geral, possuem conexões apenas com o $user_0$ não possuem qualquer relevância significativa na rede. O terceiro agrupamento, assim como na centralidade por grau, possui alguns usuários com certa relevância na rede, os quais são destacados na Figura 5b, com a retirada do $user_0$.

Em comparação com o uso das duas centralidades, no caso da centralidade por

Vértice	Centralidade por Autovetor
user_0	0.1458
user_354	0.1149
user_259	0.1145
user_56	0.1137
user_191	0.1071
user_203	0.1067
user_323	0.1065
user_367	0.1056
user_19	0.1039
user_167	0.1021

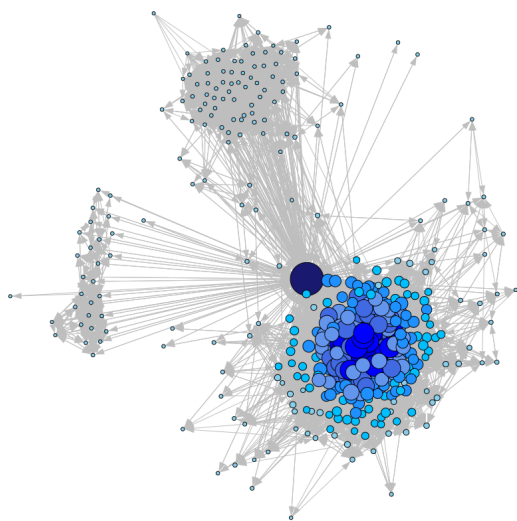
Tabela 2: Centralidade por Grau

autovetor fica explícito o quão pouco relevantes são os indivíduos que não possuem muitas conexões na rede além do $user_0$. Ao considerar a centralidade de grau, por estarem interconectados em um agrupamento menor, os pontos ainda aparecem um pouco maiores, porém, ao considerar a importância das ligações, sua relevância é minimizada.

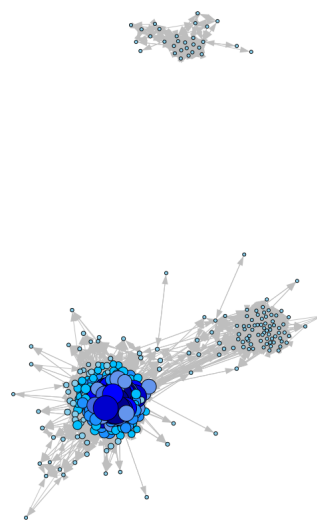
4 Conclusão

Neste trabalho, abordamos a construção e análise de uma rede de usuários do *Instagram*, conectados a partir de um usuário central pré-definido, $user_0$. Comparamos o comportamento e a topologia da rede ao considerá-la tanto com quanto sem esse usuário, analisando os impactos para a estrutura da rede.

Na análise das centralidades fica evidente a importância desse usuário nas conexões da rede, o que já era esperado. Além disso, foi possível identificar qual agrupamento continha usuários relevantes para a rede, desempenhando papéis centrais quando $user_0$ é excluído.



(a) Rede com $user_0$



(b) Rede sem $user_0$

Figura 5: Centralidade por Autovetor