

Análise numérica e topológica de uma Ego Rede pessoal

Eduardo G. R. Miranda¹, Guilherme Fonseca¹

¹Departamento de Ciência da Computação
Universidade Federal de São João del Rei (UFSJ)

Abstract. *The study of ego networks is very important because in today's society the communication between people is a subject that is in evidence due to the great use of social networks. This article will present a proposal for an algorithm that will identify the similarity between two communities, define rules for how the ego networks should be created and how the weights of each edge (relation) must be filled. To conclude, a numerical analysis of basic measures of a graph will be carried out, in addition to identifying how impactful the weight of the edges is in the generation of communities and the validation of the generation of these communities according to the authors' expectations.*

Resumo. *O estudo de ego redes tem grande importância pois na sociedade atual a comunicação entre as pessoas é um assunto que está em evidência devido o grande uso das redes sociais. Este artigo apresentará uma proposta para um algoritmo que identificará a similaridade entre duas comunidades, definirá regras de como as ego redes devem ser criadas e como os pesos de cada aresta (relação) devem ser preenchidos. Para concluir será feita uma análise numérica de medidas básicas de um grafo, além de identificarmos o quão impactante é o peso das arestas na geração de comunidades e a validação da geração destas comunidades de acordo com a expectativa dos autores.*

1. Introdução

A análise de redes sociais tem sido um campo de estudo em rápido crescimento, impulsionado pelo advento das mídias sociais e das vastas quantidades de dados geradas por essas plataformas. Nesse contexto, uma área de investigação cada vez mais relevante é a análise de grafos em redes egocêntricas. As redes egocêntricas, também conhecidas como redes pessoais, concentram-se em um indivíduo central (ego) e nas conexões estabelecidas com os outros atores (*alters*) dentro dessa rede. Essa abordagem proporciona uma perspectiva única para entender as dinâmicas sociais e a estrutura de relacionamentos em um nível mais granular.

Ao analisar grafos em redes egocêntricas, o foco é direcionado para o ego e seus *alters* imediatos, explorando aspectos como o grau de centralidade do ego, as características dos *alters* conectados e os padrões de interação entre eles. Essa abordagem permite uma compreensão mais profunda dos papéis desempenhados pelos indivíduos dentro de sua rede social imediata, bem como a identificação de grupos sociais, influenciadores-chave e estruturas de poder latentes.

A análise de grafos em redes egocêntricas tem aplicações em diversos campos, incluindo sociologia, psicologia, ciência política e marketing. Ao compreender as interações em nível individual dentro de uma rede, é possível investigar questões relacionadas à difusão de informações, propagação de comportamentos, formação de opiniões

e até mesmo impacto em processos de tomada de decisão. Além disso, essa abordagem pode ser utilizada para identificar lacunas na rede social do ego e desenvolver estratégias de intervenção ou fortalecimento de relacionamentos.

2. Referencial Teórico

2.1. Ego Rede

Em análise de redes sociais segundo [Newman 2018], uma rede egocêntrica é uma rede formada em torno de um indivíduo em particular, consistindo de seus contatos imediatos. O indivíduo em questão é denominado de ego e seus contatos imediatos de *alters*.

Redes egocêntricas são comumente obtidas através de questionários, entrevistas, ou ambos conduzidos diretamente junto aos participantes. Além disso, geralmente constroem-se diversas redes egocêntricas obtidas de diferentes egos da população do estudo em questão.

2.2. Algoritmos e Métricas

Utilizaremos algumas métricas básicas apresentadas por [Newman 2010] para identificar fenômenos que acontecem com tais métricas quando possui-se um ego-grafo.

2.2.1. Centralidade

- **Grau:** $C_{\text{grau}}(v) = \frac{\text{grau}(v)}{N-1}$, no qual:

$C_{\text{grau}}(v)$ é a centralidade de grau do vértice v ;

$\text{grau}(v)$ é o número de arestas que incidem no vértice v ;

N é o número total de vértices no grafo.

- **Betweenness:** $C_{\text{betweenness}}(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$, no qual:

$C_{\text{betweenness}}(v)$ é a centralidade de *betweenness* do vértice v ;

s é um vértice de origem diferente de v e t é um vértice de destino diferente de v e de s ;

σ_{st} é o número total de caminhos mais curtos do vértice s para o vértice t ;

$\sigma_{st}(v)$ é o número de caminhos mais curtos do vértice s para o vértice t que passam pelo vértice v .

- **Closeness:** $C_{\text{closeness}}(v) = \frac{1}{\sum_{u \neq v} d(v,u)}$, no qual:

$C_{\text{closeness}}(v)$ é a centralidade de proximidade do vértice v ;

$d(v, u)$ é a distância mais curta entre o vértice v e o vértice u no grafo;

A soma é realizada sobre todos os vértices u diferentes de v .

- **Eigenvector:** $C_{\text{eigenvector}}(v) = \frac{1}{\lambda} \sum_{u \in \Gamma(v)} A_{uv} C_{\text{eigenvector}}(u)$, no qual:

$C_{\text{eigenvector}}(v)$ é a centralidade de eigenvector do vértice v ;

λ é o maior autovalor associado ao grafo;

$\Gamma(v)$ é o conjunto de vértices adjacentes ao vértice v ;

A_{uv} é o elemento da matriz de adjacência que indica a existência de uma aresta entre os vértices u e v .

- **Pagerank:** $PR(v) = (1 - d) + d \cdot \sum_{u \in \mathcal{N}(v)} \frac{PR(u)}{\text{outdegree}(u)}$, no qual

$PR(v)$ é a centralidade de PageRank do vértice v ;

d é um fator de amortecimento (geralmente definido como 0,85);

$\mathcal{N}(v)$ é o conjunto de vértices adjacentes ao vértice v ;

$\text{outdegree}(u)$ é o grau de saída do vértice u .

2.2.2. Coeficiente de *Clustering*

- **Clustering:** $C(v) = \frac{2 \cdot E(v)}{k(v) \cdot (k(v) - 1)}$, no qual:

$C(v)$ é o coeficiente de clustering do vértice v ;

$E(v)$ é o número de arestas entre os vizinhos do vértice v ;

$k(v)$ é o grau do vértice v (número de vizinhos).

2.2.3. Comunidades

- **Método Louvain:**

De acordo com [Aires and Nakamura 2017] o algoritmo de Louvain é um método de detecção de comunidades que busca otimizar a modularidade. O algoritmo funciona em duas fases iterativas: na primeira fase, cada nó é inicialmente atribuído a uma comunidade separada e, em seguida, os nós são combinados em uma única comunidade se isso levar a um aumento significativo na modularidade. Na segunda fase, a rede é representada por uma nova rede agregada, onde os nós são as comunidades encontradas na fase anterior. Esse processo iterativo é repetido até que não haja mais melhorias na modularidade.

2.3. Trabalhos Relacionados

O trabalho de [Tsapatsoulis et al. 2019] tem como objetivo identificar se determinado perfil do twitter é um *fake influencer* fazendo a análise da ego-rede de forma mais eficiente.

Para fazer esta identificação os autores primeiramente identificam a comunidade central utilizando um algoritmo genético, após isso é feita a identificação do ego-nó utilizando como métrica o grau de entrada, para identificação de perfis *fakes* que existem só para fazer número e não estão realmente aumentando o alcance da rede é utilizada a métrica de grau médio visto que os usuários *fakes* tipicamente possuem um grau de entrada baixo e por fim são feitos cálculos de degeneração para classificar respectivamente os influenciadores reais e *fakes*.

Como maneira de validação os autores utilizaram 11 perfis legítimos e 9 perfis *fakes* que foram minerados utilizando a API do twitter, a validação contemplou grafos de mesma proporção/densidade em quantidade de vértices e arestas para a verificação da robustez do algoritmo em classificar os usuários além de eliminar o viés do funcionamento do algoritmo.

Já o trabalho de [Milard and Pitarch 2023] tem como objetivo estudar o efeito da publicação de um artigo em sua ego-rede de citações e também quais são as relações sociais que sustentam tais mudanças.

Os autores utilizaram 102 artigos que tiveram certa relevância em sua área e que foram publicados entre 2008 e 2013 para fazer a pesquisa. Para cada artigo foram criadas 2 redes de citação, uma representando os 5 anos anteriores ao artigo analisado ser publicado e a outra representando os 5 anos depois da publicação. Na rede de citação os vértices representam os artigos citados pelo artigo analisado e existe uma aresta entre 2 vértices caso exista pelo menos 3 artigos (publicados dentro do período analisado pela rede) que citem os 2 artigos.

Eles também entrevistaram os autores dos 102 artigos para tentar entender melhor sobre os autores, os impactos que os artigos tiveram em sua carreira e a relação dele com os autores dos trabalhos que ele citou. Os autores analisaram as redes geradas e separaram os artigos analisados em 4 classes de acordo com a evolução da rede de citações. Por fim o trabalho termina com os autores se aprofundando na discussão das 4 classes levando em conta as respostas das entrevistas.

3. Metodologia

Nesta seção, descreveremos a metodologia adotada para a criação das ego-redes individuais e as análises subsequentes.

3.1. Algoritmo proposto para similaridade de comunidades

Como os autores possuíam dificuldades para encontrar algum algoritmo que calcule o quão similar são duas comunidades propusemos o algoritmo 1 que basicamente passa por cada comunidade em X e para cada uma delas identifica a comunidade em Y que possui mais similaridade, acumula-se então a cardinalidade da interseção entre X e Y e por fim o valor deste somatório é dividido pela quantidade de vértices do grafo.

Algorithm 1 Similaridade de Comunidade

Ensure: similaridade(comunidades1, comunidades2, V)

```
1: similares  $\leftarrow 0$ 
2: for all comunidadeX  $\in$  comunidades1 do
3:   maiorSemelhanca  $\leftarrow 0$ 
4:   for all comunidadeY  $\in$  comunidades2 do
5:     interseccao  $\leftarrow$  comunidadeX  $\cap$  comunidadeY
6:     if  $|interseccao| > maiorSemelhanca$  then
7:       maiorSemelhanca  $\leftarrow |interseccao|$ 
8:     end if
9:   end for
10:  similares  $\leftarrow$  maiorSemelhanca
11: end for
12: return similares /  $|V|$ 
```

3.2. Criação da Base de Dados da ego rede

A priori, para a construção das ego-redes individuais, cada autor coletou dados de pessoas conhecidas. Essa base de dados consistiu em uma lista de indivíduos com os quais o pesquisador possui algum tipo de relação, levando em consideração o critério de lembrar pelo menos nome e o rosto dos conhecidos. As informações incluíram os nomes das pessoas, apelido (quando houver) e cidade, casos onde o pesquisador não lembrou ou não soube alguma destas informações o campo foi deixado em branco.

3.3. Especificação das Relações

Apos a criação da base de dados de pessoas conhecidas cada membro do grupo caracterizou suas relações pessoais com as pessoas e também as relações entres os conhecidos de sua base de dados.

Foram definidas treze categorias para classificar as relações, que são elas: Família, Melhor Amigo, Namorado(a)/Cônjuge, Amigo, Parente, Parça, Professor, Aluno, Colega, Trabalho, Consanguíneo, Ficante e Conhecido. A tabela 1 traz as categorias criadas e uma breve descrição das relações que cada categoria representa.

Para cada relação especificada, atribuímos um peso específico, representando a importância da relação em comparação com as outras. Utilizamos uma escala de 1 a 13, em que 1 representa a categoria de maior importância e 13 representa a categoria de menor importância. O peso de cada relação também pode ser visto na tabela 1.

Vale ressaltar que tanto a existência de uma relação entre pessoas da base de dados quanto a sua categorização foi feita pelo pesquisador com base na sua concepção mas tentando ao máximo representar a veracidade das relações.

Peso	Relação	Descrição
1	Família	Pessoa de parentesco direto sendo irmão, pai e mãe. Pode também ser de parentesco indireto (tio, avó, primo, etc) cujo qual possui um vínculo muito forte.
2	Melhor Amigo	Pessoa que se tem alto nível de confiança e liberdade que foi construída/confirmada através do tempo e/ou experiências.
3	Cônjuge Namorado(a)	Pessoa que se tem/teve um relacionamento amoroso não superficial no qual existe/existiu um vínculo afetivo.
4	Amigo	Pessoa que já teve vínculo afetivo maior porém não há contato regular e esse vínculo já não é tão forte quanto foi. Ou pessoa com um forte vínculo mas não há uma liberdade/confiança tão alta.
5	Parente	Pessoa que tem o laço sanguíneo, porém o vínculo não forte o suficiente para que possa ser considerado família, mas há afinidade na relação.
6	Parça	Pessoa que se conhece há pouco tempo mas há uma forte sintonia/sinergia, gostos parecidos, ideologia parecida, porém a relação não passou por tempo o suficiente para se criar confiança/liberdade.
7	Professor	Pessoa de quem você foi professor.
8	Aluno	Pessoa de quem você foi aluno.
9	Colega	Pessoa cujo qual o vínculo é o de pertencer a mesma classe, seja de escola, faculdade, curso, etc...
10	Trabalho	Pessoa cujo qual se tem uma relação puramente profissional.
11	Consanguíneo	Pessoa cujo qual o único vínculo é o de sangue.
12	Ficante	Pessoa cujo qual se teve algum tipo de relação amorosa, mas não foi criado nenhum vínculo afetivo, puramente carnal.
13	Conhecido	Pessoa cujo qual se sabe o nome e rosto, não possui vínculo significativo.

Tabela 1. Descrição das relações

3.4. Construção

Com base nos dados coletados e na especificação das relações, construímos a ego rede. A ego rede foi criada utilizando técnicas de representação de redes, em que cada pessoa conhecida é representada como um nó e as relações são representadas por arestas direcionadas, com os pesos atribuídos conforme a categoria de relação.

A imagem 1 ilustra o grafo criado através da metodologia explicada anteriormente.

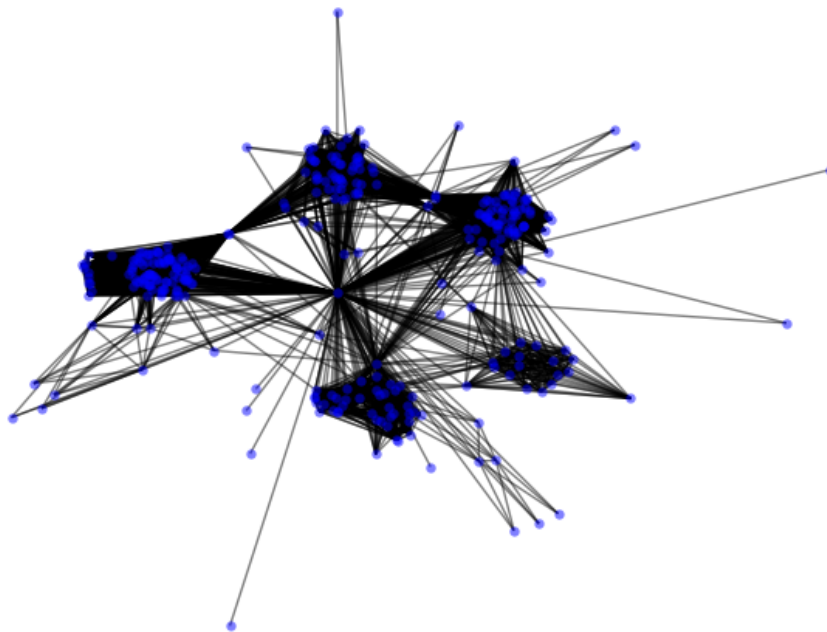


Figura 1. Grafo com o Ego Vértice

3.5. Análise das Ego-redes

A análise da ego-rede será feita da seguinte maneira:

- Serão calculadas as centralidades mencionadas na seção 2 além do coeficiente de *clustering* no grafo com a presença do ego-nó e sem ele. Com as centralidades calculadas será calculada a média e desvio padrão do resultado obtido para análise.
- Será utilizado o algoritmo de *Louvain* descrito na sub-seção 2.2.3 para calcular as comunidades no grafo. Após isso o algoritmo 1 será aplicado para verificar o quão similar são as comunidades calculadas com diferentes ponderações de arestas.
- Por fim será feita uma análise dos vértices presentes em cada uma das comunidades afim de verificar como as comunidades foram formadas e se estas estão de acordo com as expectativas dos autores.

4. Resultados

4.1. Análise numérica das medidas do Grafo

Pode ser visto na tabela 2 as médias e desvios padrão (DP) de algumas métricas de grafos calculadas do grafo da figura 1 com a presença do ego nó a esquerda e sem a presença do ego nó a direita.

As medidas de grau ou a centralidade por grau tiveram pouca mudança já que só retiramos um vértice e suas arestas, todavia as medidas de *closeness* e *betweenness* que dependem fortemente da topologia do grafo sofreram mudanças negativas significativas ao ponto de quase dobrar o valor só pela retirada do ego nó.

A centralidade por *EigenVector* quase não sofreu mudança ao retirarmos o ego nó, já a centralidade por *PageRank* manteve sua média mas teve uma alteração no DP visto que este algoritmo trabalha com os conceitos arestas de entrada e saída ao retirarmos uma aresta de entrada de todos os vértices tivemos uma alteração nesse DP. Como o coeficiente de *clustering* utiliza em seu calculo o número de nós e arestas vizinhas ao removermos o nó ego diminuimos proporcionalmente para todos os vértices e não houve mudança em seus resultados.

Medida	C\ego nó		S\ego nó	
	Média	DP	Média	DP
Grau	106.45	82.80	105.46	82.80
Cntr. Grau	0.1912	0.1141	0.1848	0.1049
Cntr. Closeness	0.5553	0.0406	0.4016	0.0785
Cntr. Betweenness	0.0026	0.0363	0.0044	0.0213
Cntr. EigenVector	0.0340	0.0455	0.0323	0.0467
Cntr. PageRank	0.0032	0.0018	0.0032	0.0013
Coef. Clustering	154.0	89.20	154.0	89.20

Tabela 2. Medidas de Centralidade

4.2. Análise das comunidades geradas

Como observado na figura 2, o algoritmo identificou na rede quatro comunidades centrais que os autores identificaram como sendo:

- Em vermelho - Vértices relacionados a UFSJ. Alunos, professores e funcionários.
- Em verde - Vértices relacionados a escola SESI e SENAI
- Em Amarelo - Vértices relacionados as escolas Aureliano Pimental e Garcia de Lima
- Em Azul - Familiares, amigos próximos da família.

Vértices que não se enquadraram nas três primeiras comunidades o algoritmo atribuiu na comunidade em azul, vértices que só tinham relação com o nó ego por exemplo foram designados na comunidade azul, que é onde o nó ego se encontra.

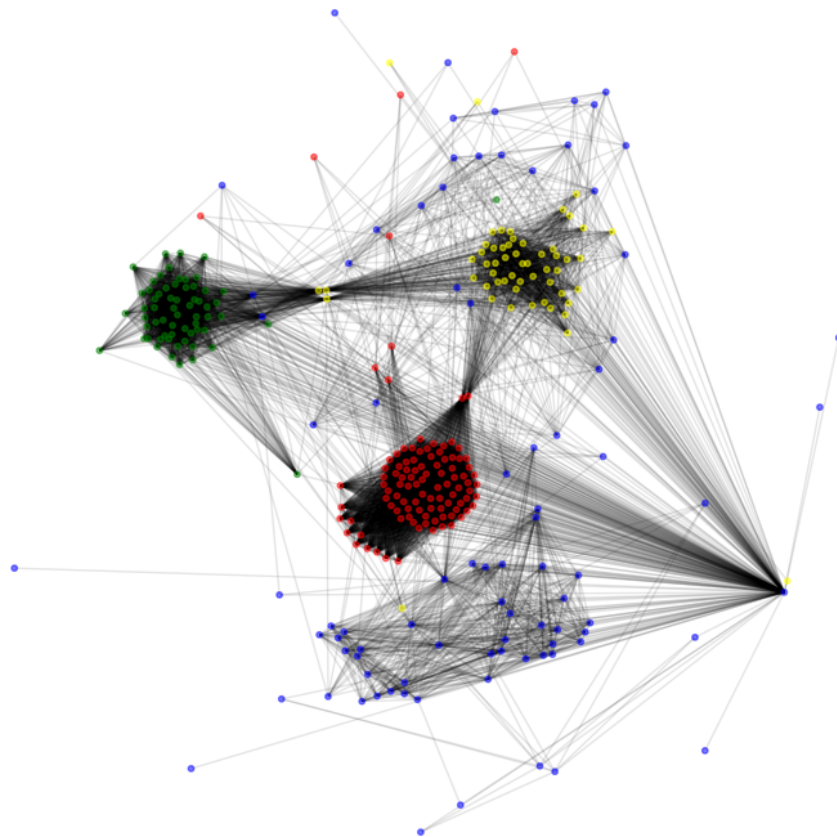


Figura 2. Grafo de comunidades com o Ego Vértice

Como observado na figura 3, onde há a ausência do ego nó o numero de comunidades encontradas é 11, isso se deve ao fato de que alguns dos nós tinham relação apenas com o ego nó, portanto após sua remoção eles ficaram desconexos do grafo e foram consideradas como suas próprias comunidades.

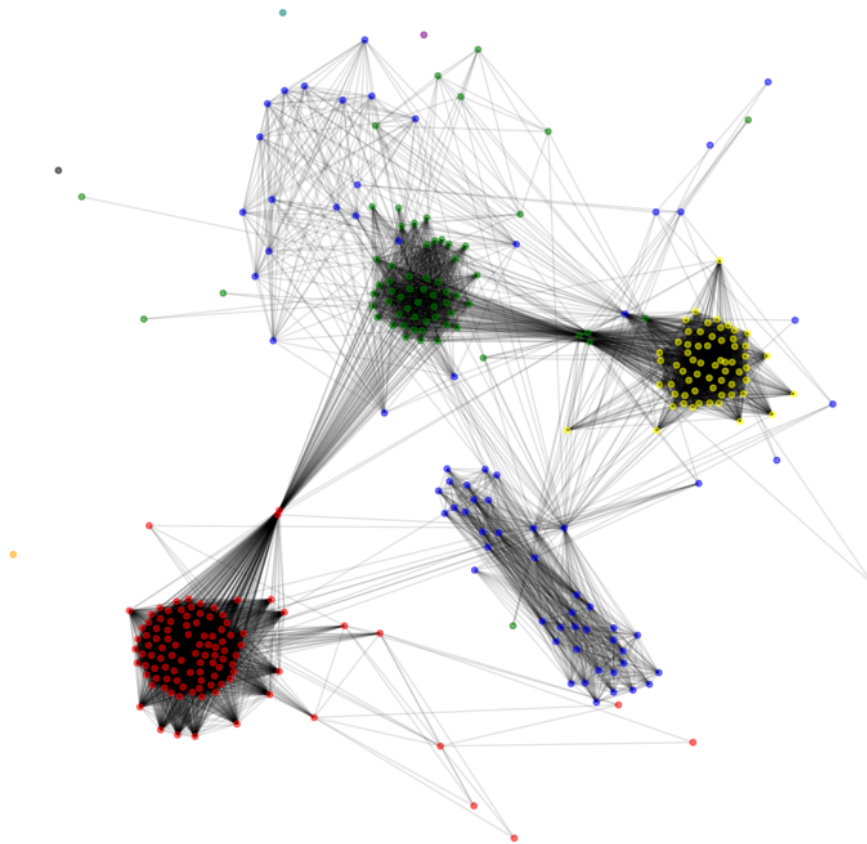


Figura 3. Grafo de comunidades sem o Ego Vértice

4.3. Análise da similaridade das comunidades geradas

Nas tabelas 3, 4 e 5 pode ser visto a similaridade entre as comunidades dos grafos que estão na vertical X horizontal. Temos as seguintes nomenclaturas para identificarmos a configuração de cada grafo: Caso haja um S\ antes da palavra ego significa que este grafo não possui o ego nó caso contrário ele possui. Quanto ao pesos das arestas se houver a palavra Normal após a palavra ego significa que este grafo possui o peso das classes definidas de 1-13, caso a palavra seja Mult esse peso de 1-13 será multiplicado por 2 e por fim caso seja Eleva o peso da aresta será igual a 2^n onde n era o peso original da aresta. Células com **X** são células onde as análises não foram realizadas, pois seriam ou a comparação de uma rede com ela mesma ou porque seria a comparação de uma rede com o ego nó e uma rede sem o ego nó mas com multiplicadores de aresta diferentes.

Na tabela 3 podemos identificar que independentemente do quanto aumente-se peso das arestas a diferença das comunidades não sofre alterações significantes. E se analisarmos em conjunto as tabelas 4 e 5 temos que o peso das arestas também não interferem de forma significativa no quão similar as comunidades são. A partir disso conseguimos identificar que o que realmente fez as comunidades mudarem de uma maneira mais signi-

ficativa foi a remoção do ego nó do grafo que causou uma diferença de aproximadamente 14% na similaridade das comunidades.

	S\Ego Normal	S\Ego Mult	S\Ego Eleva
Ego Normal	0.8640	x	x
Ego Mult	x	0.8543	x
Ego Eleva	x	x	0.8640

Tabela 3. Similaridade das comunidades com ego nó X sem ego nó

	Ego Mult	Ego Eleva
Ego Normal	0.9902	0.9644
Ego Mult	x	0.9644

Tabela 4. Similaridade das comunidades com ego nó

	S\Ego Mult	S\Ego Eleva
S\Ego Normal	1	0.9902
S\Ego Mult	x	0.9902

Tabela 5. Similaridade das comunidades sem ego nó

5. Conclusão

Como pôde ser visto as métricas de um grafo podem variar bastante com a remoção do ego nó, então para aplicações que sejam sensíveis a topologia do grafo como exemplo, redes sociais que dependam de ego nós chave para disseminação de conteúdo podem ter seu alcance comprometido.

Também foi analisado que o algoritmo de comunidades "Louvain" consegue somente montar as comunidades mais óbvias como família e pessoas que estudaram no mesmo local utilizando primordialmente a topologia do grafo, pois ao alterarmos drasticamente o peso das arestas as comunidades formadas eram muito similares as originais, logo o foco do estudo que era a informação da classe de relação entre os nós teve baixíssima relevância.

6. Trabalhos Futuros

Para trabalhos futuros sugere-se:

- Utilizar outros algoritmos de descoberta de comunidades para verificar se o peso das arestas tem maior significância.
- Tornar o grafo mais confiável através da comunicação com os nós *alters* chave para validar as relações do grafo.
- Utilizar a estrutura da ego rede para simular propagação de doenças ou de informações com base nas relações entre os nós.
- Analisar ego redes de um número maior de pessoas buscando padrões em comum entre elas.

Referências

Aires, V. P. and Nakamura, F. G. (2017). Detecção de comunidades em redes sociais: Relacionando o método louvain a medidas de centralidade. In *Anais do XXXVI Concurso de Trabalhos de Iniciação Científica da SBC*. SBC.

- Milard, B. and Pitarch, Y. (2023). Egocentric cocitation networks and scientific papers destinies. *Journal of the Association for Information Science and Technology*, 74(4):415–433.
- Newman, M. (2010). *Networks: An introduction*.-oxford, uk: Oxford university press.
- Newman, M. (2018). *Networks* second edition.
- Tsapatsoulis, N., Anastasopoulou, V., and Ntalianis, K. (2019). The central community of twitter ego-networks as a means for fake influencer detection. pages 177–184. IEEE.