Detecção de Comunidades em Redes Sociais: Relacionando o Método Louvain a Medidas de Centralidade*

Victoria P. S. Aires, Fabiola G. Nakamura

Instituto de Computação – Universidade Federal do Amazonas (UFAM)
69077-000 – Manaus – AM – Brasil

{vpsa,fabiola}@icomp.ufam.edu.br

Abstract. *In this project, the focus is on algorithms for the problem of detecting communities in social networks, especially the Louvain method. The goal is to relate the method to the concepts of centrality measures in complex networks, proposing their use to modify the greedy criterion of the method and verifying if this change increases the quality of the communities found. We showed that the construction of communities from the less central vertices improved the quality of communities obtained by the method in large and sparse networks, but did not bring significant gains in small and dense networks.*

Resumo. *Neste projeto de iniciação científica, o foco está em algoritmos para o problema de detecção de comunidades em redes sociais, em especial, no método Louvain. O objetivo é relacionar o método aos conceitos de medidas de centralidade em redes complexas, propondo a utilização das mesmas para modificar o critério guloso do método e verificando se esta mudança aumenta a qualidade das comunidades encontradas. Mostramos que a construção de comunidades a partir dos vértices menos centrais melhorou a qualidade das comunidades obtidas pelo método em redes grandes e esparsas, porém, não trouxe ganhos significativos em redes pequenas e densas.*

1. Introdução

A *rede social* de um indivíduo é resultado das interações e relacionamentos do mesmo com outros membros da sociedade [Bedi and Sharma 2016]. Alguns exemplos de redes sociais são aquelas geradas por amizades, interações em uma plataforma virtual como Facebook e Twitter, redes de citação e colaboração de cientistas e interação de personagens em um filme ou obra literária.

Redes sociais são um dos tipos de *redes complexas*. Tais redes são utilizadas para modelar aplicações do mundo real. Elas possuem algumas propriedades, como distribuição do grau dos vértices seguindo a lei de potência, alto coeficiente de agrupamento e estrutura de comunidades bem definida. Estas comunidades, no contexto de redes sociais, são oriundas da tendência de indivíduos similares estarem associados direta ou indiretamente. O processo de descobrir grupos em uma rede é conhecido como *detecção de comunidades*.

A qualidade das comunidades obtidas pelos algoritmos disponíveis na literatura pode ser comparada através da *modularidade*, uma métrica que indica o quanto os elementos de uma comunidade estão conectados. Uma vez que obter a máxima modularidade

*Apoiado pelo Programa Institucional de Bolsas de Iniciação Científica - PIBIC (UFAM e FAPEAM).

em uma rede é um problema NP-Difícil [Brandes et al. 2007], são necessários algoritmos heurísticos, como o método Louvain [Blondel et al. 2008]. Este algoritmo possui complexidade $O(m)$, onde m é o número de arestas da rede, e é capaz de identificar comunidades com alta modularidade através de um algoritmo iterativo de duas fases, onde as comunidades são construídas através da avaliação sequencial dos vértices. Porém, este critério é muito simples, não garantindo solução ótima.

Neste projeto de iniciação científica, o objetivo é estudar o impacto que uma solução gulosa aliada a métricas de redes complexas tem na qualidade das comunidades encontradas pelo método Louvain. Para isso, modificamos o critério para construção de comunidades: os vértices são avaliados segundo as *medidas de centralidade* dos mesmos. Estas indicam quais os vértices mais influentes ou centrais da rede segundo algum critério. Observamos que esta modificação tem efeito positivo em redes sociais grandes e esparsas, quando as comunidades são construídas a partir dos vértices menos centrais.

O restante do artigo está organizado da seguinte forma: a Seção 2 contém os fundamentos teóricos utilizados para o desenvolvimento do trabalho. A Seção 3 apresenta alguns algoritmos para o problema de detecção de comunidades em redes sociais. A Seção 4 descreve como foi feita a alteração no método Louvain, proposta neste trabalho. A Seção 5 traz os resultados obtidos e, por fim, a Seção 6 contém as considerações finais do artigo, seguida das referências.

2. Fundamentação teórica

Nesta seção, serão descritos os conceitos de redes complexas, medidas de centralidade e redes sociais, fundamentos teóricos que foram necessários para o desenvolvimento deste trabalho.

2.1. Redes complexas

Na natureza e sociedade existe uma variedade de sistemas complexos que podem ser modelados através de redes. Estes fenômenos incluem redes de telecomunicações, relações sociais, a *World Wide Web* e a Internet, entre outros exemplos.

Uma abordagem inicial para descrever as redes do mundo real são as *redes aleatórias*. Neste tipo de redes, um conjunto com n vértices é conectado por um número fixo de *links* (ou arestas) de maneira completamente aleatória. A distribuição do grau dos vértices segue uma distribuição de Poisson. Porém, a grande maioria de aplicações do mundo real não tem o comportamento aleatório descrito, com todos os vértices tendo igual probabilidade de estarem conectados: elas apresentam diferentes regras para conexão entre os vértices e crescimento. Em muitos casos, a distribuição de grau segue a lei da potência, sem nenhuma característica de escala [Dehmer and Emmert-Streib 2009]. Este fato levou à criação do conceito de *redes livres de escala*, também chamadas de *redes complexas*. A Figura 1 apresenta exemplos de distribuições de grau em redes aleatórias e redes livres de escala.

Redes complexas possuem propriedade de *small-world*, onde os vértices são separados por distâncias pequenas e possuem a característica de serem altamente conectados [Hofstad 2016]. Isto significa que há tendência a apresentar um alto coeficiente de agrupamento, levando à formação de *clusters* ou comunidades. Assim, define-se que redes complexas possuem as seguintes propriedades [Shen 2013]:

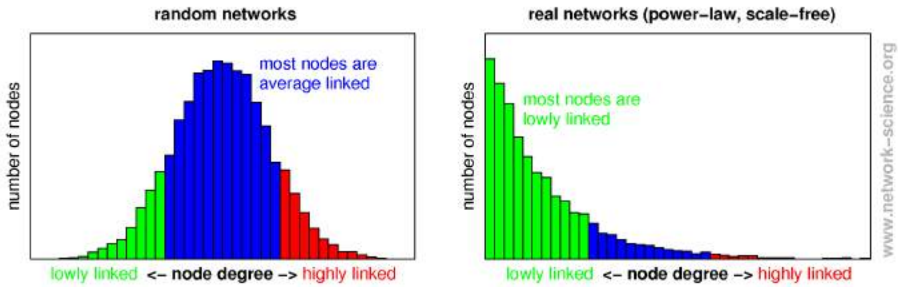

Figura 1. Exemplos de distribuições de grau dos nós de uma rede. À esquerda, em redes aleatórias e ao lado, em redes livres de escala. Na primeira, a distribuição dos graus segue uma distribuição de Poisson, enquanto na segunda, segue a lei de potência [Scholz 2009].

- distribuição de grau dos nós seguindo a lei de potência (*power-law degree*);
- características de mundo pequeno (*small-world*);
- alto coeficiente de agrupamento e estrutura de comunidades.

Em uma rede complexa, alguns nós possuem maiores valores de centralidade que os demais, sendo considerados mais influentes ou importantes. Para mensurar o quanto um nó é central na rede, são utilizadas *medidas de centralidade*. Neste trabalho, focamos nas seguintes métricas:

- Centralidade de grau ponderado (*weighted degree centrality*): a métrica mais simples, trata-se basicamente do grau de cada vértice, levando em consideração os pesos em grafos ponderados. Segundo esta medida, a centralidade de um nó é dada pelo seu número de conexões [Perra and Fortunato 2008].
- Centralidade de intermediação(*betweenness centrality*): esta métrica indica, dado um vértice, quantos caminhos mínimos entre dois nós quaisquer passam por ele. Um nó que possui alta intermediação funciona como ponte na rede, pois conecta vários elementos, adquirindo maior importância na rede [Dehmer and Emmert-Streib 2009].
- Centralidade de proximidade (*closeness centrality*): esta métrica é obtida ao computar o inverso do comprimento médio dos caminhos mínimos entre um dado nó e todos os demais na rede. Quanto maior a proximidade, mais central é o nó, por estar mais próximo em média do restante da rede [Motta et al. 2009].

2.2. Redes sociais

As *redes sociais* nasceram na Sociologia [Scott and Carrington 2011], porém, com o advento da internet e dos avanços na computação, tornou-se possível coletar um grande número de dados do mundo real para serem analisados e processados. Uma rede social é um tipo de rede complexa, ou seja, possui as propriedades características de redes complexas, como *power-law degree*, propriedade de *small-world* e estrutura de comunidades.

Assim como em outros tipos de redes complexas, uma rede social é formada por entidades e relacionamentos. Neste contexto, as entidades podem ser pessoas, instituições, *web sites* e os relacionamentos podem indicar amizade, trabalho, colaboração e outros tipos de interação social. Além disso, estes relacionamentos podem ser recíprocos ou não [Özturk 2014].

Alguns exemplos de redes sociais da vida real incluem redes com base em amizade, telefone, e-mail, e redes de colaboração e citação de cientistas. Representar tais redes como grafos permite a extração de informações relevantes sobre a relação que ela modela, através do estudo e análise das propriedades das mesmas, possuindo um grande potencial para aplicações úteis, como sistemas de recomendação.

3. Algoritmos para detecção de comunidades

Uma *comunidade* pode ser definida como um grupo de entidades mais próximas entre si em comparação com as outras na rede [Bedi and Sharma 2016]. Em outras palavras, uma comunidade é formada por indivíduos que interagem mais frequentemente uns com os outros do que com aqueles que estão fora do grupo.

Nesta seção são apresentados alguns dos principais algoritmos para o problema de detecção de comunidades, baseados no conceito de modularidade: Girvan-Newman [Girvan and Newman 2002], Clauset, Newman e Moore [Clauset et al. 2004] e método Louvain [Blondel et al. 2008].

3.1. Introdução à modularidade

Em geral, quando um método de detecção de comunidades é aplicado em uma base de dados do mundo real, as comunidades ainda não são conhecidas. Logo, uma questão importante é de que forma a qualidade das comunidades encontradas pode ser avaliada.

A *modularidade* é uma medida proposta por Girvan e Newman [Girvan and Newman 2002] e amplamente aceita como a única medida para avaliar a qualidade de comunidades. O cálculo da modularidade é mostrado na equação 1.

$$Q = \sum_i (e_{ii} - a_i^2) \tag{1}$$

Na equação 1, i representa uma comunidade, e_{ii} é a fração de arestas que estão inteiramente na comunidade i e a_i é a fração de arestas que possuem pelo menos um extremo na comunidade i. A modularidade assume um valor entre -1 e 1. Quanto mais próximo de 1, melhores são as estruturas de comunidades encontradas.

O problema de determinar a máxima modularidade é NP-Difícil [Brandes et al. 2007], ou seja, este problema pertence a uma classe de problemas (que podem ser de decisão ou otimização) para os quais não se conhece solução em tempo polinomial.

3.2. Algoritmo de Girvan-Newman

O *algoritmo Girvan-Newman (GN)* [Girvan and Newman 2002] é um algoritmo divisivo (arestas são progressivamente removidas) e tem sua base no conceito de *edge-betweenness* para um grafo não direcionado e não ponderado. O foco do algoritmo está nas arestas que estão mais "entre" (em inglês, *"between"*) as comunidades. O *betweenness* de uma aresta é calculado através do número de caminhos mínimos entre pares de vértices que utilizam aquela aresta. As comunidades são reveladas por meio da progressiva remoção das arestas de maior *betweenness* do grafo original, ou seja, através da remoção das arestas que estão entre as comunidades. Os passos do algoritmo são os seguintes:

1. Calcular o *betweenness* para todas as arestas da rede.
2. Remover da rede a aresta com o maior *betweenness*.
3. Recalcular o *betweenness* para todas as arestas que foram afetadas pela remoção.
4. Voltar ao passo 2 até que não existam mais arestas.

O resultado final é um dendrograma (árvore hierárquica) construído na abordagem *top-down*: inicia com uma comunidade geral, que é a rede original, e vai crescendo com a sucessiva remoção de conexões entre as comunidades, até as folhas, que são nós individuais. O algoritmo possui tempo de execução $O(n^3)$ no pior caso para grafos esparsos e $O(m^2n)$ para grafos densos, onde n é o número de vértices e m, o número de arestas. Apesar de apresentar um tempo de execução razoável, o algoritmo foi estudado e melhorado por diversos autores devido a sua dificuldade de se adaptar a redes muito grandes, ou seja, a escalabilidade é um problema deste método.

3.3. Algoritmo de Clauset, Newman e Moore

O algoritmo proposto por Clauset, Newman e Moore [Clauset et al. 2004] também utiliza a modularidade Q proposta por Girvan e Newman [Girvan and Newman 2002] como uma medida para otimizar a estrutura de comunidades. Trata-se de um método aglomerativo, isto é, constrói comunidades segundo a abordagem *bottom-up*: os nós são iterativamente agrupados em comunidades.

O *algoritmo CNM* funciona da seguinte forma: inicialmente cada vértice constitui por si uma comunidade (ou seja, há n comunidades). As comunidades são unidas em pares repetidamente, escolhendo a cada passo a junção que resulta no maior ganho ou na menor perda de Q. Ao final é produzido um dendrograma, revelando a estrutura de comunidades identificada.

O tempo de execução em uma rede com n vértices e m arestas é $O(md \log n)$, onde d é a profundidade do dendrograma. A maioria das redes do mundo real são esparsas e hierárquicas, com $m \approx n$ e $d \approx \log n$, caso em que a complexidade do algoritmo é $O(n \log^2 n)$. Como limitações, apesar de sua complexidade mostrar-se menor que os demais no caso médio, seu desempenho depende muito da rede dada como entrada: em um grafo denso, $m \approx n^2$, o que elevaria o tempo de execução neste caso.

3.4. Método Louvain

O *método Louvain* é um algoritmo heurístico baseado em otimização da modularidade Q [Blondel et al. 2008]. Também é um método de natureza aglomerativa. Assumindo que a entrada é uma rede ponderada com n vértices, o funcionamento do método é dividido em duas fases.

Na primeira fase, cada vértice é considerado uma comunidade. Após, para cada vértice i, considera-se cada um de seus j vizinhos e avalia-se o ganho de modularidade que ocorreria se o vértice i fosse removido de sua comunidade e colocado na comunidade de j. Ao final da avaliação em todos os vizinhos, o vértice i é colocado na comunidade onde o ganho é máximo, mas apenas se o ganho é positivo. Se não for, o vértice permanece em sua comunidade. Este processo se repete até que nenhum indivíduo possa melhorar a modularidade.

A segunda fase consiste em construir uma nova rede onde os vértices são as comunidades encontradas na primeira fase. Para isto, os pesos das conexões entre os novos

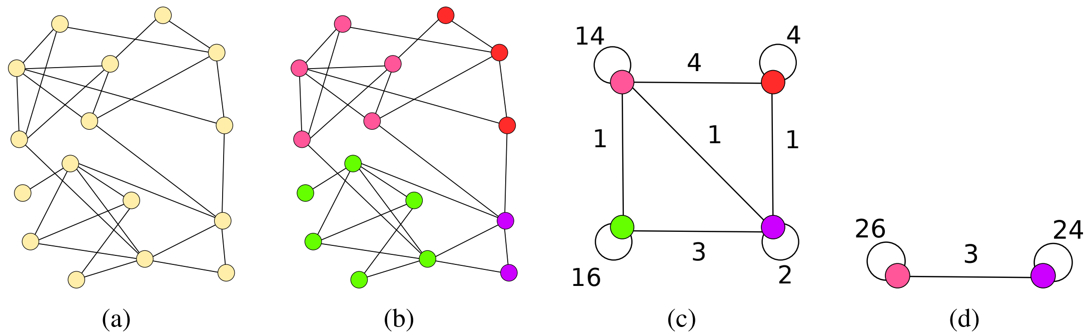

Figura 2. Exemplo de execução do método Louvain. Em 2(a), rede original. Em 2(b), após passar pela primeira fase, com as comunidades identificadas. Em 2(c), a nova rede construída na segunda etapa. Por fim, em 2(d) a rede obtida após a repetição da primeira e segunda fase na rede obtida em 2(c).

vértices são dados pela soma dos pesos das arestas entre vértices nas duas comunidades correspondentes. Conexões entre vértices na mesma comunidade tornam-se laços nesta comunidade na nova rede. A Figura 2 apresenta um exemplo destes passos em uma rede de entrada. Uma vez que a segunda fase esteja concluída, a primeira fase é reaplicada na nova rede. Cada execução de primeira e segunda fases é chamada de passo. Os passos prosseguem até que não ocorram mais mudanças e a máxima modularidade seja obtida.

Este método tem complexidade $O(m)$, onde m é o número de arestas. Isto significa que o método é linear, conseguindo um bom desempenho em redes de tamanhos variados. Além disso, o método obtém melhores valores de modularidade em comparação aos obtidos pelos algoritmos GN e CMN.

4. Proposta de alteração no método Louvain

O método Louvain possui vantagens interessantes sobre os outros métodos, como seu tempo de execução linear e comunidades com maior qualidade, de acordo com os valores de modularidade obtidos em comparação aos demais algoritmos. Porém, observamos que as comunidades são construídas de acordo com um critério muito simples: os vértices são analisados sequencialmente, ou seja, seguindo a ordem definida pelos rótulos dos vértices, de 0 a $n-1$.

Sendo o Louvain um método adequado para encontrar comunidades em redes complexas, desenvolvemos uma modificação, buscando aliar o mecanismo guloso aos conceitos de medidas de centralidade. Esta modificação ocorre apenas na primeira fase do método, e altera a ordem dos vértices a serem analisados quando as comunidades estão sendo construídas. O objetivo é verificar o impacto que a escolha do critério para o método guloso possui na modularidade final obtida.

As métricas de centralidade escolhidas foram *weighted degree* (grau ponderado), *betweenness* (intermediação) e *closeness* (proximidade), uma vez que quantificam, de maneira complementar, a conectividade de cada vértice ao resto da rede. O *weighted degree* é uma métrica local dos vértices uma vez que leva em consideração apenas seus vizinhos, retratando as interações entre estes vértices. O *betweenness* e o *closeness* avaliam cada vértice a partir de uma visão global de rede. Em função destas características achamos interessante explorar como estas métricas poderiam ser integradas ao processo de formação de comunidades.

Antes de iniciar a análise da primeira fase, as medidas de centralidade de todos os vértices da rede são calculadas e ordenadas. Após, a execução do método prossegue normalmente, porém, em lugar de analisar os vértices sequencialmente, a análise ocorre seguindo a ordem obtida a partir das medidas de centralidade. A segunda fase do método não sofreu alterações.

5. Resultados obtidos

Para observar os resultados, realizamos experimentos modificando uma implementação em Python do método Louvain [Odent and Saint-Guillain 2012], utilizando Python 3.5.2. As bases de dados utilizadas durante os testes foram as seguintes redes sociais:

- Arxiv e Hep-th-citations [Gehrke et al. 2003]: rede de citações entre cientistas no *journal Arxiv High Energy Physics*. A primeira possui 9377 vértices e 48214 arestas e a segunda, 27770 vértices e 352807 arestas.
- Facebook-ego [Leskovec and Mcauley 2012]: rede social de amizades no Facebook, coletada a partir da rede pessoal de alguns participantes anônimos. Possui 4039 vértices e 88234 arestas.
- *Zachary's karate club* [Zachary 1977]: rede de amizade entre 34 membros de um clube de karatê em uma universidade americana. Possui 34 vértices e 77 arestas.
- Os Miseráveis [Knuth 1993]: rede de coaparência dos personagens da obra literária *Os Miseráveis*. Possui 77 vértices e 254 arestas.

Os experimentos ocorreram da seguinte forma: cada medida de centralidade (*weighted degree*, *betweenness* e *closeness*) foi testada individualmente nas bases de dados. Além disso, para cada uma das medidas, na primeira execução os vértices foram ordenados em ordem crescente (do menor valor da medida para o maior), e na segunda, em ordem descrescente (do maior valor da medida para o menor). A Tabela 1 apresenta os valores de modularidade obtidos em cada teste e a Figura 3 apresenta o ganho que cada critério aliado às medidas de centralidade obteve em relação ao algoritmo original.

Tabela 1. Modularidade obtida originalmente e após as modificações no método Louvain em cada base de dados, destacada por cores. Em negrito, os melhores valores de modularidade obtidos.

	Arxiv	Hep-th-citations	Facebook-ego	Karate	Os miseráveis
Original	0,81431774	0,65047117	0,83481924	0,4298364	0,55555211
Weighted degree crescente	0,81422133	0,65468872	0,83544190	**0,42983640**	0,55552111
Weighted degree decrescente	0,81397961	0,64838742	0,83376415	**0,42983640**	**0,55827237**
Betweenness crescente	0,81470074	**0,65497587**	0,83493780	**0,42983640**	**0,55827237**
Betweenness decrescente	0,81466182	0,64465661	0,83455295	**0,42983640**	**0,55827237**
Closeness crescente	0,81438528	0,65491286	**0,83546865**	0,42696913	0,55553661
Closeness decrescenete	**0,81551700**	0,63228495	0,83393666	**0,42983640**	**0,55827237**

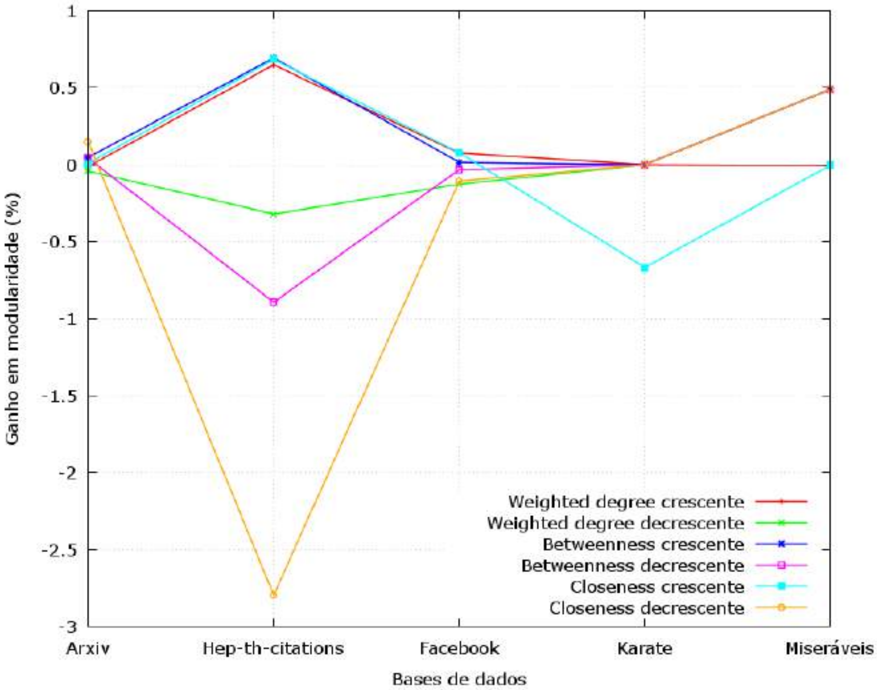

Figura 3. Ganho na modularidade obtido por cada critério através das medidas de centralidade em ordem crescente e decrescente.

Podemos observar que as alterações ocasionaram tanto valores melhores quanto piores em modularidade. Na menor e mais densa rede (Karatê), o uso dos critérios não alterou a modularidade obtida em quase todos os casos. Em Os Miseráveis (segunda menor rede), o uso das métricas trouxe melhoria, mas várias empataram no valor obtido. Este fato indica que a solução gulosa proposta possui melhor resultado em redes maiores e mais esparsas, pois em redes onde os vértices são altamente conectados, eles tendem a obter medidas de centralidade muito semelhantes.

Analisando os resultados obtidos para as três redes mais esparsas (Arxiv, Hep-th-citations e Facebook-ego) através da Figura 3, percebe-se que as estratégias de ordenação por meio das métricas *weighted degree*, *betweenness* e *closeness* em ordem descrescente levaram à perda em relação ao algoritmo original. Por outro lado, quando as métricas foram usadas em ordem crescente, houve ganho de modularidade. Este resultado indica que é uma boa estratégia classificar os vértices mais periféricos primeiro, evitando que os mesmos fiquem isolados, formando comunidades muito pequenas.

Podemos observar um indício deste resultado ao analisar o caso da métrica *betweenness*. Quando utilizada para ordenar os vértices de modo crescente, houve melhoria na modularidade de todas as redes testadas. Uma vez que esta métrica indica o quanto o vértice funciona para interligar os demais vértices na rede, vértices com alto *betweenness* são aqueles mais entre as comunidades, chamados vértices de fronteira.

De modo semelhante, a métrica *closeness* trouxe ganho de modularidade nas redes maiores ao ordenar os vértices crescentemente. Esta métrica aponta o quanto os vértices estão próximos aos demais vértices da rede. De forma similar ao *betweenness*, esta métrica classifica como mais centrais aqueles que estão nas fronteiras entre as comunidades.

Observando o resultado das duas métricas, que indicam a importância dos vértices em um contexto global, percebe-se que analisar primeiro os vértices mais periféricos leva à construção das comunidades em torno dos vértices de fronteira, aumentando a qualidade da solução. Já a métrica *weighted degree* denota a importância local do vértice na rede, ou seja, na sua própria comunidade. Utilizar a métrica para ordenar os vértices em ordem crescente trouxe resultado positivo em duas das três redes maiores, indicando que, mesmo no contexto local, analisar os vértices periféricos primeiro traz ganhos em qualidade das comunidades encontradas.

Portanto, os resultados indicam que o uso das medidas de centralidade trouxe benefícios para a solução gulosa do método Louvain em redes grandes e esparsas, ao utilizar a estratégia de construir comunidades a partir dos vértices mais periféricos. Por outro lado, a estratégia pode não ser promissora em redes pequenas e muito densas, pois nestas redes os vértices são muito conectados, tornando suas medidas de centralidade muito semelhantes.

6. Considerações finais

Neste projeto de iniciação científica, o foco esteve no estudo de algoritmos para o problema de detecção de comunidades em redes sociais, mais especificamente, no método Louvain, que utiliza heurística gulosa e é mais eficiente em tempo de execução e em modularidade das comunidades encontradas, em comparação a outros algoritmos da literatura.

O objetivo foi aliar os conceitos de medidas de centralidade em redes complexas ao método Louvain. Para isto, alteramos o critério do algoritmo de construção das comunidades, fazendo com que os vértices fossem analisados de acordo com suas medidas de centralidade, em ordem crescente e decrescente de *weighted degree*, *betweenness* e *closeness centrality*.

Descrevemos neste trabalho o impacto que estas modificações trouxeram em termos de modularidade obtida pelo algoritmo. Durante nossos experimentos, observamos que ordenar os vértices de acordo com as medidas, da menor pontuação para a maior, leva à construção de comunidades com mais qualidade, pois os vértices mais periféricos vão sendo progressivamente agrupados aos mais centrais, evitando vértices isolados e comunidades muito pequenas. Além disso, os resultados indicam que estas modificações possuem maior eficiência em redes esparsas, pois nelas as medidas de centralidade de cada vértice são mais variáveis.

Como trabalhos futuros, pretendemos fazer uso de mais de uma métrica para o critério guloso, verificando se esta mudança possui alguma influência na qualidade das comunidades encontradas pelo método Louvain. Temos interesse em verificar o impacto da geração de várias ordenações aleatórias no desempenho do algoritmo, analisando o melhor resultado e buscando relacioná-lo às medidas de centralidade. Por fim, pretendemos também analisar a relação dos resultados obtidos com outras características estruturais da rede, como o coeficiente de agrupamento global e local.

Referências

Bedi, P. and Sharma, C. (2016). Community detection in social networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(3):115–135.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. (2007). On finding graph clusterings with maximum modularity. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 121–132. Springer.

Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.

Dehmer, M. and Emmert-Streib, F. (2009). *Analysis of complex networks: from biology to linguistics*. John Wiley & Sons.

Gehrke, J., Ginsparg, P., and Kleinberg, J. (2003). Overview of the 2003 kdd cup. *ACM SIGKDD Explorations Newsletter*, 5(2):149–151.

Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.

Hofstad, R. V. D. (2016). *Random graphs and complex networks*, volume 1. Cambridge University Press.

Knuth, D. E. (1993). *The Stanford GraphBase: a platform for combinatorial computing*, volume 37. Addison-Wesley Reading.

Leskovec, J. and Mcauley, J. J. (2012). Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, pages 539–547.

Motta, R., de Alneu Andrade Lopes, and de Oliveira, M. C. F. (2009). Centrality measures from complex networks in active learning. In *International Conference on Discovery Science*, pages 184–196. Springer.

Odent, J. and Saint-Guillain, M. (2012). Automatic detection of community structure in networks. Technical report, Université catholique de Louvain.

Özturk, K. (2014). *Community detection in social networks*. PhD thesis, Middle East Technical University.

Perra, N. and Fortunato, S. (2008). Spectral centrality measures in complex networks. *Physical Review E*, 78(3):036107.

Scholz, M. (2009). Node degree distribution. Disponível em: `www.network-science.org/powerlaw_scalefree_node_degree_distribution.html`. Acessado em: 01/2017.

Scott, J. and Carrington, P. J. (2011). *The SAGE handbook of social network analysis*. SAGE publications.

Shen, H.-W. (2013). *Community structure of complex networks*. Springer Science & Business Media.

Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473.