

Geração de Topologias da Internet por Redução do Grafo Original

Mariane Raquel S. Gonçalves¹, Joylan N. Maciel, Cristina D. Murta¹

¹Departamento de Computação — CEFET-MG
Belo Horizonte, MG

Resumo. *Este artigo apresenta um método para obter topologias sintéticas da Internet no nível dos sistemas autônomos a partir de amostragem aleatória do grafo original. A abordagem mais comum para a geração de topologias sintéticas é construir um grafo a partir de modelos da topologia real. No entanto, a complexidade das interações que produzem a topologia real é difícil de ser captada pelos modelos. Além disso, devido à evolução da rede, os modelos precisam ser continuamente atualizados. A abordagem de redução do grafo elimina a necessidade de construir modelos da topologia. Os resultados indicam que reduções de até 80% no grafo original preservam suas características. Os resultados são comparados com métodos similares encontrados na literatura e com grafos gerados por geradores sintéticos.*

Abstract. *In this paper, we present a method to generate synthetic graphs that mimics the Internet topology at the autonomous systems level based on random sampling of the original graph. The most common approach to obtain a synthetic topology is to build a graph based on models of the real topology. However, the complexity of interactions that produce the real topology is difficult to be reproduced through models. Moreover, due to the evolution of the network, the models must be continually updated. The approach of graph minoring eliminates the need of building models of the topology. The results show that reductions of up to 80% in the original graph preserve their characteristics. The results are compared to similar methods and also to synthetic graphs from topology generators.*

1. Introdução

A construção de topologias sintéticas similares à topologia dos sistemas autônomos da Internet para uso em simulações e estudos ainda encontra vários problemas. Geradores de topologias sintéticas, baseados em estudos e modelagem de características da topologia real, são usados rotineiramente para este fim. No entanto, a topologia real é complexa, e capturar e reproduzir simultaneamente várias de suas características é um problema não resolvido [Tangmunarunkit et al. 2002, Chang et al. 2004, Alderson et al. 2005, Oliveira et al. 2007]. A dificuldade da tarefa é reforçada pelo fato de que a topologia está sempre em evolução e, portanto, os modelos precisam ser continuamente atualizados.

Os geradores de topologias sintéticas constroem grafos similares à topologia real a partir de um grafo vazio, adicionando progressivamente vértices e arestas, até obterem um grafo do tamanho desejado pelo usuário. A construção do grafo é guiada por modelos matemáticos e estatísticos definidos em cada gerador [Winick and Jamin 2002,

Mahadevan et al. 2006a, Maciel and Murta 2008]. Em geral, o tamanho do grafo, representado pelo número de vértices, é o principal parâmetro de entrada do gerador.

Um grafo sintético da topologia deve apresentar pelo menos duas características: deve ser pequeno e suas métricas precisam ser similares às métricas observadas na topologia real. A similaridade é essencial para que o grafo sintético possa efetivamente representar a Internet nos experimentos, o que contribuirá para a confiabilidade dos resultados. O tamanho reduzido é importante em experimentos cujo tempo para finalização depende do tamanho da rede utilizada nos testes. A função de complexidade de tempo pode ser quadrática ou cúbica em relação ao número de nodos na rede, o que indica a necessidade de redes menores.

Recentemente, uma nova abordagem para a construção de grafos sintéticos foi proposta no contexto da topologia da Internet [Krishnamurthy et al. 2005, Krishnamurthy et al. 2007]. Esta abordagem é baseada no conceito de minoração ou redução de grafos [Robertson and Seymour 1985, Lovász 2006], e consiste em aplicar operações de eliminação e/ou contração de vértices e arestas em variadas combinações para gerar grafos reduzidos da topologia original. No contexto do presente trabalho, as operações de eliminação e contração visam reduzir o grafo preservando as suas principais características. O objetivo é obter o menor grafo possível que preserve as características da topologia real.

A construção de topologias sintéticas por redução de grafos é particularmente interessante porque elimina a necessidade de produzir modelos que representem as características da topologia. A produção de modelos de topologias é uma tarefa complexa [Mahadevan et al. 2006b, Oliveira et al. 2007, Latapy and Magnien 2008, Alves and Murta 2008]. Estudos mostram que os geradores de topologias da Internet mais avançados não são capazes de capturar alguns aspectos da topologia, o que revela a complexidade do grafo dos sistemas autônomos da Internet [Winick and Jamin 2002, Mahadevan et al. 2006a, Maciel and Murta 2008]. A abordagem de redução do grafo da topologia real elimina a complexa etapa de modelagem da topologia real para cada métrica do grafo e posterior composição destes modelos nos geradores de topologias sintéticas. A construção de topologias sintéticas por redução de grafos requer que sejam definidas estratégias para excluir vértices e arestas a partir do grafo real. Uma vez definidas estas estratégias, a redução consiste em aplicá-las ao grafo original. Portanto, o custo da solução é menor.

Além disso, os modelos utilizados nos geradores de topologias sintéticas requerem atualização frequente para refletir a situação mais recente do sistema real. Assim, outra vantagem importante dos métodos de redução é a facilidade de gerar grafos reduzidos que refletem a topologia mais recente pois, para isso, basta coletar um grafo recente da topologia da Internet, disponível diariamente na Web [IRL 2009], e aplicar os métodos de redução, para obter imediatamente uma topologia sintética similar.

Seguindo as primeiras propostas para produzir grafos sintéticos da topologia dos sistemas autônomos a partir do método de redução de grafos [Krishnamurthy et al. 2005, Krishnamurthy et al. 2007], este artigo apresenta um novo método para redução do grafo da topologia real. Este método, denominado método histórico, consiste em reduzir o grafo da topologia real considerando a ordem histórica de numeração dos sistemas

autônomos, que são os vértices do grafo que representa a topologia. Este método é fundamentado no modelo de conexão preferencial (*preferential attachment*) [Barabási 2002, Newman et al. 2006].

A numeração dos sistemas autônomos agrega um contexto histórico que revela a importância específica de cada vértice na evolução do padrão de conectividade da rede, em função de seu número. O método histórico é projetado para considerar este contexto histórico e o respectivo papel de cada vértice na rede ao fazer as escolhas dos vértices a serem retirados, com o objetivo de preservar as características da topologia real no grafo da topologia reduzida.

Os resultados apresentados neste artigo indicam que o método proposto apresenta melhores resultados do que o melhor método encontrado nas tentativas anteriores. Utilizando o método histórico, podemos obter topologias sintéticas similares à topologia real dos sistemas autônomos, reduzindo o grafo real em até 80% do seu tamanho. Os resultados anteriores produziram reduções de 70%. Além disso, nossos experimentos incluem um número maior de métricas para comparação das topologias e nossos critérios de comparação são mais rigorosos. O método proposto foi comparado também com grafos produzidos por topologias sintéticas.

Este artigo está organizado em seis seções. A próxima seção apresenta os trabalhos relacionados. A Seção 3 descreve a topologia real utilizada neste trabalho, bem como as métricas utilizadas para comparar os grafos real, reduzido e os grafos sintéticos gerados. A Seção 4 apresenta o método proposto para redução da topologia e os demais métodos utilizados para comparação. A Seção 5 apresenta e discute os resultados e a última seção conclui o trabalho.

2. Trabalhos Relacionados

A topologia dos sistemas autônomos é um aspecto estrutural fundamental da rede mundial. A caracterização desta topologia e sua reprodução em escala menor são essenciais no estudo de vários aspectos da Internet [Tangmunarunkit et al. 2002, Bu and Towsley 2002, Chang et al. 2004]. A topologia dos sistemas autônomos é modelada por um grafo não dirigido cujos nodos são os sistemas autônomos e as arestas são as conexões lógicas entre eles. Para gerar topologias sintéticas para experimentos e simulações, vários geradores de topologias foram projetados, dentre eles o Inet [Winick and Jamin 2002], o Brite [Medina et al. 2001] e o NIT [Maciel and Murta 2008].

Estes geradores constroem a topologia a partir de um grafo vazio, no qual são adicionados vértices e arestas, seguindo modelos específicos em cada gerador. Estes modelos tentam reproduzir vários aspectos da topologia. A produção destes modelos requer caracterização sistemática da topologia real. Apesar dos intensos esforços de caracterização e modelagem da topologia, os geradores ainda não reproduzem nas topologias sintéticas vários aspectos da topologia real [Alderson et al. 2005, Mahadevan et al. 2006b, Mahadevan et al. 2006a, Oliveira et al. 2007, Latapy and Magnien 2008].

Uma nova abordagem para produzir topologias sintéticas foi proposta em [Krishnamurthy et al. 2005, Krishnamurthy et al. 2007]. A ideia é gerar uma topolo-

gia sintética reduzindo por amostragem o grafo da topologia original. Os métodos utilizados para a redução do grafo foram (i) retiradas aleatórias de vértices e arestas, (ii) contrações de vértices adjacentes e (iii) exploração do grafo com busca em largura e em profundidade. Os grafos reduzidos foram analisados considerando-se a distribuição dos graus dos vértices, o grau médio, o desvio padrão do grau e a análise espectral. Dentre os métodos de redução avaliados pelos autores, o método híbrido 0.8 foi o que apresentou os melhores resultados. Este método é uma combinação de dois métodos de remoção de arestas, em proporções de 20% e 80%.

Este artigo se diferencia dos trabalhos anteriores citados acima pois propõe um novo método para fazer a redução do grafo. Este novo método faz uma amostragem mais criteriosa dos vértices a serem retirados, tentando preservar as características da rede. Esta amostragem considera a ordem histórica de numeração dos sistemas autônomos, bem como o modelo de conexão preferencial, que explica a distribuição dos graus dos nodos na topologia da Internet, da Web e em várias redes complexas [Barabási 2002, Newman et al. 2006].

Outra diferença importante do nosso artigo em relação aos trabalhos anteriores [Krishnamurthy et al. 2005, Krishnamurthy et al. 2007] refere-se à maneira de verificar a similaridade entre as topologias real e reduzida. No nosso caso, comparamos as topologias reduzidas com a topologia real a partir da qual as topologias reduzidas foram obtidas. A comparação é feita por meio de métricas de grafos. Nos trabalhos citados, os autores comparam as topologias sintéticas com uma topologia real antiga, que apresente o mesmo número de vértices da topologia sintética gerada. Devido à grande evolução da rede nos últimos anos [Alves and Murta 2008], nós entendemos que esta comparação não é interessante pois, ao utilizar o grafo reduzido, o usuário estará fazendo testes com uma topologia cujas características são similares a uma instância da Internet do passado, e não à topologia atual da Internet.

3. A Topologia Real e Métricas de Análise de Grafos

Os dados da topologia real dos sistemas autônomos da Internet utilizados neste estudo foram gerados pelo *Internet Research Laboratory* da UCLA [IRL 2009]. Esta base de dados da topologia, denominada IRL nesse trabalho, foi escolhida por ser bastante completa, uma vez que a coleta de dados é feita em diversas fontes [Zhang et al. 2005]. Além disso, a atualização é diária e a base de dados fica disponível na Web. As topologias analisadas neste trabalho foram coletadas em março e outubro de 2009.

A numeração dos sistemas autônomos é conhecida como ASN (*Autonomous System Number*) e segue padrões do IETF (*Internet Engineering Task Force*) [IETF 2009]. Os números são atribuídos por organizações de alocação e registros de endereços Internet, em ordem de solicitação. Assim, os sistemas autônomos mais antigos têm números menores. A numeração segue majoritariamente a notação ASPLAIN, em que o ASN é um número inteiro positivo. Dados recentes indicam que já foram criados cerca de 40.000 sistemas autônomos, dos cerca de 65.500 disponíveis. O grafo coletado em 1 de março de 2009 apresentou 38.740 vértices e 233.942 arestas.

As métricas utilizadas para avaliar e comparar as topologias reais, reduzidas e sintéticas são as seguintes: estatísticas e distribuição dos graus dos vértices, estatísticas da distância entre pares de vértices, coeficiente de agrupamento e análise espectral.

O grau $g(v)$ de um vértice v é definido como o número de arestas que incidem sobre ele. O grau médio de um grafo é a média dos graus de todos seus vértices. A distribuição dos graus dos vértices é considerada uma propriedade estrutural característica da topologia da Internet e segue uma lei de potência [Faloutsos et al. 1999], o que indica grande variabilidade nos graus e revela a propriedade “livre de escala” [Barabási 2002, Pastor-Satorras and Vespignani 2004, Newman et al. 2006].

A distância entre dois vértices de um grafo conectado é o comprimento, em número de arestas, do menor caminho entre eles. A distância média do grafo é a média das distâncias entre todos os pares de vértices. A distância é outra métrica característica da topologia e está relacionada à latência da rede e ao desempenho dos acessos e operações [Leskovec et al. 2007]. A pequena distância observada na topologia da Internet caracteriza a propriedade conhecida como *small world*, que indica que pares de vértices estão conectados por caminhos bem curtos [Barabási 2002, Pastor-Satorras and Vespignani 2004, Newman et al. 2006].

O coeficiente de agrupamento de um vértice quantifica a conectividade de seus vizinhos, isto é, quão longe estão os vizinhos de formar um subgrafo completo (clique). O coeficiente de agrupamento $c(v)$ de um vértice v é a proporção entre o número de conexões observadas entre os vizinhos de v e o número máximo possível de conexões. O número máximo de conexões entre vizinhos de v é dado por $g(v) * (g(v) - 1)/2$. O coeficiente de agrupamento é dado por:

$$c(v) = \frac{2|E_{Nv}|}{g(v) * (g(v) - 1)}$$

em que $|E_{Nv}|$ é o número de conexões entre os vizinhos de v e $g(v)$ é o grau do vértice. O coeficiente de agrupamento global é a média dos coeficientes de agrupamento de todos os vértices de um grafo.

A análise espectral de matrizes examina os autovalores e autovetores calculados das matrizes de adjacências normalizadas dos grafos. Os maiores autovalores e seus correspondentes autovetores representam propriedades estruturais da rede [Gkantsidis et al. 2003]. A análise espectral captura aspectos da hierarquia dos nodos da rede e do agrupamento entre os nodos. Portanto, é também uma medida dos agrupamentos observados em uma rede. Estudos anteriores indicam que os resultados da análise espectral da topologia dos sistemas autônomos são pouco alterados com o crescimento da rede [Gkantsidis et al. 2003, Vukadinovic et al. 2001].

4. Métodos de Redução da Topologia

A redução ou minoração de grafos da topologia da Internet, no contexto deste trabalho, tem como objetivo gerar topologias sintéticas menores, reduzidas a partir do grafo original, preservando suas características. Os métodos de redução são baseados no mecanismo de sorteio e exclusão de vértices e/ou arestas. A escolha do elemento a ser retirado - vértice ou aresta - e a forma de sorteio deste elemento é o que diferencia os diversos métodos aplicados. As entradas de dados desses métodos são o grafo $G(V, E)$, com $|V|$ vértices e $|E|$ arestas, obtido da topologia real, e o percentual p de vértices ou arestas a ser excluído. Os métodos são aplicados a esta entrada, reduzindo o tamanho do grafo até o tamanho desejado.

Cada remoção de vértice ou aresta pode gerar componentes separados no grafo, tornando-o um grafo não conectado. Uma remoção de vértice tem como consequência a remoção de todas as arestas que o conectam a outros vértices. Assim, pode ser o caso de um ou mais destes vértices tornarem-se desconectados do maior componente. Portanto, uma remoção de vértice ou de aresta pode implicar na remoção de mais vértices e arestas. Em consequência, após cada procedimento de remoção, são executados algoritmos de busca em grafos para identificar seus componentes e filtrar o maior componente, que será a entrada para a execução da próxima retirada. Nesta etapa, o maior componente é mantido e as estatísticas de retirada são atualizadas. A execução prossegue até que o maior componente tenha alcançado o maior valor menor ou igual ao percentual p dado na entrada de dados. Cada método de redução aplicado neste trabalho é descrito a seguir.

Remoção de vértices Consiste na redução do grafo por remoção aleatória do percentual desejado de vértices. Todos os vértices têm a mesma probabilidade de serem sorteados. As arestas adjacentes a cada vértice sorteado são também excluídas do grafo. Como a distribuição dos graus é muito variável, o número de arestas removidas a cada exclusão de vértice é também variável.

Remoção de arestas Consiste na redução do grafo por remoção aleatória do percentual desejado de arestas. A escolha da aresta a ser removida é feita a partir de um vértice, da seguinte forma: um vértice é sorteado e, dentre as arestas que nele incidem, uma é sorteada e excluída. Os vértices e arestas são selecionados por escolha uniforme, isto é, todos têm a mesma probabilidade de serem escolhidos.

Remoção de arestas simples Consiste na redução do grafo por remoção aleatória do percentual desejado de arestas. As arestas são numeradas e uma é sorteada. Todas as arestas têm a mesma probabilidade de serem sorteadas. Este método é computacionalmente mais caro do que o método de redução de arestas pois, se a numeração de arestas não é alterada, pode ser necessário realizar vários sorteios até que se encontre uma aresta ainda não retirada.

Remoção híbrida Neste método, o objetivo é obter um grafo com um certo percentual do número de vértices do grafo original; porém, as remoções são baseadas em arestas. Portanto, consiste na redução do grafo por exclusão aleatória de arestas até que seja alcançada a redução desejada no número de vértices. A exclusão de arestas é feita com uma composição da aplicação dos métodos remoção de arestas e remoção de arestas simples (descritos acima). Aplica-se a remoção de arestas com probabilidade w e a remoção de arestas simples com probabilidade $1 - w$. Este método, descrito em [Krishnamurthy et al. 2007], foi o que apresentou o melhor resultado naquele trabalho, o que nos motivou a incluí-lo em nossos testes para comparação. Utilizando o método híbrido com $w = 0.8$ foi possível reduzir o grafo original em até 70% mantendo suas principais propriedades [Krishnamurthy et al. 2007].

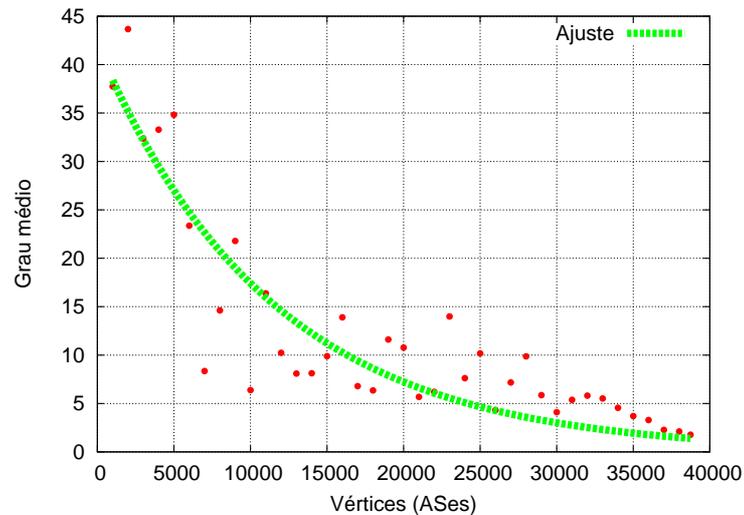


Figura 1. Grau médio a cada sequência de mil vértices na topologia real

Remoção histórica O método consiste em excluir vértices da topologia real, até o limite indicado na entrada. No entanto, a escolha do vértice a ser removido leva em consideração a sequência histórica dos registros dos sistemas autônomos. Os sistemas autônomos são numerados sequencialmente, sendo que os mais antigos possuem numeração menor. Vértices mais antigos e mais recentes têm importância distinta na rede e o objetivo da remoção histórica é preservar parte destes vértices na mesma proporção que são observados na rede original.

O método de remoção histórica considera que o crescimento da rede segue o modelo de conexão preferencial [Barabási 2002, Newman et al. 2006], composto por dois componentes: o crescimento contínuo da rede e a preferência por estabelecer conexões com os vértices mais conectados. Os vértices ganham novas conexões na proporção do número de conexões que já têm. Assim, os vértices antigos são mais conectados e os mais recentes apresentam grau menor.

A relação entre o número do vértice e seu grau, que evidencia o modelo de conexão preferencial, é apresentada no gráfico da Figura 1. Este gráfico mostra, para a topologia real coletada em março de 2009, o grau médio de conjuntos de mil vértices contados sequencialmente de acordo com o ASN. Observamos que os vértices com menor número apresentam grau significativamente maior do que os vértices atribuídos mais recentemente, que são os de numeração maior. A curva contínua apresenta um ajuste feito no Matlab para o conjunto de pontos.

O método de remoção histórica procura preservar, no grafo reduzido, vértices novos e antigos em igual proporção, com o objetivo de preservar as características da rede original no grafo minorado. Na implementação deste método, os vértices são considerados em sequências de n elementos, seguindo a ordem crescente da numeração ASN. Um vértice de cada sequência é removido por etapa. Por exemplo, para excluir 10% dos vértices, a remoção histórica sorteia um em cada sequência de dez vértices, isto é, um vértice é sorteado e excluído entre os dez primeiros da lista, mais um é sorteado na segunda dezena, e assim sucessivamente. O mesmo número de vértices é retirado em toda a faixa de vértices da rede. Assim, esperamos manter as principais características da

topologia original pois estamos eliminando vértices em igual proporção ao longo de toda a extensão de numeração dos vértices da rede.

5. Resultados

Esta seção apresenta e discute os resultados da aplicação dos métodos descritos em instâncias da topologia real. Foram testadas duas topologias reais, coletadas nos dias 1 de março e 1 de setembro de 2009. Os resultados obtidos são muito similares e, devido à restrição de espaço, apresentamos somente os resultados para a topologia coletada em março de 2009. Para verificar a estabilidade dos métodos em relação ao procedimento aleatório, os experimentos foram reproduzidos com pelo menos três dezenas de sementes diferentes. Os resultados indicaram que os métodos são robustos à escolha da semente, o que é consistente com os resultados da literatura [Krishnamurthy et al. 2007].

As reduções foram feitas retirando-se o percentual definido de vértices ou arestas do grafo, de acordo com o objetivo de cada método. Este valor variou de 10% a 90% do número inicial presente. Após a aplicação de cada método, analisamos os grafos obtidos, aplicando algoritmos para estimar as métricas descritas na Seção 3. Analisamos também o tamanho final do grafo reduzido e a proporção de vértices e arestas. Os métodos de redução aplicados ao grafo da topologia real foram os seguintes: remoção de vértices, remoção de arestas, remoção híbrida 0.8, remoção híbrida 0.5 e remoção histórica. Os resultados da aplicação dos métodos são apresentados a seguir. Foram feitas também comparações com topologias sintéticas geradas pelo gerador NIT, que são apresentadas na seção 5.2.

5.1. Análise dos Grafos Reduzidos

Os gráficos apresentados a seguir descrevem o comportamento das métricas selecionadas em função do grau da redução, para os métodos citados. Em todos os gráficos, a topologia real é a topologia que aparece com 0% de redução, que é a entrada de dados para os métodos de redução. O valor de cada métrica da topologia real é o que deve ser preservado nas reduções, exceto, obviamente, o tamanho do grafo.

A primeira análise da redução tem o objetivo de estudar o perfil dos grafos reduzidos formados, em relação à sua composição, à medida em que o grafo é reduzido. O gráfico apresentado na Figura 2 mostra os percentuais de arestas e vértices retirados em cada método. Observamos que os métodos, exceto o método de redução de arestas, experimentam maior redução de vértices do que de arestas. Por exemplo, para uma redução de 60% dos vértices, há uma redução entre 70% e 80% das arestas. O método de redução de arestas apresenta perfil claramente diferente dos demais.

A Tabela 1 apresenta o número de vértices e arestas na topologia real, o cálculo de 20% deste valor, para comparação com as reduções de 80%, e o tamanho dos grafos produzidos por cada método após redução de 80%. Observamos que o número de vértices na redução final é bastante próximo para todos os métodos, exceto remoção de arestas. No método remoção de arestas, a redução é feita sobre as arestas, e podemos observar que o número de arestas final deste método é próximo ao valor de 20% do número de arestas do grafo original. Para os demais métodos, observa-se uma variação maior quanto ao número de arestas.

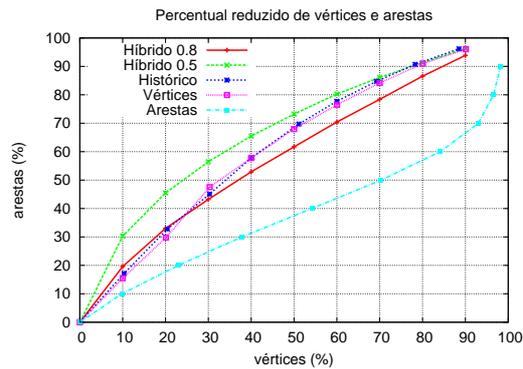


Figura 2. Composição dos grafos reduzidos: percentual de vértices e arestas retirados.

Tabela 1. Tamanho do grafo original e após reduções de 80% em cada método

Grafo	Vértices	Arestas
Original	38.740	233.942
20% do original	7.748	46.789
Híbrida 0.5	7.747	20.405
Híbrida 0.8	7.748	31.478
Histórica	7.739	18.633
Vértices	7.722	21.142
Arestas	1.327	46.776

A seguir analisamos o efeito das reduções nos graus dos vértices. O gráfico à esquerda da Figura 3 apresenta a relação entre o grau médio e o percentual reduzido. Os métodos redução de vértices e histórico apresentam queda linear do grau médio, diferentemente dos métodos híbridos. Observa-se que a redução híbrida 0.8 se difere das demais e mantém maior grau para reduções maiores que 30%. O grau médio obtido pelo método de redução de arestas é crescente e se distancia largamente dos demais valores encontrados, por isso não foi incluído no gráfico.

A métrica expoente do grau é definida como a inclinação dos pontos no gráfico da frequência dos graus, plotado em escala logarítmica em ambos os eixos. Este expoente é o parâmetro da função de distribuição dos graus dos vértices, que segue uma lei de potência. O gráfico da frequência dos graus foi feito para cada topologia gerada, a inclinação da curva foi anotada e é apresentada no gráfico à direita da Figura 3. Observamos que o método histórico mantém o valor desta métrica aproximadamente constante em relação ao grafo original (0%) até a redução de 80%. O método de redução de vértices apresenta alguns pontos em comum com o histórico mas diverge para reduções maiores. Os demais métodos divergem quanto ao valor desta métrica especialmente para reduções acima de 50%.

A distância entre pares de vértices é característica essencial nas topologias. A análise desta métrica é apresentada nos gráficos da Figura 4. O gráfico à esquerda apresenta a distância média calculada para as diversas topologias, à medida que são reduzidas. Observamos que o método histórico é o que melhor preserva a distância média para reduções de até 80%. O aumento significativo da distância média nas reduções acima de

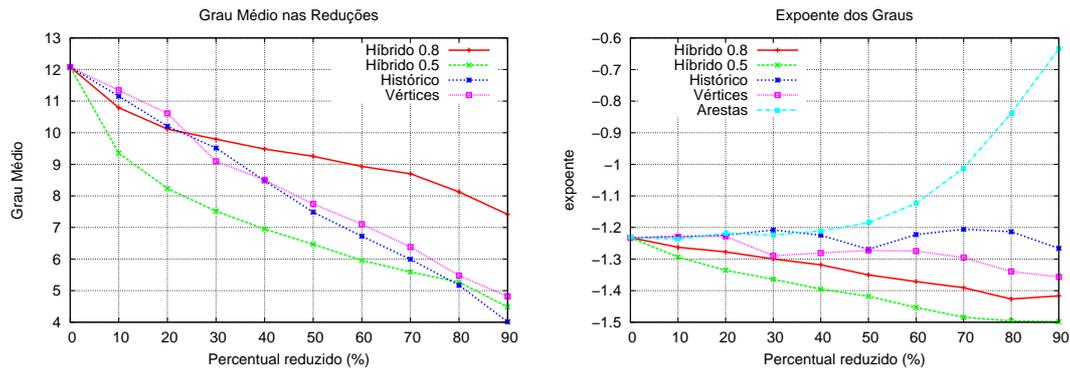


Figura 3. Grau médio (esq) e expoente do grau (dir) nas topologias.

80% deve-se à retirada de um AS específico, que é o mais conectado. Como a rede já está bem reduzida, o impacto é maior, e representa uma situação de pior caso. O método de redução de arestas apresenta um decréscimo significativo na distância média enquanto os demais métodos geram topologias com distâncias médias cada vez maiores. O gráfico à direita da mesma Figura apresenta a maior distância observada em cada topologia. O método de redução de vértices apresenta maior estabilidade nesta métrica, seguido pela redução histórica. Os demais métodos divergiram mais em relação à topologia original.

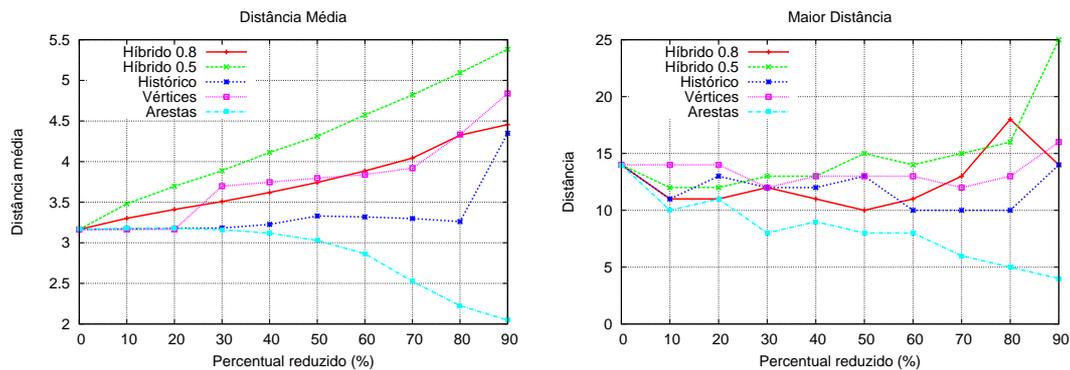


Figura 4. Impacto nas estatísticas de distância: distância média (esq) e maior distância (dir)

O coeficiente de agrupamento quantifica a formação de agrupamentos no grafo. A Figura 5 apresenta o coeficiente de agrupamento médio de acordo com o percentual reduzido. Observamos que os métodos histórico e redução por arestas são os que preservam melhor o grau de agrupamento da rede real, apresentando poucas variações para reduções de até 80%. Os métodos híbridos alteram significativamente o coeficiente de agrupamento em suas reduções, produzindo coeficientes menores, ou seja, grafos com padrões de conectividade significativamente diferentes.

Os resultados da análise espectral são apresentados a seguir. Estes resultados são apresentados de maneira diferente dos anteriores. A análise espectral foi feita para os grafos da topologia real e para os grafos reduzidos pelos métodos testados, com reduções de 70% e 80%. Estes resultados são apresentados na Figura 6. O gráfico à esquerda apresenta os resultados para as reduções de 70% e para a topologia original. O gráfico à direita apresenta os resultados para as reduções de 80% e inclui também a topologia

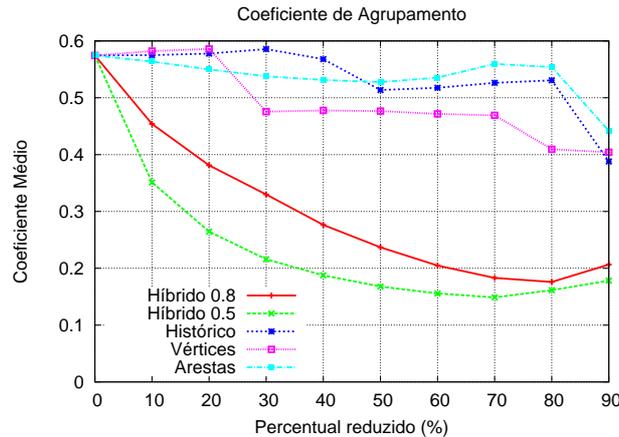


Figura 5. Coeficiente de agrupamento dos grafos reduzidos.

original. Observamos nos dois casos que tanto o método híbrido 0.8 quanto o método de redução de arestas se afastam do resultado da topologia real. As reduções histórica e de vértices acompanham a topologia original nos grafos avaliados.

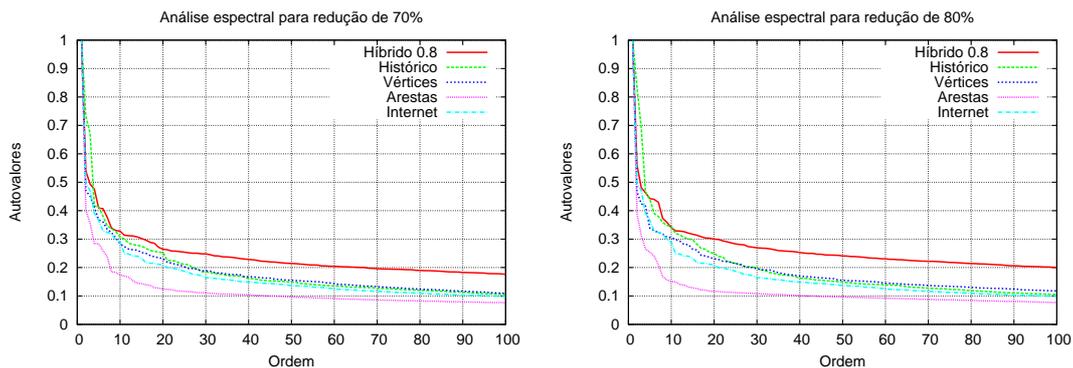


Figura 6. Análise espectral para a matriz da Internet e grafos reduzidos.

Uma análise conjunta dos resultados indica que os métodos remoção de arestas e remoção híbrida 0.5 apresentam os resultados mais distantes do objetivo, divergindo dos valores da topologia real em várias métricas. O método remoção de vértices não faz uma aproximação tão boa quanto o método histórico para as métricas expoente do grau e coeficiente de agrupamento. O método remoção híbrida 0.8, embora preserve um grau médio maior nos grafos reduzidos, apresenta resultados ruins para as métricas expoente do grau, distância média, coeficiente de agrupamento e análise espectral. O método histórico apresenta os melhores resultados no conjunto de métricas, sendo o que mais preserva as características da topologia original.

5.2. Comparação com Topologias Sintéticas

Esta seção apresenta uma comparação dos grafos reduzidos pelos métodos testados na seção anterior, com grafos sintéticos de mesmo tamanho gerados pelo gerador de topologias sintéticas NIT [Maciel and Murta 2008]. O experimento ocorreu da seguinte forma: geramos topologias NIT de tamanho similar aos grafos reduzidos e avaliamos os grafos gerados por construção (NIT) e por redução, utilizando as mesmas métricas.

A Figura 7 apresenta os resultados das métricas distância média e expoente dos graus. Observa-se que a remoção histórica preserva melhor os valores observados na rede real (redução 0%). A topologia sintética NIT e a remoção híbrida 0.8 não preservam estes valores.

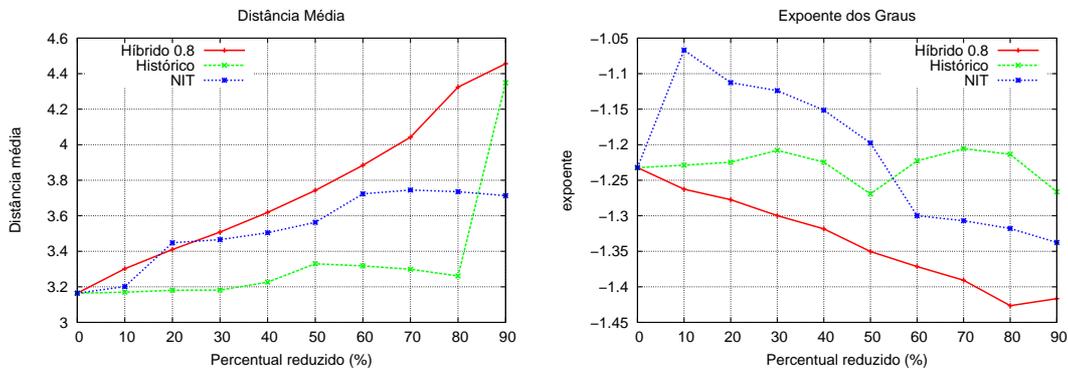


Figura 7. Comparação com topologia NIT: distância média (esq) e expoente do grau (dir).

A Figura 8 apresenta no gráfico à esquerda o resultado da métrica coeficiente de agrupamento. Da mesma forma, observamos que remoção histórica preserva melhor os valores da topologia original (redução 0%). O gráfico à direita da mesma figura apresenta os resultados da análise espectral. Novamente, a remoção histórica apresenta uma similaridade maior com os resultados da topologia real (curva Internet no gráfico).

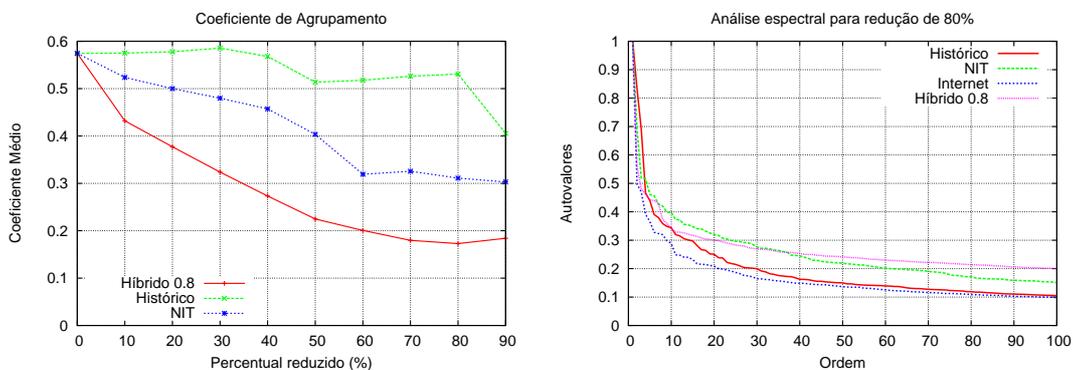


Figura 8. Comparação com topologia NIT: coeficiente de agrupamento (esq) e análise espectral (dir).

6. Conclusões

Este artigo apresentou um novo método para geração de topologias reduzidas da Internet. Estas topologias são úteis em simulações e pesquisas na área de redes de computadores. O método apresentado, denominado remoção histórica, tem o objetivo de preservar as características da rede original. O método proposto apresentou, nos testes realizados, resultados melhores do que os demais métodos testados, incluindo o melhor método descrito na literatura, a remoção híbrida 0.8. Além disso, alcançamos reduções de até 80%, em comparação com 70% dos trabalhos anteriores.

O método de remoção histórica permite gerar rapidamente uma topologia reduzida, a partir de uma instância da topologia dos sistemas autônomos. Sua implementação é simples. É possível gerar pequenos grafos similares à qualquer instância da topologia dos sistemas autônomos da Internet, tanto atual como de qualquer época.

O método de remoção histórica é baseado no modelo de conexão preferencial, que explica o crescimento da rede e o perfil da distribuição dos graus dos vértices. Este modelo de crescimento mostrou-se adequado como base teórica para a redução. Em contrapartida, os resultados apresentam mais uma evidência da validade desse modelo. Além disso, este método pode ser indicado para produzir grafos reduzidos de outras redes complexas que são modeladas pelo modelo de conexão preferencial.

Os desafios futuros incluem a proposta de novas formas de redução do grafo, bem como a ampliação do conjunto de métricas para avaliação da similaridade. Uma questão interessante é o limite de redução que pode ser alcançado. Por exemplo, podemos ter uma topologia similar à topologia da Internet que tenha apenas 1% do seu tamanho? Há um limite teórico para a redução?

Agradecimentos

Os autores agradecem ao CNPq e à FAPEMIG pelo apoio e ao *Internet Research Laboratory* da UCLA pela coleta e geração dos dados da topologia real dos sistemas autônomos.

Referências

- Alderson, D., Li, L., Willinger, W., and Doyle, J. C. (2005). Understanding Internet Topology: Principles, Models, and Validation. *IEEE/ACM Trans. Netw.*, 13(6):1205–1218.
- Alves, R. A. and Murta, C. D. (2008). Análise da Topologia dos Sistemas Autônomos: Evolução e Predição. In *Anais do Simp. Bras. de Redes de Computadores e Sistemas Distribuídos, SBRC 2008*.
- Barabási, A. L. (2002). *Linked: The New Science of Networks*. Perseus Publishing.
- Bu, T. and Towsley, D. (2002). On Distinguishing between Internet Power-law Topology Generators. In *Proceedings of IEEE INFOCOM 2002*.
- Chang, H., Govindan, R., Jamin, S., Shenker, S. J., and Willinger, W. (2004). Towards Capturing Representative AS-level Internet Topologies. *Computer Networks*, 44(6).
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On Power-law Relationships of the Internet Topology. In *Proceedings of the ACM SIGCOMM '99*, pages 251–262, New York, NY, USA. ACM Press.
- Gkantsidis, C., Mihail, M., and Zegura, E. W. (2003). Spectral Analysis of Internet Topologies. In *INFOCOM 2003*.
- IETF (2009). Internet Engineering Task Force. <http://www.ietf.org>. Acesso em 2009.
- IRL (2009). Internet Research Laboratory. <http://irl.cs.ucla.edu/topology>. Acesso em 2009.
- Krishnamurthy, V., Faloutsos, M., Chrobak, M., Cui, J.-H., Lao, L., and Percus, A. G. (2007). Sampling Large Internet Topologies for Simulation Purposes. *Comput. Netw.*, 51(15):4284–4302.

- Krishnamurthy, V., Faloutsos, M., Chrobak, M., Lao, L., Cui, J.-H., and Percus, A. G. (2005). Reducing Large Internet Topologies for Faster Simulations. In *IFIP-TC6 Networking Conference*, pages 328–341.
- Latapy, M. and Magnien, C. (2008). Complex Network Measurements: Estimating the Relevance of Observed Properties. In *INFOCOM 2008. The 27th Conference on Computer Communications.*, pages 1660 – 1668.
- Leskovec, J., Kleinberg, J., and Faloutsos, C. (2007). Graph Evolution: Densification and Shrinking Diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):2.
- Lovász, L. (2006). Graph Minor Theory. *Bulletin of the American Mathematical Society*, pages 75–86.
- Maciel, J. N. and Murta, C. D. (2008). Um Novo Gerador de Topologias da Internet. In *Anais do Simp. Bras. de Redes de Computadores e Sistemas Distribuídos, SBRC 2008*.
- Mahadevan, P., Krioukov, D., Fall, K., and Vahdat, A. (2006a). Systematic Topology Analysis and Generation using Degree Correlations. In *SIGCOMM '06*, pages 135–146, New York, NY, USA. ACM Press.
- Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., k. claffy, and Vahdat, A. (2006b). The Internet AS-level Topology: Three Data Sources and One Definitive Metric. *SIGCOMM Comput. Commun. Rev.*, 36(1):17–26.
- Medina, A., Lakhina, A., Matta, I., and Byers, J. (2001). BRIT: An Approach to Universal Topology Generation. In *9th IEEE MASCOTS*, pages 346–353.
- Newman, M., Barabási, A.-L., and Watts, D. (2006). *The Structure and Dynamics of Networks*. Princeton University Press, Princeton, NJ, USA.
- Oliveira, R. V., Zhang, B., and Zhang, L. (2007). Observing the Evolution of Internet AS Topology. *SIGCOMM Comput. Commun. Rev.*, 37(4):313–324.
- Pastor-Satorras, R. and Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press.
- Robertson, N. and Seymour, P. D. (1985). *Graph Minors — A Survey*, chapter Surveys in Combinatorics, pages 153 – 171. Cambridge University Press.
- Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S., and Willinger, W. (2002). Network Topology Generators: Degree-based vs. Structural. *SIGCOMM Comput. Commun. Rev.*, 32(4):147–159.
- Vukadinovic, D., Huang, P., and Erlebach, T. (2001). A Spectral Analysis of the Internet Topology. In *Dimacs Workshop on Internet and WWW Measurement, Mapping, and Modeling*.
- Winick, J. and Jamin, S. (2002). Inet-3.0: Internet Topology Generator. Technical Report UM-CSE-TR-456-02, EECS, University of Michigan.
- Zhang, B., Liu, R., Massey, D., and Zhang, L. (2005). Collecting the Internet AS-level Topology. *SIGCOMM Comput. Commun. Rev.*, 35(1):53–61.