Investigação Sobre Robustez de Comunidades em Redes

Vinícius da F. Vieira*†, Vitor E. do Carmo†, Carolina R. Xavier*†, Alexandre G. Evsukoff* and Nelson F. F. Ebecken*
*UFSJ - Universidade Federal de São João del Rei
São João del Rei, Minas Gerais, Brasil
†COPPE/UFRJ - Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brasil

***Resumo*—Em sistemas complexos modelados como redes, onde os nós representam os indivíduos e as arestas representam os relacionamentos entre os indivíduos, uma das principais questões a serem exploradas é a robustez, ou seja, a capacidade de uma rede suportar ataques a seus elementos. Além disso, o estudo da estrutura de comunidades de redes é bastante importante para a compreensão da topologia da rede em uma organização local. Neste trabalho, estuda-se o efeito da estrutura de comunidades na robustez de redes reais. Para isso, é feita uma investigação comparativa entre o comportamento exibido por redes e suas comunidades correspondentes submetidas a ataques. Os resultados mostram uma forte analogia entre o comportamento exibido nas redes completas e suas comunidades isoladas.**

***Keywords*—Redes, robustez, comunidades, modularidade.**

I. Introdução

Muitos sistemas reais podem ser modelados como redes, onde cada elemento pode ser modelado como nó e as interações entre esses elementos como arestas. O estudo de redes é de fundamental importância em muitas áreas, como biologia, onde pode-se compreender a interação entre proteínas [1] e sociologia, onde pode-se representar a forma como acontecem os relacionamentos entre as pessoas [2].

Através do estudo de propriedades e comportamentos de redes reais, pode-se ter uma visão bastante clara sobre o sistema complexo que está sendo modelado. Uma importante característica de redes reais é a sua capacidade de suportar falhas, sejam elas aleatórias ou originadas a partir de algum ataque deliberado. A análise da robustez da rede frente a ataques permite que se compreenda até que ponto um sistema modelado como rede permanece em funcionamento, mesmo que alguns dos seus elementos falhem ou sejam eliminados. Pode-se identificar características que levam algum nó a, ao ser removido, ter maior probabilidade de causar um grande dano à rede. Assim, é possível discutir estratégias para proteger esses nós e elaborar redes que suportem um maior número de falhas. Estudos desse tipo podem ser cruciais, por exemplo, em sistemas reais como cadeias alimentares, onde a compreensão do papel de cada espécie pode permitir a tomada de decisões mais eficientes em relação a ecossistemas de uma determinada região.

Outra propriedade muito importante em redes é a organização de seus nós em comunidades. Entende-se que a estrutura de comunidades tem grande importância na funcionalidade das redes complexas e seu estudo tem sido objeto de vários trabalhos. Uma noção consensual sobre a caracterização de uma comunidade em uma rede é um subconjunto de nós que apresenta uma grande densidade de conexões internas e uma baixa densidade de conexões externas. Diversas medidas podem ser encontradas na literatura com o objetivo de quantificar a qualidade de uma comunidade em uma rede. Entre elas, destaca-se a modularidade, proposta por Newman e Girvan [3] e que, atualmente, é a medida mais amplamente aceita para avaliação da qualidade de comunidades em redes. Para uma determinada comunidade, a modularidade pode ser entendida, de maneira geral, como a diferença entre a fração de arestas contidas na comunidade e o que se espera de uma versão aleatória da comunidade (preservando a distribuição de graus dos nós). A importância da detecção de comunidades pode ser destacada em diversos problemas reais. Em redes de citação de artigos científicos, por exemplo, onde artigos são representados por vértices e as citações entre esses artigos são representadas por arestas, as estruturas de comunidades podem auxiliar a identificação de grupos de assuntos afins.

Neste trabalho, pretende-se investigar a robustez de redes reais através da investigação do comportamento frente a ataques deliberados considerando diferentes critérios de escolha de nós. Para isso, será feita uma comparação da robustez de redes reais com a robustez observada em suas comunidades isoladamente. O objetivo principal do trabalho é investigar se uma comunidade isolada de uma rede pode ser usada para que se tenha uma visão da robustez da rede como um todo. Para isso, duas questões básicas são levantadas: 1) Quando um ataque é realizado, o comportamento em uma comunidade pode ser comparado ao comportamento exibido na rede? 2) Em caso positivo, em qual comunidade da rede o comportamento é análogo ao observado na rede como um todo?

O restante do trabalho está organizado da seguinte maneira. Na Seção II são apresentados alguns conceitos importantes para a compreensão da metodologia e dos experimentos realizados neste trabalho. A Seção III apresenta os experimentos realizados para a análise de robustez nas redes e nas comunidades. Os resultados são apresentados na Seção IV, onde é também realizada uma discussão sobre os mesmos. Uma conclusão sobre o trabalho é apresentada na Seção V.

II. Conceitos e Trabalhos Relacionados

Esta seção apresenta brevemente alguns conceitos e fundamentos teóricos necessários para uma boa compreensão do trabalho. Os conceitos principais estão organizados nas subseções

a seguir. A ideia de centralidade de nós é descrita, permitindo o entendimento dos critérios para escolha de nós a serem utilizados nos ataques realizados. Uma noção de robustez, central no presente trabalho, é apresentada em seguida. É também introduzida a ideia de comunidades, importante para a compreensão dos cenários a serem experimentados no trabalho.

Primeiramente, define-se um grafo como uma estrutura composta por um par de conjuntos que pode ser representado por $\mathbb{G} = (\mathbb{V}, \mathbb{E})$. Diz-se que $\mathbb{V} = \{v_i, i = 1...n\}$ é um conjunto discreto cujos elementos são chamados de vértices (ou nós ou pontos), e cuja cardinalidade $n = |\mathbb{V}|$ representa sua ordem. Diz-se também que $\mathbb{E}$ é um conjunto de elementos definidos em função dos elementos de $\mathbb{V}$ e representam relações de adjacência. Os elementos de $\mathbb{E}$ são, de forma geral, chamados de ligações ou arestas do grafo. O valor $m = |\mathbb{E}|$ é o tamanho do grafo. Diz-se que $\mathbf{A}$ é a matriz de adjacência de um grafo e, neste trabalho, considera-se grafos sem peso e não direcionados. Assim, $\mathbf{A}_{ij} = 1$ se os vértices v_i e v_j estão ligados e $\mathbf{A}_{ij} = 0$, caso contrário.

A. Centralidade

Em um grupo de indivíduos em qualquer contexto, o relacionamento entre as entidades evidencia características especiais de alguns indivíduos que podem destacá-los do resto do grupo. A noção de centralidade de nós pode estar relacionada a diversas características, que dão origem a diferentes medidas [4]. Quando uma lista é definida considerando os nós mais importantes organizados em ordem crescente, tem-se um *rank*. A construção desse *rank* é bastante importante no escopo deste trabalho, pois define a ordem na qual os nós serão removidos na análise de robustez.

A forma mais simples de avaliar a importância de um nó em uma rede é através do seu grau, ou seja, o número de arestas adjacentes a ele. Em casos onde a rede não tem peso nem direção, o grau $\mathbf{k}_i$ de um vértice v_i pode ser definido como

$$\mathbf{k}_i = \sum_{j=1}^{n} \mathbf{A}_{ij}. \tag{1}$$

Uma extensão da centralidade por grau, chamada de centralidade de autovetor, considera não apenas o número de vizinhos de um vértice, mas também a importância de cada um deles. A centralidade por autovetor $\mathbf{c}_i^{(autovetor)}$ de um vértice v_i pode ser definida como

$$\mathbf{c}_i^{(autovetor)} = \sum_{j} \mathbf{A}_{ij} \mathbf{c}_j^{(autovetor)}. \tag{2}$$

A centralidade de um nó pode também ser medida levando-se em consideração a proximidade na qual ele está de outros nós. Essa é a ideia da centralidade de *closeness*, que leva em conta a distância geodésica média entre um nó e todos os outros por ele alcançáveis, ou seja a média da distância entre um vértice e todos os outros através do caminho mais curto. Assim, se d_{ij} é o comprimento de um caminho geodésico do vértice v_i a um vértice v_j, a centralidade de *closeness* $\mathbf{c}_\mathbf{i}^{(closeness)}$ de um vértice $\mathbf{v}_i$ pode ser definida como

$$\mathbf{c}_\mathbf{i}^{(closeness)} = \frac{1}{n} \sum_{j=1}^{n} d_{ij}. \tag{3}$$

Uma outra ideia baseada em distâncias entre os nós é a centralidade de *betweenness*, que considera a fração de caminhos mínimos que passam por um determinado nó como uma medida de sua importância. Tomando $g_i^{(st)}$ como o número de caminhos geodésicos que partem de um vértice inicial v_s para um vértice final v_t e passam pelo vértice v_i e considerando também que n_{st} é o número total de caminhos geodésicos entre v_s e v_t, a medida de centralidade de *betweenness* $\mathbf{c}_i^{(betweenness)}$ de v_i pode ser calculada como

$$\mathbf{c}_i^{(betweenness)} = \sum_{st} \frac{g_i^{(st)}}{n_{st}}. \tag{4}$$

Diversos trabalhos podem ser encontrados na literatura com discussões mais aprofundadas a respeito da definição e a aplicação de centralidades de nós em redes para a geração de *rankings*, como os trabalhos de Freeman [5] e Opsahl [4].

B. Robustez

Uma questão fundamental no estudo de sistemas é a robustez a falhas em uma ou mais de suas partes. Um aspecto importante disso é a compreensão do efeito da falha dos componentes individuais sobre o desempenho de todo o sistema. Quando leva-se em consideração sistemas modelados como redes complexas, as falhas estão relacionadas ao não funcionamento de vértices e arestas na rede. Por exemplo, é importante entender como a falha de roteadores individuais na Internet afeta a função global da rede. É tambem crucial compreender como ocorrem falhas em redes de contatos para, por exemplo, tentar comprender como uma vacinação eficiente pode conter a propagação de uma determinada doença.

A robustez pode ser investigada através da análise de como a estrutura de uma determinada rede é afetada à medida que seus elementos (vértives ou arestas) são removidos. Essa remoção pode ser feita de maneira aleatória, ou seja, a probabilidade de escolha de todos os indivíduos é igual no caso de uma remoção, ou de maneira determinística, de acordo com algum critério. Na literatura, é bastante comum a investigação de robustez de redes a ataques em vértice e, no caso de ataques determinísticos automáticos, como é feito neste trabalho, é bastante frequente a utilização de *rankings* baseados em medidas de centralidade como critério de escolha de nós a serem removidos. Assim, para a realização de ataques, determina-se a importância dos vértices na rede através do cálculo de alguma medida de centralidade e então calcula-se o efeito sobre o tamanho da maior componente conexa da rede ao se remover uma dada fração dos vértices, escolhidos em ordem decrescente à determinada medida de centralidade. A partir das remoções, o comportamento da rede pode ser analisado e diversas propriedades podem ser utilizadas nessa etapa, como o tamanho da componente gigante, a distância média entre os nós da rede e o coeficiente de *clustering*. Albert *et al.* apresentam um estudo [6] no qual são realizados diversos ataques aleatórios e determinísticos em redes reais. Os autores constatam que redes reais são muito robustas a ataques aleatórios mas, por outro lado, são surpreendentemente vulneráveis a ataques determinísticos, mais especificamente, considerando a remoção de nós de maior grau.

Duas estratégias distintas podem ser adotadas para a remoção de vértices em uma rede. Em uma delas, a medida

de centralidade é calculada para todos os vértices da rede e, em seguida, uma fração específica dos vértices é removida em ordem decrescente pela medida centralidade. Este procedimento é chamado por Iyer *et al.* de ataque direcionado simultâneo [7]. Em outra estratégia, a medida centralidade é calculada para todos os vértices da rede inicial e o vértice com maior centralidade é removido. A remoção deste vértice resulta em uma nova rede em que as medidas de centralidade dos vértices restantes podem ser diferentes dos valores que foram calculados para eles anteriormente. Então recalcula-se as medidas de centralidade de todos os vértices na nova rede e, novamente, retira-se o mais bem classificado. Este processo de recálculo das medidas de centralidade e a remoção do vértice melhor classificado é continuado até que a fracção desejada de vértices seja removida. Este procedimento é definido por Iyer *et al.* como ataque direcionado sequencial [7]. Holme *et al.* apresentam um estudo [8] no qual redes reais e artificiais, geradas a partir de diferentes modelos, são submetidas a ataques considerando as centralidades de grau e de *betweenness*. No estudo de Holme *et al.*, as duas estratégias para cálculo das centralidades (simultânea e sequencial) são comparadas. Os autores verificam que a estrutura da rede muda muito durante os ataques e que, por isso, ataques sequenciais são muito mais prejudiciais à rede do que ataques simultâneos. Iyer *et al.* [7] fazem um estudo bastante extenso no qual comparam a robustez de redes para ataques baseados em ainda mais medidas de centralidade, como autovetor e *closeness*, considerando as duas estratégias para cálculo de centralidade.

C. Comunidades

Comunidades em redes são grupos de vértices com uma alta densidade de ligações internas e uma baixa densidade de ligações externas. A organização de redes em comunidades pode ser observada em diversos contextos e pode ser utilizada como base para a compreensão dos vértices como unidades que apresentam uma mesma função. Em um sistema biológico, por exemplo, pode-se modelar proteínas como vértices de uma rede e as interações entre as proteínas como arestas. Nesse caso, pode-se identificar comunidades de vértices e relacioná-los a determinadas patologias. Em uma rede de pessoas, modeladas como vértices, as arestas podem indicar algum tipo de amizade ou afinidade de perfis de comportamento e a análise de comunidades pode auxiliar a identificação de padrões de comportamento. Neste trabalho, as comunidades têm um papel fundamental, já que pretende-se investigar comparativamente o comportamento de redes e comunidades frente a ataques.

Diversas medidas utilizadas para avaliar a qualidade de uma divisão de rede em comunidades podem ser encontradas na literatura. Atualmente, a medida mais amplamente adotada para avaliação da qualidade de partições é a modularidade que, por esse motivo, foi utilizada como base para a metodologia proposta neste trabalho. A modularidade foi proposta por Newman [3] e tem como ideia principal a definição de um determinado subconjunto de vértices como comunidade caso o número de conexões internas entre seus vértices seja maior que o número de conexões esperadas entre esses vértices no caso de uma formação aleatória na estrutura. Assim, se confirmada essa suposição, entende-se que há uma estrutura organizada, uma comunidade. A modularidade Q de um particionamento de uma rede é calculada por

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{\mathbf{k}_i \mathbf{k}_j}{2m} \right) \delta(c_i, c_j), \tag{5}$$

onde c_i e c_j são, respectivamente, as comunidades dos vértices v_i e v_j e $\delta(.,.)$ é o delta de Kronecker, uma função que retorna 1 caso os operandos sejam iguais e 0 caso contrário.

Vários trabalhos podem ser encontrados na literatura com o objetivo de propor e estudar métodos para a identificação automática de comunidades em redes [3], [9], [10]. Este trabalho é baseado no método espectral de Newman [3], que funciona através do cálculo do autovetor dominante de uma matriz obtida a partir da matriz de adjacência. Uma discussão aprofundada sobre o método de Newman não faz parte do escopo desse trabalho e pode ser encontrada em [3].

III. Experimentos

Para a realização dos experimentos definiu-se, inicialmente, um conjunto de redes reais, abrangendo uma ampla gama de aplicações e frequentemente utilizadas como *benchmark* em estudos de redes. A Tabela I apresenta um resumo das principais características das redes.

TABELA I. Principais características das redes: número de nós (n); número de arestas (m); grau médio ($\hat{k}$); número de comunidades encontradas (nc).

Rede	n	m	$\hat{k}$	Q	nc	Tipo
Adjnoun[1]	112	425	7.59	0.29	7	Rede de palavras.
C.Elegans[2]	453	2040	9.01	0.42	7	Rede metabólica.
Cit-HepPh[3]	34546	420921	24.37	0.71	34	Rede de citação.
Cond-Mat[1]	30460	120029	7.88	0.72	117	Rede de coautoria.
Email[2]	1133	5451	9.62	0.55	10	Rede de comunicação.
Netscience[2]	1461	2742	3.75	0.94	79	Rede de coautoria.

Cada rede foi dividida em comunidades através do método espectral de Newman, seguindo a abordagem descrita por Vieira *et al.* [11]. Foi utilizada uma implementação do método de Newman, utilizando a linguagem C, livremente disponível na *web* [1], que utiliza procedimentos e estruturas de dados eficientes de forma a permitir a divisão de grandes redes em comunidades a um custo computacional baixo. A metodologia para divisão da rede em comunidades proposta por Vieira *et al.* [11] e adotada neste trabalho utiliza uma variação na etapa que Newman define como ajuste fino do método espectral [3] que permite que o método seja executado de maneira rápida em computadores pessoais sem grande redução na modularidade, ou seja, sem grandes prejuízos à qualidade da partição. A modularidade calculada para cada uma das partições geradas (Q), assim como o número de comunidades (nc) identificadas pelo método utilizado neste trabalho são também apresentados na Tabela I.

Foi também realizado o cálculo das centralidades dos vértices das redes descritas na Tabela I considerando cada uma das medidas descritas na Seção II-A: grau, autovetor,

[1]Baixado de: http://www-personal.umich.edu/\~mejn/netdata/

[2]Baixado de: http://deim.urv.cat/\~aarenas/data/welcome.htm

[3]Baixado de: http://snap.stanford.edu/data/

[1]http://www.github.com/vfvieira/

betweenness e *closeness*. O mesmo foi feito em cada uma das comunidades encontradas pelo método de Newman para cada uma das redes. O cálculo de cada centralidade permitiu a definição de um *ranking*, utilizado como base para a remoção dos vértices durante os ataques.

Os ataques foram feitos em passos de 5% e apenas sobre as componentes gigantes de cada rede. Para comunidades muito pequenas, com menos de 20 vértices, a remoção foi feita de um em um vértice. É importante mencionar que o cálculo das medidas de centralidade, assim como os ataques, foram realizados com o auxílio da biblioteca de *software* igraph, utilizando a linguagem Python. Em cada cenário estudado, foram avaliados a porcentagem de nós removidos e a porcentagem de nós restantes na componente gigante, o que foi usado como base para a análise dos resultados.

Neste trabalho, o comportamento das medidas de centralidade nas redes e suas respectivas comunidades foi verificado a partir de ataques direcionados simultâneos, devido à sua simplicidade e sua ampla adoção por outros trabalhos na literatura. Mesmo apresentando alguns pontos negativos, como o fato de não considerar as alterações realizadas na rede durante o ataque, pode-se argumentar que a utilização de ataques direcionados simultâneos não prejudica a análise dos resultados neste trabalho, uma vez que o objetivo principal é comparar o comportamento das redes frente a ataques considerando diversos critérios e não testar a vulnerabilidade das redes.

IV. Resultados e Discussão

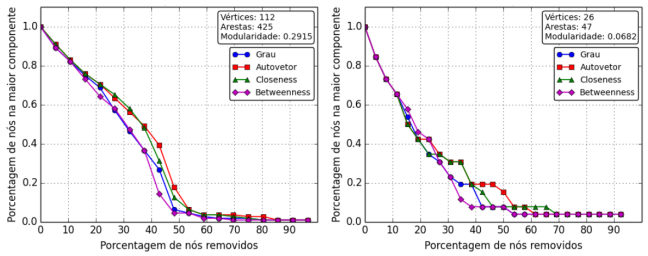
Esta seção apresenta os resultados obtidos pela aplicação da metodologia descrita na Seção III ao conjunto de seis redes reais apresentadas na Tabela I. A discussão é dividida em duas partes, considerando a comparação da robustez na rede completa e nas comunidades isoladas, após a aplicação do método de Newman, e considerando a comparação do comportamento exibido por cada comunidade encontrada.

A. Comparação da Robustez nas Redes Completas e nas Comunidades Isoladas

Foram realizados ataques direcionados simultâneos nas redes descritas na Tabela I e em suas comunidades. Os ataques foram feitos levando em consideração diferentes critérios de centralidade dos nós, como descrito na Seção III. É muito importante ressaltar que o estudo apresentado nesta seção envolve apenas as comunidades de maior modularidade de cada rede.

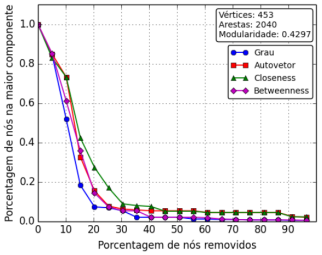
A Figura 1 apresenta a variação do tamanho da componente gigante após a realização de ataques nas redes completas (Figuras 1(a), (c), (e), (g), (i) e (k)) e em suas comunidades de maior modularidade isoladas (Figuras 1(b), (d), (f), (h), (j) e (l)). Em cada uma das figuras relacionadas à rede completa, estão indicados o número de nós, o número de arestas e a modularidade obtida pela partição realizada pelo método de Newman. Nas figuras relacionadas às comunidades, estão indicados o número de nós e arestas de cada grupo identificado, assim como a modularidade medida para a comunidade isolada.

A observação da Figura 1 permite que sejam analisados, inicialmente, alguns aspectos relacionados à robustez das redes

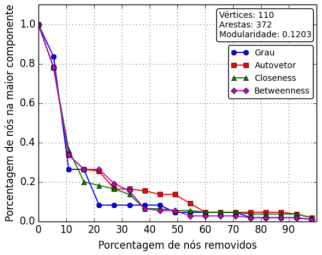


(a) Adjnoun (rede completa)

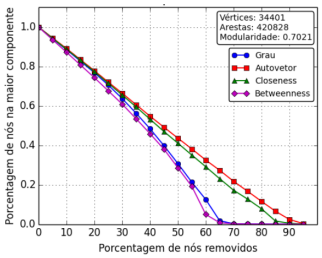
(b) Adjnoun (comunidade isolada)



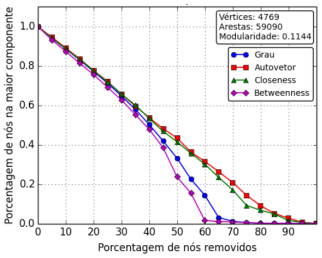
(c) C.Elegans (rede completa)



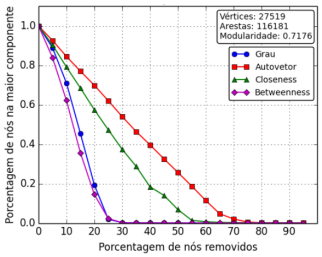
(d) C.Elegans (comunidade isolada)



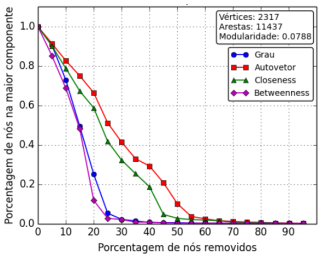
(e) Cit-HepPH (rede completa)



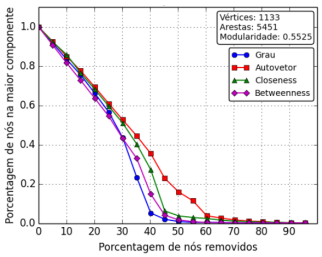
(f) Cit-HepPH (comunidade isolada)



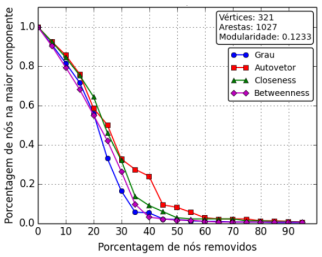
(g) Cond-Mat (rede completa)



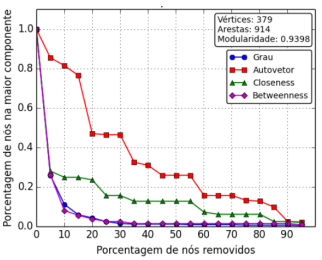
(h) Cond-Mat (comunidade isolada)



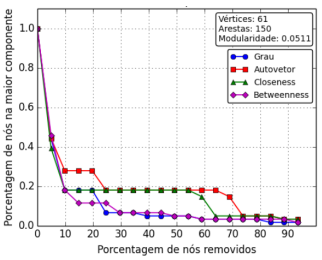
(i) Email (rede completa)



(j) Email (comunidade isolada)



(k) Netscience (rede completa)



(l) Netscience (comunidade isolada)

Fig. 1. Robustez frente a ataques considerando diferentes medidas de centralidade: (a)(c)(e)(g)(i)(k) redes completas; (b)(d)(f)(h)(j)(l) comunidades isoladas.

ao serem atacadas considerando comparativamente os diferentes critérios adotados. Pode-se perceber que as diferentes redes se comportam de diferentes maneiras quando submetidas a ataques direcionados. Em algumas das redes estudadas, a resposta das redes aos ataques realizados considerando todos os critérios de escolhas de nós foi bastante semelhante, como é o caso das redes Adjnoun, C. Elegans e Email (Figuras 1(a), 1(c) e 1(i), respectivamente). Em outro caso, o decaimento do tamanho da componente gigante apresenta semelhanças em alguns intervalos de porcentagem de nós removidos. É o que acontece com a rede Cit-HePH (Figura 1(e)), na qual o comportamento apresentado para as diferentes centralidades é bastante semelhante até a remoção de cerca de 40% dos nós. Na rede Netscience e Cond-Mat (Figuras 1(k) e 1(g)), o comportamento observado quando utiliza-se a centralidade de autovetor é bastante diferente do comportamento observado para as outras centralidades.

Ainda considerando a robustez das redes, de maneira geral as centralidades de grau e *betweenness* apresentaram maior eficiência quando utilizadas como critérios para remoção de nós e, em alguns casos, foram capazes de reduzir o tamanho da componente conexa das redes drasticamente, bastando remover cerca de 20% dos nós para que a rede fosse praticamente destruída. Por outro lado, um comportamento padrão que pôde ser observado é a baixa eficiência da centralidade de autovetor como critério de escolha de nós a serem removidos.

Considerando o escopo deste trabalho, algumas observações importantes podem ainda ser feitas quando compara-se o decaimento do tamanho da componente gigante na rede completa e nas comunidades isoladas. Na maior parte dos casos, observa-se uma grande semelhança nos comportamentos exibidos pelas redes e pelas comunidades ao serem submetidos a ataques por diferentes critérios de centralidade. Na rede Cit-HepPh, observa-se um comportamento quase linear no decaimento do tamanho da componente gigante considerando os diferentes critérios de escolha de nós e esse comportamento é refletido na comunidade de maior modularidade. A melhor eficiência da centralidade de *betweenness* na rede Cit-HepPh é também observada na comunidade atacada, como pode-se observar na comparação das Figura 1(e) e 1(f). Uma grande semelhança entre o comportamento exibido frente aos ataques realizados nas redes completas e nas comunidades isoladas é também observada nas bases Cond-Mat e Email.

Essa grande semelhança nos padrões encontrados para os ataques realizados nas redes e nas comunidades indica que as comunidades selecionadas (especificamente, as comunidades de maior modularidade) representam com bastante eficiência a estrutura das redes completas. Assim, uma análise de robustez pode ser feita apenas localmente e, ainda assim, fornecer uma visão bastante rica sobre o comportamento que seria exibido quando realiza-se uma análise na rede completa, trazendo uma grande simplificação ao processo.

Em algumas das redes estudadas (Netscience e Adjnoun), o comportamento do decaimento não apresenta grande semelhança em todos os critérios de escolha de nós mas, mesmo nesses casos, é possível observar vários padrões. Na rede Netscience (Figura 1(k)), a centralidade de autovetor é bem mais eficiente na comunidade isolada do que na rede completa, o que também acontece na rede Adjnoun (Figura 1(a)), embora de maneira menos acentuada. Entretanto, em todos os casos estudados, mesmo naqueles em que há distinção entre os valores observados nos ataques às redes e às comunidades, observa-se uma forte analogia entre o comportamento exibido pelas redes e pelas suas comunidades ao terem nós atacados, reforçando a noção de que comunidades podem ser utilizadas como base para a análise de robustez de redes.

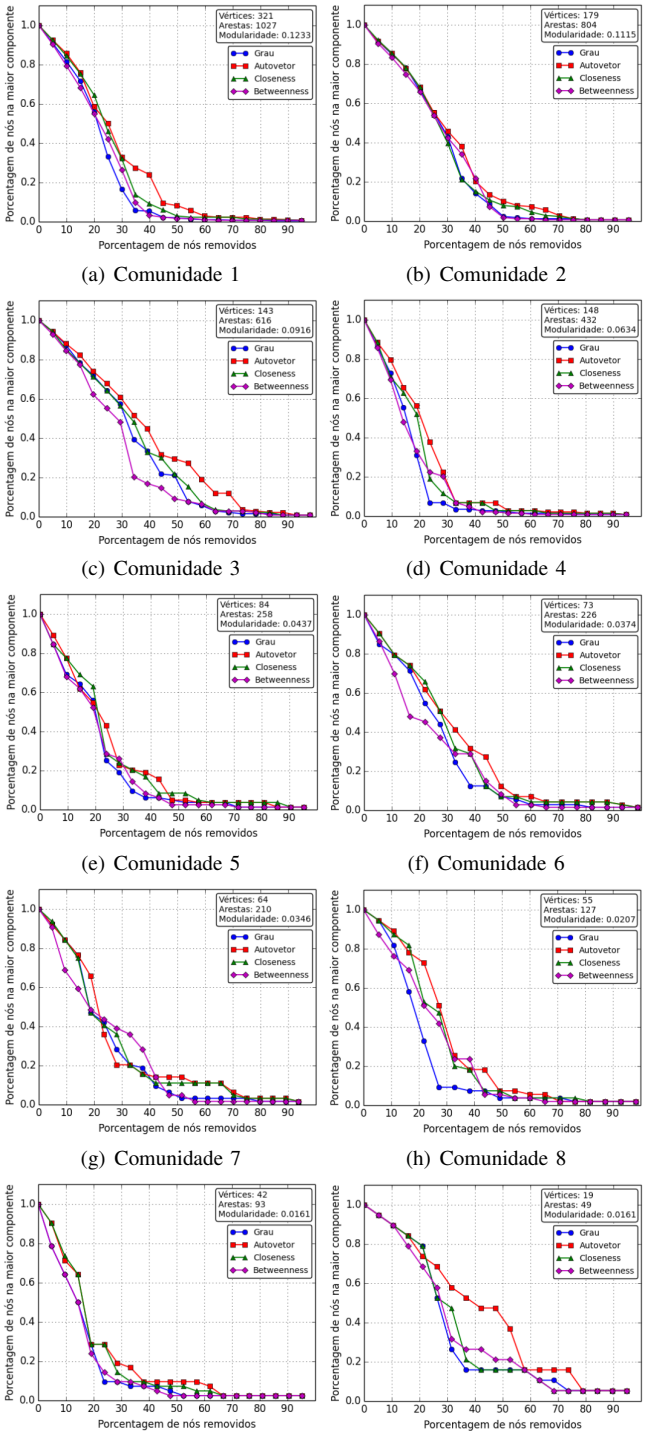
B. Comparação da Robustez em Comunidades de uma Mesma Rede

Considerando a comparação do uso de comunidades isoladas e redes completas para análise de robustez, uma questão importante que surge é qual comunidade a ser utilizada. Nesse sentido, se faz necessária uma análise da relação entre a qualidade de cada uma das comunidades e o comportamento exibido frente aos ataques realizados. Neste trabalho, as redes foram divididas em comunidades utilizando o método de Newman, conforme descrito na Seção III, e foram escolhidas as comunidades de maior modularidade. Nesta seção, a análise de robustez é feita tomando como base diversas comunidades identificadas em uma mesma rede.

A Figura 2 apresenta os gráficos de decaimento utilizando os diferentes critérios de escolha de nós para as diferentes comunidades da rede Email, que exibe, ao mesmo tempo, um grande número de nós, se aproximando de grandes redes encontradas em cenários reais e apenas 10 comunidades, o que permite que a análise seja realizada em cada uma delas. Os gráficos de ataques são apresentadas na Figura 2 em ordem decrescente da modularidade observada para as comunidades extraídas. Novamente são apresentados o número de nós e arestas e a modularidade calculada para cada uma das comunidades nas subfiguras da Figura 2.

Como discutido na Seção IV-A, a comunidade 1, que possui maior modularidade (Figura 1(j)), exibe comportamento bastante similar ao exibido pela rede completa (Figura 1(i)). A comunidade 1 pode ser interpretada, então, como uma boa representante da rede em relação a robustez. Uma investigação rápida permite observar que a medida em que a modularidade das comunidades diminui, o comportamento da comunidade pode se distanciar do observado para a rede completa. Quando isso ocorre, pode-se dizer que as comunidades perdem a capacidade de representar a rede na análise de robustez.

Uma visão mais clara é obtida quando leva-se em consideração os valores de modularidade de cada comunidade. Da mesma forma que a comunidade 1, a comunidade 2 também captura o comportamento da rede completa. De fato, o valor da modularidade obtido para a comunidade 2 (0.1115) é próximo do valor obtido para a comunidade 1 (0.1233), $\sim$90%. Nas comunidades restantes, os valores de modularidade são bastante inferiores aos valores calculados para a comunidade 1 (por exemplo, $\sim$74% para a comunidade 3) e, ainda assim, em alguns casos há comunidades que poderiam ser utilizadas para representar o comportamento da rede, como é o caso da comunidade 5. Por esse motivo, é razoável considerar a comunidade mais modular como tendo um bom potencial para representar o comportamento da rede completa no estudo de robustez em redes e a modularidade é uma medida capaz de capturar essa noção, embora haja outras comunidades em que o comportamento se assemelha ao comportamento da rede.



(a) Comunidade 1 (b) Comunidade 2

(c) Comunidade 3 (d) Comunidade 4

(e) Comunidade 5 (f) Comunidade 6

(g) Comunidade 7 (h) Comunidade 8

(i) Comunidade 9 (j) Comunidade 10

Fig. 2. Robustez das 10 comunidades identificadas para a rede Email.

V. Conclusões e Trabalhos Futuros

Neste trabalho foi realizado um estudo da robustez de redes de diferentes contextos frente a ataques de nós considerando diversas medidas de centralidade: grau, autovetor, *betweenness* e *closeness*. Comparou-se o comportamento exibido pelas redes com o exibido pelas comunidades quando atacadas e, a partir disso, pôde-se perceber uma forte analogia entre a robustez das redes e das comunidades, o que indica uma potencial utilidade das comunidades como forma de representarem a rede completa nesse contexto. Além disso, foi feito um estudo com todas as comunidades de uma mesma rede e verificou-se que comunidades mais modulares têm maior semelhança com a rede completa. Assim, a modularidade pode ser entendida como uma medida de qualidade para a avaliação da analogia de uma comunidade com uma rede.

É importante ressaltar que este trabalho tem caráter preliminar e realizou apenas uma análise qualitativa sobre o comportamento exibido pelas redes e suas comunidades frente a ataques. Um estudo quantitativo pode ser realizado a fim de relacionar o decaimento do número de nós exibidas pelas redes e pelas comunidades com diferentes valores de modularidade a fim de melhor compreender o impacto da estrutura de comunidades na robustez. Ainda com o objetivo de trazer uma visão mais clara sobre a representatividade das redes por suas comunidades, é importante, nos próximos passos, realizar o mesmo experimento em uma gama ainda maior de redes reais e sintéticas, geradas a partir de modelos capazes de controlar a formação de comunidades em redes.

Agradecimentos

Os autores agradecem à Capes, ao CNPq e à FAPEMIG pelo apoio financeiro.

Referências

[1] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.

[2] J. P. Scott, *Social Network Analysis: A Handbook*. SAGE Publications, Jan. 2000.

[3] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks." *Physical Review E: Statistical, nonlinear and soft matter physics*, vol. 69, no. 2, Feb. 2004.

[4] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, no. 3, pp. 245–251, Jul. 2010.

[5] L. Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.

[6] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[7] S. Iyer, T. Killingback, B. Sundaram, and Z. Wang, "Attack robustness and centrality of complex networks," *PloS one*, vol. 8, no. 4, 2013.

[8] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, no. 5, 2002.

[9] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, Jun. 2006.

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, 2008.

[11] V. da Fonseca Vieira, C. R. Xavier, N. F. F. Ebecken, and A. G. Evsukoff, "Performance evaluation of modularity based community detection algorithms in large scale networks," *Mathematical Problems in Engineering*, vol. 2014, 2014.