

Classificação de Relações Sociais para Melhorar a Detecção de Comunidades

Jeancarlo C. Leão, Michele A. Brandão, Pedro O. S. Vaz de Melo,
Alberto H. F. Laender

¹Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte – MG

{jcl,micheleabrandao,olmo,laender}@dcc.ufmg.br

Abstract. *Social relationships can be divided into different classes based on the regularity with which they occur and the similarity among them. In this context, we propose a process to handle social network data that exploits temporal features to improve the detection of communities by existing algorithms. By removing random interactions, we observe that social networks converge to a topology with more purely social relationships and more modular communities.*

Resumo. *Relacionamentos sociais podem ser separados em diferentes classes pela regularidade com que ocorrem e pela similaridade entre eles. Neste contexto, propomos um processo para tratamento de dados de redes sociais que explora as características temporais para melhorar a detecção de comunidades por algoritmos existentes. Por meio de um processo de remoção de interações aleatórias, observamos que as redes sociais convergem para uma topologia com interações mais puramente sociais e comunidades com maior modularidade.*

1. Introdução

Redes sociais dinâmicas geralmente passam por alterações em sua estrutura decorrente das interações entre seus membros ao longo do tempo. Tais redes sociais podem ser representadas por um grafo temporal, no qual os vértices representam suas entidades e as arestas, as interações, associadas ao tempo em que ocorreram. Nesse tipo de rede, a detecção de comunidades tem sido pouco estudada. Em geral, a própria rede temporal é utilizada como entrada para algoritmos que encontram as comunidades explorando diretamente a sua dimensão temporal. A detecção de comunidades em tais redes difere de como é feito em uma rede social estática, onde a dimensão temporal é adequada ou simplesmente ignorada.

Muitos estudos sobre detecção de comunidades utilizam grafos estáticos em razão da maior dificuldade em considerar o aspecto temporal. Entretanto, não considerar o aspecto temporal gera perda de informação em relação à ordem e proximidade das interações [Holme & Saramäki 2012]. Essa simplificação causa ruído informacional nas relações sociais, o que pode gerar erros na associação de membros às suas respectivas comunidades. Por exemplo, considere um grupo de pessoas que não se conheciam e trocam muitas mensagens eletrônicas em um único dia, mas depois não voltam a se comunicar novamente. Agora considere um outro grupo de pessoas que trocam muitas mensagens regularmente ao longo de anos. Apesar dos relacionamentos entre os membros de ambos os grupos possuírem uma mesma topologia quando se considera uma rede estática, a dimensão temporal permite diferenciar a existência ou não de uma comunidade.

Uma propriedade fundamental compartilhada pelas diferentes definições de comunidade¹ em redes sociais é a presença de relações sociais reais [Newman 2006], que geralmente ocorrem ao longo do tempo [Vaz de Melo et al. 2015]. Isso motiva o uso da dimensão temporal na identificação de comunidades a partir de redes puramente sociais. Assim, ao invés de desenvolver novos algoritmos para detecção de comunidades, neste artigo propomos um método para favorecer aqueles já existentes explorando-se propriedades temporais.

Mais especificamente, propomos um processo que avança a detecção de comunidades em redes estáticas a partir do uso da dimensão temporal. Para isso, removemos as interações aleatórias que representam ruído quando mapeadas em redes estáticas. Então, a rede resultante dessa filtragem é constituída apenas por relações sociais regulares e, assim, as suas estruturas sociais tornam-se mais aparentes. Isso naturalmente implica em um aumento da modularidade² da rede, que é uma medida da força da divisão de uma rede em módulos (ou comunidades). Redes com alta modularidade têm conexões densas entre os nós dentro de módulos, mas escassas conexões entre nós em diferentes módulos. Mostramos que o método proposto aumenta significativamente a modularidade da rede quando essa é dividida em comunidades por algoritmos de detecção de comunidades do estado da arte. Assim, a contribuição principal deste trabalho é um processo de mineração de relações sociais que promove a obtenção de comunidades com maior modularidade por algoritmos já existentes na literatura.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta uma breve revisão de trabalhos relacionados. A seguir, a Seção 3 apresenta a modelagem adotada para representação das redes sociais consideradas e os conjuntos de dados utilizados nos experimentos, bem como descreve os algoritmos selecionados para a avaliação na detecção de comunidades. Em seguida, a Seção 4 descreve o processo proposto para melhorar a detecção de comunidades. A Seção 5 apresenta os resultados experimentais obtidos. Finalmente, a Seção 6 apresenta as conclusões principais e considerações sobre trabalhos futuros.

2. Trabalhos Relacionados

Em redes complexas, uma comunidade pode ser vista como grupos de nós densamente interconectados, mas são esparsamente conectados com o restante da rede [Abrahao et al. 2012, Brandão & Moro 2017a, Newman 2004]. Existem diferentes algoritmos de detecção de comunidades em grafos [Abrahao et al. 2012, Blondel et al. 2008, Clauset et al. 2004, Newman 2004, Newman & Girvan 2004, Raghavan et al. 2007]. Entretanto, tais algoritmos não consideram o aspecto temporal das interações.

Considerar o aspecto temporal no estudo de interações sociais é importante, pois revela propriedades e padrões que não podem ser percebidos sem considerar tal característica. De fato, Holme e Saramäki [2012] mostram como o aspecto temporal pode ser representado em grafos estáticos e demonstram as implicações desse mapea-

¹Não há uma definição universalmente aceita para a estrutura de comunidade [Fortunato 2010, Palla et al. 2005] que provou ser difícil de se definir, quantificar e extrair [Abrahao et al. 2012].

²A modularidade caracteriza a existência de uma estrutura de comunidade em uma rede. Tal métrica leva em consideração a proporção de arestas entre os nós em uma mesma comunidade e em comunidades diferentes [Newman 2006, Newman & Girvan 2004].

mento, enquanto Vaz de Melo et al. [2015] abordam a classificação de relações baseada em propriedades temporais em redes de mobilidade. Outros estudos consideram a agregação temporal para analisar a evolução das redes de colaboração [Alves et al. 2013, Barabási et al. 2002]. Abrahao et al. [2012] e Xie et al. [2013] apresentam uma análise abrangente das propriedades das comunidades obtidas por diversos algoritmos e mencionam a necessidade de melhorias na tarefa de detecção.

Especificamente, este trabalho baseia-se na técnica proposta por Vaz de Melo et al. [2015] que obtém classes de relações sociais em grafos temporais. Com o uso dessa técnica, investigamos os efeitos da presença de interações aleatórias em redes sociais reais de domínios distintos. Para isso, removemos o ruído causado nas redes pela presença de tais relações aleatórias. Esse processo tornou as relações sociais reais disponíveis para aplicação de técnicas existentes de detecção de comunidades³. Após essa etapa de mineração das relações, avaliamos os resultados de diferentes algoritmos de detecção de comunidades. Tal avaliação vai além da simples comparação dos resultados obtidos na detecção de comunidades por algoritmos distintos, mas também mede as melhorias obtidas por cada um deles.

Neste artigo, tal melhoria é medida com a métrica *modularidade*, que é comumente usada no processo de validação de métodos de detecção de comunidades [Fortunato 2010, Newman 2006, Orke et al. 2013, Yang et al. 2016]. A modularidade tem sido utilizada para comparar a qualidade de algoritmos de detecção de comunidades em diferentes redes e em diversas áreas de aplicação [Lambiotte et al. 2008, Newman & Girvan 2004, Radicchi et al. 2004, Raghavan et al. 2007, Sah et al. 2014, Šubelj & Bajec 2011, Wang et al. 2015]. Além disso, muitos algoritmos de detecção de comunidades usam a modularidade em sua lógica como métrica a ser maximizada no processo de detecção [Clauset et al. 2004, Barber & Clark 2009, Liu & Murata 2010, Schuetz & Caffisch 2008].

Dessa forma, nosso trabalho difere dos existentes por não apenas detectar comunidades [Abrahao et al. 2012, Blondel et al. 2008, Clauset et al. 2004, Newman 2006, Yang et al. 2016], mas também por investigar como a mineração das relações influencia na qualidade (medida com a modularidade) do resultado de tais algoritmos. Ademais, apresentamos análises para diferentes tipos de redes sociais.

3. Materiais e Métodos

As redes complexas temporais utilizadas para detecção de comunidades podem ser modeladas por grafos de agregação temporal [Vaz de Melo et al. 2015]. Dado um grafo não direcionado $G = (V, E)$, em que $V = \{v_1, \dots, v_n\}$ é o conjunto de vértices e $E = \{e_1, \dots, e_m\}$ o conjunto de arestas que representam uma interação entre dois vértices, o problema de detecção de comunidades consiste em encontrar o conjunto de comunidades $C = \{c_1, c_2, \dots, c_k\}$ em que cada vértice v_i associa-se a uma comunidade $c_i \in C$. Cada grafo temporal $G_t(V_t, E_t)$ em G representa a agregação das interações em períodos discretos de tempo t . Assim, para um dado valor de t , V_t são todos os vértices que interagiram no t -ésimo período de tempo. As arestas no conjunto E_t representam o emparelhamento das interações entre os pares de vértices (v_i, v_j) durante o período de tempo t .

³São muitas as abordagens utilizadas por diversos algoritmos de detecção de comunidades [Brandes et al. 2008, Sah et al. 2014, Wang & Hopcroft 2010]

Tabela 1. Propriedades das redes estudadas

<i>Rede</i>	<i>Período</i>	<i>V</i>	<i>E</i>	<i>D</i>	<i>CC</i>	$\Delta(G)$	$\alpha(G)$
<i>PubMed</i>	00-16(1)	4×10^5	6×10^6	6×10^{-5}	0.4	3246	25.0
<i>DBLP</i>	00-15(1)	9×10^5	5×10^6	9×10^{-6}	0.2	1413	8.3
<i>Dartmouth</i>	2(2)	1×10^3	3×10^4	4×10^{-2}	0.5	236	46
<i>USC</i>	2(2)	3×10^3	2×10^5	5×10^{-2}	0.5	652	128

V: número de vértices; *E*: número de arestas;

D: densidade; *CC*: coeficiente de clusterização;

$\Delta(G)$: grau máximo; $\alpha(G)$: grau médio;

DBLP e *PubMed*: redes de colaboração científica;

Dartmouth e *USC*: redes de mobilidade em *campi* universitários;

(1): ano inicial - ano final; (2): número de meses.

Esse modelo de grafo temporal foi aplicado a redes sociais reais de colaboração científica derivadas de dados da DBLP e PubMed [Brandão & Moro 2017b]⁴, e de redes *WiFi* de mobilidade de *campi* universitários (Dartmouth e USC) [Vaz de Melo et al. 2015]. A Tabela 1 apresenta as propriedades de cada uma dessas redes, onde, respectivamente, os nós representam pesquisadores ou usuários do sistema e as arestas, coautoria ou pontos de acesso de rede sem fio. Essas redes temporais foram submetidas ao processo de mineração de relações sociais e também foram utilizadas na avaliação desse processo. Em tais redes, as interações são agregadas e associadas à data em que ocorreram.

Utilizando essas redes sociais temporais, aplicamos algoritmos de detecção de comunidades e comparamos a medida de modularidade das comunidades detectadas na rede original com a modularidade das comunidades da rede filtrada. Note que a métrica modularidade foi escolhida por ser mais utilizada na avaliação de algoritmos de detecção de comunidades [Fortunato 2010] e ser objeto de maximização por grande parte de tais algoritmos no estado da arte [Blondel et al. 2008, Clauset et al. 2004, Newman 2004, Raghavan et al. 2007]. Adicionalmente, foi feita uma comparação com a modularidade das comunidades obtidas de amostras da rede original, ou seja, construídas pela remoção aleatória de arestas até atingir o tamanho da rede filtrada. Estas redes amostrais permitiram comparar o processo proposto com um método estocástico de remoção de arestas.

Baseando-se no estado da arte em detecção de comunidades, selecionamos diferentes algoritmos para a etapa de comparação entre a modularidade das comunidades antes e após a remoção do ruído. Esses algoritmos são brevemente descritos a seguir.

Louvain Modularity. É um algoritmo guloso popular para otimização de modularidade. Inclui cada nó da rede em um grupo que resulte na melhoria da medida de modularidade local [Abraham et al. 2012, Blondel et al. 2008, Newman 2004].

⁴PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>): constitui de colaborações científicas da MEDLINE; Conferências da DBLP (<http://dblp.org/>): rede de coautoria científica em ciência da computação, descrita por [Brandão & Moro 2017b]. Compõe-se de colaborações ocorridas no período de 2000 a 2016.

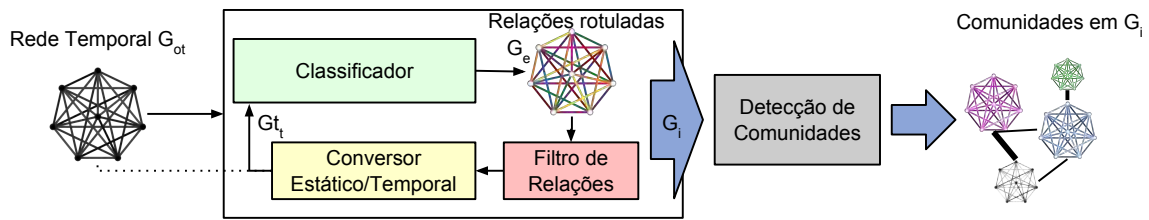


Figura 1. Visão geral do processo de mineração de relações sociais que resulta em um grafo estático filtrado G_i , para a tarefa de detecção de comunidades.

Edge Betweenness. Este algoritmo descobre comunidades em redes pelo particionamento de nós em subgrupos densamente conectados [Newman & Girvan 2004]. Baseia-se na remoção iterativa de arestas da rede a partir de uma medida de intermediação de arestas (*betweenness*) que é recalculada após cada remoção.

Greedy Optimization of Modularity (GOM). Este algoritmo baseia-se na maximização da modularidade usando uma abordagem gulosa [Clauset et al. 2004]. Em uma primeira etapa, ele identifica uma estrutura hierárquica de comunidades. As comunidades são então extraídas do corte dessa estrutura que maximiza globalmente a modularidade.

Label Propagation. Este algoritmo utiliza etiquetas que identificam as comunidades. Os nós são etiquetados e por propagação iterativa cada nó tem sua etiqueta atualizada consensualmente pela que prevalece entre os vizinhos. Ao final, os nós com as mesmas etiquetas representam uma mesma comunidade [Raghavan et al. 2007].

Um cálculo simples foi utilizado para identificar o erro na detecção de comunidades pelos algoritmos selecionados, quando a rede possui interações aleatórias. Para as comunidades obtidas de um mesmo algoritmo, é feita a contagem do número de nós que trocaram de comunidades após a filtragem. Dada uma rede original R_o e uma rede filtrada R_n , onde n é a iteração em que houve convergência no processo de mineração de relações sociais. Cada comunidade cn_i detectada em R_n , é comparada com a comunidade mais semelhante⁵ co_j em R_o . A contagem do número de membros que estão em R_n e não estão em R_o corresponde ao erro local à comunidade cn_i . Do somatório dos erros locais é obtido o erro global da rede R_o . Nesta comparação, também foi calculada a precisão e revocação [Baeza-Yates et al. 1999] das comunidades detectadas para análise de relevância da filtragem.

4. Mineração de Relações

Nesta seção, detalhamos o processo de mineração de relações sociais que permite remover o ruído das redes, causado pela presença de interações aleatórias e, conseqüentemente, diminuir o erro na associação dos nós às comunidades. A Figura 1 detalha as etapas para realização desse processo de mineração que são: (i) classificação das relações; (ii) remoção de arestas aleatórias; e (iii) reconstrução do grafo temporal que é a entrada para a iteração seguinte. Ao final das iterações, quando não existem mais arestas aleatórias,

⁵Dado um grafo A e um grafo B obtido da filtragem de A pelo processo de mineração de relações sociais. A medida de semelhança utilizada entre duas comunidades C_A e C_B baseia-se no número de nós de C_B que estão em C_A .

Algorithm 1 Processo de mineração de relações sociais**Require:** Go_t : Grafo Temporal Original**Input:** Go_t, k

```

1:  $Gt_t \leftarrow Go_t$ 
2: for  $i = 1, i < k$  do
3:    $G_e \leftarrow recast(G1_t)$ 
4:   if  $converge(G_e, Gt_t)$  then
5:     break
6:   end if
7:    $Gt_t \leftarrow remove\_aleatorias(G_e, Gt_t)$ 
8: end for

```

Tabela 2. Percentuais de alteração nas métricas das redes sociais

Rede	I	V	E	C1	C2	C3	C4	D	CC	Δ	α
<i>PubMed</i>	6 ^a	91	68	96	111	91	0	82	173	35	75
<i>DBLP</i>	3 ^a	60	38	215	20	80	0	104	98	14	63
<i>Dartmouth</i>	6 ^a	76	13	21	72	1	0	22	102	15	17
<i>USC</i>	4 ^a	12	1	5	68	1	0	55	92	7	6

 I : iteração em que obteve o percentual em relação à rede original. V : vértices, E : arestas. $C1 - C4$: classes de relações (amizade, ponte, conhecido e aleatório) D : densidade, CC : coeficiente de clusterização. $\Delta(G)$: grau máximo, $\alpha(G)$: grau médio.

obtem-se o grafo estático constituído apenas por relações sociais. Tal grafo estático permite que comunidades mais modulares sejam detectadas por algoritmos desenvolvidos para essa finalidade.

A visão geral do processo de mineração de relações sociais é apresentada no Algoritmo 1, onde os dados de entrada são as interações do grafo temporal original Go_t . Em cada iteração do algoritmo, primeiro são classificadas as arestas (linha 3), resultando em um grafo estático de relações rotuladas G_e . Em seguida, é feita a verificação de convergência e se ainda existirem arestas aleatórias para serem removidas, o procedimento que faz a separação e reconstrução do grafo temporal Gt_t é chamado (linha 7).

Para classificar as arestas a partir de interações temporais, foi utilizado o algoritmo *RECAST* (*Random Relationship Classifier Strategy*) que é uma técnica de classificação de arestas a partir de interações temporais. Esse classificador foi proposto por Vaz de Melo et al. [2015] e é capaz de caracterizar claramente interações aleatórias além de identificar diferentes tipos de interação social (amizade, ponte e conhecido). O *RECAST* baseia-se em duas características que estão sempre presentes nas relações sociais: a regularidade das interações e a similaridade entre os pares, isto é, o número de indivíduos conhecidos em comum. A partir dessas características, o *RECAST* atribui, para pares únicos de vértices (u, v) que se interagem, uma das classes de relações.

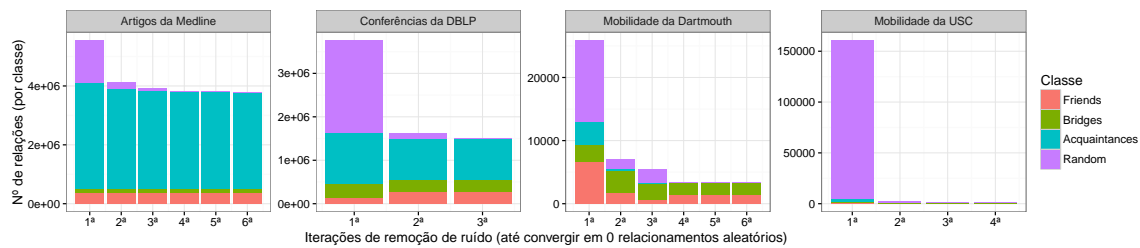


Figura 2. Classes de relações ao final de cada iteração de confirmação de convergência do processo de mineração de relações sociais.

Tabela 3. Percentuais obtidos com a remoção do ruído - Algoritmo GOM

Rede	Remoções		Correções/Aprimoramentos		Ganho	
	Ruído	Erro	Precisão	Revocação	$F1$	Modularidade
<i>PubMed</i> ₁₋₆	32	54	55	43	48	+26
<i>DBLP</i> ₁₋₃	62	41	49	30	37	+32
<i>Dartmouth</i> ₁₋₆	87	9	10	3	4	+38
<i>USC</i> ₁₋₄	99	2	13	3	5	+38

Ruído: removido, representado por interações aleatórias (percentual de arestas)

Erro: na associação de nós em comunidades (percentual de reposicionamentos).

5. Resultados Experimentais

As comunidades obtidas por cada combinação de algoritmo e rede selecionados foram caracterizadas pela sua modularidade dentre outras métricas de análise de redes sociais. Essa análise também foi feita sobre a versão filtrada da rede, isto é, após a remoção completa das arestas aleatórias. Conforme mostra a Tabela 2, ocorreram diferentes percentuais de alteração entre as redes selecionadas. Analisando separadamente cada rede, é possível distingui-las pela aleatoriedade total das relações.

As redes de mobilidade são as que possuem maior proporção de relações aleatórias. Além disso, são da classe aleatória a maioria das suas relações (coluna E da Tabela 3). Nessas redes, prevalece também o número de nós em que todas as suas relações são aleatórias. Como consequência disso, esses nós são desconectados da rede por não participarem socialmente de uma comunidade. Assim, ao comparar as comunidades obtidas das redes filtradas com as obtidas das redes originais, o erro está presente em pequena quantidade, conforme mostra Tabela 3.

Também na Tabela 3, verifica-se que a relevância das comunidades filtradas é maior para as redes de colaboração científica. A métrica proposta para cálculo de erro parte do pressuposto que cada comunidades da rede filtrada é um subconjunto de uma comunidade nas redes originais e que estas podem ser subdivididas. Em geral, as redes de mobilidade e de colaboração científica são distintas em quantidade de ruído e no impacto que este ruído provoca na formação das comunidades. Assim, as redes de colaboração científica possuem menor proporção de ruído contudo, este ruído provoca uma proporção maior de associações incorretas dos membros (coautores) às suas respectivas comunidades. De forma oposta, as redes de mobilidade possuem maior aleatoriedade nas relações e menor erro nas comunidades detectadas.

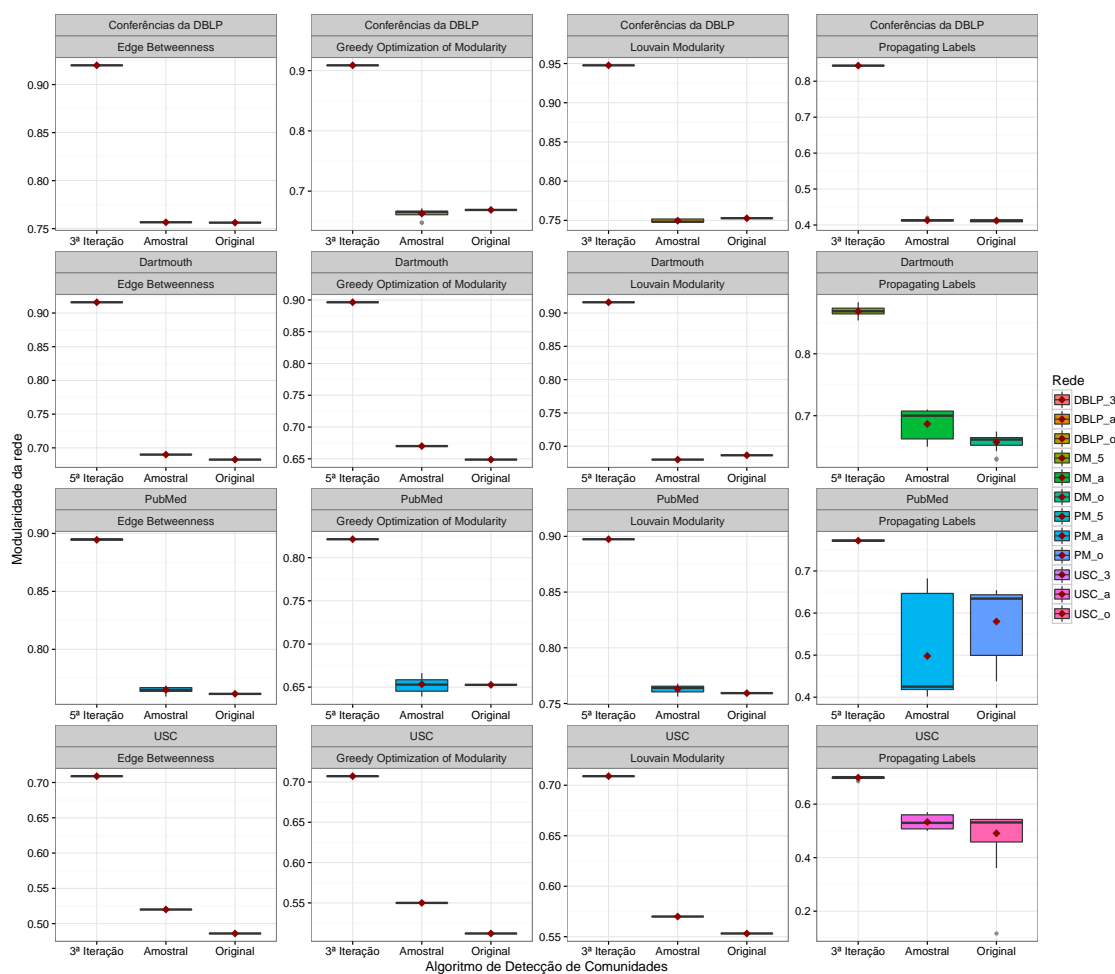


Figura 3. Comparação entre a modularidade das comunidades na rede original (o), na sua amostra (a) e na rede filtrada até a n -ésima iteração do processo de mineração de relações sociais.

Com base nas redes sociais estudadas, foi possível verificar o comportamento do algoritmo *RECAST* em redes reais. Conforme mostra a Figura 2, a maioria das relações aleatórias foram removidas na primeira iteração do processo de mineração de relações sociais. Ademais, houve um ganho em modularidade nas comunidades detectadas a cada iteração. O grau máximo das redes sociais também foi afetado na proporção inversa à quantidade de ruídos removidos. Além disso, o grau dos *hubs* foi reduzido em proporção maior do que a dos demais nós da rede.

Após a remoção das arestas aleatórias, houve um aumento significativo na modularidade das comunidades obtidas pelos algoritmos selecionados em cada uma das quatro redes estudadas, conforme mostra a Figura 3. O ganho médio em modularidade das comunidades detectadas pelo algoritmo GOM - *Greedy Optimization of Modularity*⁶ é apresentado na Tabela 3. Apesar de tal algoritmo obter um dos menores ganhos em modularidade, há uma correlação positiva desse ganho com a proporção de ruído removido. Essa correlação também ocorre nas demais redes, conforme médias de modularidade apresentadas na Figura 3 e da proporção de arestas removidas apresentada na Figura 2.

⁶GOM foi exemplificado por ser determinístico, simplificar a apresentação e por obter menor ganho.

6. Conclusões

A principal contribuição deste trabalho é representada por um processo de mineração de relações sociais que remove ruído em grafos temporais. Esse processo baseia-se na classificação de arestas aleatórias e na construção de um grafo estático composto apenas por interações sociais. Especificamente, desenvolvemos um arcabouço que utiliza o algoritmo *RECAST* para classificação das relações. Esse processo foi aplicado e avaliado sobre redes temporais reais de colaboração científica e de mobilidade. Analisamos diferentes métricas de redes sociais sobre cada rede social e para as redes resultantes do processo de mineração.

A aplicação do processo convergiu na remoção total do ruído das redes e foi comprovada melhoria na detecção de comunidades por algoritmos do estado da arte. Além disso, demonstramos que a quantidade de ruído presente nas redes sociais afeta negativamente a tarefa de detecção de comunidades, pois conclui em erros na estrutura das comunidades detectadas.

Como trabalhos futuros, pretendemos detalhar a interpretação das alterações topológicas nas redes sociais com base no respectivo domínio de aplicação. Por exemplo, nós que tiveram suas relações classificadas como aleatórias poderão ser analisados qualitativamente a fim de explicar as propriedades que estão associadas à aleatoriedade e às demais classes de relações.

Agradecimentos. Este trabalho é apoiado pelos projetos InWeb (processo MCT/CNPq 573871/2008-6) e MASWeb (processo FAPEMIG/PRONEX APQ-01400-14)), e por auxílios individuais de pesquisa concedidos aos autores pelo CNPq, Fapemig e CAPES. Particularmente, o primeiro autor agradece ao Instituto Federal do Norte de Minas Gerais - IFNMG, pela concessão de uma bolsa de estudos no âmbito do Programa de Bolsas para Qualificação de Servidores (PBQS).

Referências

- [Abrahao et al. 2012] Abrahao, B., Soundarajan, S., Hopcroft, J., and Kleinberg, R. (2012). On the separability of structural classes of communities. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD12, pages 624–632, New York, NY, USA. ACM.
- [Alves et al. 2013] Alves, B. L., Benevenuto, F., and Laender, A. H. (2013). The role of research leaders on the evolution of scientific communities. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13 Companion*, pages 649–656, New York, NY, USA. ACM.
- [Baeza-Yates et al. 1999] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463.
- [Barabâsi et al. 2002] Barabâsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3):590–614.
- [Barber & Clark 2009] Barber, M. J. and Clark, J. W. (2009). Detecting network communities by propagating labels under constraints. *Phys. Rev. E*, 80:026129.
- [Blondel et al. 2008] 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.
- [Brandão & Moro 2017a] Brandão, M. A. and Moro, M. M. (2017a). Social professional networks: A survey and taxonomy. *Computer Communications*, 100:20 – 31.
- [Brandão & Moro 2017b] Brandão, M. A. and Moro, M. M. (2017b). The strength of co-authorship ties through different topological properties. *Journal of the Brazilian Computer Society*, 23(1):5.
- [Brandes et al. 2008] Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., and Wagner, D. (2008). On modularity clustering. *IEEE Trans. on Knowl. and Data Eng.*, 20(2):172–188.
- [Clauset et al. 2004] Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 70:066111.
- [Fortunato 2010] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5):75–174.
- [Holme & Saramäki 2012] Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.
- [Lambiotte et al. 2008] Lambiotte, R., Delvenne, J.-C., and Barahona, M. (2008). Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*.
- [Liu & Murata 2010] Liu, X. and Murata, T. (2010). Advanced modularity-specialized label propagation algorithm for detecting communities in networks. *Physica A: Statistical Mechanics and its Applications*, 389(7):1493–1500.
- [Newman 2006] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.
- [Newman 2004] Newman, M. E. J. (2004). Detecting community structure in networks. *The European Physical Journal B*, 38(2):321–330.
- [Newman & Girvan 2004] Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):26113.
- [Orke et al. 2013] Orke, R. G., Maillard, P., Schumm, A., Staudt, C., Wagner, D., Görke, R., Maillard, P., Schumm, A., Staudt, C., and Wagner, D. (2013). Dynamic graph clustering combining modularity and smoothness. *Journal of Experimental Algorithmics (JEA)*, 18(1):1–5.
- [Palla et al. 2005] Palla, G., Derenyi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.
- [Radicchi et al. 2004] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D., and Fisica, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663.
- [Raghavan et al. 2007] Raghavan, U. N., Albert, R., and Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):1–12.
- [Sah et al. 2014] Sah, P., Singh, L. O., Clauset, A., and Bansal, S. (2014). Exploring community structure in biological networks with random graphs. *BMC Bioinformatics*, 15(1):220.
- [Schuetz & Cafilisch 2008] Schuetz, P. and Cafilisch, A. (2008). Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement. *Physical Review E*, 77(4):046112.

- [Šubelj & Bajec 2011] Šubelj, L. and Bajec, M. (2011). Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction. *Physical Review E*, 83(3):036103.
- [Vaz de Melo et al. 2015] Vaz de Melo, P. O. S., Viana, A. C., Fiore, M., Jaffrès-Runser, K., Mouël, F. L., Loureiro, A. A. F., Addepalli, L., and Guangshuo, C. (2015). RECAST: Telling Apart Social and Random Relationships in Dynamic Networks. *Performance Evaluation*, 87:19–36. "Special Issue: Recent Advances in Modeling and Performance Evaluation in Wireless and Mobile Systems".
- [Wang & Hopcroft 2010] Wang, L. and Hopcroft, J. (2010). Community structure in large complex networks. In *International Conference on Theory and Applications of Models of Computation*, pages 455–466. Springer.
- [Wang et al. 2015] Wang, M., Wang, C., Yu, J. X., and Zhang, J. (2015). Community detection in social networks: an in-depth benchmarking study with a procedure-oriented framework. *Proceedings of the VLDB Endowment*, 8(10):998–1009.
- [Xie et al. 2013] Xie, J., Kelley, S., and Szymanski, B. K. (2013). Overlapping Community Detection in Networks : The State-of-the-Art and Comparative Study. *ACM Computing Surveys (csur)*, 45(4):43.
- [Yang et al. 2016] Yang, Z., Algesheimer, R., and Tessone, C. J. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Nature Publishing Group*, (August):1–16.