

Um Novo Gerador de Topologias da Internet

Joylan Nunes Maciel¹, Cristina Duarte Murta²

¹Departamento de Informática, UFPR
Bolsista do PDTA/FPTI-BR

²Departamento de Computação, CEFET-MG

Resumo. *Geradores de topologia são essenciais em várias pesquisas na área de redes de computadores. Este artigo apresenta um novo gerador de topologias dos sistemas autônomos da Internet. Comparações dos resultados desse gerador, em relação a diversas métricas, com topologias reais e com topologias sintéticas produzidas pelo gerador Inet mostram que o gerador proposto gera topologias mais similares às coletadas na Internet do que as topologias criadas pelo gerador Inet, para todas as métricas analisadas. Esse trabalho pretende contribuir para a geração de topologias sintéticas mais realísticas.*

Abstract. *Internet topology generators play an essential role in computer network research. This paper presents a new autonomous system level Internet topology generator. Synthetic topologies generated by the proposed generator are compared to actual Internet topologies and to the results from the Inet topology generator. The results show that, for all the considered metrics, the proposed generator is able to produce more realistic topologies. This work aims to contribute to the generation of more realistic synthetic topologies.*

1. Introdução

A topologia dos sistemas autônomos da Internet evolui rapidamente e o padrão de evolução é alterado em função do estágio de desenvolvimento da rede. O conhecimento adquirido na análise dessa evolução é importante em vários aspectos. Por exemplo, uma caracterização revelou grande variabilidade na distribuição dos graus dos vértices da topologia [Faloutsos et al. 1999], conhecimento que provocou impacto em várias áreas de pesquisa em redes, incluindo a geração de topologias sintéticas. Para conhecer essa evolução é necessário caracterizar continuamente a topologia, confrontar os resultados com os padrões revelados e aceitos e reformulá-los se necessário.

Um dos principais objetivos da análise da topologia é a construção de geradores de topologias sintéticas [Bu and Towsley 2002, Alderson et al. 2005], que são utilizados em simuladores e em testes de laboratório de novos protocolos e aplicações. A construção de grafos sintéticos realísticos contribui para a eficácia dos testes e dos experimentos. Vários geradores de topologia da Internet são conhecidos, por exemplo, o Inet [Winick and Jamin 2002], o BRITE [Medina et al. 2001], o GLP [Bu and Towsley 2002], o RMAT [Chakrabarti and Faloutsos 2006] e o Orbis [Mahadevan et al. 2006]. Há geradores mais antigos, tais como o Tiers [Doar 1996] e o Transit-Stub [Zegura et al. 1997] que têm caráter histórico bastante importante, mas que não reproduzem aspectos observados atualmente na topologia dos sistemas autônomos da Internet.

Provavelmente o gerador de topologia dos sistemas autônomos mais utilizado atualmente é o Inet 3.0, que é também o primeiro recomendado na página do simulador NS2¹. A versão atual do Inet (3.0) foi construída com base na análise de topologias coletadas no período entre novembro de 1997 e fevereiro de 2002 [Winick and Jamin 2002]. No entanto, desde a implementação do Inet 3.0, em 2002, a coleta de dados sobre os sistemas autônomos e a geração de bases de dados da topologia foram objeto de grande debate por parte da comunidade [Zhang et al. 2005, Alderson et al. 2005, Mahadevan et al. 2006, Chang et al. 2004, Chang and Willinger 2006, Cohen and Raz 2006]. Em consequência, bases de dados mais completas e confiáveis foram construídas.

Além disso, tendo em vista a velocidade das mudanças em diversos aspectos da Internet, da infra-estrutura a novos serviços e empresas, o período de cinco anos (desde a última atualização do Inet até o presente ano) parece ser suficiente para revelar novos padrões de evolução. Diariamente surgem novos sistemas autônomos e outros desaparecem, e mudam também as conexões entre estes sistemas. Dados recentes [Oliveira et al. 2007] indicam que novos sistemas autônomos surgem à taxa de 10,3 por dia, enquanto a taxa de desaparecimento é de 2,87 por dia. As conexões (*links*) nessa topologia surgem à taxa de 67,3 por dia e desaparecem à taxa de 45,7 por dia. Esta evolução diária requer caracterização frequente da topologia.

Este artigo apresenta um novo gerador de topologias similares à dos sistemas autônomos da Internet. A necessidade de um novo gerador é revelada comparando-se topologias coletadas recentemente com topologias sintéticas equivalentes geradas pelo Inet. Os resultados indicam discrepâncias em várias métricas essenciais à caracterização da topologia. O novo gerador é capaz de gerar topologias mais similares às topologias reais do que o Inet, considerando as mesmas métricas.

As contribuições desse artigo são:

- uma caracterização da evolução de topologias reais dos sistemas autônomos da Internet coletadas no período de 2004 a 2007;
- uma comparação das topologias reais analisadas com topologias equivalentes geradas pelo Inet 3.0; esta comparação sugere que o Inet não reproduz precisamente características essenciais das topologias atuais, tais como a distribuição dos graus dos vértices, as distâncias entre os vértices e a conectividade;
- uma análise da ocorrência de cliques nas topologias atuais, bem como da evolução das cliques na topologia ao longo do período analisado;
- a proposta de um novo gerador de topologias, baseado no Inet, porém com alterações que permitem reproduzir características atuais da topologia quanto a distribuição dos graus dos vértices, as distâncias entre os vértices e a conectividade;
- resultados de caracterização das topologias geradas pelo novo gerador proposto, em comparação com topologias reais coletadas recentemente.

Os resultados indicam que o novo gerador produz topologias mais próximas das topologias reais atuais do que o gerador Inet 3.0.

Este artigo está organizado em seis seções. A Seção 2 apresenta os trabalhos

¹<http://www.isi.edu/nsnam/ns/ns-topogen.html>

relacionados. A Seção 3 apresenta resultados da análise de topologias reais coletadas no período de 2004 a 2007, bem como resultados da análise de topologias equivalentes às reais, geradas pelo Inet 3.0. A Seção 4 apresenta o novo gerador de topologias proposto nesse trabalho. A Seção 5 apresenta resultados comparando as topologias sintéticas geradas pelo novo gerador com as topologias reais. A última seção conclui o trabalho.

2. Topologia dos Sistemas Autônomos: Métricas, Geradores, Bases de Dados

Diversos trabalhos tratam da análise da topologia da Internet sob diferentes aspectos e em períodos distintos [Oliveira et al. 2007, Pastor-Satorras and Vespignani 2004, Dorogovtsev and Mendes 2003, Alderson et al. 2005]. Um dos trabalhos mais completos [Magoni and Pansiot 2001] analisa topologias no período de novembro de 1997 a maio de 2000 utilizando várias métricas.

A análise da topologia de redes complexas como a Internet pode ser feita por meio de um número grande de métricas [Barabasi 2002, Pastor-Satorras and Vespignani 2004, Dorogovtsev and Mendes 2003]. Não há consenso na comunidade em relação às métricas que melhor representem a topologia dos sistemas autônomos [Tangmunarunkit et al. 2002, Chang et al. 2004, Park et al. 2004, Mahadevan et al. 2006]. No entanto, desde a descoberta de que a frequência dos graus dos vértices segue uma distribuição de cauda pesada [Faloutsos et al. 1999], essa métrica tem sido analisada nos vários estudos já citados. Além da distribuição dos graus dos nodos, dentre as métricas mais utilizadas para caracterizar a topologia estão a distribuição das distâncias, e métricas relacionadas à conectividade do grafo tais como maior clique e coeficiente de agrupamento.

A descoberta de grande variabilidade na distribuição dos graus dos vértices motivou a construção de geradores baseados nessa característica [Tangmunarunkit et al. 2002, Winick and Jamin 2002, Medina et al. 2001]. Para realizar o presente trabalho, foram analisados vários geradores de topologias, com o objetivo de identificar os que poderiam reproduzir melhor as características dos sistemas autônomos da Internet. O objetivo inicial foi analisar e comparar resultados de vários geradores. Os geradores BRITE [Medina et al. 2001], R-MAT [Chakrabarti and Faloutsos 2006], Orbis [Mahadevan et al. 2006] e GLP [Bu and Towsley 2002], além do Inet, foram considerados.

O gerador BRITE foi eliminado porque já havia sido comparado ao Inet [Jin et al. 2000], com melhores resultados para o Inet. Além disso, o BRITE requer a configuração de um conjunto grande de parâmetros, o que amplia as possibilidades de geração de topologias, porém torna extremamente difícil o ajuste dos parâmetros para uma topologia específica.

O gerador R-MAT [Chakrabarti and Faloutsos 2006] divide recursivamente a matriz de adjacências que representa o grafo em quatro partes, gerando conexões entre essas partes. Esse aspecto fundamental de seu modelo torna difícil o mapeamento dos parâmetros em aspectos da topologia dos sistemas autônomos. Além disso, experimentos realizados com o R-MAT indicam que ele gera grafos desconexos, com arestas duplicadas e *self-loops*, características que não ocorrem no grafo alvo desse trabalho. O gerador Orbis [Mahadevan et al. 2006] foi proposto recentemente e, no início dessa pesquisa, não estava disponível para testes.

Por fim, analisamos os geradores Inet 3.0 e GLP [Bu and Towsley 2002], que são desenvolvidos especificamente para representar a topologia dos sistemas autônomos da Internet [Chakrabarti and Faloutsos 2006]. Os parâmetros do GLP são extraídos de modelos elaborados a partir da análise de topologias reais coletadas entre 2000 e 2002. Infelizmente o código do GLP não estava disponível para testes. Segundo os autores², há uma implementação do GLP no BRITE. Analisamos essa e outra suposta implementação do GLP mas não conseguimos nos certificar da autenticidade da implementação, por isso não prosseguimos essa análise.

O Inet 3.0 [Winick and Jamin 2002] é um projeto mais elaborado, com código fonte aberto e vários documentos e versões de código produzidos. Sua configuração é simples e seus resultados são superiores aos de vários outros geradores [Jin et al. 2000, Winick and Jamin 2002]. Além disso, o gerador Inet 3.0 é utilizado em inúmeros estudos pois é implementado no simulador NS-2.

A base de dados da topologia real utilizada nesse artigo é a gerada pelo *Internet Research Laboratory* da UCLA [IRL 2006] e descrita em [Zhang et al. 2005]. Essa base de dados, denominada aqui IRL, é resultado de um projeto que visa reconstruir a topologia dos sistemas autônomos da Internet da forma mais completa e atualizada, coletando e adicionando dados de várias fontes, obtidos por diversos métodos, incluindo bases BGP, dados dos projetos RouteViews [RouteViews 2006] e RIPE [RIPE 2006], e das bases de dados *Looking glasses* e IRR. Os autores projetaram e implementaram um método para atualizar a topologia de forma diária e automática [Zhang et al. 2005]. Há dados disponíveis a partir de 2004.

3. Análise das Topologias Reais e Sintéticas Geradas pelo Inet 3.0

3.1. Descrição do Inet 3.0

O Inet é um gerador cuja estrutura básica é fundamentada na distribuição dos graus dos vértices. O Inet produz topologias a partir do número de vértices do grafo a ser gerado, que é o principal dado de entrada do programa. O algoritmo pode ser dividido em 4 fases: geração da distribuição dos graus, construção do componente conexo, conexão dos vértices de grau 1, e por fim, conexão das arestas dos vértices com graus livres. A conexão das arestas é feita com base no modelo de conexão preferencial [Winick and Jamin 2002].

Na primeira fase, a de geração da distribuição dos graus, o Inet utiliza o número de vértices para inferir vários aspectos da rede, a começar pela sua “idade”. A referência básica é a seguinte: em novembro de 1997, número de sistemas autônomos na Internet era 3037. Esse é o número mínimo de vértices que o Inet aceita como entrada de dados.

A partir do número de vértices desejado, o Inet estima, usando um modelo, a “idade” da topologia, isto é, o número de meses da topologia a ser gerada a partir do valor de referência. Em seguida são estimados os parâmetros da distribuição de cauda pesada, utilizada para gerar os graus dos vértices, com base no número de meses calculado anteriormente. A próxima etapa é gerar os graus de todos os vértices da topologia. Um parâmetro opcional é a fração de vértices cujo grau é igual a 1, cujo valor *default* é 29%, segundo caracterizações realizadas na época da proposta do gerador.

²Comunicação pessoal.

A segunda fase é a construção do componente conexo. Inicialmente é construída uma árvore geradora mínima conectando todos os vértices de grau maior que um, começando com o vértice de maior grau e respeitando a distribuição dos graus gerada anteriormente. O objetivo é garantir a criação de um único componente conexo, como é o caso do grafo da topologia da Internet.

Na terceira fase são conectados os vértices de grau 1 à árvore, assegurando a construção de um único componente conexo. Esta fase contribui para aumentar a altura da árvore. Na última fase são feitas conexões dos demais graus livres, começando com o nodo de maior grau, e respeitando o grau definido para cada vértice. Ao final desta fase todos os vértices terão o número de conexões equivalente ao grau definido para cada um.

3.2. Evolução da Topologia Real e Resultados do Inet 3.0

A proposta do Inet é construir topologias a partir do número de vértices informado pelo usuário do gerador. Esta seção apresenta resultados de análises de topologias sintéticas geradas pelo Inet 3.0 e sua comparação com topologias reais do mesmo tamanho. As topologias são modeladas por grafos.

Os dados da topologia real são os registrados na base IRL, descrita na Seção 2. Para analisar a evolução de características da topologia, foram coletadas duas topologias por ano, correspondentes ao meses de janeiro e junho dos anos de 2004 a 2007. Portanto, são analisadas oito topologias que demonstram a evolução semestral dos sistemas autônomos da Internet. Para cada uma destas topologias utilizamos o número de vértices e a porcentagem de vértices de grau 1 como parâmetros de entrada para geração das topologias sintéticas equivalentes no Inet 3.0.

Topologias reais e sintéticas foram analisadas quanto às métricas número de arestas, distribuição dos graus, distribuição das distâncias, excentricidade e cliques. Os resultados são descritos a seguir. O número de arestas foi escolhido pois é gerado pelo Inet e é dado essencial em qualquer topologia. A distribuição dos graus foi escolhida por ser uma característica fundamental do Inet, e considerada também essencial na caracterização da topologia dos sistemas autônomos. A distribuição das distâncias indica aspectos globais da rede tais como o raio e o diâmetro, derivados da excentricidade. O número e o tamanho das cliques caracteriza a conectividade da rede.

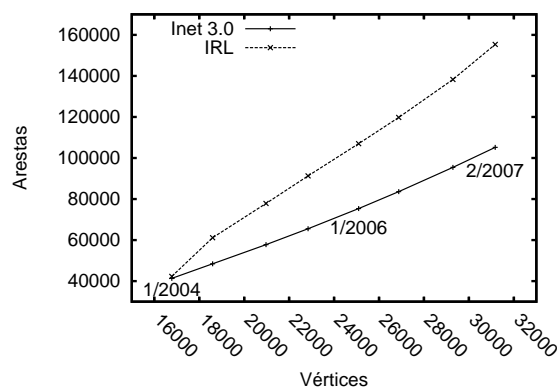


Figura 1. Crescimento do número de arestas nas topologias IRL e Inet 3.0.

A Figura 1 apresenta o crescimento do número de arestas nas topologias reais

e sintéticas. Observamos que o número de vértices (eixo x) é o mesmo em cada data, pois foi coletado da topologia real e dado como entrada no gerador. Verificamos que a topologia real (IRL) apresenta um crescimento maior do número de arestas do que o crescimento modelado pelo Inet, indicando que o gerador Inet 3.0, implementado em 2002, não modela satisfatoriamente o aumento das conexões entre sistemas autônomos observado em tempos mais recentes. Este resultado - menor número de arestas - tem implicações em outras características relacionadas à conectividade da rede.

A Figura 2 apresenta dois gráficos, a frequência acumulada dos graus dos vértices (à esquerda) e o complemento dessa frequência, exibindo a cauda da distribuição (à direita). Os dados são para a topologia real mais recente analisada, de junho de 2007, com 31.170 vértices, e a topologia sintética equivalente. A distribuição dos graus dos vértices é o fundamento para o gerador Inet e, segundo os autores, uma métrica que o gerador modelou de forma praticamente idêntica aos dados reais quando foi proposto [Winick and Jamin 2002]. No entanto, observamos uma nítida diferença entre os dados reais e os dados sintéticos tanto no corpo da distribuição quanto na cauda.

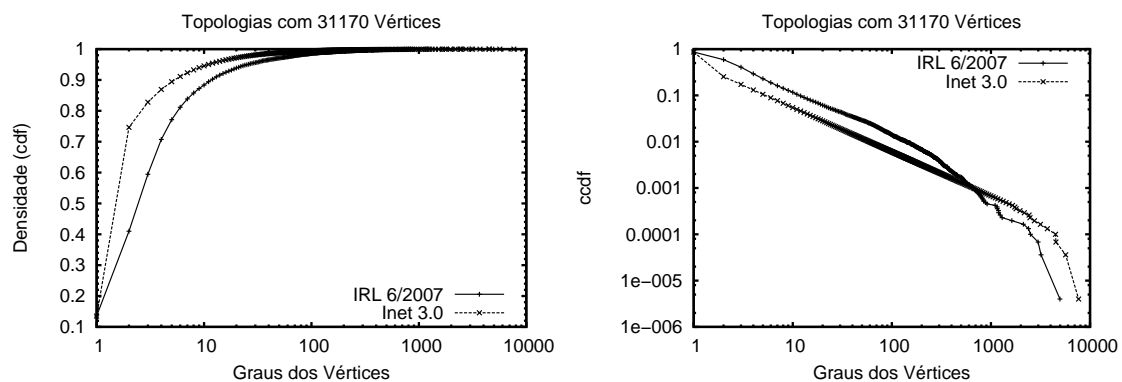


Figura 2. Graus nas topologias IRL e Inet 3.0.

Em grafos, a distância é o comprimento, em número de arestas, do menor caminho entre dois vértices. A Figura 3 apresenta os resultados da análise da distribuição das distâncias em duas topologias, de janeiro de 2004, com 16.782 vértices, e de junho de 2007, 31.170 vértices. Os gráficos indicam que o Inet 3.0 não modelou adequadamente as distâncias observadas na topologia real. O Inet gera uma quantidade muito maior de distâncias pequenas (até 3) e distância máxima igual a 7, enquanto as topologias reais apresentam distância máxima igual a 9 e melhor distribuição das distâncias no intervalo.

A excentricidade é definida a partir da distância. A excentricidade de um vértice v é a distância para o vértice mais longe de v . A excentricidade é uma medida do desvio ou afastamento do centro, ou seja, indica quão centrais ou periféricos são os vértices. Como as distâncias não são bem modeladas no Inet, o mesmo ocorre com a excentricidade, que é uma métrica baseada na distância, como podemos constatar nos gráficos da Figura 4.

A medida do número e tamanho de cliques encontrados na topologia é uma das características mais interessantes da rede. No entanto, o Inet 3.0 reconhecidamente não modela bem o padrão de cliques da topologia [Winick and Jamin 2002, Chakrabarti and Faloutsos 2006]. A Figura 5 apresenta a ocorrência de cliques nas topologias reais avaliadas ao longo do período, bem como a mesma medida para as

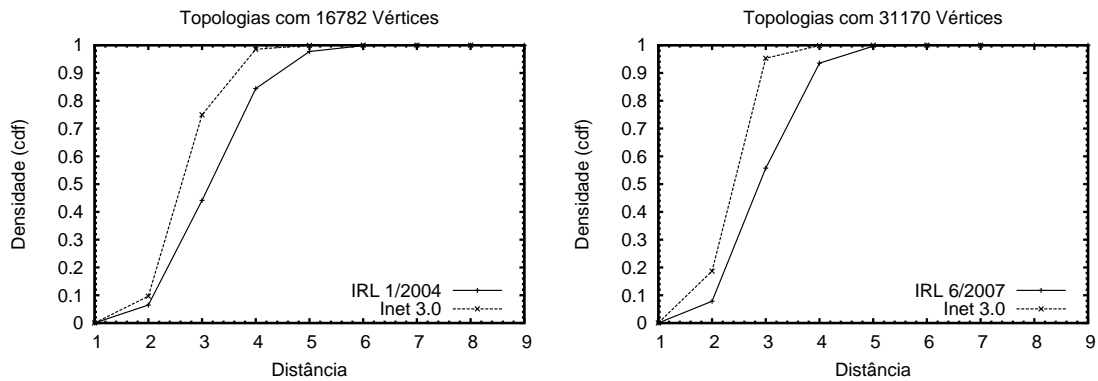


Figura 3. Distâncias nas topologias IRL e Inet 3.0.

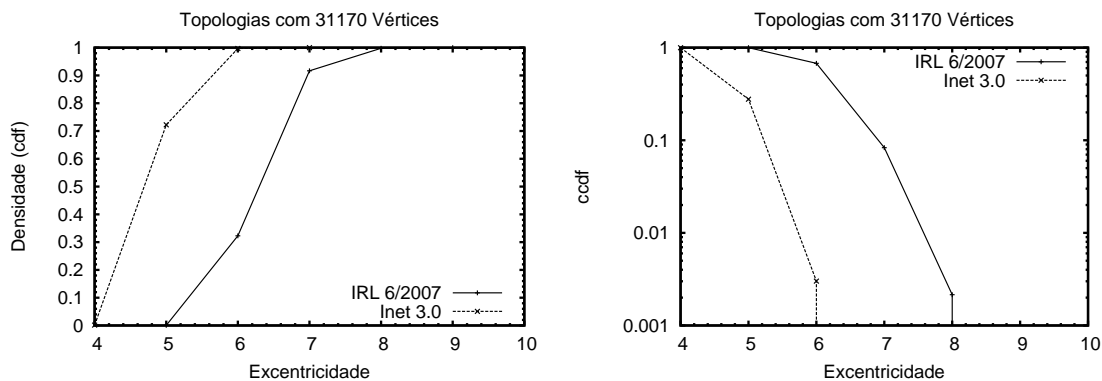


Figura 4. Excentricidades nas topologias IRL e Inet 3.0.

topologias sintéticas equivalentes. Os resultados demonstram padrões bastante distintos. A ocorrência de cliques na topologia real tem evoluído para uma distribuição de Gauss (normal) com tamanhos médios das cliques cada vez maiores (atualmente entre 30 e 35). As topologias geradas pelo Inet não apresentam nenhum clique maior do que 25 e a média é menor do que dez. Observa-se também que a topologia real de janeiro de 2004 apresenta comportamento mais próximo das topologias geradas pelo Inet. A algoritmo utilizado para analisar e contabilizar os cliques é descrito em [Bron and Kerbosch 1973].

Os resultados apresentados nessa seção sugerem que, para essas métricas, o Inet 3.0 não se mostra adequado para modelar topologias reais e mais recentes dos sistemas autônomos. A próxima seção apresenta a proposta de um novo gerador de topologias, baseado no Inet, porém com modificações importantes visando tornar mais representativas as topologias sintéticas geradas.

4. Descrição do Novo Gerador

O algoritmo do Inet 3.0 pode ser segmentado em quatro fases, conforme apresentado na seção 3.1. Com o objetivo de modelar de forma mais precisa as topologias reais, foram alteradas as duas primeiras fases do algoritmo e foi inserida uma nova fase destinada à construção de cliques. As duas fases alteradas são a de geração dos graus dos vértices e de construção do componente conexo. As duas fases restantes foram mantidas sem alteração.

Para alterar a primeira fase, modelamos a evolução do percentual de nodos de grau

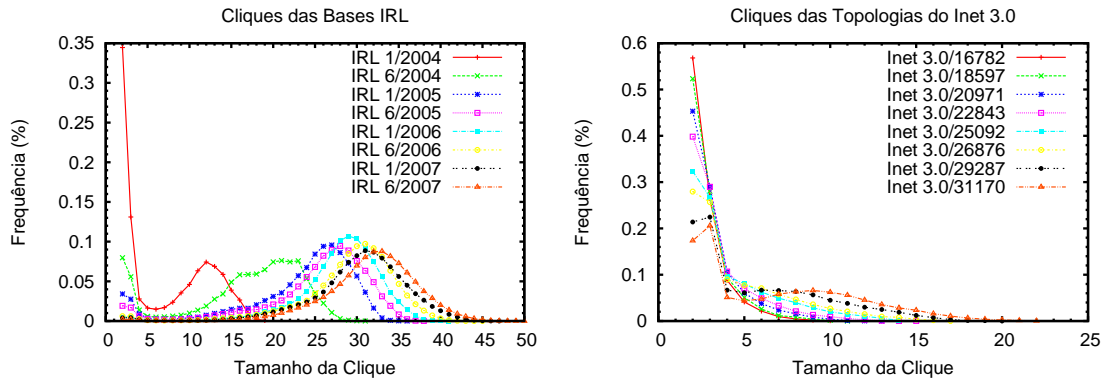


Figura 5. Cliques nas topologias IRL e Inet 3.0.

1 na topologia real no período entre 2004 e 2007. O modelo é apresentado no gráfico à esquerda da Figura 6 e é calculado por uma equação de decaimento exponencial obtida com ajuste $R^2 = 0,99969$. Assim, quando não fornecido pelo usuário, o percentual de nodos de grau 1 é calculado por esse modelo.

Para modelar a distribuição dos graus foi utilizada a distribuição de Pareto limitada (*Bounded Pareto*). Essa distribuição apresenta a característica de cauda pesada e é mais apropriada para modelar conjuntos de dados reais que apresentem limites superior e inferior. O complemento da função de distribuição cumulativa (CCDF) dessa distribuição é dado por:

$$\bar{F}(x) = \frac{j^\alpha \times \frac{1}{x^\alpha} - \left(\frac{j}{k}\right)^\alpha}{1 - \left(\frac{j}{k}\right)^\alpha} \quad (1)$$

em que o parâmetro j é o menor valor (grau = 1) e k é o maior valor, obtido por modelo de ajuste dos maiores graus para todas as bases de dados da topologia real analisadas. O parâmetro α indica o grau de variabilidade da distribuição e foi obtido pela seguinte equação:

$$E[x] = \frac{\alpha}{1 - \alpha} \times \frac{k\left(\frac{j}{k}\right)^\alpha - j}{1 - \left(\frac{j}{k}\right)^\alpha} \quad (2)$$

em que $E[x]$ expressa o valor médio da distribuição, conhecido dos dados reais, assim como são conhecidos os valores de j e k . Os valores de α para todas as topologias analisadas no período de 2004 a 2007 são apresentados no gráfico à direita da Figura 6. No mesmo gráfico é apresentado o modelo para a evolução dos valores de α à medida que o grafo da topologia cresce e evolui. Esse modelo foi obtido por um ajuste logarítmico com grau de confiança $R^2 = 0,99918$.

Portanto, a partir dos modelos para o percentual de nodos de grau 1, para o valor de α e para o maior grau, podemos calcular a distribuição dos graus dos nodos e atribuir a todos os vértices um grau específico, finalizando a primeira fase do algoritmo.

A segunda fase consiste na construção do componente conexo. Foi mantida a criação da árvore geradora mínima, incluindo todos os vértices com grau maior ou igual a

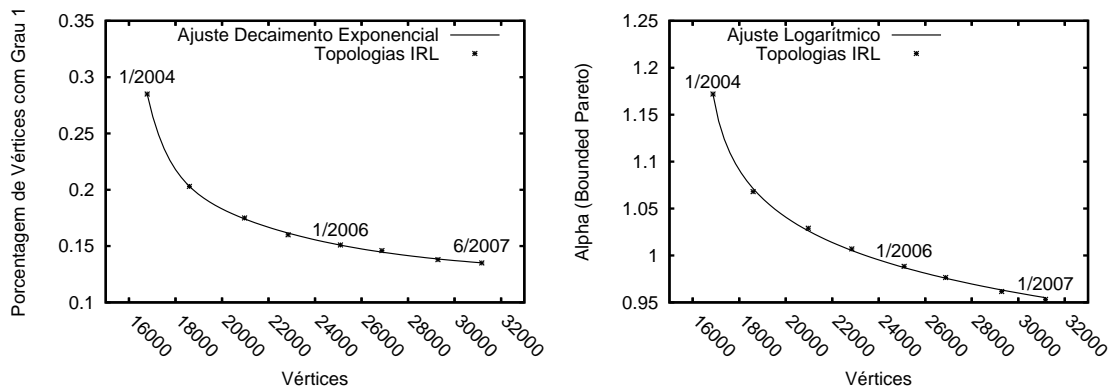


Figura 6. Ajustes da porcentagem de grau 1 e do parâmetro α para topologias IRL.

3. Após a formação da árvore, alteramos o modo de conexão dos vértices de graus 2 e 3. Cada vértice de grau 2 tem a primeira aresta conectada a um vértice aleatório da árvore. Para o restante das arestas desses vértices, 70% são conectadas novamente à árvore, 25% a outros vértices aleatórios de grau 2, e 5% fazem a conexão dos vértices de grau 1. Os vértices de grau 3 foram adicionalmente conectados a outro vértice de grau 3 ou 2, escolhido aleatoriamente. Esta estratégia e as porcentagens utilizadas visam diminuir a alta fração de cliques de tamanho pequeno presentes no Inet 3.0 (Figura 5).

A terceira fase foi criada com o objetivo de tornar o grafo mais denso, modelando melhor a distribuição das cliques e, em consequência, as métricas de distância e excentricidade. Nessa fase são computados os tamanhos e a quantidade de cliques para posterior inserção no grafo. Para essa fase foi necessário modelar o tamanho da clique máxima e da clique média em função do número de vértices, resultando em uma equação logarítmica ($R^2 = 0,99835$) para a clique média e uma equação linear ($R^2 = 0,93873$) para a clique máxima, conforme a Figura 7.

Portanto, na terceira fase do novo algoritmo, estimamos o tamanho médio das cliques e da clique máxima em função do número de vértices. Sabendo que as conexões das fases posteriores aumentam o tamanho das cliques, é necessário nesta fase gerar cliques com tamanhos menores do que os fornecidos pelas equações de ajuste. Sendo assim, definimos 90% dos tamanhos da clique média e máxima calculados. Note que, em várias métricas analisadas, a topologia mais antiga (IRL 1/2004) apresentou maior diferença das demais. Há duas hipóteses para explicar este fato. A primeira deve-se ao melhoramento do método de coleta de dados da topologia; a segunda é devida à própria evolução da rede.

Para finalizar, utilizamos a distribuição normal para calcular a quantidade de cliques de cada tamanho a ser inserida ao grafo. O algoritmo consiste em construir cliques distintas por ao menos um vértice, a partir do tamanho máximo, até cliques de tamanho 3, iniciando a partir dos vértices de maior grau para os de menor grau.

As fases 4 e 5 foram mantidas sem alteração, empregando a conexão preferencial na conexão das arestas dos vértices com graus não exauridos, tal como apresentado em [Winick and Jamin 2002]. Uma vez construído e implementado o novo algoritmo, realizamos comparações entre as topologias obtidas da base IRL, as sintéticas do Inet 3.0

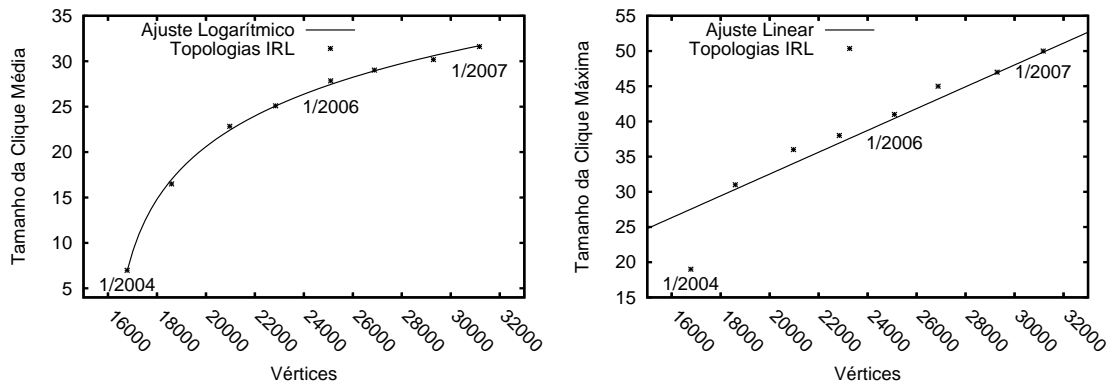


Figura 7. Modelos para o tamanho da clique média e máxima das topologias IRL.

e as geradas pelo novo algoritmo. Os resultados são apresentados na próxima seção.

5. Resultados do Novo Gerador

Esta seção apresenta resultados de análise das topologias sintéticas geradas pelo novo algoritmo proposto. A análise compara esses resultados com as mesmas medidas para a topologia real e para as topologias sintéticas geradas pelo Inet.

A Figura 8 apresenta a evolução do número de arestas nas topologias analisadas. A distribuição de Pareto limitada, utilizada no novo algoritmo, gera uma quantidade de arestas muito mais próxima da topologia real dos sistemas autônomos do que os valores gerados pelo Inet 3.0, embora ainda um pouco acima dos valores reais.

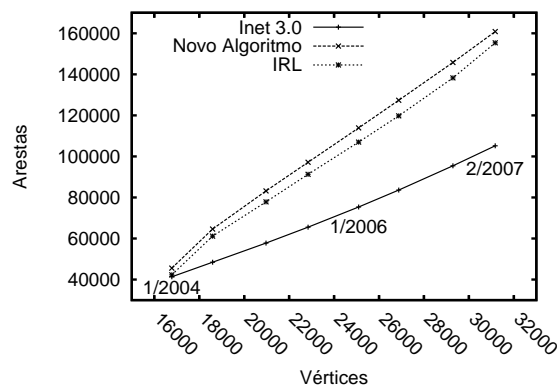


Figura 8. Crescimento do número de arestas das topologias Inet 3.0, Novo Algoritmo e IRL.

A Figura 9 apresenta a distribuição dos graus dos vértices para os três conjuntos de dados. Observamos que a distribuição de Pareto limitada modela melhor a topologia real, principalmente para os graus menores que ~ 200 . A partir desse valor, observa-se uma alteração acentuada na curva CCDF da topologia real (IRL), afastando a curva do modelo de cauda pesada.

A distribuição das distâncias para os três casos é analisada nos gráficos da Figura 10, comparando com a topologia real coletada em janeiro de 2004 (esquerda) e com a coletada em junho de 2007 (direita). Observa-se que o algoritmo proposto gerou

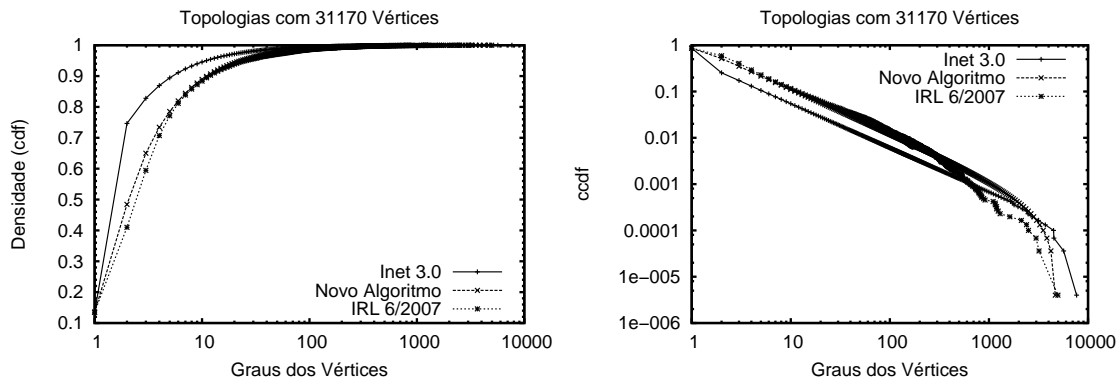


Figura 9. Graus nas topologias Inet 3.0, IRL e Novo Algoritmo.

distâncias mais representativas da Internet do que o Inet 3.0. Em particular, a análise da topologia mais recente indica que o modelo proposto é bastante próximo do real.

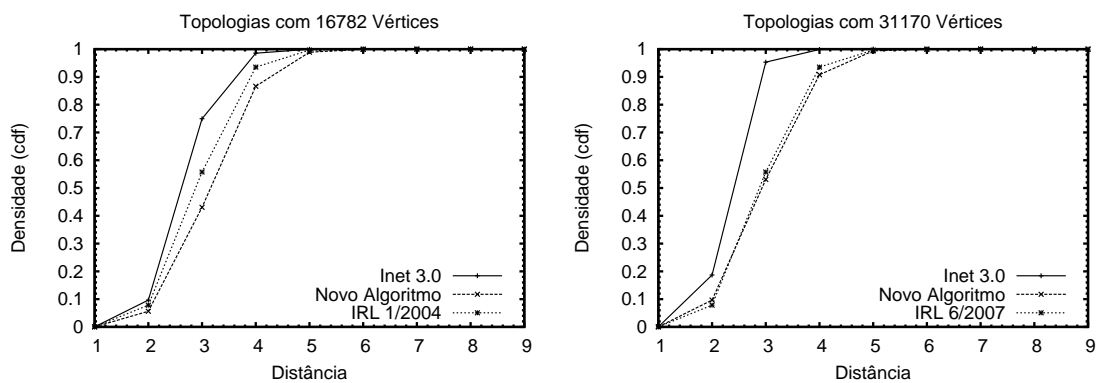


Figura 10. Distâncias nas topologias Inet 3.0, IRL e Novo Algoritmo.

Os bons resultados obtidos para modelar as distâncias estão relacionados a dois fatores. O primeiro é a alteração da fase de geração dos graus, que produziu mais arestas do que o Inet 3.0, com conseqüente aumento da conectividade do grafo. O segundo fator é a inclusão da fase de construção de cliques, que contribui para a ajustar a representação das distâncias, principalmente para as topologias mais recentes.

A Figura 11 apresenta os resultados para a métrica de excentricidade. A análise demonstra que o novo algoritmo modela melhor essa métrica, quando comparado ao Inet. A excentricidade mínima (raio) e máxima (diâmetro) apresentaram os mesmos valores das bases IRL, 5 e 9, respectivamente. O fato da excentricidade ser uma métrica derivada da distância contribui para esses resultados.

Para comparar a geração de cliques, três tamanhos de topologias foram selecionados, com 16782, 25092 e 31170 vértices. A Figura 12 apresenta os resultados para os três casos. Observa-se que o perfil das cliques geradas pelo novo gerador é mais similar à topologia real do que o perfil gerado pelo Inet. A geração de cliques feita pelo algoritmo proposto apresenta o mesmo perfil dos cliques na topologia real, embora com parâmetros distintos. Além disso, o tamanho médio das cliques e o tamanho da maior clique gerados pelo novo algoritmo são mais próximos das mesmas medidas para a base real IRL

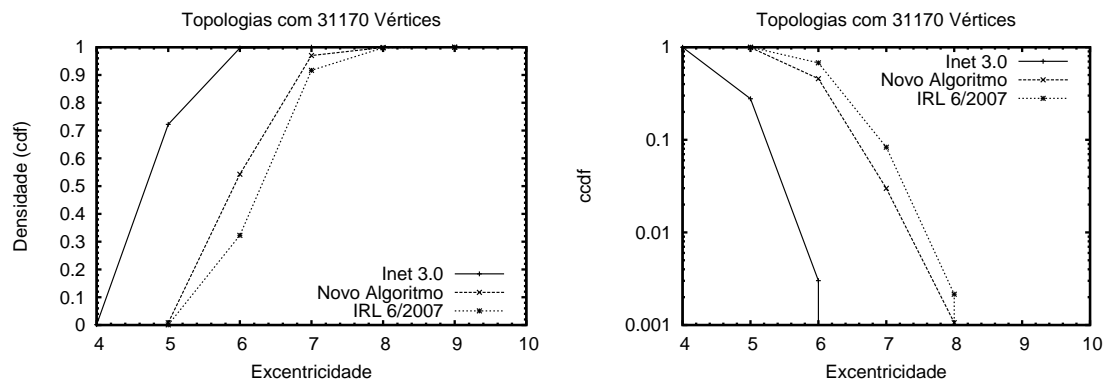


Figura 11. Excentricidades nas topologias Inet 3.0, IRL e Novo Algoritmo.

do que os gerados pelo Inet. Embora a geração das cliques não tenha sido plenamente satisfatória, a inserção dessa fase no novo algoritmo tem fundamental importância para a modelagem das distâncias.

6. Conclusão

Este artigo apresentou uma proposta de um novo gerador de topologias dos sistemas autônomos da Internet, bem como uma comparação dos resultados desse gerador, em relação a diversas métricas, com topologias reais e com topologias sintéticas geradas pelo gerador Inet 3.0. Os resultados mostraram que o gerador proposto gera topologias mais próximas às coletadas na Internet do que as topologias geradas pelo gerador Inet, para todas as métricas analisadas.

O uso de geradores de topologia, em particular do Inet, é bastante significativo nas pesquisas em redes de computadores. A validade dos resultados dessas pesquisas na rede real pode depender do desenho da topologia usada nos testes. Esse trabalho pretende contribuir para a geração de topologias sintéticas mais realísticas.

A observação de que um número grande de topologias pode ser construído a partir de uma distribuição dos graus [Li et al. 2004] torna a abordagem baseada em graus insuficiente para modelar a topologia da Internet. Acreditamos que a associação de métricas de distância e de conectividade podem constituir um novo paradigma para construção de geradores de topologias sintéticas da Internet, substituindo a geração baseada na distribuição dos graus dos nodos.

Os trabalhos futuros incluem o aprimoramento do algoritmo proposto e testes com um conjunto maior de métricas. Observa-se também que todos os geradores sintéticos geram redes de tamanhos similares à topologia real. Outra abordagem para modelar a topologia, que está sendo avaliada, consiste em gerar um modelo menor, por exemplo, com aproximadamente 1% do tamanho real da topologia. Este modelo reduzido apresenta várias vantagens em relação ao modelo maior, como menor custo para geração e testes.

Agradecimentos

Este trabalho foi parcialmente financiado pelo CNPq e pela Fundação Parque Tecnológico Itaipu (PDTA/FPTI-BR). Agradecemos ao grupo do *Internet Research Laboratory* da UCLA pela coleta e por tornar disponível os dados da topologia da Internet na Web.

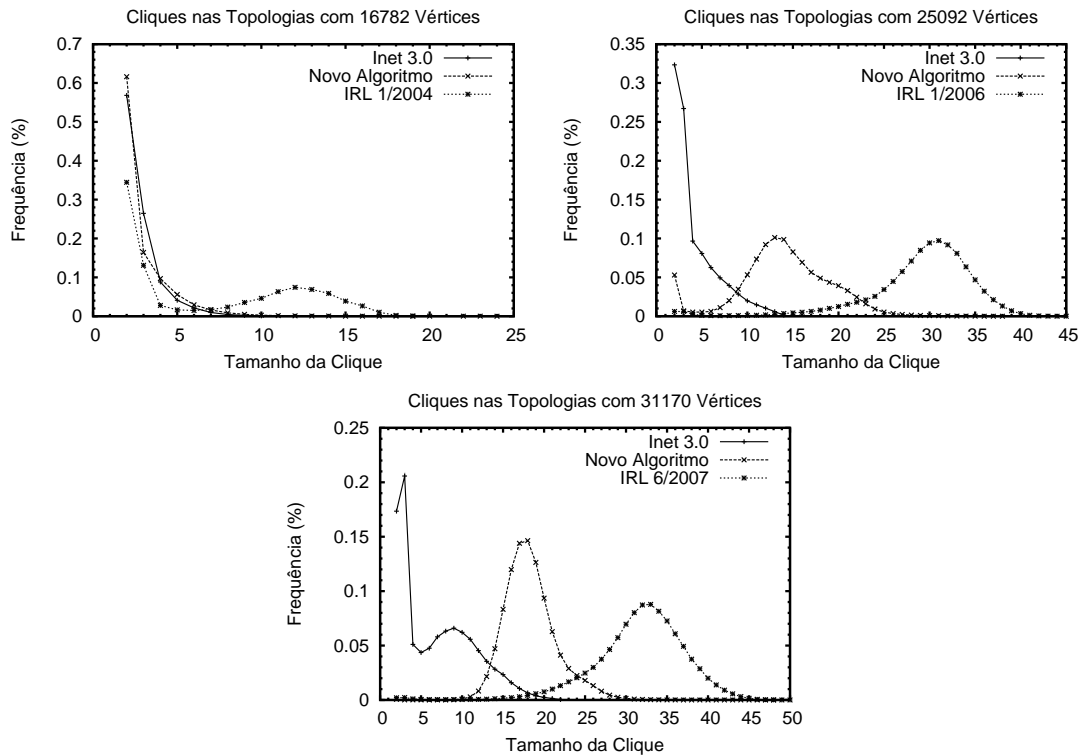


Figura 12. Cliques das topologias Inet 3.0, IRL e Novo Algoritmo.

Referências

- Alderson, D., Li, L., Willinger, W., and Doyle, J. C. (2005). Understanding Internet Topology: principles, models, and validation. *IEEE/ACM Transactions on Networking*, 13(6):1205–1218.
- Barabasi, A. L. (2002). *Linked: The New Science of Networks*. Perseus Publishing.
- Bron, C. and Kerbosch, J. (1973). Algorithm 457: Finding All Cliques of an Undirected Graph. *Commun. ACM*, 16(9):575–577.
- Bu, T. and Towsley, D. (2002). On Distinguishing between Internet Power-law Topology Generators. In *Proceedings of IEEE INFOCOM*.
- Chakrabarti, D. and Faloutsos, C. (2006). Graph mining: Laws, generators, and algorithms. *ACM Comput. Surv.*, 38(1):2.
- Chang, H., Govindan, R., Jamin, S., Shenker, S. J., and Willinger, W. (2004). Towards Capturing Representative AS-level Internet Topologies. *Computer Networks*, 44(6).
- Chang, H. and Willinger, W. (2006). Difficulties Measuring the Internet's AS-Level Ecosystem. In *Proceedings of IEEE INFOCOM*. Invited Paper.
- Cohen, R. and Raz, D. (2006). The Internet Dark Matter – on the Missing Links in the AS Connectivity Map. In *Proceedings of IEEE INFOCOM*.
- Doar, M. B. (1996). A Better Model for Generating Test Networks. In *Globecom '96*.
- Dorogovtsev, S. and Mendes, J. (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press.

- 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.
- IRL (2006). Internet Research Laboratory. <http://irl.cs.ucla.edu/topology>.
- Jin, C., Chen, Q., and Jamin, S. (2000). Inet: Internet Topology Generator. Technical report, Technical Report CSE-TR-443-00, University of Michigan.
- Li, L., Alderson, D., Willinger, W., and Doyle, J. (2004). A first-principles approach to understanding the Internet's router-level topology. In *Proc. ACM SIGCOMM*, pages 3–14.
- Magoni, D. and Pansiot, J. J. (2001). Analysis of the Autonomous System Network Topology. *SIGCOMM Comput. Commun. Rev.*, 31(3):26–37.
- Mahadevan, P., Krioukov, D., Fall, K., and Vahdat, A. (2006). 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., claffy, k., and Vahdat, A. (2006). 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.
- Oliveira, R. V., Zhang, B., and Zhang, L. (2007). Observing the Evolution of Internet AS Topology. *SIGCOMM Comput. Commun. Rev.*, 37(4):313–324.
- Park, S.-T., Pennock, D. M., and Giles, C. L. (2004). Comparing Static and Dynamic Measurements and Models of the Internet's AS Topology. In *Proc. IEEE INFOCOM*.
- Pastor-Satorras, R. and Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press.
- RIPE (2006). Réseaux IP Européens. <http://www.ripe.net>.
- RouteViews (2006). The Route Views Project. <http://www.routeviews.org>.
- 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.
- Winick, J. and Jamin, S. (2002). Inet-3.0: Internet Topology Generator. Technical Report UM-CSE-TR-456-02, EECS, University of Michigan.
- Zegura, E. W., Calvert, K. L., and Donahoo, M. J. (1997). A Quantitative Comparison of Graph-based Models for Internet Topology. *IEEE/ACM Trans. Netw.*, 5(6):770–783.
- Zhang, B., Liu, R., Massey, D., and Zhang, L. (2005). Collecting the Internet AS-level Topology. *SIGCOMM Comput. Commun. Rev.*, 35(1):53–61.