

Agrupamento em grafos direcionados pela maximização da modularidade ajustada

Camila Pereira dos Santos

Universidade Federal de São Paulo

Av. Cesare Monsueto Giulio Lattes, 1211 - Jardim Santa Ines I, São José dos Campos - SP
santos.camila@unifesp.br

Mariá Cristina Vasconcelos Nascimento Rosset

Universidade Federal de São Paulo

Av. Cesare Monsueto Giulio Lattes, 1211 - Jardim Santa Ines I, São José dos Campos - SP
mcv.nascimento@unifesp.br

RESUMO

Agrupamento em grafos é um problema que visa identificar uma partição do conjunto de vértices, de forma que os vértices intra-grupos sejam altamente relacionados. Dentre as diversas estratégias existentes para definir tais partições, destacam-se aquelas baseadas na maximização da modularidade. Na literatura, não foram identificados estudos que apresentem o comportamento da modularidade em grafos direcionados, já que a mesma é conhecida por ter um limite de resolução em grafos não direcionados. Para tratar o limite de resolução, existe um ajuste na medida que, apesar de seu potencial, poucos algoritmos o exploram. Neste trabalho, propõe-se a maximização da modularidade ajustada em grafos direcionados por uma heurística construtiva semi-gulosa. Os resultados dos experimentos com grafos artificiais e reais indicam um melhor desempenho da heurística proposta em relação a outros algoritmos de agrupamento em grafos encontrados na literatura. A aluna desenvolveu o projeto completo, sob orientação da orientadora.

PALAVRAS CHAVE. Agrupamento em grafos direcionados, Maximização da modularidade ajustada.

Área Principal: TAG - Teoria e Algoritmos em Grafos, MH - Metaheurísticas.

ABSTRACT

Graph clustering is a problem that aims at identifying groups of highly related vertices. Among the existing strategies to define such clusterings, we underline those based on the modularity maximization problem. In the literature, studies that present the behaviour of the modularity on directed graphs cannot be found. To deal with the resolution limit, there is a way to adjust the measure that, even though it seems to be a good alternative, a few algorithms effectively explore it. In this paper, we propose a semi-greedy constructive heuristic for the adjusted modularity maximization in directed graphs. The results of the experiments with real and artificial networks indicate a better performance of the proposed strategy over the most used algorithm for clustering directed networks in the literature. The student developed the whole project, under the supervision of her advisor.

KEYWORDS. Clustering in directed networks, Adjusted modularity maximization.

Main Area: TAG - Theory and Algorithms in Graphs, MH - Metaheuristics.

1. Introdução

Muitas redes reais apresentam uma propriedade de estrutura de comunidade (Girvan e Newman, 2002), que é a característica de possuir grupos de vértices altamente relacionados. Em redes direcionadas, essa propriedade estabelece que existem mais arcos entre vértices de um mesmo grupo do que entre vértices de grupos distintos. Em particular, muitas aplicações são caracterizadas por redes direcionadas, ou grafos direcionados (Malliaros e Vazirgiannis, 2013).

Existem diversas medidas na literatura para avaliar a qualidade de uma partição, ou seja, avaliar quão bem relacionados estão os vértices dos grupos em relação ao restante do grafo. A medida de modularidade, proposta por Girvan e Newman (2002) e estendida para grafos direcionados em (Newman, 2004), avalia o número de conexões intra-grupo em relação ao número esperado de arestas/arcos entre o mesmo grupo de vértices, porém em um grafo aleatório com a mesma sequência de graus. Algoritmos guiados por essa medida apresentam resultados próximos ao esperado para um número razoável de instâncias e possuem a vantagem de não precisarem de definição do número de grupos do grafo, a depender da topologia da rede em questão.

Por meio da modularidade, o problema de agrupamento em grafos se resume a um problema de otimização, cujo o objetivo é encontrar partições que maximizem a modularidade. Por ser \mathcal{NP} -Completo (Brandes et al., 2008) e por muitas aplicações serem caracterizadas por instâncias muito grandes, o problema de maximização da modularidade é frequentemente resolvido por métodos heurísticos na literatura. O problema visando a medida direcionada, entretanto, não tem sido abordado na literatura. De acordo com Malliaros e Vazirgiannis (2013), ainda que o direcionamento das arestas de redes direcionadas represente características importantes, a maior parte dos métodos de agrupamento em grafos ignora as direções dos arcos de redes direcionadas, de forma a tratá-las como redes não direcionadas.

A modularidade, contudo, apresenta um problema chamado de limite de resolução, demonstrado por Fortunato e Barthélemy (2007). Segundo Fortunato e Barthélemy (2007), a solução ótima do problema tem os seus *clusters* com um número máximo de vértices limitado em relação ao total de conexões do grafo. Reichardt e Bornholdt (2006) propõem a inclusão de um parâmetro na medida de modularidade para controlar o limite de resolução. Contudo, existem poucos trabalhos que explorem essa medida ajustada ou visem maximizá-la (Santos e Nascimento, 2013). Carvalho et al. (2014) propõem uma rede neural para identificar os valores ideais do parâmetro de ajuste para cada grafo.

Este trabalho de iniciação científica (IC) teve duas fases. Na primeira fase (primeiro ano), a aluna apresentou um algoritmo para redes não direcionadas e os resultados indicaram um excelente desempenho para centenas de grafos artificiais testados e comparando com resultados de algoritmos *benchmark* da literatura. Na segunda fase, detalhada em mais detalhes neste artigo, aprimorou-se o algoritmo de Santos e Nascimento (2013) para controlar o limite de resolução, de forma a obter melhores partições e visou-se sua adaptação para grafos direcionados. Para isso, valores do parâmetro da medida de modularidade ajustada de Reichardt e Bornholdt (2006) foram definidos por meio do método de Carvalho et al. (2014). Nos experimentos, empregam-se os melhores algoritmos de agrupamento em grafos direcionados e usam-se um grafo real e um conjunto de grafos gerados artificialmente. Segundo resultados, a estratégia proposta tem resultados mais consistentes do que os algoritmos da literatura, em termos de proximidade com a partição esperada.

2. Problema de maximização da modularidade

Uma abordagem para lidar com o problema de agrupamento em grafos é defini-lo como um problema de otimização inteira, no qual o objetivo é encontrar a partição do conjunto de vértices que melhor represente a estrutura de comunidade do grafo. A medida

de modularidade (Girvan e Newman, 2002), o *clustering coefficient* (Nascimento, 2013) e o *map equation* (Rosvall e Bergstrom, 2007) são algumas medidas utilizadas para a função objetivo avaliar uma partição.

Seja $G = (V, E, \phi)$ um grafo direcionado ponderado, em que $V(G)$ é seu conjunto de vértices, $E(G)$ é seu conjunto de arcos e $\phi : E(G) \rightarrow \mathbb{R}$ é uma função que atribui pesos aos arcos. Um arco $e \in E(G)$ é uma tupla ordenada (v, u) , em que $v, u \in V(G)$, assim, v é a cauda do arco e u a cabeça. Um vértice u é adjacente a v se $(v, u) \in E(G)$. A cada arco $e = (u, v)$ é atribuído um peso $\phi(e)$. Neste trabalho, um grafo não ponderado é definido como um grafo ponderado no qual os pesos de todos os arcos têm valor unitário. Considera-se m o peso total dos arcos do grafo, ou seja, $m = \sum_{e \in E(G)} \phi(e)$. A vizinhança de entrada de um vértice v , $\mathcal{N}^-(v)$, e de saída, $\mathcal{N}^+(v)$, são os conjuntos de vértices que são, respectivamente, cauda e cabeça dos arcos incidentes em v . Os graus de entrada e saída de um vértice v são dados, respectivamente, por $d_g^-(v) = \sum_{u \in \mathcal{N}^-(v)} \phi(u, v)$ e $d_g^+(v) = \sum_{u \in \mathcal{N}^+(v)} \phi(v, u)$. Uma partição $\mathcal{C} : v \rightarrow \mathbb{N}$ associa um índice para cada vértice, de forma que dois vértices v e z estão no mesmo *cluster* se e somente se $\mathcal{C}(v) = \mathcal{C}(z)$.

A medida de modularidade, proposta por Girvan e Newman (2002) e adaptada para grafos direcionados por (Leicht e Newman, 2008), avalia o peso total dos arcos existentes entre vértices de um mesmo grupo em relação ao peso total de arcos esperado em um grafo aleatório que apresente a mesma sequência de graus dos vértices do grafo em análise. Uma formulação para a medida de modularidade de uma partição \mathcal{C} de um grafo direcionado é apresentada na Equação (1).

$$Q(\mathcal{C}) = \frac{1}{m} \sum_{C \in \mathcal{C}} \sum_{i, j} (\phi(i, j) - p(i, j)) \quad (1)$$

em que $p(i, j) = \frac{d_G^-(i)d_G^+(j)}{m}$ é o peso esperado do arco (i, j) , $i, j \in V(G)$, em um grafo aleatório com a mesma sequência de graus de G . Assim, se o peso de um arco entre dois vértices for maior que o peso esperado no grafo aleatório definido anteriormente, então há indícios de que os vértices pertencem a um mesmo *cluster*.

Brandes et al. (2008) demonstraram que os valores possíveis para a medida de modularidade pertencem ao intervalo $[-\frac{1}{2}, 1)$. A medida de modularidade, entretanto, apresenta um problema chamado de limite de resolução, que é explicado na seção seguinte.

2.1. Limite de Resolução

Fortunato e Barthélemy (2007) identificaram um limite no número máximo de vértices dos *clusters* que na solução ótima do problema de maximização da modularidade, chamado de limite de resolução. Estudando grafos não direcionados, os autores demonstraram a dependência entre o número máximo de vértices possível nos *clusters* da solução ótima da maximização da modularidade e o número total de arestas do grafo, no caso de grafos não ponderados (ou pelo peso total das arestas do grafo, no caso de grafos ponderados). Dessa forma, a maximização da modularidade falha em identificar *clusters* menores que esse tamanho máximo.

Reichardt e Bornholdt (2006) propuseram a inclusão de um parâmetro de ajuste λ na formulação da medida modularidade. Essa nova formulação é chamada de modularidade ajustada por Carvalho et al. (2014). A medida de modularidade ajustada (Reichardt e Bornholdt, 2006) para grafos direcionados é apresentada na Equação (2).

$$Q(\mathcal{C}) = \frac{1}{m} \sum_{C \in \mathcal{C}} \sum_{i, j \in C} (\phi(i, j) - \lambda p(i, j)) \quad (2)$$

Nessa medida ajustada, o parâmetro λ ajusta o peso de arcos esperado entre os pares de vértices. Assim, o valor de λ determina a escala dos *clusters* encontrados

pela maximização da modularidade. Valores de λ menores diminuem o peso esperado de arcos entre vértices, de forma que mais vértices são considerados como tendo tendência de agrupamento e, conseqüentemente, *clusters* maiores são identificados. Analogamente, valores de λ fornecem como soluções partições com *clusters* menores.

Carvalho et al. (2014) propõem um método baseado em redes neurais para identificar os valores do parâmetro λ adequados para um dado grafo em consideração. Dessa forma, pode-se abordar o agrupamento em grafos pela maximização da modularidade ajustada, com valor λ definido pelo método.

3. Método proposto

A heurística proposta neste trabalho visa maximizar o valor de modularidade ajustada para grafos direcionados. Os valores do parâmetro λ da modularidade ajustada são definidos para cada grafo por meio do método proposto por Carvalho et al. (2014). Além disso, o algoritmo *Greedy Randomized Search Procedure* (GRASP) que foi proposto por Santos e Nascimento (2013) para a maximização da modularidade ajustada foi adaptado para grafos direcionados¹.

O GRASP é uma meta-heurística composta por duas fases que executam iterativamente: a de construção, na qual uma solução inicial é construída, e a de busca local, na qual a solução é refinada a fim de melhorar sua qualidade. Neste trabalho, a fase de construção de Santos e Nascimento (2013) é adaptada para o caso direcionado e consiste em operações de contração de arcos. Uma contração de arco $e = (v, u)$ em um grafo G_i é definida como a remoção do arco e do grafo, seguida pelo agrupamento dos vértices v e u , pontas de e , em um vértice z , chamado supernó. Essa operação também agrupa os *clusters* de v e u . A partir do grafo original G_0 , em que cada vértice está em um *cluster* isolado, $C_v = \{v\}$ para todo vértice v de $V(G)$, cada contração de arco em um grafo G_i gera um grafo G_{i+1} , com um vértice a menos. O último grafo gerado, G_k é chamado de grafo base e, conseqüentemente, é resultante de k contrações de arcos.

A fase construtiva, de forma semi-gulosa, constrói uma solução por meio de contrações de arcos em supernós, que correspondem aos grupos do agrupamento. Para isso, seja M o conjunto de todas as contrações de arcos possíveis em um grafo. Construa uma Lista de Candidatos Restrita (LCR), composta pelas melhores $\alpha|M|$ contrações, em que $\alpha \in [0, 1]$ define o grau de aleatoriedade na escolha dos arcos a serem contraídos. Para avaliar a qualidade da contração de um arco $e = (v, s)$, considera-se o ganho de modularidade ajustada em agrupar os *clusters* C_v e C_s . O ganho de modularidade (qualidade da contração de um arco) é formulado de acordo com a Equação (3).

$$\Delta Q(C_v, C_s) = \Delta Q(v, s) = \frac{1}{m} (\phi(v, s) - \lambda p(v, s)) \quad (3)$$

em que $\alpha = 0$ resulta em uma solução puramente gulosa e $\alpha = 1$ em uma solução aleatória. Os arcos a serem contraídos são escolhidos aleatoriamente da LCR, até que não exista mais nenhuma contração de arco que resulte em um ganho de modularidade ajustada. A abordagem de Santos e Nascimento (2013) considera os nós de G_k como os *clusters* da partição. Portanto, nessa abordagem, cada *cluster* é composto pelos vértices de cada supernó do grafo da última contração.

Adicionalmente, Santos e Nascimento (2013) aplicam uma estratégia de contração definitiva de arcos. Nessa estratégia, a cada 3 iterações, são contraídos arcos incidentes a vértices que estejam na mesma partição em pelo menos 50% das melhores partições obtidas

¹Essa meta-heurística foi desenvolvida no primeiro ano de Iniciação Científica da aluna, quando a mesma focou em grafos não direcionados. A aluna aperfeiçoou o trabalho em seu trabalho de conclusão de curso, e propôs uma variação para o problema em grafos direcionados. Não se reporta toda a abordagem e estudo realizado para o artigo não ficar extenso.

pelo algoritmo até a iteração corrente. Dessa forma, as próximas iterações irão partir do grafo contendo as contrações de arcos definitivas efetuadas até a iteração anterior. Ao contrair os arcos, os vértices são agrupados de maneira definitiva e, portanto, essa estratégia constitui um mecanismo de memória para guiar a construção das próximas soluções, além de reduzir o número de vértices e arestas do grafo ao longo das iterações.

4. Estratégia Consensual

A rede neural proposta por Carvalho et al. (2014) foi utilizada para determinar o intervalo mais apropriado para definir o parâmetro de ajuste da modularidade. Os seguintes intervalos de valores foram considerados: $[1; 1, 5]$, $[1, 6; 2]$, $[2, 1; 2, 5]$, $[2, 6; 3]$, $[3, 1; 3, 5]$, $[3, 6; 4]$, $[4, 1; 4, 5]$ e $[4, 6; 5]$.

Para um determinado intervalo $[l; r]$, são obtidas partições pelo algoritmo proposto para os valores do parâmetro de ajuste $l, l+0, 1, \dots, r$. As partições obtidas são combinadas por meio da estratégia de *clustering* consensual, na qual vértices que estiverem em um mesmo *cluster* em mais de 50% das partições a serem combinadas estarão em um mesmo *cluster* na partição final, saída do algoritmo.

5. Experimentos

O algoritmo proposto, chamado de CONCLUS, foi validado por meio de dois experimentos com grafos direcionados. No primeiro experimento, um grafo real direcionado é estudado e, no segundo experimento, um conjunto de grafos *benchmark* artificiais é utilizado para validar o algoritmo proposto. O melhor algoritmo para agrupar grafos direcionados, o *Infomap* (Rosvall e Bergstrom, 2007), que é um dos poucos propostos para grafos direcionados foi considerado. Além disso, o *Label Propagation* (LP) (Raghavan et al., 2007), que ignora as direções dos arcos, também foi considerado. Em ambos os experimentos, as partições foram avaliadas por meio da medida *Normalized Mutual Information* (NMI) (Dannon et al., 2005), que avalia a correlação, nesse contexto, entre duas partições. Portanto, avalia-se a correlação entre a partição encontrada e a esperada. O valor do NMI varia de 0 a 1, sendo 0 nenhuma correlação e 1 a igualdade entre as partições.

5.1. Experimento I

Nesse experimento, considera-se o grafo direcionado *polBlogs* (Adamic e Glance, 2005). O *polBlogs* é uma rede que representa os *hyperlinks* entre *blogs* relacionados à política americana, na forma de um grafo direcionado, no qual os vértices são os *blogs* e os arcos são os *hyperlinks*. Essa rede é caracterizada por *blogs* liberais ou conservativos. Portanto, a partição ideal desse grafo é composta por dois *clusters*, cada um representando uma dessas categorias.

A partição obtida pelo CONCLUS é ilustrada na Figura 1. Nessa figura, as componentes desconexas menores foram desconsideradas e os *clusters* são representados como vértices. Quanto maior o tamanho do vértice e a espessura dos arcos, maior o número de vértices no *cluster* e arcos entre os *clusters*, respectivamente. As duas categorias de *blogs* são identificadas na figura com os rótulos 1 e 3. O valor de NMI obtido para o CONCLUS foi de 0,381536, superior ao *Infomap*, que atingiu NMI 0,314973 e ao LP, que obteve 0,2949371. A grande quantidade de vértices isolados no grafo é uma das razões para explicar os baixos valores de NMI, ainda que os dois *clusters* principais tenham sido identificados.

5.2. Experimento II

Nesse experimento, foram gerados grafos direcionados artificiais, caracterizados por *clusters* pequenos, no *software* de Lancichinetti et al. (2008). Os parâmetros dos grafos foram os sugeridos pelos autores, com exceção do coeficiente de mistura μ , que define o grau de mistura dos *clusters*, isto é, quão definidos são. Foram gerados grafos com 500, 1000, 2000, 3000 e 4000 vértices. Para cada tamanho, foram gerados 5 grafos para cada valor de $\mu \in \{0, 1; 0, 2; \dots; 0, 8\}$. Os resultados obtidos para os grafos são apresentados na Tabela 1.

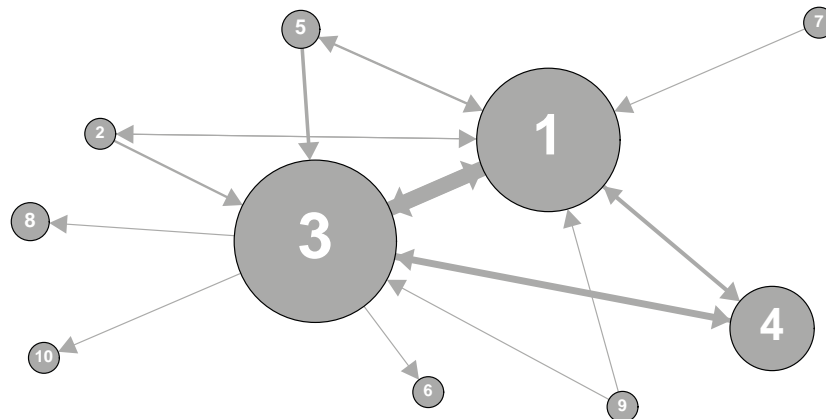


Figura 1: Partição encontrada pelo CONCLUS para o grafo direcionado *polBlogs*.

O número de execuções independentes do CONCLUS e do LP foram 5, em contrapartida ao *Infomap*, para o qual, devido à sugestão dos autores e ao seu baixo tempo computacional, foram feitas 100 execuções. Nessa tabela, as colunas 1 e 2 indicam o número de vértices do grafo e o valor de μ , respectivamente. Nas colunas seguintes são apresentados, para cada algoritmo, os valores de NMI médios em execuções independentes ("NMI"), a variação padrão entre os valores de NMI obtidos nessas execuções (" σ_{ex} ") e o desvio padrão dos valores de NMI de grafos com o mesmo μ (" σ_{μ} "). O tempo computacional médio das execuções é reportado apenas para o CONCLUS, entretanto, tanto o *Infomap* quanto o LP tem complexidade de custo computacional linear no tamanho do grafo, enquanto o CONCLUS é quadrático.

Os resultados apresentados na tabela permitem observar que o CONCLUS foi o único método que obteve partições com NMI maior que 0 para $\mu = 0,8$ e, além disso, obteve valores maiores que 0,65. Além disso, para $\mu = 0,7$ os poucos valores de NMI do *Infomap* maiores que 0 ainda são inferiores aos obtidos pelo CONCLUS. Para os valores de μ menores que 0,7, o CONCLUS foi competitivo com os demais. Uma característica positiva do CONCLUS é sua robustez, pois, ao contrário do *Infomap*, que apresenta alguns valores isolados de NMI ruins até para baixos valores de μ , e do LP, para o qual os valores decrescem conforme μ aumenta, os resultados do CONCLUS são consistentes.

6. Conclusão

Neste trabalho de IC foram propostas uma meta-heurística GRASP e uma heurística construtiva para, respectivamente, o problema de maximização da modularidade em grafos não direcionados e direcionados. Este artigo reporta os resultados e o método desenvolvido para tratar grafos direcionados, para os quais, a literatura é bastante escassa. O algoritmo proposto denominado de CONCLUS, neste trabalho, visa o problema da maximização da medida de modularidade ajustada. O método de identificação do parâmetro de ajuste na modularidade é o responsável por reduzir o impacto do limite de resolução do CONCLUS.

Dois experimentos computacionais realizados atestam a superioridade do algoritmo proposto em relação aos dois algoritmos comparados na literatura, sendo um deles o melhor algoritmo de agrupamento reportado na literatura para grafos direcionados. Em especial, no segundo experimento, os elevados valores de NMI obtidos pelo CONCLUS para os maiores coeficientes de mistura indicam que é adequado para grafos que apresentem um nível

Tabela 1: Resultados do CONCLUS, *Infomap* e LP para grafos artificiais gerados.

| Tamanho | μ | CONCLUS | | | | <i>Infomap</i> | | | LP | | |
|---------|-------|---------|---------------|----------------|-----------|----------------|---------------|----------------|-------|---------------|----------------|
| | | NMI | σ_{ex} | σ_{μ} | Tempo (s) | NMI | σ_{ex} | σ_{μ} | NMI | σ_{ex} | σ_{μ} |
| 500 | 0,1 | 0,978 | 0,000 | 0,002 | 2,634 | 0,828 | 0,000 | 0,154 | 0,998 | 0,000 | 0,001 |
| 500 | 0,2 | 0,947 | 0,001 | 0,009 | 3,547 | 1,000 | 0,000 | 0,000 | 0,996 | 0,001 | 0,002 |
| 500 | 0,3 | 0,981 | 0,001 | 0,008 | 4,640 | 1,000 | 0,000 | 0,000 | 0,995 | 0,001 | 0,002 |
| 500 | 0,4 | 0,989 | 0,000 | 0,002 | 5,521 | 1,000 | 0,000 | 0,000 | 0,983 | 0,002 | 0,004 |
| 500 | 0,5 | 0,987 | 0,000 | 0,002 | 6,280 | 0,999 | 0,000 | 0,000 | 0,785 | 0,032 | 0,136 |
| 500 | 0,6 | 0,977 | 0,001 | 0,005 | 7,243 | 0,994 | 0,000 | 0,002 | 0,000 | 0,000 | 0,000 |
| 500 | 0,7 | 0,762 | 0,002 | 0,030 | 9,881 | 0,493 | 0,000 | 0,180 | 0,000 | 0,000 | 0,000 |
| 500 | 0,8 | 0,662 | 0,002 | 0,009 | 10,875 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 1000 | 0,1 | 0,990 | 0,000 | 0,002 | 22,578 | 0,687 | 0,000 | 0,172 | 1,000 | 0,000 | 0,000 |
| 1000 | 0,2 | 0,965 | 0,001 | 0,008 | 34,365 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 1000 | 0,3 | 0,996 | 0,000 | 0,002 | 47,996 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 1000 | 0,4 | 1,000 | 0,000 | 0,000 | 58,886 | 1,000 | 0,000 | 0,000 | 0,995 | 0,001 | 0,002 |
| 1000 | 0,5 | 1,000 | 0,000 | 0,000 | 70,581 | 1,000 | 0,000 | 0,000 | 0,992 | 0,001 | 0,003 |
| 1000 | 0,6 | 1,000 | 0,000 | 0,000 | 81,136 | 1,000 | 0,000 | 0,000 | 0,302 | 0,040 | 0,192 |
| 1000 | 0,7 | 0,977 | 0,001 | 0,005 | 76,725 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 1000 | 0,8 | 0,732 | 0,001 | 0,013 | 89,129 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 2000 | 0,1 | 1,000 | 0,000 | 0,000 | 217,145 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 2000 | 0,2 | 0,987 | 0,000 | 0,004 | 394,828 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 2000 | 0,3 | 1,000 | 0,000 | 0,000 | 579,786 | 1,000 | 0,000 | 0,000 | 0,999 | 0,000 | 0,000 |
| 2000 | 0,4 | 1,000 | 0,000 | 0,000 | 794,380 | 1,000 | 0,000 | 0,000 | 0,999 | 0,000 | 0,001 |
| 2000 | 0,5 | 1,000 | 0,000 | 0,000 | 996,105 | 1,000 | 0,000 | 0,000 | 0,995 | 0,001 | 0,002 |
| 2000 | 0,6 | 1,000 | 0,000 | 0,000 | 1096,392 | 1,000 | 0,000 | 0,000 | 0,660 | 0,070 | 0,189 |
| 2000 | 0,7 | 1,000 | 0,000 | 0,000 | 1148,630 | 0,600 | 0,000 | 0,219 | 0,000 | 0,000 | 0,000 |
| 2000 | 0,8 | 0,787 | 0,003 | 0,038 | 1542,796 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 3000 | 0,1 | 1,000 | 0,000 | 0,000 | 1017,708 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 3000 | 0,2 | 0,992 | 0,000 | 0,003 | 2044,569 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 3000 | 0,3 | 1,000 | 0,000 | 0,000 | 2843,455 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 3000 | 0,4 | 1,000 | 0,000 | 0,000 | 3801,536 | 1,000 | 0,000 | 0,000 | 0,999 | 0,000 | 0,001 |
| 3000 | 0,5 | 1,000 | 0,000 | 0,000 | 4613,192 | 1,000 | 0,000 | 0,000 | 0,996 | 0,001 | 0,002 |
| 3000 | 0,6 | 1,000 | 0,000 | 0,000 | 5289,309 | 1,000 | 0,000 | 0,000 | 0,903 | 0,044 | 0,072 |
| 3000 | 0,7 | 1,000 | 0,000 | 0,000 | 5381,116 | 1,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 3000 | 0,8 | 0,724 | 0,007 | 0,033 | 9626,669 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 4000 | 0,1 | 1,000 | 0,000 | 0,000 | 2882,299 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 4000 | 0,2 | 0,997 | 0,000 | 0,001 | 5613,915 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 4000 | 0,3 | 1,000 | 0,000 | 0,000 | 9049,245 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 4000 | 0,4 | 1,000 | 0,000 | 0,000 | 11041,230 | 1,000 | 0,000 | 0,000 | 1,000 | 0,000 | 0,000 |
| 4000 | 0,5 | 1,000 | 0,000 | 0,000 | 13817,134 | 1,000 | 0,000 | 0,000 | 0,994 | 0,001 | 0,003 |
| 4000 | 0,6 | 1,000 | 0,000 | 0,000 | 17461,386 | 1,000 | 0,000 | 0,000 | 0,822 | 0,036 | 0,141 |
| 4000 | 0,7 | 1,000 | 0,000 | 0,000 | 19034,504 | 1,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 4000 | 0,8 | 0,745 | 0,006 | 0,026 | 28671,003 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |

alto de limite de resolução. Além disso, é possível observar a consistência dos resultados encontrados pelo método. Entretanto, o custo computacional do CONCLUS é consideravelmente superior ao dos demais métodos. Portanto, para lidar com grafos com dezenas de milhares de nós, um trabalho futuro é desenvolver um algoritmo com custo computacional inferior. Uma sugestão é adaptar o *Infomap*, substituindo a função objetivo *map equation* pela modularidade ajustada conforme o estudo aqui realizado.

O trabalho completo desta IC está submetido em periódico internacional.

Agradecimentos

As autoras agradecem a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro.

Referências

- Adamic, L. A. e Glance, N.** (2005). The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM.
- Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikolosk, Z. e Wagner, D.** (2008). On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20:172–188.
- Carvalho, D. M., Resende, H. e Nascimento, M. C. V.** (2014). Modularity maximization adjusted by neural networks. In *21st International Conference in Neural Information Processing (ICONIP)*, volume 8834, pages 287–294. Springer.
- Danon, L., Diaz-Guilera, A., Duch, J. e Arenas, A.** (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008–[11 pages].
- Fortunato, S. e Barthélemy, M.** (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36.
- Girvan, M. e Newman, M. E. J.** (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99:7821–7826.
- Lancichinetti, A., Fortunato, S. e F, R.** (2008). Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78:046110–[5 pages].
- Leicht, E. A. e Newman, M. E.** (2008). Community structure in directed networks. *Physical review letters*, 100(11):118703.
- Malliaros, F. D. e Vazirgiannis, M.** (2013). Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4):95–142.
- Nascimento, M. C.** (2013). Community detection in networks via a spectral heuristic based on the clustering coefficient. *Discrete Applied Mathematics*, 176(0):89–99. ISSN 0166-218X. doi:http://dx.doi.org/10.1016/j.dam.2013.09.017.
- Newman, M.** (2004). Analysis of weighted networks. *Physical Review*, 70:056131.
- Raghavan, U. N., Albert, R. e Kumara, S.** (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3 Pt 2):036106.
- Reichardt, J. e Bornholdt, S.** (2006). Statistical mechanics of community detection. *Physical Review E*, 74:016110.
- Rosvall, M. e Bergstrom, C. T.** (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences U.S.A*, 104:7327–[5 pages].
- Santos, C. e Nascimento, M.** (2013). Clustering consensual para a detecção eficiente de comunidades em redes pela modularidade ajustada. *Aceito para publicação na Revista de Iniciação Científica da SBC e publicado nos anais do CSBC 2013*.