

JOYLAN NUNES MACIEL

UM NOVO GERADOR DE TOPOLOGIAS DA INTERNET

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientadora: Profa. Dra. Cristina Duarte Murta.

CURITIBA

2008

JOYLAN NUNES MACIEL

UM NOVO GERADOR DE TOPOLOGIAS DA INTERNET

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientadora: Profa. Dra. Cristina Duarte Murta.

CURITIBA

2008

JOYLAN NUNES MACIEL

UM NOVO GERADOR DE TOPOLOGIAS DA INTERNET

Dissertação aprovada como requisito parcial para a obtenção do grau de Mestre em Informática no Programa no Pós-Graduação em Informática da Universidade Federal do Paraná, pela comissão formada pelos professores:

Orientadora: Profa. Dra. Cristina Duarte Murta.
Departamento de Informática, UFPR

Prof. Dr. Alcides Calsavara
Pontifícia Universidade Católica do Paraná, PUC-PR

Prof. Dr. Luis Carlos Erpen De Bona
Departamento de Informática, UFPR

Curitiba, 29 de Julho de 2008

AGRADECIMENTOS

A toda minha família, em especial a minha mãe Lucy, meu pai José Alberi e minha irmã Jocylaine por todo apoio, carinho, incentivo, confiança e dedicação que foram indispensáveis conclusão desse mestrado.

A minha orientadora Cristina Duarte Murta pela excelente orientação, dedicação e paciência na condução deste trabalho. Agradeço por ela ter sido minha orientadora ensinando coisas que serão úteis para toda minha vida.

Aos meus orientadores e eternos mestres Feng Chung Wu (Paulo), Hwei Diana Lee e Renato Bobsin Machado, pela inestimável dedicação para a realização deste trabalho. Agradeço pelo apoio, paciência, incentivo e dedicação em todos os momentos e pelos preciosos ensinamentos que levarei durante a minha vida.

Aos meus amigos da família LABI Richardson, Rafael, Daniel, Andrés, André, Maks, Burin, Everton, Newton, Neymar, Shiki e Bianca por todo o auxílio, apoio e incentivo.

A minha querida amiga, extrovertida, animada e as vezes falante Fabiany, pelas inúmeras horas de estudo durante os créditos do mestrado, almoços, passeios e apoio nos momentos complexos.

A Simone, uma pessoa muito especial, pelo grande apoio e incentivo fundamental na fase final do mestrado.

A minha amiga Cibele pelo ajuda e incentivo durante o mestrado.

Ao Programa de Desenvolvimento Tecnológico Avançado do Parque Tecnológico Itaipu (PDTA/FPTI-BR), pelo apoio financeiro na concessão da bolsa de estudos.

Ao Laboratório de Bioinformática (LABI) da UNIOESTE e Fundação Parque Tecnológico Itaipu (FPTI-BR) pela disponibilização da ótima infraestrutura de estudos.

Aos todos os amigos e colegas do mestrado pela amizade, estudos e festas durante o período do mestrado.

A todas as pessoas que direta ou indiretamente contribuíram para a realização deste trabalho.

SUMÁRIO

LISTA DE FIGURAS	vi
LISTA DE TABELAS	viii
RESUMO	ix
ABSTRACT	x
1 INTRODUÇÃO	1
2 GERADORES DE TOPOLOGIAS	5
2.1 Histórico	5
2.2 Evolução dos Geradores de Topologias	7
2.2.1 O Gerador Waxman	7
2.2.2 O Gerador BRITE	8
2.2.3 O Gerador <i>Internet Topology Generator</i> (Inet)	10
2.2.3.1 Inet 2.2	10
2.2.3.2 Inet 3.0	11
2.2.4 O Gerador GLP	12
2.2.5 O Gerador R-MAT	14
2.2.6 O Gerador Orbis	16
2.3 Estado da Arte	18
2.4 Síntese	19
3 CARACTERIZAÇÃO DE TOPOLOGIAS REAIS E SINTÉTICAS	21
3.1 Coleta de Dados da Topologia Real da Internet	21
3.2 Descrição das Métricas para Caracterização das Topologias	23
3.2.1 Distâncias, Excentricidade, Raio e Diâmetro	24
3.2.2 Distribuição dos Graus dos Vértices	25

3.2.3	Conectividade: Cliques e Agrupamentos	26
3.3	Caracterização das Topologias Coletadas na Internet	27
3.3.1	Composição dos Grafos e Estatísticas dos Graus dos Vértices	27
3.3.2	Componentes dos Grafos e Árvores nas Bordas	29
3.3.3	Distância entre Pares de Vértices	30
3.3.4	Excentricidade	31
3.3.5	Conectividade: Cliques e Agrupamentos	34
3.4	Caracterização das Topologias Sintéticas do Inet 3.0	35
3.4.1	Composição dos Grafos e Estatísticas dos Graus dos Vértices	37
3.4.2	Componentes dos Grafos e Árvores nas Bordas	39
3.4.3	Distância entre Pares de Vértices	39
3.4.4	Excentricidade	40
3.4.5	Conectividade: Cliques e Agrupamentos	43
3.5	Síntese	45
4	DESCRIÇÃO DO ALGORITMO DO GERADOR NIT	47
4.1	Fase 1: Alteração da Geração dos Graus dos Vértices	48
4.2	Fase 2: Alteração da Construção do Componente Conexo	49
4.3	Fase 3: Construção das Cliques	51
4.4	Fase 4: Conexão dos Vértices de Grau 1	53
4.5	Fase 5: Conexão dos Vértices com Graus não Exauridos	54
4.6	Síntese	54
5	RESULTADOS	56
5.1	Composição dos Grafos e Estatísticas dos Graus dos Vértices	56
5.2	Distância entre Pares de Vértices e Excentricidades	58
5.3	Conectividade: Cliques e Agrupamentos	61
5.4	Variabilidade do Gerador NIT	64
5.5	Síntese	67
6	CONCLUSÕES E TRABALHOS FUTUROS	69

REFERÊNCIAS BIBLIOGRÁFICAS

LISTA DE FIGURAS

2.1	Categorias dos geradores de grafos. Adaptado de [9].	6
2.2	O modelo R-MAT [9].	15
3.1	Excentricidades de um grafo.	25
3.2	Exemplo de uma clique de tamanho 4 em um grafo G	26
3.3	Número de vértices por grau (esquerda) e evolução das arestas em função do número de vértices (direita) nas topologias IRL.	28
3.4	Distribuição dos graus nas topologias IRL.	29
3.5	Distribuição das distâncias entre pares de vértices nas topologias IRL.	31
3.6	Distribuição das excentricidades nas topologias IRL.	32
3.7	Grau médio em função da excentricidade nas topologias IRL.	33
3.8	Distribuição das cliques (esquerda) e cauda da distribuição das cliques (direita) nas topologias IRL.	34
3.9	Coeficiente de agrupamento global (esquerda) e local (direita) nas topologias IRL.	35
3.10	Número de vértices por grau (esquerda) e evolução das arestas em função do número de vértices (direita) nas topologias Inet e IRL.	38
3.11	Distribuição dos graus nas topologias Inet e IRL.	38
3.12	Distribuição das distâncias entre pares de vértices nas topologias Inet e IRL.	40
3.13	Distribuição das excentricidades nas topologias Inet e IRL.	41
3.14	Grau médio em função das excentricidades nas topologias Inet e IRL.	42
3.15	Distribuição das cliques nas topologias Inet (esquerda) e comparação com as topologias IRL (direita).	44
3.16	Evolução da clique máxima (direita) e coeficiente de agrupamento global (direita) nas topologias Inet e IRL.	44
3.17	Coeficiente de agrupamento médio em função do grau nas topologias Inet e IRL.	45

4.1	Ajustes da porcentagem de grau 1 e do parâmetro α para as topologias IRL.	49
4.2	Ajuste para o tamanho da clique média (esquerda) e da clique máxima (direita) das topologias IRL.	51
4.3	Ajuste para a quantidade total de cliques nas topologias IRL.	53
5.1	Evolução das arestas em função do número de vértices nas topologias Inet, NIT e IRL.	56
5.2	Distribuição dos graus nas topologias Inet, NIT e IRL.	57
5.3	Distribuição das distâncias entre pares de vértices nas topologias Inet, NIT e IRL.	59
5.4	Excentricidades nas topologias Inet, NIT e IRL.	60
5.5	Distribuição das cliques das topologias Inet, NIT e IRL.	61
5.6	Evolução da clique média (esquerda) e máximas (direita) nas topologias Inet, NIT e IRL.	62
5.7	Coefficiente de agrupamento global nas topologias Inet, NIT e IRL.	63
5.8	Coefficiente de agrupamento médio por grau nas topologias Inet, NIT e IRL.	63
5.9	Variabilidade da distância média (direita) e da excentricidade média (esquerda) nas topologias NIT.	65
5.10	Variabilidade do Raio (esquerda) e do diâmetro (direita) nas topologias NIT.	66
5.11	Variabilidade do grau médio (esquerda) e coeficiente de agrupamento (direita) nas topologias NIT.	67

LISTA DE TABELAS

3.1	Composição dos grafos e estatísticas dos graus dos vértices nas topologias IRL.	28
3.2	Estatísticas das distâncias entre pares de vértices nas topologias IRL.	30
3.3	Estatísticas das excentricidades nas topologias IRL.	31
3.4	Raio, diâmetro, centro e periferia nas topologias IRL.	33
3.5	Estatísticas das cliques nas topologias IRL.	34
3.6	Composição dos grafos e estatísticas dos graus dos vértices nas topologias Inet.	37
3.7	Estatísticas das distâncias nas topologias Inet.	39
3.8	Estatísticas das excentricidades nas topologias Inet.	41
3.9	Raio, diâmetro, centro e periferia nas topologias Inet.	42
3.10	Estatísticas das cliques nas topologias Inet.	43
4.1	Fração de vértices com grau 2 das árvores periféricas nas topologias IRL.	50
4.2	Desvio padrão das distribuições das cliques nas topologias IRL.	52
5.1	Composição dos grafos e estatísticas dos graus dos vértices nas topologias NIT.	57
5.2	Estatísticas das distâncias nas topologias NIT.	58
5.3	Estatísticas das excentricidades nas topologias NIT.	59
5.4	Raio, diâmetro, centro e periferia nas topologias NIT.	60
5.5	Estatísticas das cliques nas topologias NIT.	62
5.6	Estatística das distâncias médias nas topologias NIT.	64
5.7	Estatísticas das excentricidades médias nas topologias NIT.	65
5.8	Estatísticas dos raios nas topologias NIT.	65
5.9	Estatísticas dos diâmetros nas topologias NIT.	66
5.10	Estatísticas do grau médio nas topologias NIT.	67

5.11 Estatísticas do coeficiente de agrupamento global nas topologias NIT. . . .	68
--	----

RESUMO

A geração de topologias sintéticas da Internet tem aplicação fundamental em experimentos com novos protocolos e serviços, propostos com frequência pela comunidade de pesquisa na área de redes de computadores e sistemas distribuídos. As novas propostas são testadas em laboratório, usando simuladores de redes, sobre topologias sintéticas geradas para o experimento. Os resultados obtidos e sua exatidão em relação ao comportamento do sistema proposto na rede real dependem da similaridade entre a rede real e a topologia sintética utilizada nos testes. Além disso, as topologias sintéticas auxiliam em atividades de planejamento da rede e no estudo de sua evolução, pois podemos utilizá-las para simular o crescimento da rede no futuro.

Este trabalho apresenta uma caracterização da evolução da topologia dos sistemas autônomos da Internet, bem como uma comparação da topologia real com topologias sintéticas geradas pelo gerador de topologias Inet. A comparação é feita com base em métricas de distância em grafos e suas derivadas, métricas de conectividade da rede e a distribuição dos graus dos vértices. Os resultados indicam que as métricas baseadas em distância são invariantes e muito apropriadas para caracterizar a topologia. Este trabalho também propõe um novo gerador de topologias sintéticas da Internet denominado NIT. Este gerador é baseado no gerador Inet, porém utiliza modelos mais atuais para as métricas, além de incluir novas métricas no processo de geração da topologia. Os resultados do gerador NIT são melhores do que os do Inet em todas as métricas avaliadas. Em consequência, o NIT produz topologias sintéticas mais similares às da Internet atual do que o Inet. Os resultados evidenciam que a associação de métricas de distância, conectividade e distribuição dos graus dos vértices pode ser utilizada para fundamentar uma nova geração de geradores de topologias que representem de forma mais exata as características da Internet atual.

ABSTRACT

The generation of synthetic topologies that resemble the Internet autonomous systems graph has applications in studies of new protocols and services, often proposed by the research community in computer networks and distributed systems areas. The new proposals are tested in experiments using network simulators that generate synthetic topologies, and test the proposals over these topologies. The similarity between the actual network and the synthetic topologies used in the tests is crucial to the accuracy of the results. Moreover, synthetic topologies also help in network planning and network evolution studies, since they can be used to simulate the Internet growth.

This dissertation presents a characterization of the evolution of the Internet autonomous systems topology, and a comparison of the actual topology with synthetic topologies generated by the Inet topology generator. This comparison is based on metrics of distance in graphs and its derivatives, metrics of connectivity and the distribution of vertex degrees. The results indicate that the distance-based metrics are invariant and appropriate to characterize the topology. This work also proposes a new topology generator called NIT. This generator is based on the Inet generator, but it takes advantage of the new models for the metrics, described in this work. It also includes new metrics in the topology generation process. The results of the NIT topology generator are better than those of the Inet in all evaluated metrics. In this way, the NIT produces synthetic topologies more similar to the current Internet than the Inet does. The results show that the composition of distance and distance-based metrics, connectivity metric and degree distribution may be used to build a new generation of topology generators that resemble the current Internet characteristics more accurately.

CAPÍTULO 1

INTRODUÇÃO

A presença ubíqua e pervasiva da Internet na sociedade atual é uma medida de sua importância. No entanto, seu crescimento ocorreu rapidamente e muitos aspectos de sua estrutura ainda não são bem entendidos. A topologia da rede pode ser apontada como a sua principal característica. Estudos sobre a estrutura topológica da Internet e sua evolução ao longo do tempo podem contribuir para o desenvolvimento de novos algoritmos, protocolos e aplicações para a Internet atual e futura.

A topologia da Internet pode ser estudada e analisada em dois níveis. O primeiro nível é composto pelo conjunto de Sistemas Autônomos (SAs) e suas interconexões. Um sistema autônomo consiste em uma ou mais redes de computadores controladas por uma entidade autônoma tal como uma empresa ou organização. Cada SA está sob um único domínio administrativo e possui um número específico denominado *Autonomous System Number* (ASN). O segundo nível de representação da topologia da Internet é o nível dos roteadores. Cada SA pode ter um ou vários roteadores. Em ambos os casos, a topologia é modelada por grafos, nos quais os vértices representam os roteadores ou os SAs, e as arestas são as interconexões entre os vértices. Neste trabalho estamos interessados no estudo de grafos que representem a topologia da Internet no nível dos SAs.

A topologia dos Sistemas Autônomos (SAs) da Internet evolui rapidamente e é uma das principais características de sua estrutura. Seu conhecimento é importante tanto em atividades de pesquisa quanto nas operações diárias da rede, relativas ao desempenho de aplicações, protocolos e serviços, roteamento, proteção e segurança, provisionamento de recursos e geração de topologias sintéticas [19, 25, 41, 46].

Para conhecer a evolução da Internet é necessário caracterizar continuamente a topologia, confrontar os resultados com os padrões revelados e aceitos e reformulá-los se necessário. Diariamente surgem novos sistemas autônomos e outros desaparecem, e mudam também

as conexões entre estes sistemas. Dados recentes [35] 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 freqüente da topologia.

Um dos principais objetivos da análise da topologia é a construção de geradores de topologias sintéticas [7, 25, 41], que são utilizados em simuladores e em testes de laboratório de novos protocolos e aplicações. Nesse contexto, os geradores de topologia sintéticas são ferramentas essenciais em várias pesquisas na área de redes de computadores. Novos protocolos e aplicações para a Internet são testados e experimentados em laboratório. Para isso, é necessário gerar inúmeras redes sintéticas, executar os protótipos sobre estas topologias e avaliar os resultados.

A construção de grafos sintéticos realísticos contribui para a eficácia dos testes e dos experimentos. Em conseqüência, os geradores de topologias sintéticas precisam também ser continuamente atualizados para produzirem topologias sintéticas que imitem a rede real do melhor modo possível. Os geradores sintéticos têm ainda uma função especial, que é a de gerar topologias sintéticas futuras, com o objetivo de prever o comportamento futuro da rede. Gerando topologias sintéticas que imitam a Internet futura, podemos prever o comportamento dos protocolos e aplicações em laboratório. Diante desse contexto, a pesquisa na geração de topologias sintéticas da Internet torna-se uma área de pesquisa atual e necessária.

A topologia da Internet tem sido estudada também no contexto de redes complexas [37], numa área de pesquisa recentemente denominada ciência das redes [14]. Redes complexas são redes grandes e dinâmicas tais como a Internet. Há uma relação intrínseca entre a estrutura, as funções e as propriedades de uma rede complexa. Um dos objetivos dos estudos nessa área é descobrir como as características topológicas da rede afetam a sua função. Assim, conhecer a topologia da Internet é o primeiro passo para identificar relações entre sua estrutura, suas funções e seu desempenho.

O objetivo deste trabalho é propor um novo gerador de topologias sintéticas similares à dos sistemas autônomos da Internet. Para isso, é necessário conhecer a topologia atual da

rede e sua evolução em período recente. Assim, como parte deste trabalho, foi feita uma caracterização extensiva da topologia da Internet. Diversas instâncias da topologia dos SAs, coletadas nos últimos quatro anos, foram analisadas, considerando-se as principais métricas descritas na literatura.

O gerador proposto é denominado NIT (*New Internet Topology Generator*), e constrói grafos representativos da estrutura topológica dos SAs. As topologias sintéticas do NIT contemplam características identificadas por meio da análise de várias métricas, incluindo a distância e a distribuição dos graus [19]. Os resultados do NIT são comparados com a topologia real e topologias sintéticas geradas pelo gerador Inet [47], que é um dos principais e mais utilizados geradores de topologias dos SAs.

De modo resumido, as métricas quando aplicadas à topologia real da Internet apresentam determinados valores. O gerador proposto deve ser capaz de construir topologias que, quando submetidas a essas mesmas métricas, apresentem valores similares àqueles obtidos com a topologia real. Assim, os grafos gerados pelo gerador proposto deverão apresentar as mesmas características topológicas do grafo dos SAs da Internet, de acordo com as métricas analisadas.

A necessidade de um novo gerador é revelada comparando-se topologias coletadas recentemente com topologias sintéticas equivalentes geradas pelo Inet [47]. 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 trabalho de dissertação são:

- Uma caracterização da evolução de topologia real dos sistemas autônomos da Internet, a partir de instâncias da topologia, coletadas no período de 2004 a 2007; esta caracterização é feita com base em várias métricas, por exemplo, a distribuição dos graus dos vértices, a métrica de distância e suas derivadas, o coeficiente de agrupamento do grafo da topologia, e a ocorrência de cliques nas topologias atuais;
- Uma comparação das topologias reais analisadas com topologias equivalentes geradas pelo gerador de topologias sintéticas Inet; esta comparação sugere que o Inet

não reproduz precisamente características essenciais da topologia atual, tais como a distribuição dos graus dos vértices, as distâncias entre os vértices e a conectividade da rede;

- A proposta e implementação do NIT, 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 da rede;
- Resultados de caracterização das topologias sintéticas geradas pelo NIT, em comparação com topologias sintéticas do Inet e topologias reais coletadas recentemente.

Esta dissertação está organizada em seis capítulos. No próximo capítulo são apresentados os trabalhos relacionados, com ênfase para os geradores de topologias sintéticas. No capítulo 3 são definidas as métricas para caracterização da topologia e é apresentada a caracterização da topologia real da Internet e das topologias sintéticas obtidas com o gerador Inet. O capítulo 4 descreve o algoritmo do gerador NIT, cujo código fonte é baseado no Inet. No capítulo 5 são demonstrados os resultados desse novo gerador e, por fim, no capítulo 6 são apresentadas as conclusões sobre o trabalho, seus resultados e as propostas de trabalhos futuros.

CAPÍTULO 2

GERADORES DE TOPOLOGIAS

A estrutura topológica da Internet tem sido cada vez mais estudada por parte da comunidade científica, pois sua evolução é constante e a utilização da Internet tem crescido acentuadamente, principalmente na última década. Na próxima seção será descrito um breve histórico dessa área de pesquisa, e na seção 2.2 são apresentados os principais geradores de grafos no contexto deste trabalho, sendo alguns desenvolvidos especificamente para representar a topologia dos SAs da Internet. Na seção 2.3 é descrito um breve resumo do estado da arte, com as possibilidades e oportunidades de pesquisa nesta área.

2.1 Histórico

A topologia dos Sistemas Autônomos (SAs) 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 [19], conhecimento que provocou impacto em várias áreas de pesquisa em redes, incluindo a geração de topologias sintéticas.

Para a construção de topologias sintéticas da Internet, representadas por grafos, é necessário comparar características entre tais topologias, processo realizado por meio de métricas. Existem várias métricas que podem ser empregadas na caracterização de topologias. Trabalhos recentes analisam diferentes métricas para caracterização topológica da Internet, tal como o uso de métricas baseadas em conectividade [26]. Outra abordagem consiste em analisar com maior nível de detalhes a topologia dos SAs [34].

Contudo, após o trabalho de Faloutsos et. al [19], as métricas baseadas em grau dos vértices tornaram e continuam sendo as mais aceitas e utilizadas nas análises topológicas dos SAs da Internet. Isso motivou a construção de geradores principalmente fundamenta-

dos na distribuição dos graus. A construção desses geradores é um dos principais objetivos da análise topológica [7, 3, 41], pois, são utilizados em simuladores e em testes de laboratório de novos protocolos e aplicações.

De modo genérico, segundo Chakrabarti e Faloutsos [9], os geradores de grafos podem ser amplamente classificados em cinco categorias, conforme apresentado na Figura 2.1, ou então apresentar características híbridas. Essas categorias são discutidas a seguir:

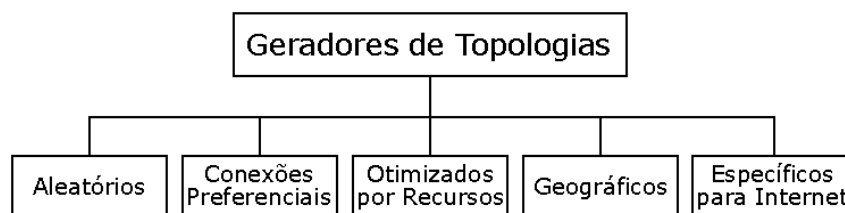


Figura 2.1: Categorias dos geradores de grafos. Adaptado de [9].

- **Geradores Aleatórios:** Nessa categoria os grafos são construídos por meio de conexões aleatórias entre os vértices;
- **Geradores com Conexões Preferenciais:** Nesses geradores a adição de novas arestas e vértices ocorre com maior probabilidade nos vértices de alto grau, gerando uma distribuição dos graus que pode ser modelada por funções que seguem leis de potência (*power laws*), conforme descrito em [5];
- **Geradores de Grafos Otimizados por Recursos:** Nesses modelos as *power laws* surgem a partir da otimização de recursos, ou seja, compromissos entre lucro, recursos e tolerância a riscos, conforme modelos apresentados em [18] e [8];
- **Geradores Baseados em Informações Geográficas:** Nesses geradores são considerados aspectos geográficos, por exemplo, a distância e posição dos vértices, para a conexão dos vértices na construção da topologia;
- **Geradores Específicos para Internet:** Enquanto as quatro categorias anteriores são aplicáveis para gerar grafos que representem qualquer rede, geradores de propósitos especiais têm sido desenvolvidos para modelar algumas características

específicas da Internet. Geralmente são híbridos e utilizam idéias de outras categorias para satisfazer alguns requisitos da Internet.

Vários geradores de topologia da Internet são conhecidos. Por exemplo, há os mais antigos, tais como o Tiers [16] e o Transit-Stub [48] 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 SA mais utilizado atualmente é o Inet 3.0 [47], que é também o primeiro recomendado na página do simulador NS2¹.

Nesse contexto, na próxima seção são apresentados os principais geradores de grafos, dentre os quais alguns específicos para caracterização topológica dos SAs na Internet.

2.2 Evolução dos Geradores de Topologias

Neste trabalho estamos interessados na aplicação dos geradores de grafos para modelar topologias no nível de SAs da Internet. A seguir serão apresentados alguns importantes geradores de topologias que se enquadram em uma ou mais das categorias de geradores descritas na seção anterior.

Para cada gerador serão apresentados seu propósito, principais características, alguns detalhes de implementação e pontos positivos e negativos. Muitos geradores podem enquadrar-se em mais de uma categoria e, dependendo dos parâmetros de configuração, as topologias geradas podem conter características híbridas.

2.2.1 O Gerador Waxman

O primeiro e mais conhecido gerador de grafos aleatórios é o gerador Waxman [44], o qual utiliza uma variação do modelo de Erdős e Rényi [17] com informações geográficas. Nesse gerador os grafos são criados probabilisticamente, ou seja, os vértices da rede são distribuídos uniformemente no plano bidimensional e as arestas são adicionadas de acordo com a probabilidade dependente da distância Euclidiana entre os vértices.

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

O usuário fornece como parâmetro o número de vértices e as coordenadas do espaço geográfico onde estarão esses vértices. A probabilidade de haver uma aresta entre os vértices u e v é dada pela seguinte equação:

$$P(u, v) = \beta e^{\left(\frac{-d(u, v)}{L^\alpha}\right)}, \quad (2.1)$$

onde β e α são parâmetros entre 0 e 1, $d(u, v)$ é a distância Euclidiana de u até v , e L é a distância máxima entre quaisquer vértices. Os parâmetros α e β controlam as limitações geográficas. Um aumento no parâmetro β produz grafos com alta densidade de arestas, enquanto um aumento em α produz uma maior taxa de arestas longas do que arestas curtas [44].

O gerador Waxman pode ser aplicado na modelagem de diversas topologias complexas, porém, não é capaz de reproduzir leis de potência na distribuição dos graus, conforme aparenta evidenciar o grafo dos SAs da Internet [19].

2.2.2 O Gerador BRITE

Um gerador de topologias que foi bastante utilizado pela comunidade científica de redes para simulação da estrutura da Internet é o BRITE (*Boston University Representative Internet Topology Generator*) [32, 31, 33]. Esse gerador foi desenvolvido com o objetivo de permitir a geração de diversos modelos de topologias, adição de novos modelos, interoperabilidade com outros geradores e visualizadores de grafos [32].

As principais características do BRITE são:

- **Adição dos vértices:** Há duas possibilidades para adição de arestas no BRITE. No modelo geográfico é assumida uma grade quadrada na qual os vértices são colocados aleatoriamente ou pode-se adicionar os vértices seguindo uma distribuição de cauda pesada formando agrupamentos;
- **Número de arestas por vértice:** A quantidade de arestas por vértice é um dos parâmetros de entrada do BRITE e, dependendo do modelo de topologia a ser gerada, esse valor é configurado pelo usuário;

- **Crescimento incremental:** O BRITE fornece duas opções para o crescimento da rede. A primeira abordagem é colocar todos os vértices e então adicionar as arestas (modelo Waxman). A segunda é adicionar vértices e arestas simultaneamente (modelo BA);
- **Conexão das arestas:** Os autores fornecem três possibilidades: (1) as arestas podem conectar vértices escolhidos aleatoriamente, (2) pode-se ter a conectividade preferencial original do modelo BA, e (3) pode-se combinar a conectividade preferencial com restrições geográficas.

Além de informações de conectividade, o gerador pode criar topologias com atributos referentes à largura de banda, identificador de AS, dentre outros. Por gerar diversos tipos de topologias, o BRITE possui diversos parâmetros de configuração. As topologias geradas podem combinar propriedades geográficas do gerador Waxman com o crescimento incremental (*incremental growth*) e métodos de conexões preferenciais.

Os detalhes específicos referentes à geração das topologias dependem do modelo de geração que é utilizado. De modo amplo, o processo de criação das topologias segue quatro passos:

1. Colocação dos vértices no plano bidimensional;
2. Interconexão dos vértices por meio das arestas;
3. Associação dos atributos topológicos aos componentes (atrasos e largura de banda para arestas, identificadores dos SAes para vértices roteadores, etc.);
4. Geração do resultado para uma saída de formato específico.

No último *release* do BRITE, em 2002, foi implementado uma versão do gerador GLP que é descrito na seção seguinte. Após isso, o projeto BRITE foi desativado por seus autores.

Por ser um gerador cujo objetivo é construir modelos de topologias genéricos, o BRITE necessita de diversos parâmetros de configuração. Desse modo, o grande número de

parâmetros e a falta de estudo da inter-relação entre esses parâmetros dificulta a utilização desse gerador.

2.2.3 O Gerador *Internet Topology Generator* (Inet)

Um dos mais importantes e utilizados geradores de topologias específicas da Internet é o Inet (*Internet Topology Generator*). A seguir são apresentadas as características desse gerador nas versões 2.2 [22] e 3.0 [47]. Ambas recebem como parâmetros de entrada da topologia o número de vértices, que é obrigatório, e opcionalmente a porcentagem de vértices com grau 1.

2.2.3.1 Inet 2.2

O gerador Inet versão 2.2 constrói topologias em nível de SA seguindo características da Internet observadas entre Novembro de 1997 e Fevereiro de 2002, fornecendo somente informação de conectividade e nenhum outro tipo de informação, tal como latência ou largura de banda.

O Inet 2.2 utiliza uma variação do modelo de conexões preferenciais, apresentado em [5], para representar a topologia da Internet. A topologia é construída por meio das quatro seguintes fases:

Fase 1: Para cada vértice é atribuído um grau definido por uma distribuição que segue uma *power law* com um corte (*cutoff*) exponencial. A probabilidade de um vértice ter k arestas é dada por $= Ck^{-\gamma}e^{-k/\kappa}$, onde C , γ e κ são constantes. Este modelo apresenta uma *power law* ($k^{-\gamma}$) multiplicada por um corte exponencial ($e^{-k/\kappa}$). Acredita-se que o corte exponencial está presente em algumas redes sociais e biológicas, e reduz o comportamento de cauda pesada da distribuição dos graus da lei de potência original;

Fase 2: Uma árvore geradora mínima é formada a partir de todos os vértices com grau maior que 1. Isto assegura um só componente conexo, pois somente existirão os vértices com grau 1 para serem conectados posteriormente à árvore;

Fase 3: Todos os vértices com grau 1 são conectados à árvore geradora mínima por meio do método de conexão preferencial descrito em [22], o qual foi modificado na versão 3.0 do Inet. Este passo aumenta a profundidade da árvore;

Fase 4: Todos os vértices na árvore geradora mínima, iniciando com os de maior grau, recebem arestas extras por meio do método de conexão preferencial, até que sejam alcançados seus graus associados na Fase 1.

A principal vantagem desta seqüência de passos é assegurar que o grafo final permaneça conectado. Contudo, Winick e Jamin [47] analisaram as topologias geradas por este método e constataram que muitos vértices com grau baixo são conectados a outros vértices de grau baixo. Por exemplo, na topologia do Inet 2.2, 35% dos vértices de grau 2 tem adjacentes com grau menor ou igual a 3. Na Internet isso acontece somente para 5% dos vértices de grau 2.

Inversamente, os vértices de grau alto no Inet 2.2 não conectam muitos vértices de grau baixo como na Internet. Por exemplo, aproximadamente 45% dos vértices vizinhos ao segundo vértice de maior grau na topologia Inet 2.2 tem o grau menor ou igual a 3. Na topologia da Internet 75% dos vértices vizinhos ao segundo vértice de maior grau possuem grau menor ou igual a 3. Para corrigir isto, Winick e Jamin [47] propuseram gerador Inet versão 3.0 descrito a seguir.

2.2.3.2 Inet 3.0

O gerador Inet 3.0 é utilizado como base neste trabalho e possui as mesmas fases da versão 2.2. Contudo, nas fases 2, 3 e 4 é utilizada uma equação modificada baseada no modelo de conexão preferencial do Inet 2.2, a qual é descrita a seguir.

Nessa nova versão a equação possui um fator de peso que utiliza os graus dos vértices em ambas as extremidades de qualquer aresta. Seja d_i e d_j os graus dos vértices i e j , respectivamente, e $f(d_i)$ e $f(d_j)$ a freqüência destes graus. Então, w_i^j é o valor do peso de d_j com relação a d_i , onde:

$$w_i^j = \text{MAX} \left(1, \sqrt{\left(\log \frac{d_i}{d_j} \right)^2 + \left(\log \frac{f(i)}{f(j)} \right)^2} \right). \quad (2.2)$$

Na fórmula acima, $f(i)$ e $f(j)$ são o número de vértices com graus i e j , respectivamente, e que podem ser obtidos pela equação de distribuição dos graus. A probabilidade $P(i, j)$ de um vértice i com grau i conectar um vértice j com grau j é dada por:

$$P(i, j) = \frac{w_i^j}{\sum_{k \in G} w_i^k}. \quad (2.3)$$

Intuitivamente, este método faz o seguinte: quando os graus i e j estão próximos, a equação de conexão preferencial permanece linear. Contudo, quando existe uma grande diferença em graus, o peso é a distância Euclidiana entre os graus i e j no logaritmo do plano, e este peso cresce com o aumento da diferença em graus. Assim, arestas conectando vértices com uma grande diferença em graus têm maior preferência.

Segundo Winick e Jamin [47], um problema do Inet 3.0 é relacionado a conectividade, pois o coeficiente de agrupamento e o tamanho da clique máxima, explicados na seção 3.2.3, possuem valores significativamente abaixo do que os encontrados na topologia real da Internet em 2002.

2.2.4 O Gerador GLP

Bu e Towsley [7] propuseram um gerador, denominado GLP (*Generalized Linear Preference*), para construir topologias específicas da Internet quanto a distribuição dos graus. Esse gerador utiliza uma variação do modelo de conexão preferencial AB proposto em [2]. Observou-se que, no modelo AB, a preferência de conexão nos vértices de grau alto é menor do que a encontrada nas topologias da Internet [13, 7].

Para construção do gerador GLP, o modelo AB foi alterado removendo-se o processo de religação das arestas e modificando-se a equação de conexão preferencial, com o objetivo de aumentar a preferência de conexão para vértices com maiores graus.

Seja $p(i)$ a probabilidade de escolher um vértice i . No gerador GLP, $p(i)$ é dada pela

seguinte equação de conexão preferencial:

$$p(i) = \frac{d(i) - \beta}{\sum_j (k(j) - \beta)}, \quad (2.4)$$

onde $p(i)$ representa a probabilidade do vértice i ser um vértice de conexão (*endpoint*), $k(j)$ representa o grau do vértice j , e $\beta \in (-\infty, 1)$ é um parâmetro ajustável, que indica a preferência de um novo vértice conectar os vértices de maior grau. Com um pequeno valor em β , menor preferência é dada aos vértices com alto grau. Note que β deve ser menor do que 1 para que haja uma probabilidade não 0 de vértices com grau 1 adquirirem novas arestas.

O modelo GLP captura dois eventos correspondentes à adição de novos vértices e arestas. Inicia-se com a criação de um *backbone* inicial com m_0 vértices conectados por $m_0 - 1$ arestas e, em cada estágio, executa-se uma das seguintes operações:

1. **Com probabilidade p adiciona-se $m \leq m_0$ novas arestas:** Para cada extremidade de cada aresta, um vértice i é escolhido com probabilidade $p(i)$, definindo a característica de novas arestas conectarem preferencialmente vértices de maior grau;
2. **Com probabilidade $1 - p$ adiciona-se um novo vértice:** O novo vértice tem m novas arestas. Cada aresta é conectada ao vértice i existente no sistema com probabilidade $p(i)$.

De modo geral, os grafos deste gerador são por meio de funções que seguem leis de potência na distribuição dos graus. Em seu trabalho [7], Bu e Towsley analisaram as topologias do GLP, Inet e Internet por meio diversas métricas, com melhores resultados para as duas métricas descritas a seguir.

A primeira métrica é a *Characteristic Path Length* L de um grafo, a qual é dada pela mediana das médias dos tamanhos dos menores caminhos que conectam cada vértice v pertencente ao grafo a todos os outros vértices. A segunda métrica é o coeficiente de agrupamento apresentada na seção 3.2.3. Segundo Bu e Towsley, de acordo com estas métricas o GLP constrói grafos mais semelhantes aos da Internet do que o Inet.

2.2.5 O Gerador R-MAT

Outro gerador proposto, que constrói topologias gerando distribuição dos graus com cauda pesada, é o gerador de grafos R-MAT (*Recursive Matrix*) [10]. Este foi desenvolvido para contemplar a geração de diversos modelos de topologias pois, segundo os autores, muitos geradores baseavam-se na geração de grafos com modelos únicos, como por exemplo, a distribuição dos graus dos vértices.

O gerador R-MAT é recursivo e necessita de poucos parâmetros de entrada. Estes parâmetros são a quantidade de vértices n , arestas m e as probabilidades a , b , c e d , detalhadas adiante. A idéia básica do R-MAT é recursivamente subdividir a matriz de adjacências ($n \times n$) em quatro partes de tamanhos iguais e distribuir as arestas entre estas partes com uma probabilidade desigual. A partir dos parâmetros de entrada o R-MAT constrói grafos com 2^N vértices, geralmente $N = \log_2(n)$.

Na construção do gerador R-MAT o objetivo foi atender as seguintes características:

- **Economia:** Uso de poucos parâmetros para configuração do gerador;
- **Realismo:** Gerador capaz de construir grafos direcionados, não direcionados, bi-partidos, obedecer leis de potência e combinar propriedades dos grafos do mundo real (por exemplo graus dos expoentes e diâmetro) com os valores apropriados em seus parâmetros;
- **Velocidade:** Algoritmo de geração escalável e eficiente, idealmente linear quanto ao número dos vértices e arestas.

O algoritmo inicia com uma matriz de adjacências vazia, dividindo-a em quatro partes com tamanhos iguais. Uma das quatro partes é escolhida com probabilidade a , b , c , d , respectivamente ($a + b + c + d = 1$) conforme representado na Figura 2.2.

A parte escolhida é novamente subdividida em quatro partes menores, e esse procedimento se repete até atingir uma célula simples ($= 1 \times 1$). O vértice (linha e coluna) correspondente a essa célula é conectado ao grafo com uma aresta. É necessário escolher 2 vértices para colocar uma aresta. Este processo é repetido E vezes para gerar um

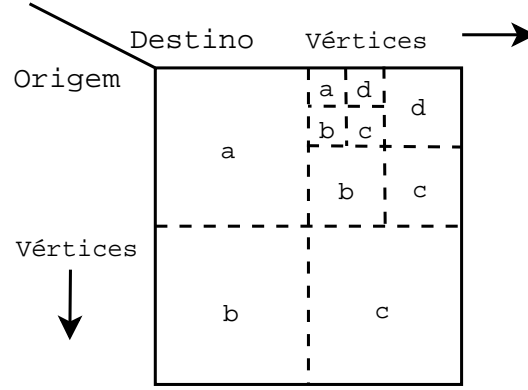


Figura 2.2: O modelo R-MAT [9].

grafo completo. Existe um ponto delicado nesse processo: arestas duplicadas são geradas, porém somente uma delas é mantida quando gera-se um grafo sem pesos (*unweighted*).

Para analisar as alterações externas na distribuição dos graus, alguns ruídos são adicionados aos valores a , b , c , d em cada estágio da recursão, seguido pela renormalização (de modo que $a + b + c + d = 1$). Geralmente, $a \leq b$, $a \leq c$ e $a \leq d$.

Assim, somente três parâmetros são definidos, as probabilidades das partições a , b , e c , pois $d = 1 - a - b - c$. Alterações nesses parâmetros ($a \geq d$) conduzem a distribuições como lognormal e leis de potência. Esta técnica gera agrupamentos ou grupos no grafo, limitados a grupos e subgrupos de quatro. As partes a e d representam agrupamentos separados de vértices que correspondem aos grupos. As partes b e c são conexões transversais aos grupos a e d . A natureza recursiva das partes significa que automaticamente são formados subgrupos dentro dos grupos existentes.

No modelo R-MAT pode-se calcular o número esperado de vértices c_k com grau de saída k :

$$c_k = \binom{E}{k} \sum_{i=0}^n \binom{n}{i} [p^{n-i}(1-p)]^k [1 - p^{n-i}(1-p)]^{E-k}, \quad (2.5)$$

onde 2^n é o número de vértices no grafo R-MAT e $p = a + b$.

Os grafos gerados nesse modelo são grafos direcionados. Para gerar grafos não-direcionados é gerada uma matriz de adjacências simétrica, por meio da cópia dos elementos acima da diagonal principal para abaixo da diagonal principal.

De modo geral, o gerador R-MAT gera naturalmente leis de potência na distribuição

dos graus e conduz a grafos com características de mundo pequeno (*small world graphs*), conforme apresenta a topologia da Internet. Este gerador é econômico e veloz, mas apesar dessas vantagens, é difícil mostrar se são geradas topologias realísticas, isto é, o mapeamento das probabilidades de entrada (a, b, c e d) em características de redes reais torna-se algo complexo.

Adicionalmente, segundo Chakrabarti et al. [10], o R-MAT pode ainda gerar topologias com vários componentes desconexos e apesar de receber como entrada o número de vértices especificado pelo usuário, experimentos realizados mostraram que o maior componente conectado possuía entre 80% e 90% dos vértices do valor do parâmetro de entrada. Outra característica é que as topologias geradas podem-se apresentar *self-loops* e múltiplas arestas entre vértices, aspectos que dificultam o uso do gerador.

2.2.6 O Gerador Orbis

Em um dos mais recentes trabalho, Mahadevan et al. [29] apresentaram uma nova e sistemática abordagem para analisar topologias de rede, introduzindo a distribuição da probabilidade *dK series*, que especifica a completa correlação dos graus de subgrafos de tamanho d contidos em um determinado grafo G . Aumentando o valor de d captura-se progressivamente mais propriedades de G ao custo de uma mais complexa representação da distribuição de probabilidade.

O modelo de grafos *dK series* foi proposto para satisfazer os seguintes requisitos:

- **Viabilidade de Construção:** Construção de grafos com séries de propriedades P_d com $d = 0, 1, 2, \dots$;
- **Inclusão:** Qualquer propriedade P_d assume todas as propriedades P_i , com $i = 0 \dots d - 1$, ou seja, grafos com propriedades P_d possuem todas as propriedades P_i , com $i < d$;
- **Convergência:** Conforme d aumenta, o conjunto de grafos tendo propriedades P_d “converge” para G , ou seja, existe um valor do índice $d = D$ tal que todos os grafos tendo propriedades P_D são isomórficos a G [29].

Por questão de notação, o termo “grafos dK ” representa o conjunto K de todos os grafos com propriedades P_d . Conforme d aumenta, mais propriedades de G são capturadas ao custo de maior complexidade na representação da distribuição de probabilidade. Para comparação, o grafo G deve ser passado como parâmetro para o gerador Orbis. Em seu trabalho Mahadevan et al. [29] utilizou como G a topologia (*Heuristically Optimized Tradeoffs*) gerada no trabalho de Li et al. [25].

Mahadevan et al [29] definiram as seguintes propriedades: os grafos $0K$ ($d = 0$) com propriedade (P_0), que representam o grau médio de um grafo, reproduzem o mesmo valor do grafo G . Os grafos $1K$, com propriedade P_1 , reproduzem a distribuição dos graus $P(k) = n(k)/n^1$, onde $n(k)$ é o número de vértices com grau k e n é o número de vértices do grafo. Os grafos com propriedade P_1 apresentaram a mesma forma do grafo G .

Para grafos $2K$, a propriedade P_2 descreve a distribuição conjunta dos graus ou o número de arestas total $m(k, k')$ entre vértices de grau k e k' . Assim, P_2 reproduz a mesma forma do grafo G e descreve a correlação dos graus para os pares de vértices conectados.

Para entender os grafos $3K$ é necessário definir duas geometrias dentre as possíveis na interconexão de d vértices. A primeira é uma corrente de três vértices conectados por duas arestas, e a segunda são os triângulos, cliques de três vértices. Estas duas geometrias ocorrem com diferentes frequências entre vértices com diferentes graus e cada configuração requer uma distribuição de probabilidade separada. Assim, estes dois componentes juntamente formam a distribuição $3K$, ou propriedade P_3 [29].

Para construção dos grafos dK , com $d = 0$ e $d = 1$, diversas abordagens são apresentadas. A primeira é o algoritmo *stochastic*, que gera grafos $1K$. A segunda é o algoritmo *pseudograph*, que gera grafos $2K$ com *self-loops*. Uma extensão é o algoritmo *matching*, que também gera grafos $2K$, porém, sem a presença de *self-loops*. Por fim, a abordagem *rewiring* é generalizável para qualquer d , e a principal idéia é fazer religações aleatórias preservando a existente distribuição dK ou propriedade P_d .

Conforme apresentado em [29], $d = 2$ é o suficiente para reproduzir muitas métricas de interesse e $d = 3$ reproduz exatamente todas as métricas conhecidas pelos autores na lite-

ratura, tais como por exemplo o a distribuição dos graus, os agrupamentos, a distribuição da distância, e centralidade. Contudo, por ser um gerador recentemente proposto, estudos adicionais desse gerador estão sendo realizados, tal como apresentado em [26].

Além disso, o fato de receber como parâmetro de entrada um exemplo de topologia a ser reproduzida pela distribuição *dK series* é a principal desvantagem que dificulta o uso desse gerador por parte dos usuários.

2.3 Estado da Arte

De modo geral, os principais geradores específicos para a topologia da Internet utilizam abordagens fundamentadas na distribuição dos graus dos vértices, principalmente gerando distribuições que seguem cauda pesada. Estas distribuições, criadas ou não por modelos de conexões preferenciais, continuam sendo as métricas mais estudadas e utilizadas até o momento.

A partir da publicação do artigo de Faloutsos et al. [19], no qual a distribuição dos graus dos vértices poderia ser modelada por equações que seguem leis de potência, diversos esforços foram realizados na tentativa de construir geradores que construíssem topologias sintéticas com a distribuição dos graus geradas por funções que seguem leis de potência [10, 7, 22, 32].

Apesar destes esforços, o fato de que grafos com a mesma distribuição dos graus dos vértices podem apresentar diferentes estruturas topológicas, demonstra que somente a distribuição dos graus dos vértices não é suficiente para caracterizar a topologia da Internet [25]. Além disso, cada tipo de métrica, seja baseada em distância, conectividade ou graus dos vértices, é definida e analisada de maneira independente e algumas vezes empírica, sendo, em poucos casos, diretamente mapeada nas propriedades e funções da rede.

Assim, o estudo e a definição de novas métricas associadas que caracterizem a topologia da Internet de modo preciso, mais completo, e tenham mapeamento direto nas funções intrínsecas da rede, é o primeiro passo em busca da solução para o problema de modelagem da topologia da Internet. Essa área possui um amplo potencial de pesquisa já que até o

momento as soluções, na maioria baseadas somente na distribuição dos graus dos vértices, não são satisfatórias e a inter-relação entre essas métricas não é analisada.

De modo geral, as soluções ainda não são satisfatórias por diversos motivos. O primeiro é que as métricas são analisadas de forma independente umas das outras, prejudicando a correlação entre seus resultados. Em segundo, as métricas não possuem fácil mapeamento nas funções da rede. Por fim, as métricas baseadas em somente uma abordagem, tal como o grau dos vértices, podem não representar corretamente a topologia da Internet, pois, nesse caso, diferentes topologias podem ser construídas com a mesma distribuição dos graus dos vértices.

Uma vez que a estrutura estática (fotografia) da topologia da Internet pode ser caracterizada por meio de métricas, outra área com grande potencial de pesquisa é o estudo do aspecto evolutivo da Internet. Em diversas topologias propostas aplica-se o modelo de conexão preferencial [5]. A medida que a Internet evolui, este modelo perde capacidade de modelagem desta rede [7] e necessita ser constantemente adaptado às características da Internet [2, 47]. Assim, o estudo dessa evolução constitui um problema ainda não solucionado e proporciona mais uma área com oportunidade de pesquisa.

A partir da definição de métricas, ou conjunto de métricas, que caracterizem de modo mais completo a topologia da Internet, o passo final é a construção de geradores que construam topologias sintéticas contemplando estas métricas. Nesse sentido, três pontos desejáveis em relação a esse gerador são: a adequação ao conjunto das principais métricas aplicadas atualmente na análise da topologia Internet; a possibilidade de comparar os resultados da métrica desenvolvida com outras métricas já propostas e; fundamentalmente, o mapeamento de seus significados em termo das propriedades e funções da rede.

2.4 Síntese

Nesse capítulo apresentou-se algumas das categorias de geradores de topologias definidas em [9]. Posteriormente foi descrito um breve histórico dos trabalhos relacionados e os principais e mais utilizados geradores de grafos, dentre o quais, alguns específicos para gerar topologias dos SAs da Internet.

Para cada gerador analisado na seção 2.2, foram abordados principalmente aspectos de implementação, parametrização, categorização e pontos positivos e negativos. Dentre os quais, o GLP, Inet e Orbis se destacaram por serem específicos para Internet, este último é o mais recente e, até o momento, o menos estudado pela comunidade. De modo geral, esses geradores são principalmente fundamentados na distribuição dos graus dos vértices, e utilizam diferentes abordagens para gerar a distribuição com o padrão de cauda pesada, tal como apresenta a Internet [19]. Adicionalmente, a parametrização feita pelo usuário não é simples na maioria dos geradores analisados, exceto o Inet.

A topologia da Internet está em constante evolução [49], o que requer caracterização contínua e cada vez mais completa. Uma possível abordagem é a união das principais métricas para uma caracterização comum que permita comparações entre métricas na busca de uma solução mais completa. Nesse sentido, no próximo capítulo são apresentadas as principais métricas de caracterização topológica, bem como a caracterização da topologia relativa à Internet (bases IRL) e topologias do gerador Inet, o qual foi escolhido por motivos descritos na seção 3.4.

CAPÍTULO 3

CARACTERIZAÇÃO DE TOPOLOGIAS REAIS E SINTÉTICAS

A coleta, análise e caracterização da topologia da Internet no nível de Sistemas Autônomos (SA) são temas discutidos em vários artigos [19, 11, 29, 49]. De modo geral, a abordagem destes trabalhos consiste em analisar o processo de coleta e estudar as topologias capturadas para tentar caracterizar o maior número de métricas possível, a partir dos dados das topologias reais.

Na próxima seção são apresentadas questões referentes ao processo de coleta de dados da topologia da Internet. Na seção 3.2 são apresentadas as principais métricas para caracterização de topologias da Internet, do ponto de vista deste trabalho. Por fim, nas seções 3.3 e 3.4 são apresentadas as caracterizações da topologia coletada da Internet e topologias sintéticas da Internet, construídas com o gerador Inet 3.0. Essas topologias são analisadas por meio de métricas baseadas em graus dos vértices, agrupamentos, cliques e principalmente distâncias.

3.1 Coleta de Dados da Topologia Real da Internet

A estrutura topológica da Internet tem atraído a atenção de pesquisadores nos últimos anos, pois o seu conhecimento é importante para a construção e o estudo de protocolos e algoritmos. Existem diversos trabalhos que analisam e apresentam soluções para os problemas dos processos de coleta de dados referentes à topologia real da Internet [11, 12, 23, 38].

O processo de coleta de informações sobre a topologia da Internet, no nível de SA, pode ser realizado de diversas maneiras, embora não haja um específico para essa finalidade. Os métodos mais utilizados são os baseados na tabela de roteamento do protocolo *Border Gateway Protocol* (BGP), no *Internet Routing Registry* (IRR) ou bases WHOIS e na

ferramenta de rede *traceroute*. A seguir são apresentadas as principais características referentes a estes métodos de coleta.

O método BGP utiliza informações contidas nas tabelas BGP, as quais são empregadas no roteamento entre SAs na Internet. O projeto *Route View* [40] baseia-se nessas tabelas para construir dois tipos de grafos da topologia da Internet no nível de SA. O primeiro é uma visão estática da topologia, baseada nas tabelas de roteamento BGP, e o outro representa uma visão dinâmica, baseada em informações de atualização das tabelas de roteamento BGP. Esses métodos representam a visão da topologia da Internet do ponto de vista do roteamento. Sendo assim, estas bases podem conter inconsistências e ocultar enlaces que não são publicados nas tabelas de roteamento, seja por razões técnicas ou econômicas [38].

O método IRR, que origina as bases WHOIS, fornece dados contendo políticas de roteamento dos *Internet Service Providers* (ISP) e informações da conectividade de SAs registrados, entre outros. Estas bases são manipuladas por uma linguagem denominada *Routing Policy Specification Language* (RPSL) [1] cujo propósito é permitir que operadores de rede descrevam políticas de roteamento global [11, 38, 49]. As bases de dados IRR ou WHOIS são mantidas manual e voluntariamente por operadores, fato que pode gerar informações incorretas, incompletas ou desatualizadas. A mais atualizada e completa base IRR pública é a RIPE Whois mantida pelo projeto RIPE [39], que contém informações da infra-estrutura de rede européia.

Outros conjunto de métodos são os baseados na ferramenta *traceroute*, tal como os projetos *skitter* da associação CAIDA [42] e o DIMES [15]. Ambos utilizam a ferramenta *traceroute* para coletar dados referentes à topologia da Internet. De modo geral, este método captura topologias da Internet do ponto de vista de fluxo de dados real enviado pela rede. O *skitter* captura seqüências de endereços IP (*Internet Protocol*) ao longo do caminho de uma origem até o destino. A associação CAIDA armazena endereços entre diversas origens e destinos para representar a cobertura do espaço de endereçamento IPv4. O projeto DIMES utiliza *traceroute*, e mais recentemente a ferramenta *ping*, para capturar a topologia da Internet. Neste método, milhares de agentes distribuídos geograficamente

executam *traceroutes* referentes aos caminhos dos fluxos de dados no nível de roteadores, obtendo informações que permitem o mapeamento no nível de AS.

Considerando a existência de diversos métodos de coleta, estudos recentes apontam que o modo mais completo para coleta da topologia da Internet baseia-se na união desses métodos/ferramentas. Nesse sentido, o projeto *UCLA Internet Research Lab* [21] fornece bases, aqui denominadas IRL, que são descritas em [49]. Esse projeto 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 *Route Views* [40] e *RIPE* [39], 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 [49]. Há dados disponíveis a partir de 2004.

O fato de não existirem ferramentas ou protocolos específicos para captura da topologia da Internet, acrescido de limitações técnicas e legais, são fatores que dificultam a captura do grafo que representa a conectividade real e completa da Internet. Os resultados acabam representando amostras da topologia, seja no nível de SAs ou de roteadores. Segundo Cohen e Raz [38], na topologia no nível de SA há evidências de que pelo menos 35% das conexões (*links*) não são capturadas nas bases atuais. Desse modo, as topologias coletadas podem apresentar inconsistências e ambigüidades.

No entanto, acredita-se que há pouca probabilidade de existir diferença significativa entre as topologias coletadas e a real quanto ao número de SAs. O fato de protocolos ou ferramentas desenvolvidos para outros objetivos serem aplicados na coleta da topologia da Internet é um dos principais motivos que promove a falta de confiabilidade das topologias coletadas [12].

3.2 Descrição das Métricas para Caracterização das Topologias

Muitos artigos têm proposto métricas para caracterizar a topologia da Internet [19, 28, 30, 36]. De modo geral, essas métricas são aplicadas em bases de dados referentes à topologia real da Internet e em topologias geradas sinteticamente para fins comparativos.

Dentre as métricas mais analisadas estão as baseadas na distribuição dos graus dos

vértices, distribuição conjunta de graus de vértices vizinhos, conectividade tal como formação de cliques e agrupamentos, centralidade, métricas relativas a distâncias e métricas definidas por matrizes. Todas essas métricas foram estudadas e as mais importantes na atualidade, do ponto de vista deste trabalho, são descritas nas próximas subseções.

3.2.1 Distâncias, Excentricidade, Raio e Diâmetro

A distância é uma métrica essencial em um grafo e, em especial, em uma rede, podendo servir de bases para diversas métricas. Dentre elas, a distribuição das distâncias e a excentricidade dos vértices são as mais importantes para o contexto deste trabalho. A seguir são apresentadas algumas métricas baseadas na distância.

Seja um grafo $G(V, A)$, onde V é conjunto com n vértices (por exemplo, SAs da Internet) e A é o conjunto com m arestas (ou enlaces) correspondentes as conexões lógicas entre ASes. A distância é o comprimento, em número de arestas, do menor caminho entre dois vértices: $d(u, v) | v, u \in V$. O número de pares de vértices (e de distâncias) em um grafo é dado por $P = n(n - 1)/2$.

A distribuição dos tamanhos dos caminhos mais curtos ou simplesmente a distribuição da distância $d(x)$ é a probabilidade de que um par aleatório de vértices esteja a uma distância de x arestas entre si. Em outras palavras, $d(x)$ é o número de pares de vértices que estão a uma distância x , dividido pelo número total de pares P (excluindo *self-loops*) do grafo G .

Duas medidas estatísticas básicas associadas a distribuição das distâncias em um grafo são a distância média \bar{d} e o desvio padrão σ . Seja $d(u, v)$ a distância entre os vértices u e v no grafo G , a distância média do grafo G é dada por

$$\bar{d} = \sum_{\forall u, v} \frac{d(u, v)}{P}, \quad (3.1)$$

onde $u, v \in V$ e P representa o número de distâncias em G .

Uma métrica relacionada à distância é a excentricidade. A excentricidade $e(v)$ de um vértice v é a distância para o vértice mais longe de v : $e(v) = \max\{d(u, v) | u, v \in V\}$. A

excentricidade é uma medida do desvio ou afastamento do centro, ou seja, indica quão centrais ou periféricos são os vértices. Na Figura 3.1 é apresentado um grafo simples, com os valores das excentricidades para cada vértice do grafo.

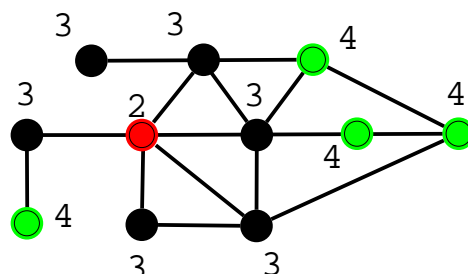


Figura 3.1: Excentricidades de um grafo.

A partir da excentricidade pode-se definir a centralidade de um grafo. O raio $R(G)$ de um grafo G corresponde à excentricidade mínima de seus vértices: $R(G) = \min\{e(v)|v \in V\}$. O diâmetro $D(G)$ de um grafo G corresponde à excentricidade máxima de seus vértices: $D(G) = \max\{e(v)|v \in V\}$. Além disso, um vértice é centro se $e(v) = R(G)$. O centro de G é o conjunto de vértices de menor excentricidade. Um vértice é periférico se $e(v) = D(G)$. A periferia de G é o conjunto de vértices do grafo com a maior excentricidade. No exemplo da Figura 3.1, o vértice central desenhado na cor vermelha (raio com valor igual a 2) e os vértices periféricos com cor verde (diâmetro com valor igual a 4).

Muitos aspectos do desempenho de protocolos e aplicações dependem da distribuição da distância [28, 30]. Em particular, distâncias pequenas e pouco variáveis são desejáveis em algoritmos de roteamento, porém, tornam a rede mais vulnerável a ataques baseados em propagação de agentes. A distância também caracteriza o denominado efeito “mundo pequeno” (*small world*), que é o fato de que grande parte dos pares de vértices na rede estão conectados por caminhos bem curtos.

3.2.2 Distribuição dos Graus dos Vértices

Seja uma rede representada por um grafo $G(V, A)$ onde V é o conjunto de vértices e A é o conjunto de arestas. As duas mais importantes características são o número de vértices n e o número de arestas m . O grau médio do grafo G é dado por $\bar{k} = 2m/n$, e é uma

métrica utilizada para caracterizar um grafo G .

A distribuição dos graus dos vértices (DGV) é a característica mais estudada e utilizada das topologias nas caracterizações. Seja $n(k)$ o número de vértices com grau k , a DGV do grafo G é a probabilidade de que um vértice aleatoriamente selecionado seja de grau k , isto é, $P(k) = n(k)/n$.

A observação de que as DGV seguem leis de potência ou distribuições de cauda pesada [19] teve um importante significado na pesquisa em topologias de rede, redirecionando os estudos nessa área de pesquisa. Apesar de esta métrica ser a mais estudada, redes com a mesma DGV podem apresentar estruturas topológicas muito diferentes, como foi observado em [25].

3.2.3 Conectividade: Cliques e Agrupamentos

Dois métricas de conectividade são analisadas neste trabalho, a distribuição das cliques e o coeficiente de agrupamento (*clustering*). Uma clique em um grafo $G(V, A)$ é um subgrafo de G que é completo, ou seja, o subgrafo é um grafo simples em que todo vértice é adjacente a todos os outros vértices [6].

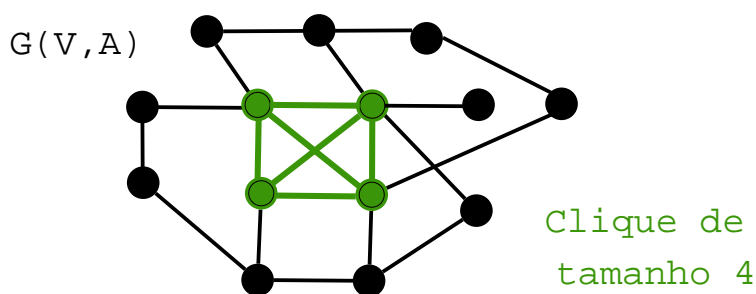


Figura 3.2: Exemplo de uma clique de tamanho 4 em um grafo G .

Na Figura 3.2 é apresentado o exemplo de uma clique, de tamanho 4, em um grafo G . Os vértices e arestas que formam o subgrafo completo, ou a clique, estão desenhados na cor verde. Por sua vez, as métricas de agrupamento fornecem informações sobre a interconexão dos vizinhos de um determinado vértice, representando a medida de quão próximos esses vizinhos estão para formar uma clique.

Há duas principais medidas estatísticas associadas ao agrupamento. A primeira é a *Local Clustering Coefficient* (CC_i), que consiste na razão entre o número de arestas existentes entre os vizinhos dos vértices i e o número máximo possível de arestas entre esses vizinhos. Seja V_i o conjunto de vértices vizinhos do vértice i , e μ_i o número de arestas entre os vértices em V_i . O CC_i para o vértice i é dado por:

$$CC_i = \frac{\mu_i}{(|V_i| * (|V_i| - 1)/2)}, \quad (3.2)$$

Na prática, o CC_i expressa a robustez local em um grafo. Quanto maior coeficiente de agrupamento de um vértice, mais interconectados são seus vizinhos, aumentando assim a diversidade de caminhos locais em torno desse vértice.

$$CC = \frac{1}{n} \sum_{i=1}^n CC_i \quad (3.3)$$

A segunda medida, o *Clustering Coefficient* (CC), fornece informação sobre a característica global do grafo, sendo calculada como a média do CC_i para cada vértice i do grafo [45]. O CC é empregado na caracterização de grafos *small world*.

3.3 Caracterização das Topologias Coletadas na Internet

Para realizar a análise e a caracterização topológica da Internet utilizou-se oito bases IRL coletadas semestralmente, referentes à topologia real da Internet, e descritas em [49]. Conforme apresentado no capítulo 3.1, estas bases agregam diversas técnicas de coleta tornando-se mais completas, além de serem atualizadas automática e diariamente.

3.3.1 Composição dos Grafos e Estatísticas dos Graus dos Vértices

As topologias analisadas são representadas por grafos, nos quais um vértice corresponde a um SA na Internet, e as arestas correspondem a conexões lógicas (caminhos) entre sistemas autônomos. A seguir será apresentada uma caracterização básica dos grafos, relativa ao número de vértices e arestas, componentes conexos e composição da borda do

grafo.

A Tabela 3.1 apresenta a data de coleta, a composição dos grafos que representam a topologia da Internet e estatísticas relacionadas aos graus dos vértices. As 8 topologias foram coletadas no décimo quinto dia de cada mês.

Os índices estatísticos dos graus dos vértices, média, mediana, maior grau e coeficiente de variação (COV) também são apresentados na Tabela 3.1. O coeficiente de variação expressa a variabilidade do conjunto de dados e é a divisão do desvio padrão pela média.

Na Figura 3.3 é apresentado o grau em função do número de vértices (esquerda) e a evolução da quantidade de arestas em função do número de vértices (direita). Pode-se notar que existem muitos vértices com grau baixo e poucos com grau alto, caracterizando a grande variabilidade nos graus dos vértices.

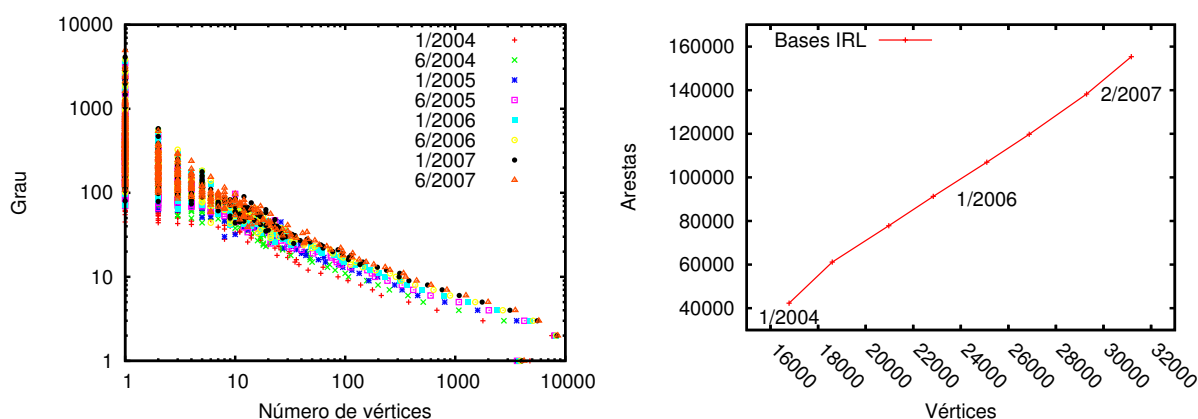


Figura 3.3: Número de vértices por grau (esquerda) e evolução das arestas em função do número de vértices (direita) nas topologias IRL.

Na Figura 3.4 é apresentada uma comparação da distribuição dos graus entre as topo-

Data	Vértices	Arestas	Mediana	Média	Maior	Grau 1 (%)	COV
1/2004	16782	42327	2	5,04	2466	28,5	7,12
6/2004	18597	61158	2	6,58	2887	20,3	6,38
1/2005	20971	77840	2	7,42	3228	17,5	6,22
6/2005	22843	91279	2	7,99	3480	16,0	6,10
1/2006	25092	106975	3	8,53	3724	15,1	6,017
6/2006	26876	119776	3	8,91	3886	14,6	5,96
1/2007	29287	138246	3	9,44	4135	13,8	5,90
6/2007	31170	155352	3	9,97	4969	13,5	5,99

Tabela 3.1: Composição dos grafos e estatísticas dos graus dos vértices nas topologias IRL.

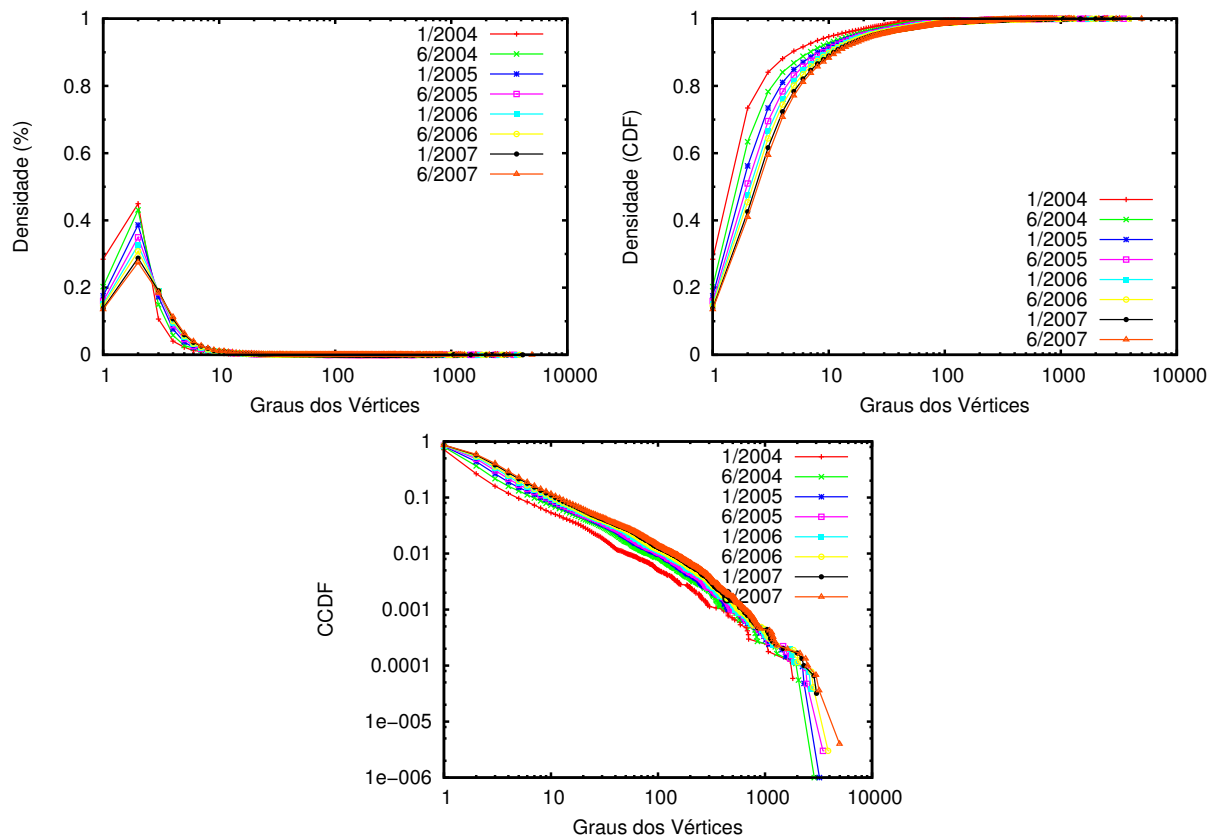


Figura 3.4: Distribuição dos graus nas topologias IRL.

gias analisadas. Nota-se que, entre as distribuições de cauda pesada, a porcentagem de vértices com grau 1 está diminuindo e 80% a 90% dos vértices, aproximadamente, têm grau abaixo de 10.

3.3.2 Componentes dos Grafos e Árvores nas Bordas

Um grafo pode ter mais de um componente conectado, isto é, subgrafos que não estão conectados entre si. O número de componentes conectados de cada grafo foi analisado com um algoritmo de pesquisa em profundidade. A análise revelou que todos os grafos são compostos por um único e grande componente conectado.

Outra análise realizada foi a pesquisa por árvores que poderiam estar nas bordas dos grafos. O objetivo desta pesquisa foi verificar se há ramificações longas em forma de árvores e a fração de vértices incluída nestas bordas. Para isso, eliminam-se as árvores da seguinte forma. Numa primeira etapa, são eliminados todos os vértices de grau 1 e as

Data	Maior	Média	Moda	COV
1/2004	9	3,67	4	0,15
6/2004	9	3,57	3	0,15
1/2005	9	3,53	2	0,13
6/2005	9	3,49	3	0,12
1/2006	9	3,48	3	0,11
6/2006	9	3,47	3	0,10
1/2007	10	3,46	3	0,09
6/2007	9	3,43	3	0,09

Tabela 3.2: Estatísticas das distâncias entre pares de vértices nas topologias IRL.

arestas correspondentes. Este procedimento possivelmente gera novos vértices de grau 1. A seguir, esse procedimento foi repetido até retirarmos todos os vértices de grau 1. Este grafo reduzido é denominado *2-core* e contém apenas vértices de grau dois ou superior.

Observa-se que, para todos os grafos, foram necessárias apenas três iterações para que todos os vértices de grau 1 fossem eliminados, o que demonstra que as árvores nas bordas dos grafos têm altura máxima igual a três. Além disso, a porcentagem de vértices nestas árvores varia entre 13% (IRL 6/2007) e 28% (IRL 1/2004), e a porcentagem de arestas retiradas está entre 3% (IRL 6/2007) e 11% (IRL 1/2004).

3.3.3 Distância entre Pares de Vértices

Conforme apresentado na seção 3.2.1, a distância é uma métrica essencial em um grafo e, em especial, em uma rede. Na Tabela 3.2 são apresentadas as estatísticas para a métrica de distância entre pares de vértices.

Podemos observar que os vértices estão a distâncias muito pequenas e pouco variáveis de todos os demais vértices. A distância máxima varia entre 9 e 10, a média está em torno de 3,5 e tem diminuído com o tempo. Isto significa que os vértices estão, em média, a uma distância entre 3 e 4 arestas de qualquer outro vértice. Os resultados são bastante homogêneos em todos os grafos, apresentando um comportamento previsível o que torna a distância uma métrica invariante na topologia dos SAs da Internet.

Na Figura 3.5 é apresentada frequência das distâncias para cada umas das topologias IRL da Internet. Observa-se que entre as duas distâncias com maior frequência, a distância 4 tem diminuído enquanto a distância 3 aumentado com o tempo.

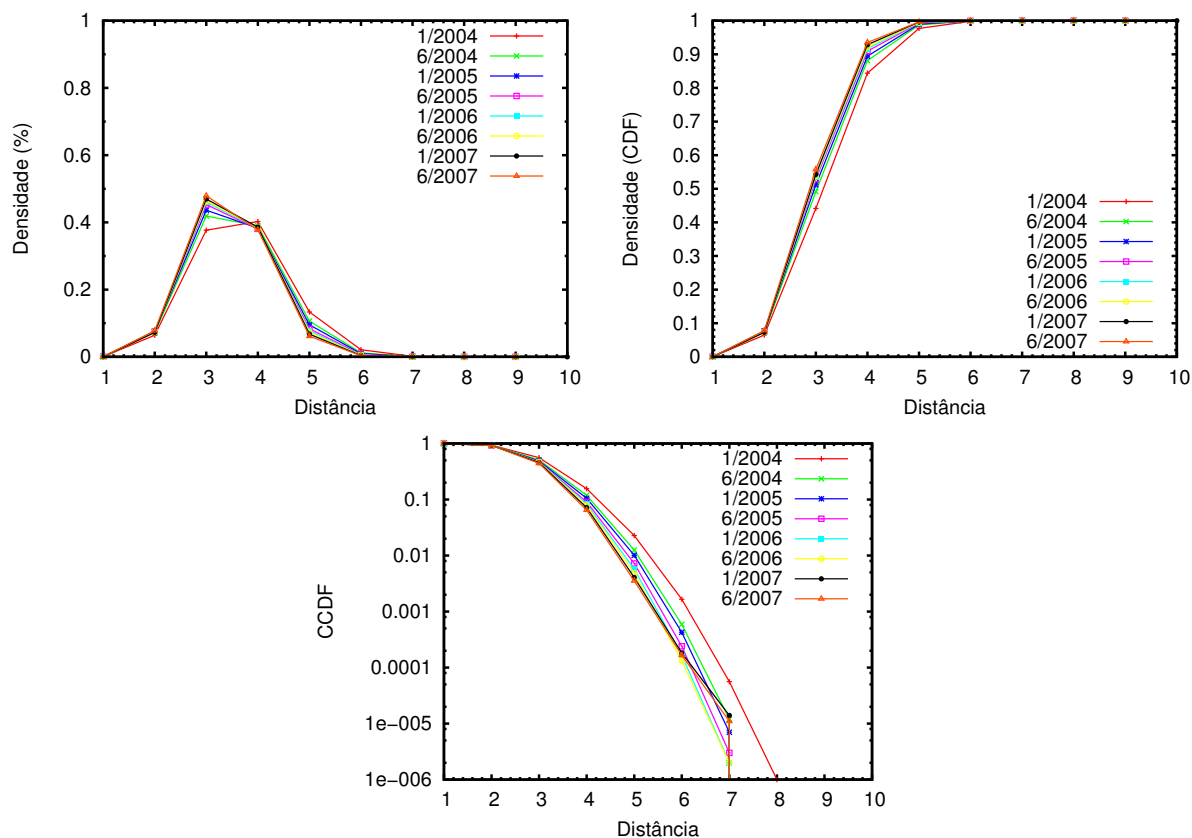


Figura 3.5: Distribuição das distâncias entre pares de vértices nas topologias IRL.

3.3.4 Excentricidade

Os resultados anteriores indicam que a distância é uma métrica característica da rede, apresentando comportamento bastante regular e invariante. Sendo a excentricidade de um vértice v a distância para o vértice mais longe de v (seção 3.2.1), a Tabela 3.3 apresenta estatísticas relativas à excentricidade das topologias IRL.

Data	Mínima	Moda	Média	Mediana	Máxima	COV
1/2004	5	6	6,52	6	9	0,72
6/2004	5	6	6,35	6	9	0,08
1/2005	5	6	6,29	6	9	0,08
6/2005	5	6	6,23	6	9	0,07
1/2006	5	6	6,20	6	9	0,08
6/2006	5	6	6,21	6	9	0,08
1/2007	5	7	6,98	7	10	0,08
6/2007	5	7	6,76	7	9	0,08

Tabela 3.3: Estatísticas das excentricidades nas topologias IRL.

Os vértices mais longínquos estão, em média, à distância 6, e, no máximo, à distância

9 ou 10. Estes resultados indicam que a rede é caracterizada por caminhos curtos e é bastante compacta. O coeficiente de variação (COV) mantém-se praticamente semelhante em todas as bases.

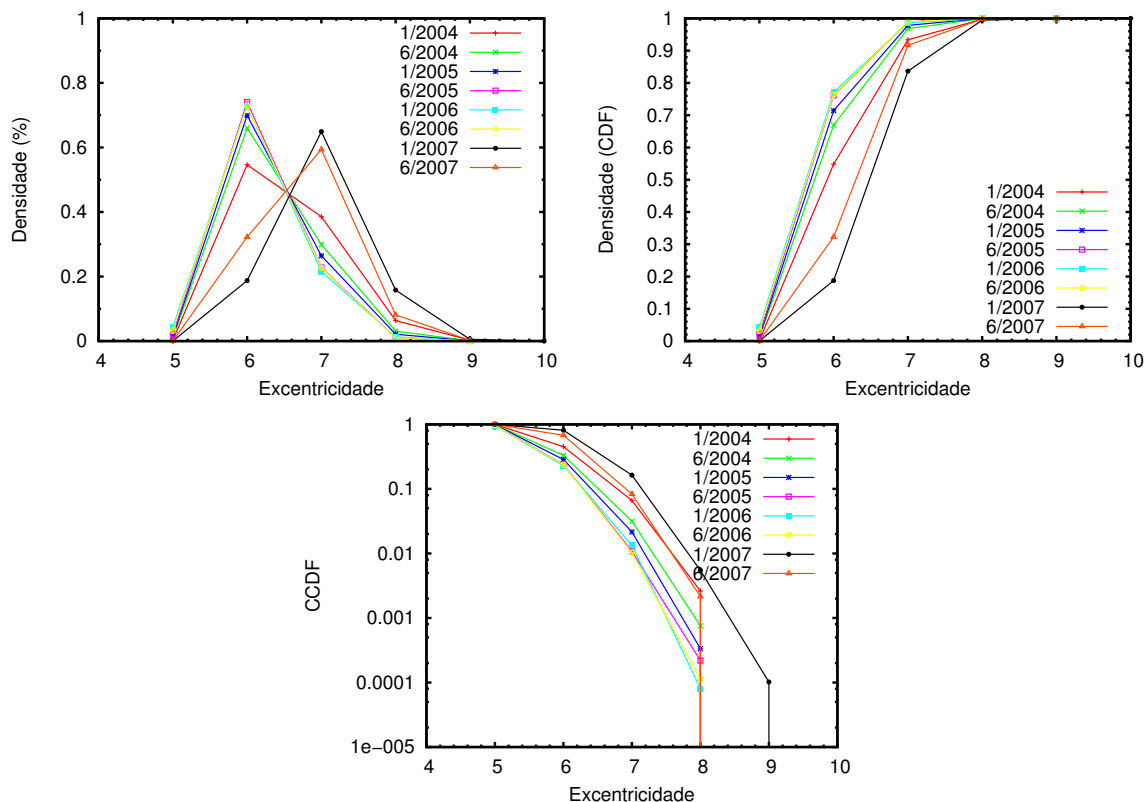


Figura 3.6: Distribuição das excentricidades nas topologias IRL.

Na Figura 3.6 é apresentada a distribuição das excentricidades para as bases IRL. Pode-se notar que a maior fração das excentricidades dos vértices está entre 6 e 7 respectivamente, e a partir de 2007, a excentricidade com maior frequência passou a ser 7.

De modo geral, percebe-se dois grupos de distribuições semelhantes, um entre 6/2004 e 6/2006 e outro referente a coleta no ano de 2007. A topologia IRL 1/2004 é bastante distinta das demais, possivelmente por ser a primeira que foi coletada e que, com o tempo, o método de coleta das bases IRL tem sido cada vez mais aprimorado.

A Figura 3.7 exibe o grau médio dos vértices em função das excentricidades para todas as topologias analisadas. Observa-se que com o tempo o grau médio tem aumentado e os vértices de maior grau são mais centrais, isto é, apresentam menor excentricidade,

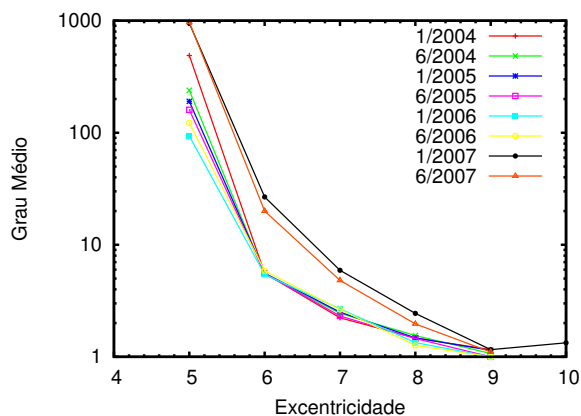


Figura 3.7: Grau médio em função da excentricidade nas topologias IRL.

Data	Raio	Diâmetro	Centro	Periferia
1/2004	5	9	31	44
6/2004	5	9	170	14
1/2005	5	9	311	7
6/2005	5	9	474	5
1/2006	5	9	1067	2
6/2006	5	9	886	3
1/2007	5	10	6	3
6/2007	5	9	16	67

Tabela 3.4: Raio, diâmetro, centro e periferia nas topologias IRL.

enquanto que os vértices periféricos têm menor grau.

De acordo com a seção 3.2.1, a partir da excentricidade podemos caracterizar o raio, o diâmetro, os vértices centrais e periféricos de um grafo. A Tabela 3.4 demonstra o resultado destas métricas para as topologias IRL analisadas.

Em todas as bases, o raio é 5 e esta é a distância máxima entre os vértices do centro do grafo e os demais vértices. O diâmetro varia entre 9 e 10, e indica a maior distância entre vértices periféricos. A Tabela 3.4 apresenta também o número de vértices no centro e na periferia de cada grafo. Observa-se que o centro e a periferia são compostos por uma pequena fração dos vértices do grafo. Grande parte dos vértices, mais de 90%, está entre o centro e a periferia. Estes resultados indicam que as métricas raio e diâmetro são também pouco variáveis e característicos da rede.

3.3.5 Conectividade: Cliques e Agrupamentos

Outras métricas bastante utilizadas são as relacionadas a conectividade da rede. Nesse sentido analisamos as topologias IRL por meio da formação de cliques e do coeficiente de agrupamento, conforme apresentado na seção 3.2.3. Na Tabela 3.5 são apresentadas as estatísticas das cliques referentes as bases IRL. Nota-se que a clique média tem tamanho entre 30 e 33, aproximadamente, e as cliques máximas têm tamanho 50.

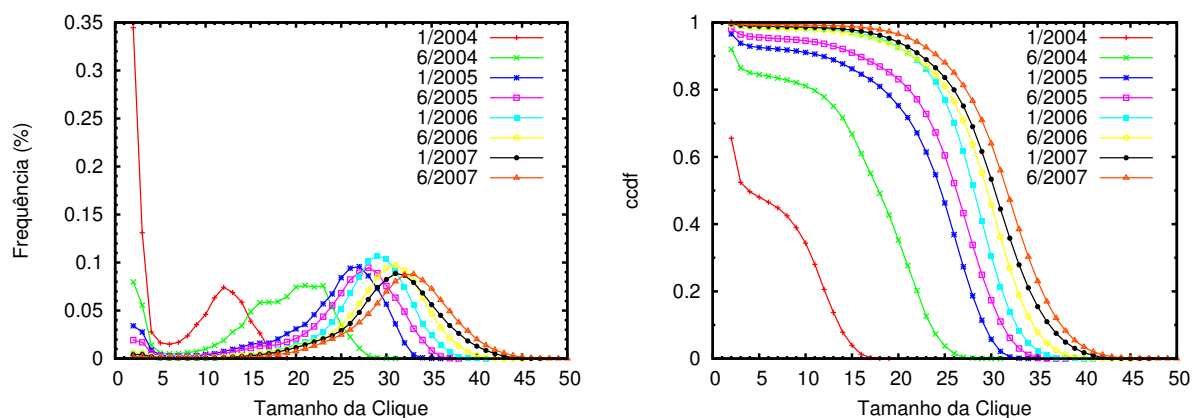


Figura 3.8: Distribuição das cliques (esquerda) e cauda da distribuição das cliques (direita) nas topologias IRL.

A Figura 3.8 apresenta a ocorrência de cliques nas topologias IRL avaliadas ao longo do período. Os resultados demonstram um padrão bastante definido principalmente para as bases mais atuais. A distribuição das cliques na topologia real tem evoluído para uma distribuição de Gauss (normal) com tamanho médio das cliques cada vez maiores (atualmente entre 30 e 35). A algoritmo utilizado para analisar e contabilizar os cliques é descrito em [6].

Data	Média	Mediana	Moda	Máxima	Número de Cliques Máximas
1/2004	6,98	4	2	19	25
6/2004	16,49	18	2	31	8
1/2005	22,84	25	27	36	6
6/2005	25,09	27	28	38	64
1/2006	27,84	29	29	41	28
6/2006	29,02	30	31	45	12
1/2007	30,17	31	31	47	64
6/2007	31,62	32	33	50	5

Tabela 3.5: Estatísticas das cliques nas topologias IRL.

Analisando o coeficiente de agrupamento ou *Clustering Coefficient* (CC) global das topologias IRL, obtivemos os valores apresentados na Figura 3.9 (esquerda). Nota-se que o comportamento CC global das topologias é previsível e tem aumentado lentamente com o tempo.

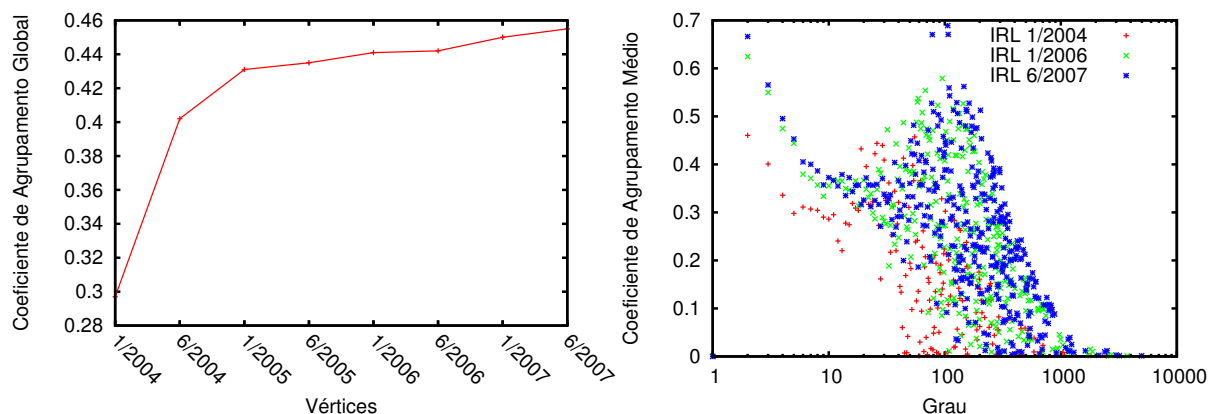


Figura 3.9: Coeficiente de agrupamento global (esquerda) e local (direita) nas topologias IRL.

Na Figura 3.9 (direita) é apresentado o coeficiente de agrupamento local médio em função do grau para as topologias IRL mais antiga, intermediária e recente. Observa-se que, tal como nas 3 topologias apresentadas, em todas as outras analisadas os vértices com graus baixos (menor ou igual a 10) possuem um CC alto que tende a diminuir em função do grau. Outro resultado interessante é que o CC médio de vértices com grau em torno de 100 aumenta em todas as bases, padrão percebido em todas as 8 topologias.

3.4 Caracterização das Topologias Sintéticas do Inet 3.0

Diversas análises e experimentos foram realizados com os geradores BRITE [32], R-MAT [10], Orbis [29], GLP [7] e Inet 3.0 [47], com o objetivo de identificar os quais poderiam reproduzir melhor as características dos sistemas autônomos da Internet.

O resultados do gerador BRITE [32] não são apresentados porque já haviam sido comparados ao Inet 2.2 [22], com melhores resultados para o Inet 2.2. Além disso, o BRITE requer a configuração de um conjunto grande de parâmetros, o que amplia as possibilidades de geração de topologias tornando-o mais genérico, porém dificulta o ajuste

dos parâmetros para uma topologia específica.

O gerador R-MAT [10] 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, por meio da implementação *GT-Graph* [4], indicam que os grafos gerados são desconexos, com arestas duplicadas e *self-loops*, características que não ocorrem no grafo alvo desse trabalho.

Dentre os geradores analisados, o gerador Orbis [29] é o mais recente e, no início dessa pesquisa, não estava disponível para testes. Segundo os autores, os resultados do Orbis caracterizam muitas das métricas propostas na literatura [29], contudo, além de outras análises estarem sendo realizadas com esse gerador [26], uma desvantagem para o usuário é que o Orbis necessita de uma topologia de referência como parâmetro para gerar as topologias, podendo ou não ter diferentes escalas. Em trabalhos futuros pode ser analisado no contexto deste trabalho.

Por fim, analisamos os geradores GLP [7] e Inet 3.0 (Inet) [47], que são desenvolvidos especificamente para representar a topologia dos sistemas autônomos da Internet. 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 autores¹, 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. Outra necessidade é a de calcular os parâmetros de entrada do GLP, além do número de vértices e porcentagem de grau 1, de acordo com as topologias IRL da Internet para possibilitar comparações posteriores. Um exemplo é o valor empírico dado para o *backbone* inicial criado no GLP [7].

O Inet [47] 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 [22]. Além disso, o gerador Inet 3.0 é utilizado

¹Comunicação pessoal.

Data de Referência	Vértices	Arestas	Mediana	Média	Maior	Grau 1 (%)	COV
Inet 1/2004	16782	41350	2	4,93	2739	28,5	8,48
Inet 6/2004	18597	48478	2	5,21	2977	20,2	8,83
Inet 1/2005	20971	57770	2	5,51	3429	17,5	9,44
Inet 6/2005	22843	65548	2	5,74	4053	16,0	9,96
Inet 1/2006	25092	75381	2	6,01	4889	15,1	10,62
Inet 6/2006	26876	83623	2	6,22	5623	14,5	11,16
Inet 1/2007	29287	95452	2	6,52	6722	13,7	11,90
Inet 6/2007	31170	105194	2	6,75	7666	13,5	12,50

Tabela 3.6: Composição dos grafos e estatísticas dos graus dos vértices nas topologias Inet.

em inúmeros estudos, pois é implementado no simulador NS-2². Desse modo, a seguir são apresentadas as análises de topologias sintéticas do gerador Inet e comparações com algumas das respectivas bases IRL.

3.4.1 Composição dos Grafos e Estatísticas dos Graus dos Vértices

Para gerar as topologias sintéticas e comparar com as bases IRL, definimos os parâmetros de entrada do gerador Inet em função de cada respectiva base IRL. No Inet esses parâmetros são somente o número de vértices (n) e a porcentagem de vértices com grau 1 (n_1).

Na Tabela 3.6 é apresentada a composição e estatísticas dos graus dos vértices das topologias sintéticas do Inet.

Na Figura 3.10 (esquerda) é apresentada o grau em função do número de vértices para as topologias sintéticas. Na Figura 3.10 (direita) demonstra-se a comparação da quantidade de arestas em função do número de vértices entre as topologias sintéticas e IRL. Pode-se notar que, principalmente para as topologias mais recentes, há um surgimento maior de arestas na topologia IRL que não é modelado satisfatoriamente pelo gerador o Inet 3.0.

Com o intuito de apresentar os resultados das topologias sintéticas, e realizar comparações com as topologias IRL, foram selecionadas 3 topologias sintéticas e suas equivalentes IRL para apresentação dos resultados.

Na Figura 3.4 é demonstrada a comparação da distribuição dos graus entre as topolo-

²<http://www.isi.edu/nsnam/ns/ns-topogen.html>

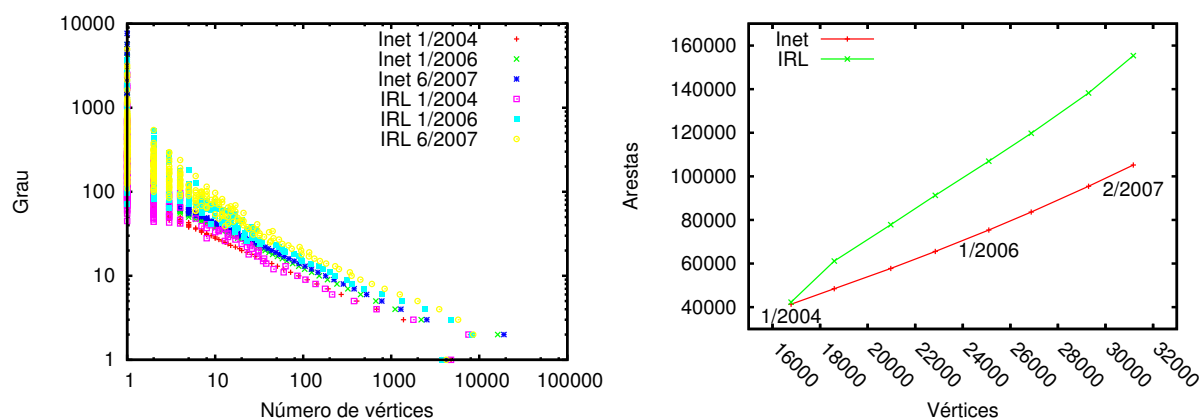


Figura 3.10: Número de vértices por grau (esquerda) e evolução das arestas em função do número de vértices (direita) nas topologias Inet e IRL.

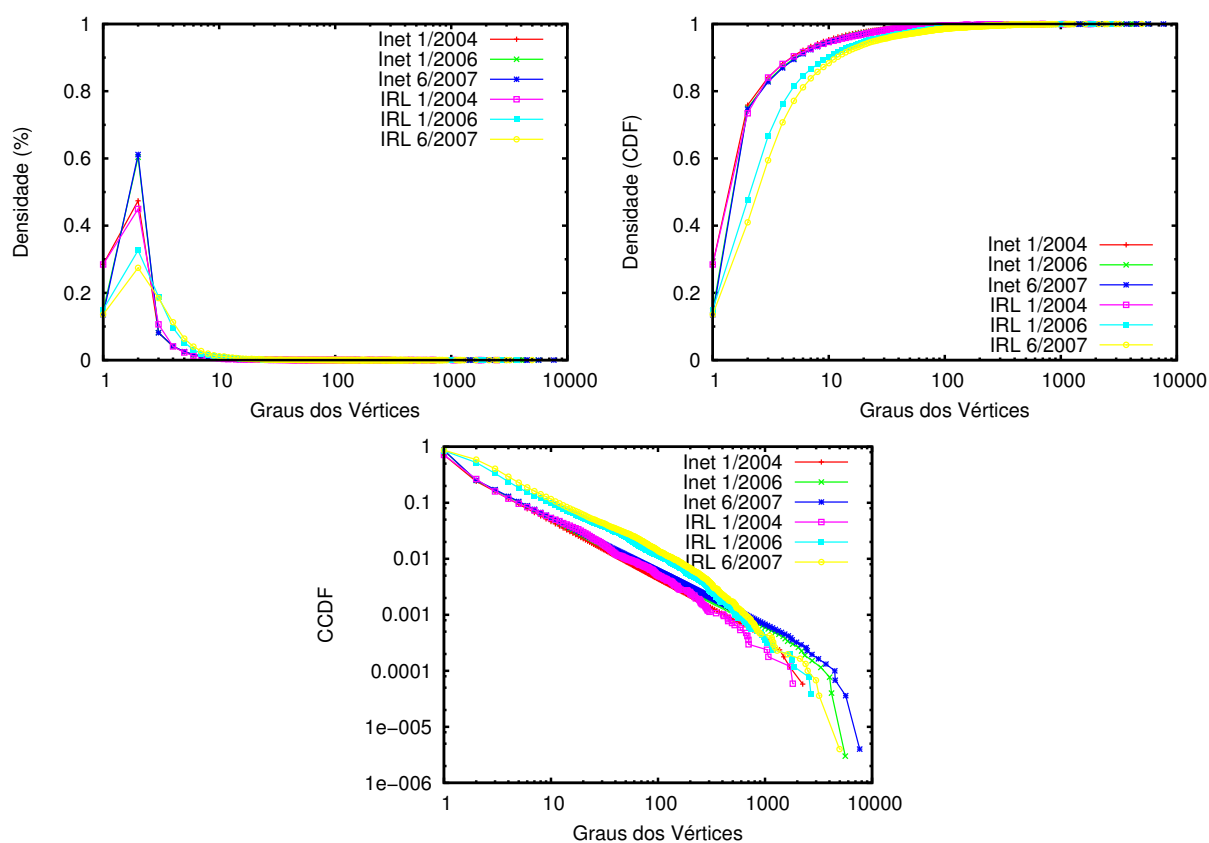


Figura 3.11: Distribuição dos graus nas topologias Inet e IRL.

gias sintéticas e IRL. Nota-se que a porcentagem de vértices com grau 2 das bases IRL é menor do que aquelas obtidas nas topologias sintéticas do Inet, e as duas topologias IRL mais recentes apresentam um padrão distinto das demais topologias sintéticas. A topologia IRL 1/2004 apresenta, praticamente, o mesmo comportamento das topologias

geradas com o Inet.

A partir desses resultados, pode-se inferir que a distribuição dos graus tem apresentado comportamento distinto daquele gerado no Inet, tendo maior frequência os graus abaixo 200, aproximadamente. Para graus acima desse valor, a cauda da distribuição tende a cair mais rapidamente, conforme o gráfico CCDF da Figura 3.4.

3.4.2 Componentes dos Grafos e Árvores nas Bordas

A análise revelou que todos os grafos são compostos por um único e grande componente conectado. Foi realizada também a pesquisa por árvores que poderiam estar nas bordas dos grafos, verificando se existem ramificações longas em forma de árvores e a fração de vértices incluída nestas bordas.

Observa-se que, para todos as topologias sintéticas, foram necessárias duas ou três iterações para que todos os vértices de grau 1 fossem eliminados, o que demonstra que as árvores nas bordas dos grafos têm altura máxima igual a três, tal como nas bases IRL. Além disso, a porcentagem de vértices nestas árvores varia aleatoriamente entre 13% e 29%, e a porcentagem de arestas retiradas está entre 5% e 12%.

3.4.3 Distância entre Pares de Vértices

As estatísticas das métricas de distâncias relativas às topologias sintéticas são apresentadas na Tabela 3.7.

Data	Maior	Média	Moda	COV
Inet 1/2004	8	3,17	3	0,13
Inet 6/2004	7	3,10	3	0,11
Inet 1/2005	8	3,04	3	0,10
Inet 6/2005	8	3,00	3	0,09
Inet 1/2006	8	2,95	3	0,08
Inet 6/2006	7	2,93	3	0,08
Inet 1/2007	8	2,89	3	0,07
Inet 6/2007	7	2,86	3	0,07

Tabela 3.7: Estatísticas das distâncias nas topologias Inet.

A distância máxima varia entre 7 e 8 com média em torno de 3. Isto significa que os vértices estão, em média, a uma distância entre 3 arestas de qualquer outro vértice. Os

resultados são bastante homogêneos em todos os grafos, apresentando um comportamento previsível do Inet. Nota-se que a distância máxima, assim como a média, são menores do que os valores obtidos com as topologias IRL da Internet.

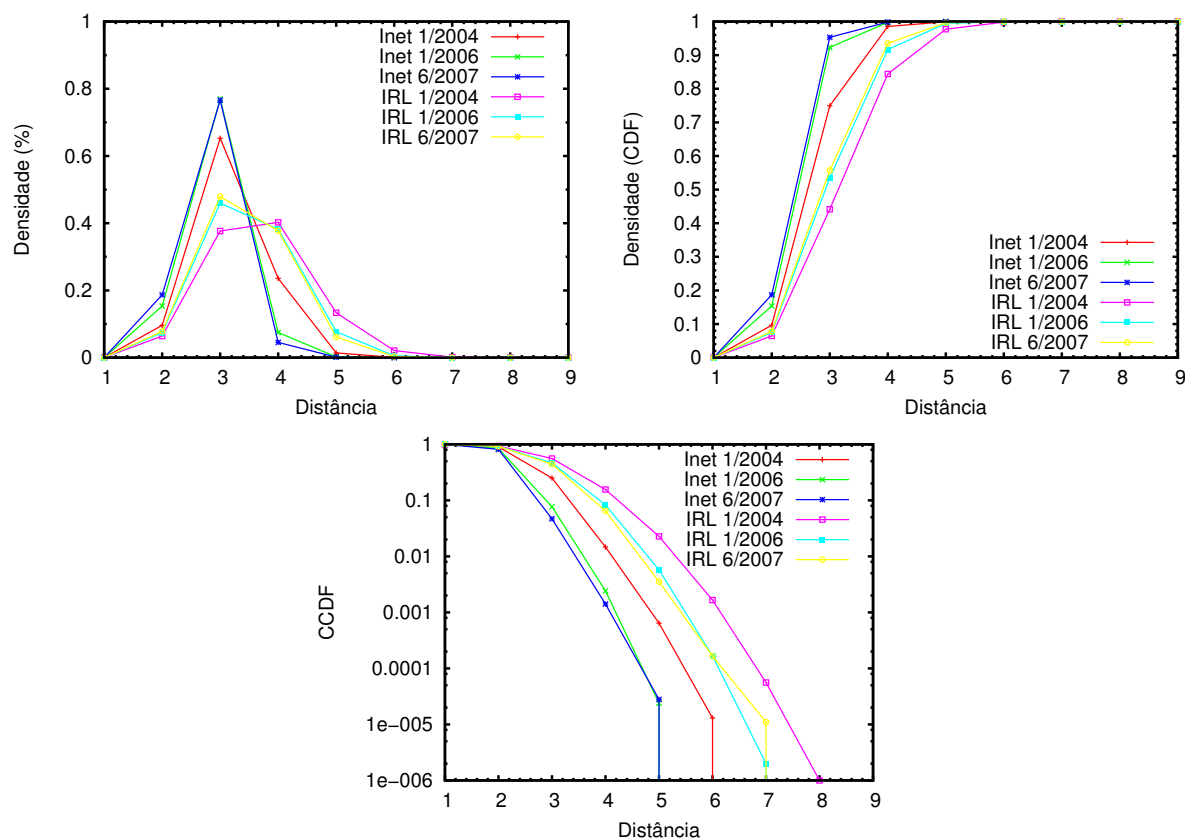


Figura 3.12: Distribuição das distâncias entre pares de vértices nas topologias Inet e IRL.

Na Figura 3.12 é apresentada a comparação da frequência das distâncias entre as topologias Inet e IRL. Observa-se que entre os dois conjuntos de topologias, existem diferenças na distribuição das distâncias, ou seja, o Inet gera uma fração maior de distâncias 3 do que existem nas topologias IRL, as quais têm maiores frações de distâncias 3 e 4, respectivamente.

3.4.4 Excentricidade

A Tabela 3.8 apresenta as estatísticas relativas às excentricidades nas topologias IRL. Os vértices mais longínquos estão, em média, à distância 5 e, no máximo, à distância 7 ou 8.

Na Figura 3.13 é apresentada a distribuição das excentricidades nas topologias sintéticas

Data	Mínima	Moda	Média	Mediana	Máxima	COV
Inet 1/2004	4	6	5,74	6	8	0,08
Inet 6/2004	4	5	5,37	5	7	0,09
Inet 1/2005	4	6	5,80	6	8	0,07
Inet 6/2005	4	5	5,28	5	8	0,08
Inet 1/2006	4	5	5,24	5	8	0,08
Inet 6/2006	4	5	5,07	5	7	0,05
Inet 1/2007	4	5	5,32	5	8	0,08
Inet 6/2007	4	5	5,28	5	7	0,08

Tabela 3.8: Estatísticas das excentricidades nas topologias Inet.

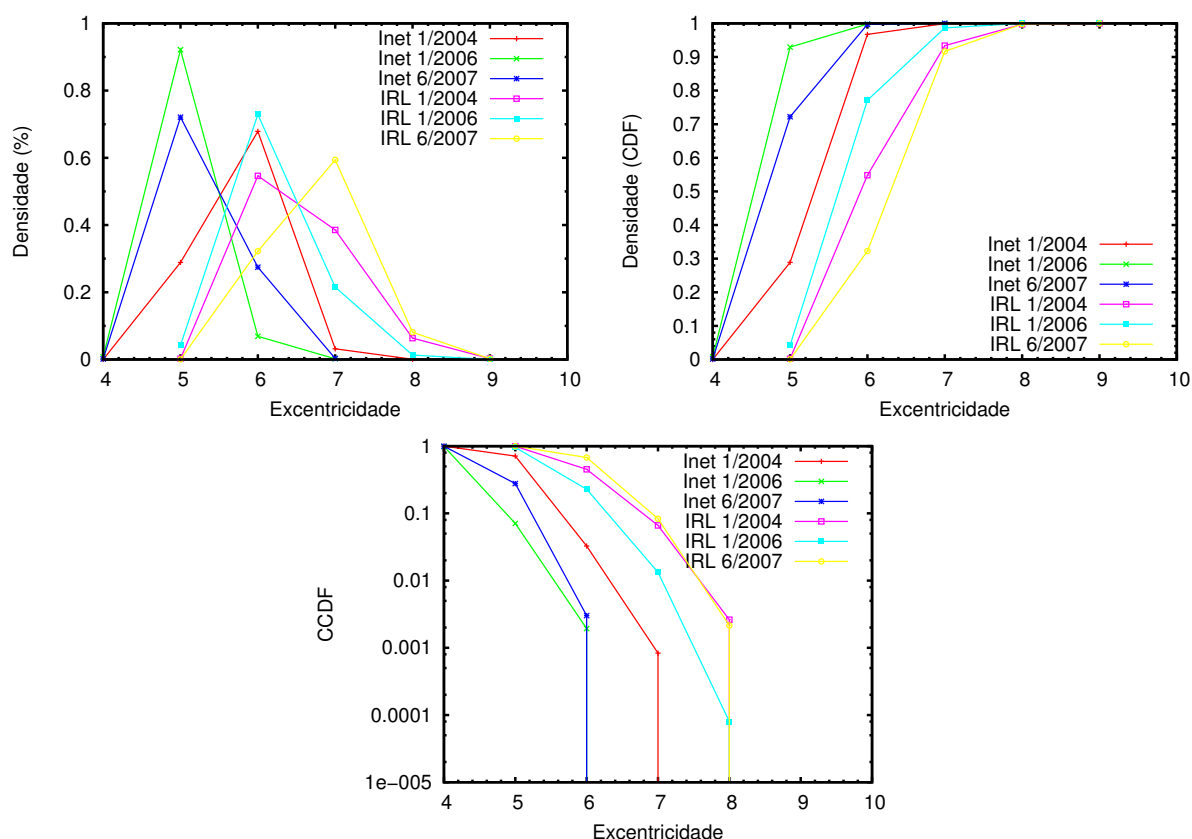


Figura 3.13: Distribuição das excentricidades nas topologias Inet e IRL.

e nas respectivas bases IRL. Percebe-se um comportamento não uniforme, tanto nas topologias sintéticas, quanto nas topologias IRL. A maior fração das excentricidades dos vértices está entre 5 e 6 para a maioria das topologias, excetuando-se para a topologia mais recente (IRL 6/2007) que possui valor 7.

A Figura 3.14 exhibe o grau médio dos vértices em função das excentricidades para todas as topologias analisadas. Observa-se que nas topologias sintéticas do Inet, o grau médio tem menor do que o encontrado para as topologias IRL. Possivelmente isso se deve

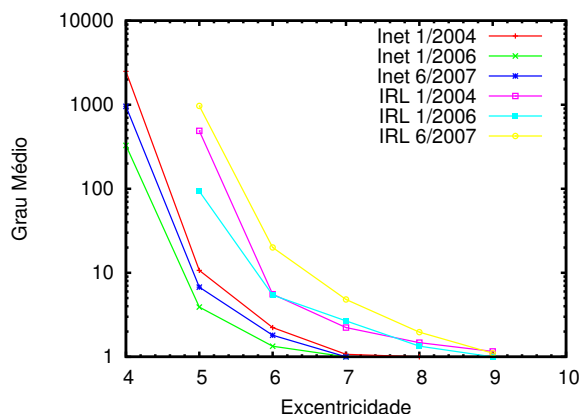


Figura 3.14: Grau médio em função das excentricidades nas topologias Inet e IRL.

Data	Raio	Diâmetro	Centro	Periferia
Inet 1/2004	4	8	2	14
Inet 6/2004	4	7	11	129
Inet 1/2005	4	8	1	12
Inet 6/2005	4	8	24	2
Inet 1/2006	4	8	25	2
Inet 6/2006	4	7	207	52
Inet 1/2007	4	8	16	2
Inet 6/2007	4	7	45	94

Tabela 3.9: Raio, diâmetro, centro e periferia nas topologias Inet.

a dois fatores: o primeiro é o modo como estão conectados os vértices e o segundo, é a maior quantidade de arestas presentes nas topologia real da Internet.

A Tabela 3.9 demonstra o raio, diâmetro, número de vértices centrais e periféricos para as topologias do Inet analisadas. O raio é 4 e esta é a distância máxima entre os vértices do centro do grafo e os demais vértices. O diâmetro varia entre 7 e 8, valor menor do que os obtidos nas topologias IRL (9 e 10).

A Tabela 3.9 apresenta também o número de vértices no centro e na periferia de cada grafo. Observa-se que o centro e a periferia são compostos por uma pequena fração dos vértices do grafo. Assim como nas bases IRL, grande parte dos vértices, mais de 90%, está entre o centro e a periferia.

3.4.5 Conectividade: Cliques e Agrupamentos

Outro conjunto de métricas bastante utilizadas são as relacionadas a conectividade da rede. Nesse sentido, analisamos as topologias sintéticas por meio das cliques e do coeficiente de agrupamento (CC). Na Tabela 3.10 é apresentada a estatística das cliques referentes as bases IRL. Nota-se que a clique média tem tamanho atual em torno de 5, aproximadamente, e as cliques máximas têm tamanho 22. Estes valores são bem abaixo daqueles encontrados nas bases IRL.

Data	Média	Mediana	Moda	Máxima	Número de Cliques Máximas
Inet 1/2004	2,74	2	2	10	22
Inet 6/2004	2,87	2	2	11	2
Inet 1/2005	3,14	3	2	13	2
Inet 6/2005	3,45	3	2	15	4
Inet 1/2006	4,11	3	2	14	74
Inet 6/2006	4,59	3	2	17	8
Inet 1/2007	5,75	4	3	20	2
Inet 6/2007	6,87	6	3	22	32

Tabela 3.10: Estatísticas das cliques nas topologias Inet.

Na Figura 3.15 (esquerda) é apresentada a ocorrência de cliques das topologias sintéticas do Inet. A Figura 3.15 (direita) apresenta a ocorrência de cliques nas topologias reais avaliadas ao longo do período, bem como a mesma medida para as topologias sintéticas equivalentes.

Nota-se uma grande fração de cliques com tamanhos pequenos para as topologias sintéticas, enquanto que as maiores frações de cliques encontradas nas topologias reais têm tamanhos 30 a 35, aproximadamente. A topologia IRL 1/2004 ainda apresentou o mesmo padrão das cliques geradas no Inet, porém, para as demais comparações os resultados demonstram padrões bastante distintos entre as topologias IRL e sintéticas. A algoritmo utilizado para analisar e contabilizar os cliques é descrito em [6].

Outra análise realizada é relativa à evolução do tamanho das clique máxima nas topologias reais e sintéticas do Inet. Conforme já observado em [47], o Inet apresenta um tamanho menor de clique máxima menor do que a topologia real. Esse fato ainda acontece para as topologias reais, de acordo com o gráfico da evolução do tamanho da clique máxima apresentado na Figura 3.16 (direita).

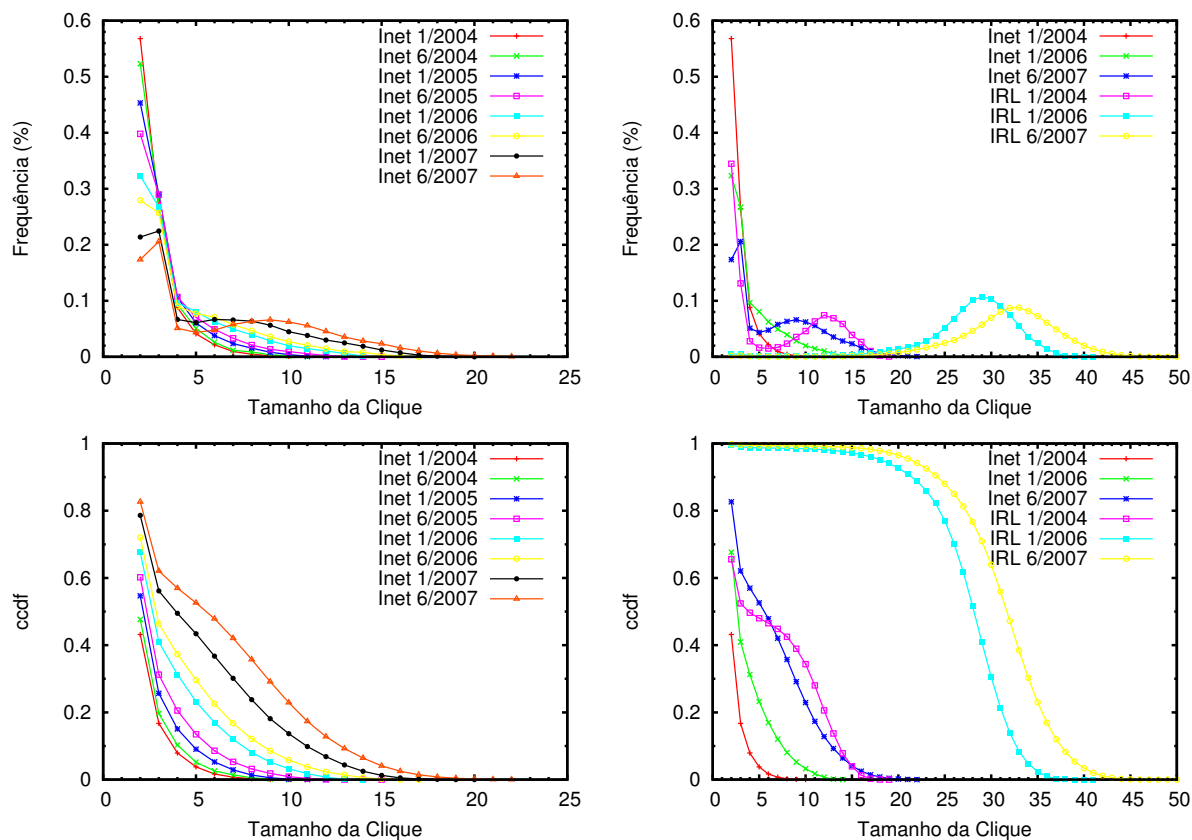


Figura 3.15: Distribuição das cliques nas topologias Inet (esquerda) e comparação com as topologias IRL (direita).

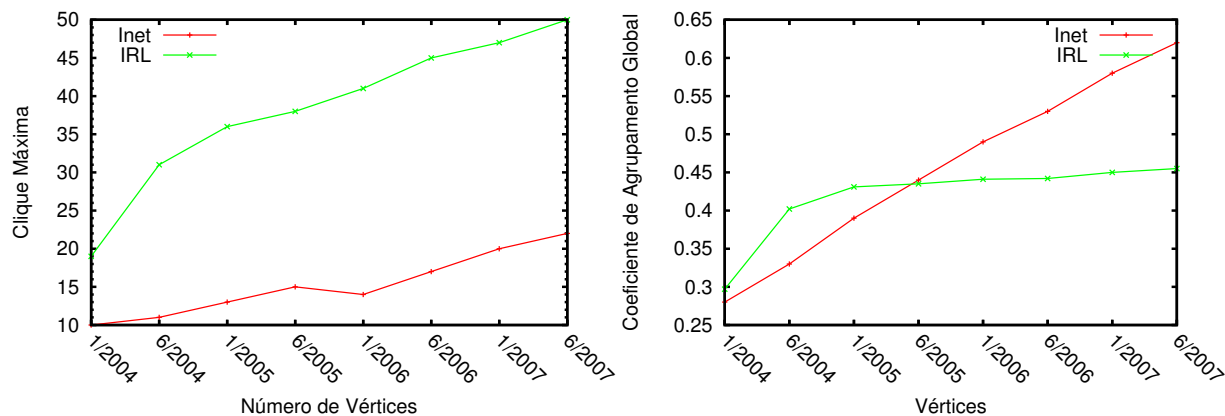


Figura 3.16: Evolução da clique máxima (direita) e coeficiente de agrupamento global (direita) nas topologias Inet e IRL.

Na análise comparativa do coeficiente de agrupamento global das topologias sintéticas com as bases IRL, conforme Figura 3.16 (esquerda), percebe-se um comportamento bastante distinto.

Os CC globais são bem diferentes entre as topologias sintéticas e IRL, tendo o Inet

apresentado um padrão distinto daquele revelado em [47]. O maior número de vértices das topologias atuais, aliado ao modo de conexão desses vértices e, principalmente, a diminuição da fração de vértices com grau 1, que possuem CC igual a 0, podem ser fatores que contribuem para o aumento maior do CC nas topologias sintéticas.

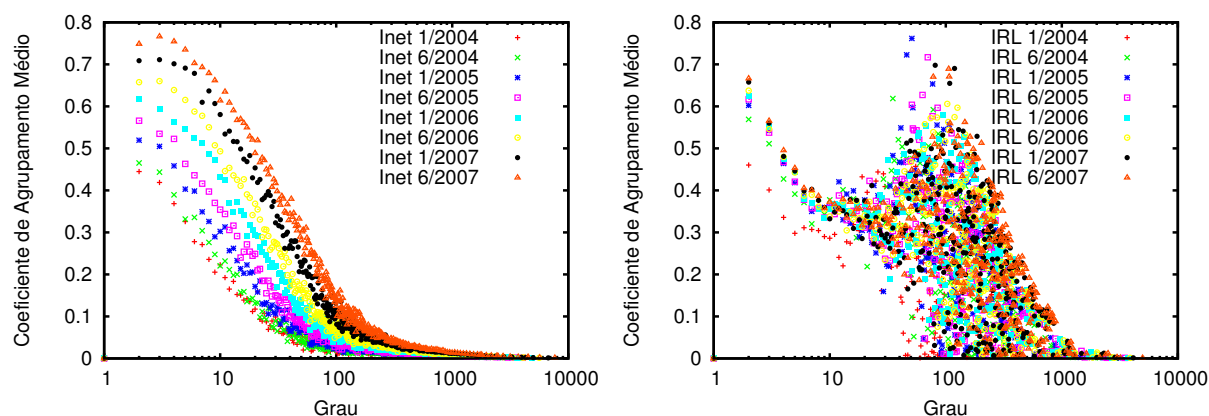


Figura 3.17: Coeficiente de agrupamento médio em função do grau nas topologias Inet e IRL.

Na Figura 3.17 (esquerda) é apresentado o coeficiente de agrupamento local médio em função do grau para as topologias sintéticas do Inet, que pode ser comparada com as topologias IRL Figura 3.17 (direita). Observa-se que, o CC das sintéticas apresenta um comportamento uniforme, enquanto que a IRL possui maior variabilidade. Nas topologias sintéticas o CC tende a diminuir em função do aumento do grau, fato que não ocorre do mesmo modo nas topologias IRL, pois existem vértices com graus próximos a 100 também com alto CC .

3.5 Síntese

Nesse capítulo foram apresentadas as principais métricas para análise de topologias da Internet no contexto deste trabalho (seção 3.2). Posteriormente, foram avaliadas e comparadas as topologias do gerador Inet e bases IRL por meio das métricas descritas.

Nas análises efetuadas nas seções 3.3 e 3.4, observa-se que a distribuição dos graus das bases IRL tem exibido características distintas em relação ao gerador Inet, principalmente para as topologias mais recentes. Os resultados demonstram que o Inet não modela

adequadamente as métricas de distâncias e derivadas, bem como a distribuição das cliques e o coeficiente de agrupamento.

Outra observação relevante é o fato da topologia mais antiga da Internet, IRL 1/2004, apresentar resultados com padrões distintos das demais topologias IRL, conforme pode ser observado nos gráficos durante todas as análises. Uma explicação para este resultado é a possível evolução dos algoritmos de coleta de dados da topologia ao longo do período, com alteração significativa no início do período de coleta. Outra possível explicação é a própria evolução da topologia, que pode ter apresentado um momento de mudança real e acentuado neste período.

Diante desse contexto, para abordar esse problema e buscar uma melhor solução para modelagem da topologia da Internet, no próximo capítulo é apresentada a descrição de um algoritmo baseado no código fonte do gerador Inet 3.0, no qual as fases foram alteradas para melhorar a modelagem nos resultados relativos as métricas analisadas.

CAPÍTULO 4

DESCRIÇÃO DO ALGORITMO DO GERADOR NIT

No capítulo anterior foram analisadas as caracterizações das topologias reais e sintéticas relativas à Internet, bem como a definição das métricas empregadas nessas análises. Observou-se que, para essas métricas, as topologias sintéticas do Inet apresentaram resultados que diferem da topologia da Internet, a qual é representada pelas bases IRL.

Desse modo, nesta seção será apresentado o algoritmo de geração de topologias em nível de SA do gerador denominado NIT, o qual é baseado no gerador Inet. O objetivo do algoritmo NIT é construir topologias com características mais semelhantes à da Internet em relação às métricas de distância e suas derivadas, apresentadas na seção 3.2.1, considerando também a distribuição dos graus dos vértices e as métricas relacionadas à conectividade.

Diversas equações e parâmetros utilizados nas fases do algoritmo NIT, detalhadas nas próximas seções, foram obtidos por meio do método de ajustes de curvas. A avaliação das equações obtidas com os pontos reais foi realizada por meio do Coeficiente de Determinação (R^2), que é o quadrado do Coeficiente de Correlação de Pearson [20]. O R^2 é uma medida da proporção da variabilidade em uma variável que é explicada pela variabilidade da outra. Ou seja, quando R^2 é igual a 1, todos os pontos da equação ajustada são coincidentes com os pontos reais, e conforme esse valor diminui, as duas variáveis tendem a ter pontos mais distantes.

O algoritmo proposto neste trabalho tem código fonte baseado no Inet 3.0 [47] e, desse modo, recebe como parâmetro de entrada obrigatório o número de vértices (n) e, opcionalmente, o número de vértices com grau 1 (g_1). Essa configuração opcional facilita o uso do gerador, visto que o usuário pode somente informar o número de vértices da topologia a ser gerada. Os demais parâmetros são internos do algoritmo NIT, ou seja, não perceptíveis ao usuário. Estes parâmetros são principalmente calculados em função

do número de vértices especificado pelo usuário do gerador, utilizando equações obtidas com o ajuste de curvas sobre as bases IRL da Internet.

O algoritmo original do Inet 3.0 pode ser segmentado em quatro fases, conforme apresentado na seção 2.2.3. Com o objetivo de modelar de forma mais precisa as topologias reais da Internet, de acordo com as métricas descritas na seção 3.2, foram alteradas as fases do algoritmo original do Inet.

Desse modo, o algoritmo NIT é composto pelas seguintes fases: geração dos graus (1), componente conexo (2), construção das cliques (3), conexão de grau 1 (4) e preenchimento de todos os graus (5). Dessas, as fases 1 e 2 foram modificadas e fase 3 foi criada com o objetivo de construir as cliques. As duas fases restantes não foram alteradas e nas próximas seções são descritas todas as alterações. Com estas modificações objetivou-se a melhoria dos resultados do Inet 3.0 em relação a modelagem das topologias mais recentes da Internet.

4.1 Fase 1: Alteração da Geração dos Graus dos Vértices

A primeira fase do algoritmo, que consiste na geração da distribuição dos graus, foi completamente modificada na nova versão. Inicialmente, modelamos a evolução do percentual de vértices de grau 1 na topologia real no período entre 2004 e 2007. O modelo é apresentado no gráfico à esquerda da Figura 4.1 e é dado por uma equação de decaimento exponencial obtida com ajuste R^2 igual a 0,99969. Assim, quando não fornecido pelo usuário, o percentual de vértices com grau 1 é dado por essa equação.

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:

$$\bar{F}(x) = \frac{j^\alpha \times \frac{1}{x^\alpha} - \left(\frac{j}{k}\right)^\alpha}{1 - \left(\frac{j}{k}\right)^\alpha} \quad (4.1)$$

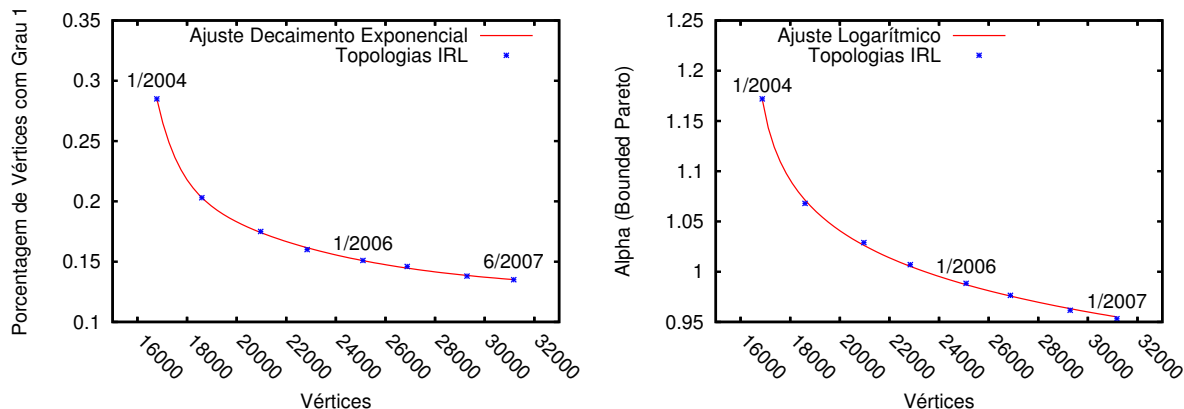


Figura 4.1: Ajustes da porcentagem de grau 1 e do parâmetro α para as topologias IRL.

em que o parâmetro j é o menor valor (grau igual a 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:

$$E[x] = \frac{\alpha}{1 - \alpha} \times \frac{k(\frac{j}{k})^\alpha - j}{1 - (\frac{j}{k})^\alpha} \quad (4.2)$$

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 4.1. 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 igual a 0,99918.

Portanto, a partir dos modelos para o percentual de vértices de grau 1, para o valor de α e para o maior grau, calcula-se a distribuição dos graus dos vértices e é atribuído a todos os vértices um grau específico, finalizando a primeira fase do algoritmo.

4.2 Fase 2: Alteração da Construção do Componente Conexo

A segunda fase consiste na criação do componente conexo, realizada por meio de uma árvore geradora mínima constituída com os vértices de grau maior ou igual a 3 e não 2 como empregado no algoritmo do Inet 3.0. Esta estratégia visa diminuir a alta fração de

cliques de tamanhos pequenos geradas no Inet (Figura 3.15).

Após a formação da árvore, alteramos o modo de conexão dos vértices de graus 3 e 2 do seguinte modo: tendo que todos os vértices de grau 3 possuem exatamente uma de suas arestas conectada à árvore, para 50% desses vértices criamos triângulos utilizando 2 vértices com grau igual a 2 escolhidos aleatoriamente. Então, estes dois vértices além de conectados entre si, são conectados ao vértice de grau 3, formando cliques de tamanho 3. Esta estratégia visa aumentar o coeficiente de agrupamento global do grafo e também diminuir a alta fração de cliques de tamanho 2 presentes no Inet, conforme apresentado na seção 3.3.5.

Posteriormente, para cada vértice com grau 2 ainda não conectado ao grafo, conectamos aproximadamente 99% desses vértices em vértices aleatórios do grafo, assegurando um só componente conexo. O restante 1% dos vértices de grau 2 são utilizados na construção das árvores nas bordas da topologia, as quais possuem no máximo altura 3 conforme apresentado na seção 3.3.2.

O valor de 1% foi calculado a partir das bases IRL do seguinte modo: para cada topologia IRL calculou-se o número de vértices (g_{1_r}) que formam as árvores nas bordas do grafo, utilizando a remoção recursiva de vértices com grau 1. Sabendo que o número de vértices com grau 1 do grafo é (g_1), calculamos o número de vértices (g_x) com grau maior que 1 que pertencem a essas árvores nas bordas utilizando $g_x = g_{1_r} - g_1$. Desse modo, podemos calcular a fração de vértices com grau 2 ($frac_{g_2}$) que compõe as árvores das topologias IRL por meio da equação $frac_{g_2} = g_x/g_2$, onde g_2 é o número de vértices com grau 2.

Topologia	Porcentagem (%)
1/2004	0,011
6/2004	0,008
1/2005	0,009
6/2005	0,007
1/2006	0,007
6/2006	0,005
1/2007	0,006
6/2007	0,004

Tabela 4.1: Fração de vértices com grau 2 das árvores periféricas nas topologias IRL.

Desse modo, obtemos as porcentagens conforme apresentado na Tabela 4.1. Fazendo a média desses valores obtemos 0,0057, que corresponde ao valor inteiro mais próximo de 1%.

4.3 Fase 3: Construção das Cliques

A terceira fase foi inserida ao NIT com o objetivo tornar o grafo mais denso, melhorar a distribuição das cliques e, em consequência, as métricas de distância e excentricidade. Adicionalmente, essa fase contribui para o aumento do coeficiente de agrupamento global das topologias.

Nessa fase são computados os tamanhos e a quantidade de cliques para posterior inserção no grafo. Nesse sentido, foi necessário estimar o tamanho da clique máxima e da clique média em função do número de vértices, resultando em uma equação logarítmica (R^2 igual a 0,99835) ajustada para o cálculo da clique média e outra equação linear (R^2 igual a 0,93873) para o cálculo da clique máxima, conforme a Figura 4.2.

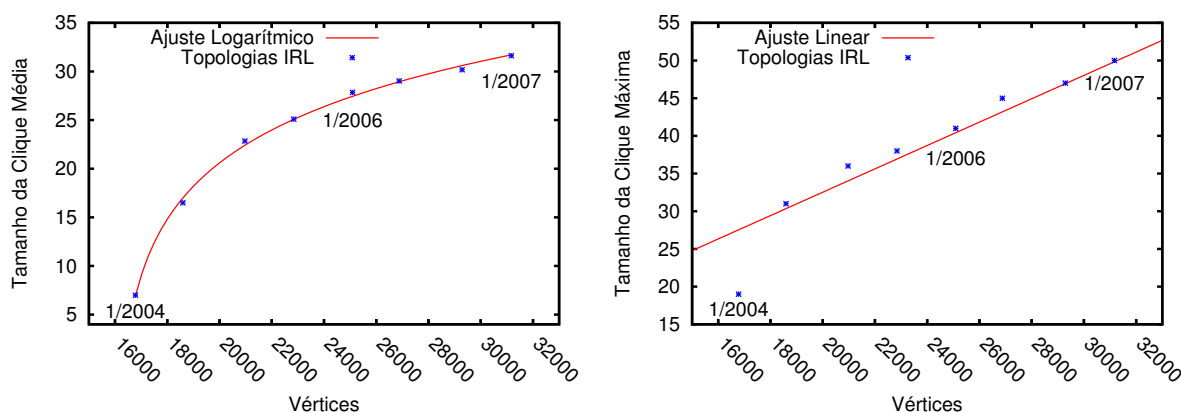


Figura 4.2: Ajuste para o tamanho da clique média (esquerda) e da clique máxima (direita) das topologias IRL.

Portanto, na terceira fase do novo algoritmo, estimamos o tamanho médio das cliques e da clique máxima em função do número de vértices e utilizamos. Sabendo que as conexões das fases posteriores aumentam o tamanho das cliques, temos a necessidade de gerar cliques com tamanhos menores do que os fornecidos pelas equações de ajuste. Sendo assim, a partir de testes realizados, utilizamos 90% dos tamanhos da clique média e máxima calculados como parâmetros para criar as cliques.

Para finalizar, utilizamos a distribuição normal (equação a seguir) para calcular a quantidade de cliques de cada tamanho a ser inserida ao grafo. O valor x é o tamanho da clique, μ corresponde à média do tamanho das cliques, obtido pelo ajuste apresentado na Figura 4.2 (esquerda).

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.3)$$

Por fim, σ corresponde a variância das distribuições das cliques é calculada a com o quadrado das médias dos desvios padrão. Ou seja σ é igual a 38,533 considerando todas as topologias.

Topologia	Desvio padrão
1/2004	5,08
6/2004	7,18
1/2005	7,48
6/2005	6,94
1/2006	5,27
6/2006	6,01
1/2007	6,04
6/2007	5,68

Tabela 4.2: Desvio padrão das distribuições das cliques nas topologias IRL.

Para o uso da distribuição normal é necessário conhecer o número total de cliques (C) a ser gerado. Desse modo, realizou-se um ajuste (R^2 igual a 0,95867) sobre as topologias IRL para obter esse valor em função do número de vértices, conforme apresentado na Figura 4.3.

Sendo assim, tendo C e a fração de cada tamanho a serem geradas ($f(x, \mu, \sigma)$), calculamos o número de cliques de cada tamanho x (C_x) em função do valor de C , por meio da equação a seguir:

$$C_x = C \times f(x, \mu, \sigma) \quad (4.4)$$

Tendo C_x previamente calculado, utilizamos aproximadamente 1% desse valor para efetuar a construção das cliques de tamanho x . Esses valor definido em virtude das conexões posteriores que aumentam o número de cliques de tamanho x . O algoritmo

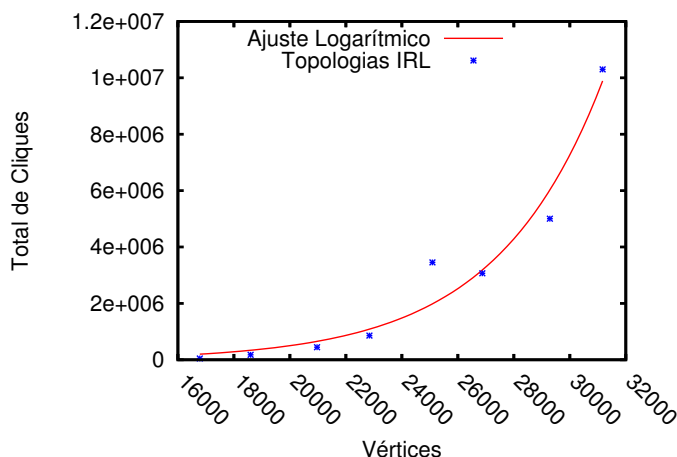


Figura 4.3: Ajuste para a quantidade total de cliques nas topologias IRL.

da construção consiste em formar cliques distintas por ao menos um vértice, a partir da clique máxima, até cliques de tamanho 3, iniciando a partir dos vértices de maior grau para os de menor grau.

Por exemplo, dado que a maior clique da topologia tem tamanho x igual a 50, e pela distribuição normal que existem C_x igual a 2 cliques desse tamanho, selecionamos os x vértices de maior grau, enumerados de 1 a 50, e criamos uma clique com estes vértices. A segunda clique de tamanho 50 é construída conectando o vértice 51 aos vértices de 1 a 49. Desse modo obtemos 2 cliques de tamanho x com, ao menos, um vértice distinto. A próxima clique a ser criada, de tamanho $x - 1 = 49$, é construída a partir do vértice numerado 52 até o vértice $(52 + 49) = 101$, aplicando a mesma abordagem descrita para as cliques de tamanho 50.

4.4 Fase 4: Conexão dos Vértices de Grau 1

A fase 3 do gerador Inet 3.0, agora denominada fase 4, foi mantida sem qualquer alteração. Nessa fase, é empregado o modelo de conexão preferencial apresentado na seção 2.2.3.2 na página 11.

Nessa etapa todos os vértices com grau associado maior que 1 já estão conectados por um só componente. Sendo assim, todos os vértices com grau 1 são conectados a esse componente por meio do modelo de conexão preferencial. Esse método gera uma

probabilidade proporcional de conexão considerando os graus de ambos os vértices a serem conectados [47].

De modo resumido, essas conexões aumentam temporariamente o raio e o diâmetro da topologia e, ao fim dessa fase, todos os vértices formam um só componente conexo 2.2.3.

4.5 Fase 5: Conexão dos Vértices com Graus não Exauridos

A fase 5 do gerador NIT, antiga fase 4 do gerador Inet 3.0, foi mantida sem alteração. Conforme descrito em [47], nessa fase são realizadas as conexões entre os vértices com graus não exauridos até preencher seus respectivos graus associados na fase 1.

Nesse sentido, as conexões são iniciadas à partir do vértice de maior grau associado até os vértices com grau igual a 2. Nessa conexões, bem como na fase anterior, é empregado o modelo de conexão preferencial e alguns vértices podem não ter seus graus completamente exauridos, conforme apresentado em [47].

Em geral, nessa fase são somente adicionadas novas arestas entre pares de vértices, diminuindo as distâncias e definindo os valores relativos ao diâmetro e ao raio das topologias geradas.

4.6 Síntese

Nesse capítulo foram apresentadas as alterações realizadas nas fases do gerador Inet 3.0, as quais originaram o novo algoritmo do gerador NIT proposto neste trabalho. A seguir são brevemente descritos os objetivos das alterações realizadas em cada fase alterada.

A alteração da fase 1 teve o propósito de melhorar a modelagem da distribuição dos graus mantendo o padrão de cauda pesada por meio da equação Pareto Limitada (seção 4.1). Na fase 2 foi alterado o modo de conexão dos vértices de grau baixo (2 e 3), que constituem no mínimo aproximadamente 40% do total, objetivando melhorar o coeficiente de agrupamento (Figura 3.17 da página 45) e diminuir a alta fração de cliques de tamanhos pequenos do Inet (Figura 3.15 da página 44). A fase 3 foi adicionada ao algoritmo com o objetivo primário de modelar as cliques, tal como nas topologias IRL. Essa mode-

lagem tem implicações diretas nas métricas de distância e suas derivadas, bem como no coeficiente de agrupamento.

De modo geral, o processo de construção de uma topologia pode ser realizado de inúmeras maneiras, e distintos modos de construção podem, no fim, gerar praticamente a mesma estrutura topológica. O desenvolvimento de um algoritmo que contemple simultaneamente as principais métricas da topologia da Internet é uma tarefa bastante complexa, pois pequenas alterações nas fases do algoritmo podem interferir não somente em uma mas em todas as métricas analisadas na topologia final. Os resultados apresentados para o gerador NIT correspondem à versão 5 da implementação. Testes similares aos apresentados neste trabalho foram realizados também com as quatro primeiras versões.

Diante dessas dificuldades, ainda assim os resultados do gerador NIT, os quais serão apresentados no próximo capítulo, foram satisfatórios na caracterização das topologias IRL para todas as métricas analisadas, principalmente em relação ao gerador Inet 3.0 considerando as topologias mais recentes.

CAPÍTULO 5

RESULTADOS

Na seção anterior foi descrito o algoritmo do gerador NIT, ou seja, as alterações realizadas nas fases do gerador Inet 3.0 [47]. Nesta seção são apresentados os resultados dos experimentos executados com este algoritmo por meio das métricas descritas na seção 3.2, bem como uma análise da variabilidade do novo algoritmo.

Os parâmetros de entrada, para cada topologia do gerador NIT, foram calculados em função das bases IRL, ou seja, o número de vértices e a porcentagem de vértices com grau 1. Os resultados são apresentados para três topologias das oito analisadas, compreendendo as topologias IRL 1/2004, IRL 1/2006 e IRL 6/2007, em ordem cronológica de coleta da topologia da Internet.

5.1 Composição dos Grafos e Estatísticas dos Graus dos Vértices

Considerando as alterações realizadas, principalmente na fase de geração da distribuição dos graus, é apresentada na Figura 5.1 a evolução do número de arestas nas topologias analisadas.

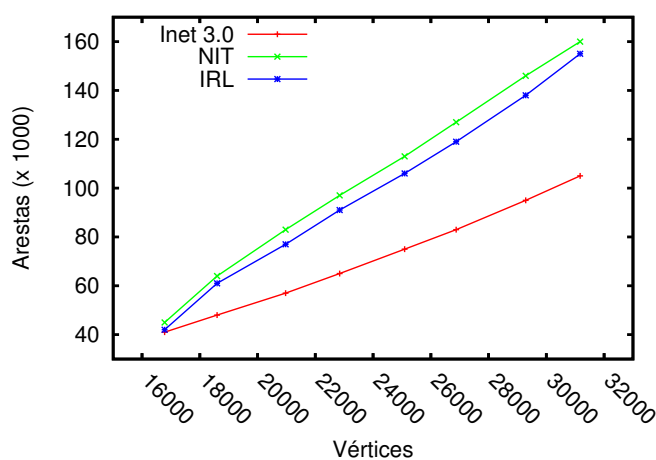


Figura 5.1: Evolução das arestas em função do número de vértices nas topologias Inet, NIT e IRL.

A distribuição de Pareto limitada, utilizada no gerador NIT, gera uma quantidade de arestas muito mais próxima da topologia real dos sistemas autônomos do que os valores gerados pelo Inet 3.0, embora com valores ainda um pouco acima dos ideais.

Na Tabela 5.1 são apresentadas a composição dos graus das topologias criadas com o gerador NIT, a partir de uma única semente aleatória (valor 3) escolhida aleatoriamente.

Data	Vértices	Arestas	Mediana	Média	Maior	Grau 1 (%)	COV
NIT 1/2004	16782	45563	2	5,43	2529	28,4	6,18
NIT 6/2004	18597	64632	2	6,95	2800	20,2	6,46
NIT 1/2005	20971	83204	2	7,94	3153	17,4	6,77
NIT 6/2005	22843	91279	2	8,51	3431	16,0	6,98
NIT 1/2006	25092	97152	3	9,08	3766	15,9	7,23
NIT 6/2006	26876	127390	3	9,48	4032	14,5	7,41
NIT 1/2007	29287	146103	3	9,98	4392	13,8	7,64
NIT 6/2007	31170	160844	3	10,32	4673	13,5	7,82

Tabela 5.1: Composição dos grafos e estatísticas dos graus dos vértices nas topologias NIT.

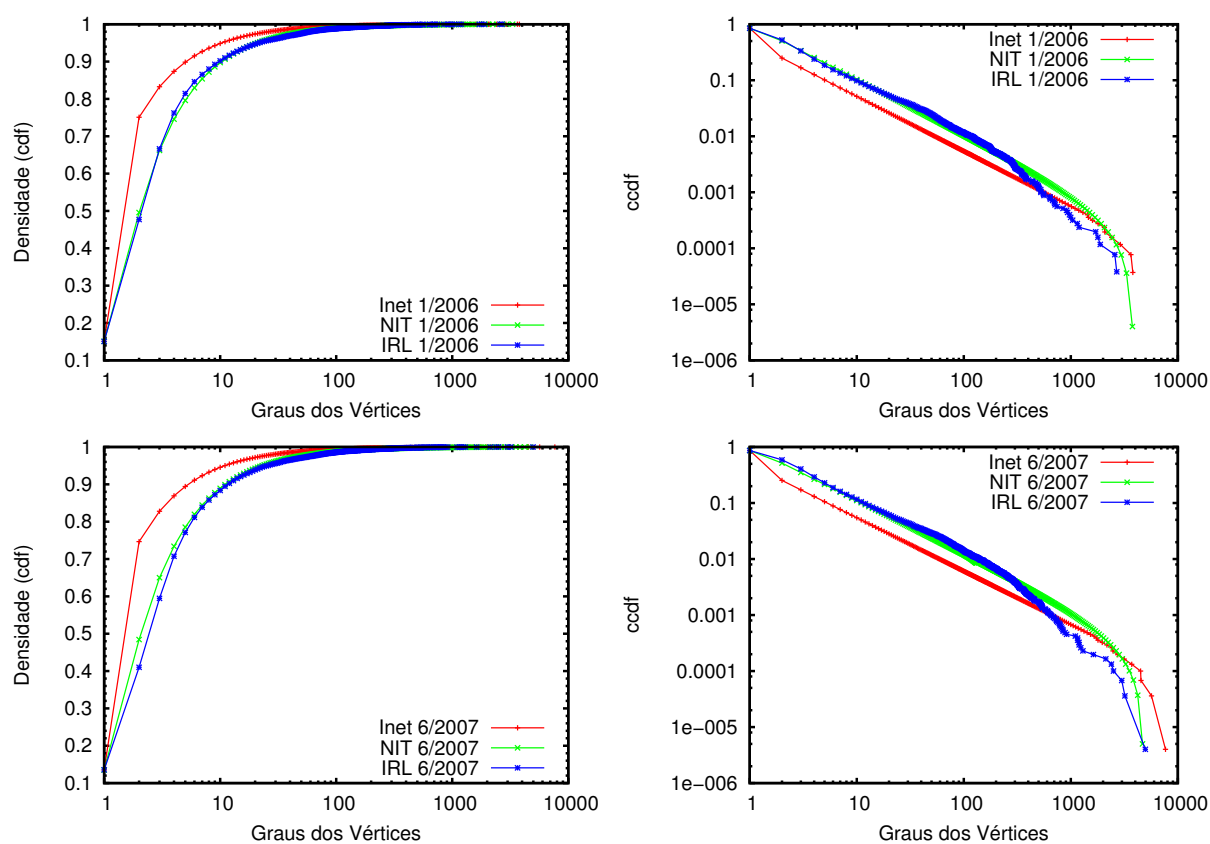


Figura 5.2: Distribuição dos graus nas topologias Inet, NIT e IRL.

Nas análises da distribuição dos graus foram selecionadas duas topologias (IRL 1/2006

e IRL 6/2007) para realizar as comparações. Não apresentamos a comparação dos graus da topologia IRL mais antiga (IRL 1/2004) por esta possuir uma distribuição semelhante a topologia do Inet 3.0, conforme os gráficos da Figura 3.11 na página 37.

Na Figura 5.2 é apresentada a distribuição dos graus dos vértices para as duas topologias selecionadas. Observa-se que a distribuição de Pareto limitada é mais semelhante a topologia real, principalmente para os graus menores que aproximadamente 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.

5.2 Distância entre Pares de Vértices e Excentricidades

A distribuição das distâncias para as três topologias é analisada nos gráficos da Figura 5.3. Compara-se o NIT com a topologia da Internet de 1/2004 (esquerda), 1/2006 (direita) e 6/2007 (centro). Observa-se que o algoritmo NIT gerou distâncias mais representativas e semelhantes da Internet do que o Inet. Em particular, as análises das topologias mais recentes indicam que o modelo proposto é bastante próximo do real em relação ao Inet.

Data	Maior	Média	Moda	COV
NIT 1/2004	9	3,73	3	0,15
NIT 6/2004	9	3,59	3	0,17
NIT 1/2005	9	3,53	3	0,16
NIT 6/2005	8	3,51	3	0,15
NIT 1/2006	9	3,49	3	0,14
NIT 6/2006	9	3,47	3	0,13
NIT 1/2007	8	3,45	3	0,12
NIT 6/2007	9	3,45	3	0,11

Tabela 5.2: Estatísticas das distâncias nas topologias NIT.

Os bons resultados da modelagem das 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 ajustar a representação das distâncias, principalmente para as topologias IRL mais recentes.

Na Tabela 5.3 e Figura 5.4 são apresentadas as estatísticas e resultados relativos às excentricidades das topologias geradas pelo NIT. A análise demonstra que o gerador

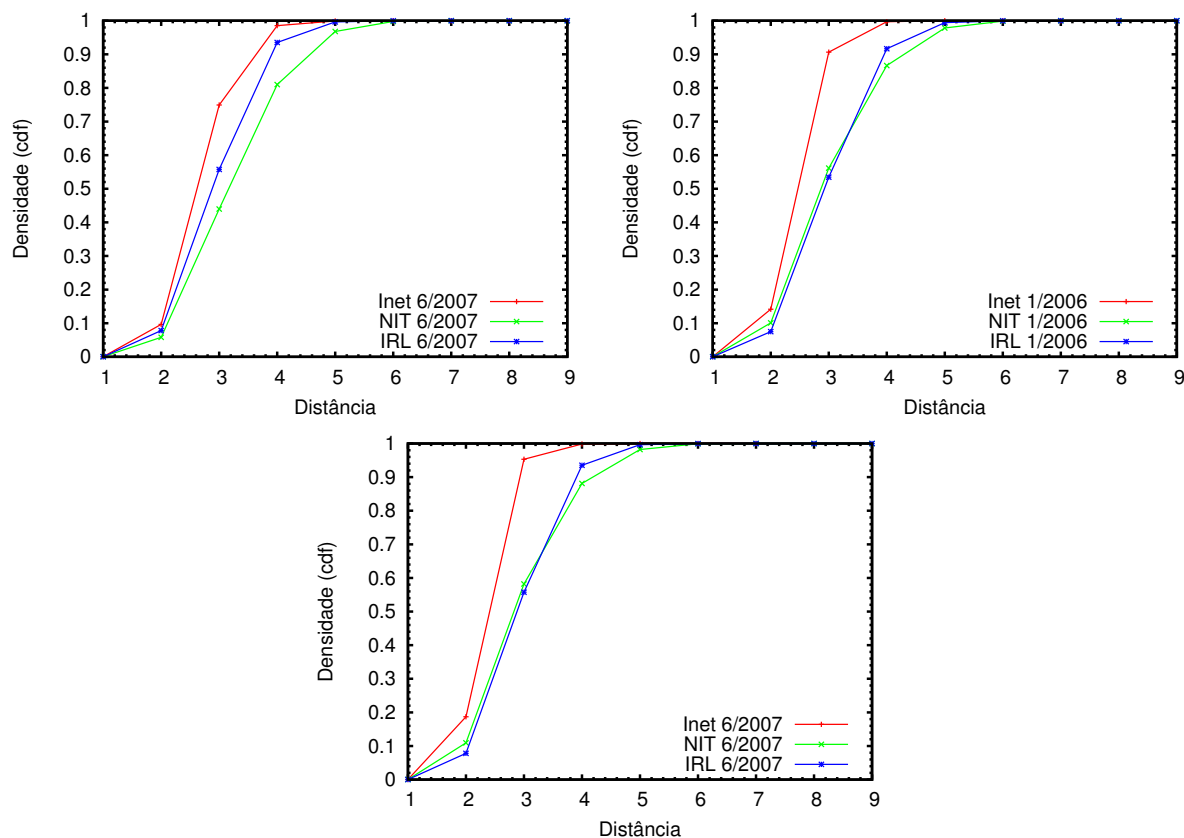


Figura 5.3: Distribuição das distâncias entre pares de vértices nas topologias Inet, NIT e IRL.

NIT apresenta melhores resultados nessa métrica quando comparado ao gerador 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 contribuiu para esses resultados.

Data	Mínima	Moda	Média	Mediana	Máxima	COV
NIT 1/2004	5	6	6,47	6	9	0,10
NIT 6/2004	5	6	6,34	6	9	0,09
NIT 1/2005	5	6	6,50	6	9	0,10
NIT 6/2005	5	6	6,14	6	9	0,10
NIT 1/2006	5	6	6,17	6	8	0,09
NIT 6/2006	5	6	6,34	6	9	0,09
NIT 1/2007	5	6	6,02	6	8	0,10
NIT 6/2007	5	6	6,14	6	9	0,09

Tabela 5.3: Estatísticas das excentricidades nas topologias NIT.

Na Tabela 5.4 são descritos os valores obtidos para o raio, diâmetro, número de vértices periféricos e centrais das topologias geradas com o NIT. Nota-se a variabilidade no número de vértices centrais e periféricos tal como ocorre nas topologias reais (ver Tabela 3.4 da

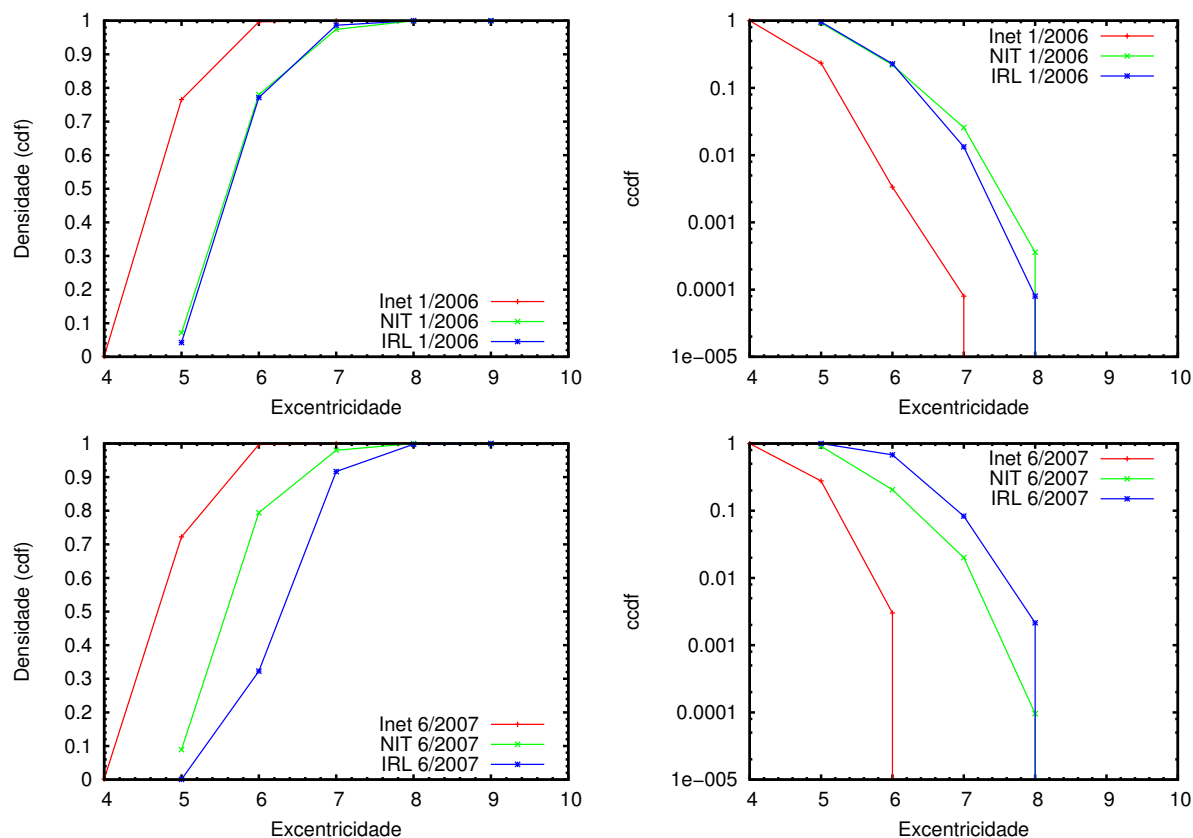


Figura 5.4: Excentricidades nas topologias Inet, NIT e IRL.

página 33).

Data	Raio	Diâmetro	Centro	Periferia
NIT 1/2004	5	9	171	23
NIT 6/2004	5	9	418	11
NIT 1/2005	5	9	141	36
NIT 6/2005	5	9	2031	521
NIT 1/2006	5	8	1789	9
NIT 6/2006	5	9	649	22
NIT 1/2007	5	8	5162	343
NIT 6/2007	5	9	2783	3

Tabela 5.4: Raio, diâmetro, centro e periferia nas topologias NIT.

Neste exemplo, verifica-se que os valores de raio e diâmetro possuem nenhuma ou baixa variação, respectivamente. É importante ressaltar que essas oito topologias foram geradas a partir de uma só semente aleatória, e os valores os valores apresentados podem variar de acordo com a semente utilizada.

5.3 Conectividade: Cliques e Agrupamentos

As duas métricas de conectividade analisadas são relacionadas as cliques e ao coeficiente de agrupamento, conforme apresentado na seção 3.2.3. Para a análise e comparação das cliques foram selecionados três conjuntos de topologias, sendo a IRL 1/2004, a IRL 1/2006 e a IRL 6/2007, com 16782, 25092 e 31170 vértices, respectivamente. A Figura 5.5 e a Tabela 5.5 apresentam os resultados para as três topologias.

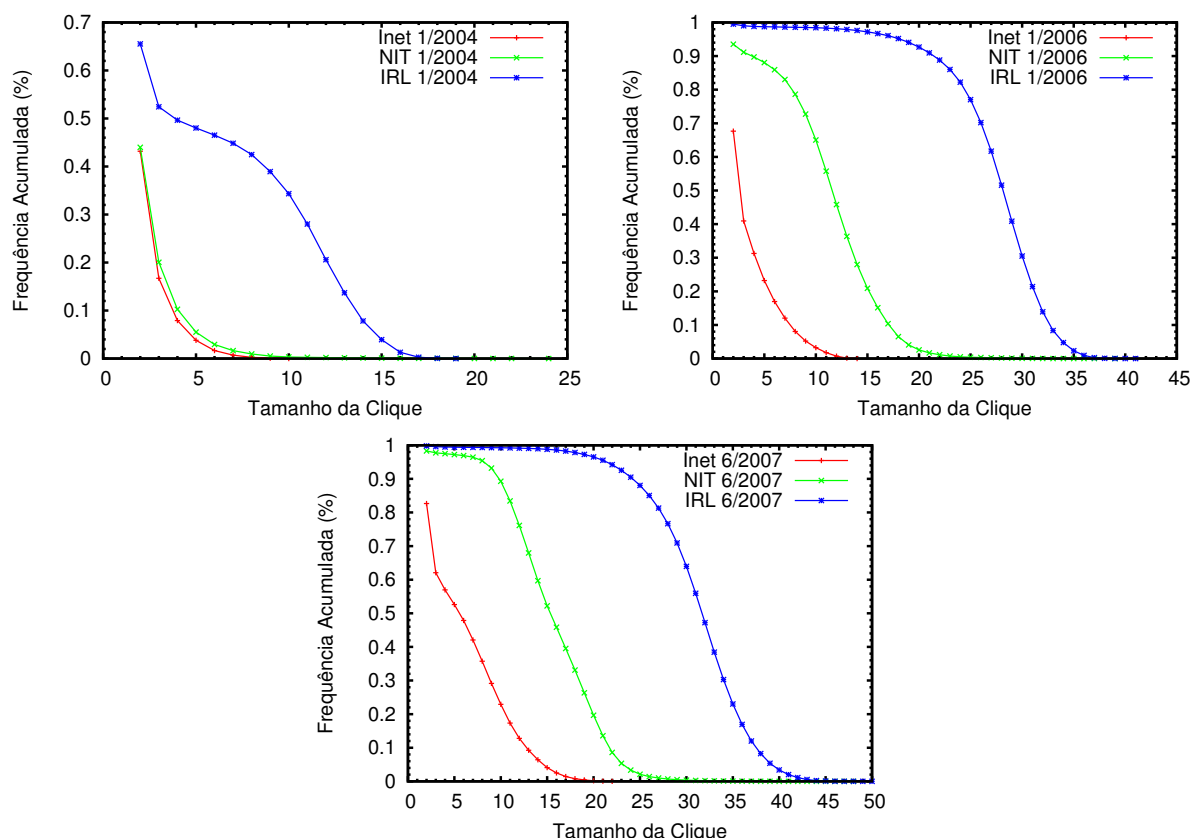


Figura 5.5: Distribuição das cliques das topologias Inet, NIT e IRL.

Conforme apresentado na Figura 5.5, observa-se que o perfil das cliques do gerador NIT é mais similar à topologia real do que o perfil do Inet, embora as duas distribuições não sejam idênticas. Na Tabela 5.5 são descritas características estatísticas das cliques de todas as topologias geradas.

A evolução do tamanho médio das cliques e o tamanho máximo das cliques são comparados dos gráficos da Figura 5.6. Observa-se que, embora distintas, o gerador NIT constrói topologias com tais medidas mais próximas da base IRL do que o gerador Inet,

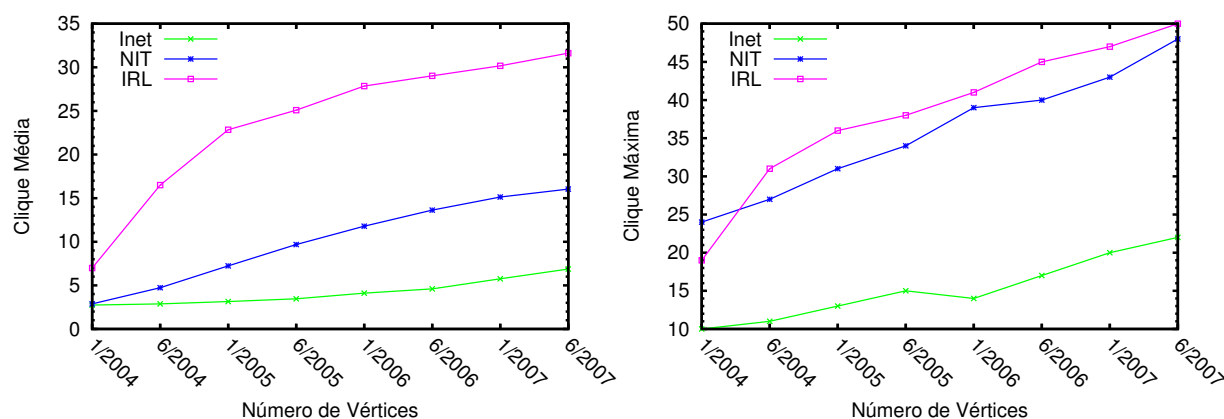


Figura 5.6: Evolução da clique média (esquerda) e máximas (direita) nas topologias Inet, NIT e IRL.

Data	Média	Mediana	Moda	Máxima	Número de Cliques Máximas
1/2004	2,8722	2	2	24	3
6/2004	4,73162	4	2	27	1
1/2005	7,2383	7	2	31	2
6/2005	9,68034	10	2	34	4
1/2006	11,7889	12	12	39	2
6/2006	13,6281	14	13	40	1
1/2007	15,1269	15	15	43	8
6/2007	16,0395	16	14	48	2

Tabela 5.5: Estatísticas das cliques nas topologias NIT.

principalmente para o tamanho da clique máxima (gráfico à direita).

Mesmo considerando que a distribuição das cliques não tenha sido semelhante as topologias IRL, a inserção da fase de geração de cliques no algoritmo do gerador NIT tem fundamental importância para a modelagem das distâncias, que consiste em uma das principais métricas a serem modeladas.

Outras métricas relativas a conectividade analisadas são o coeficiente de agrupamento médio por grau (CC_m) e o global (CC_g) das topologias do gerador NIT. Na Figura 5.7 é apresentado o CC_g das topologias analisadas e também a comparação com as topologias IRL e do gerador Inet.

Observa-se que as topologias do gerador NIT apresentam CC_g bem mais próximo da topologia da Internet do que o Inet. Essa medida expressa a conectividade global do grafo. Outra análise é apresentada nos gráficos da Figura 5.8, representando os coeficientes de agrupamento médio CC_m por grau das três topologias analisadas.

Nota-se que as topologias IRL apresentam a grande variabilidade para o CC_m , prin-

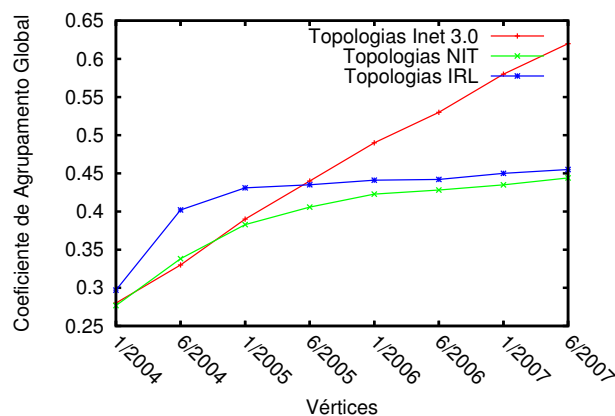


Figura 5.7: Coeficiente de agrupamento global nas topologias Inet, NIT e IRL.

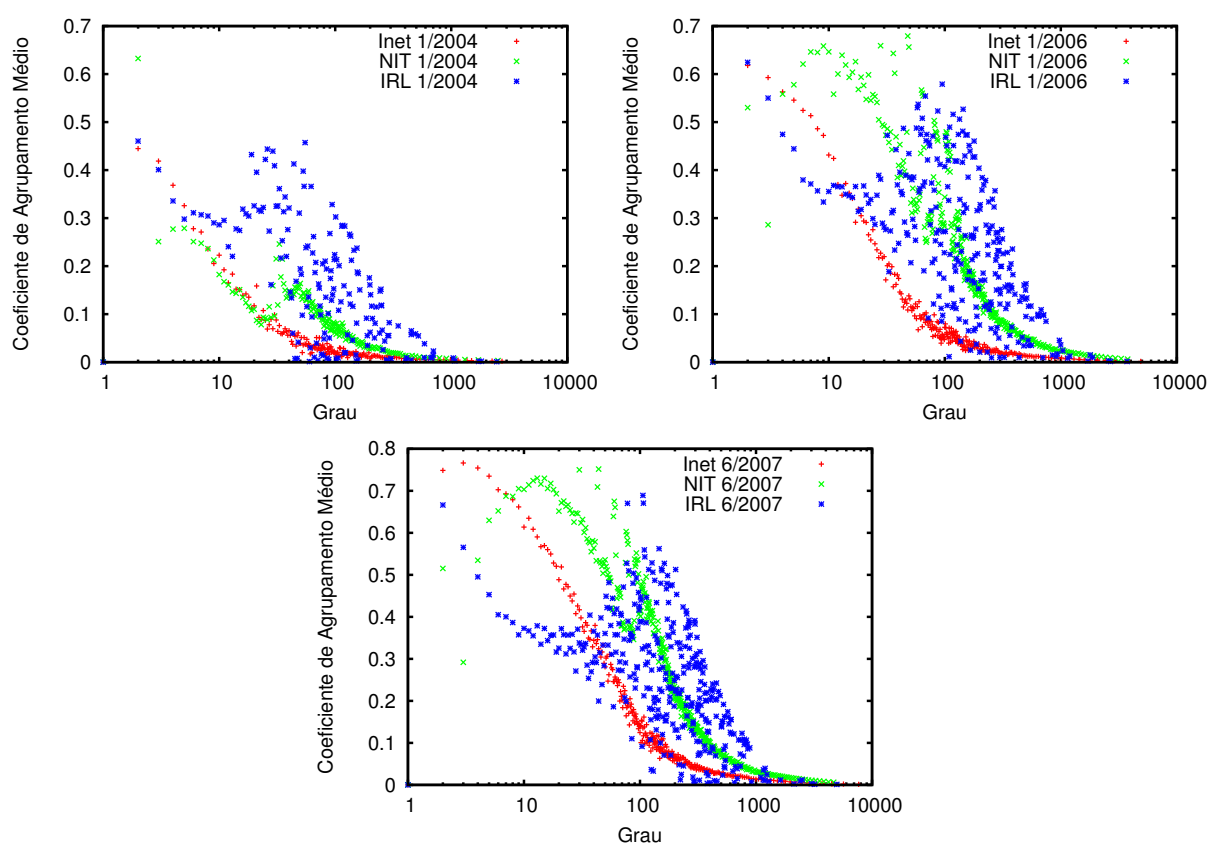


Figura 5.8: Coeficiente de agrupamento médio por grau nas topologias Inet, NIT e IRL.

principalmente para graus próximos a 100, enquanto o Inet demonstra um comportamento mais previsível. Embora não seja semelhante, o gerador NIT apresentou características de variabilidade intermediárias entre o Inet e a IRL, principalmente para as topologias mais atuais, neste caso de 6/2007.

5.4 Variabilidade do Gerador NIT

Nesta seção são apresentados os resultados relativos a variabilidade do gerador NIT. Cada topologia pode ser construída a partir da especificação, opcional, de uma semente (*seed*) que é utilizada para geração de valores aleatórios empregados nas conexões dos vértices.

Para avaliar a influência de diferentes valores para essa semente, nos resultados anteriormente apresentados neste capítulo, foram construídas com topologias distintas com cem sementes diferentes que compreendem os cem primeiros números primos. Ou seja, por exemplo para comparar a variabilidade da topologia NIT relativa a topologia IRL 6/2007, foi gerado um conjunto de 100 topologias NIT com cem sementes compreendendo os cem primeiros números primos.

Desse modo, para cada um dos oito conjuntos de topologias gerados, a seguir são apresentadas a análise e as estatísticas da variabilidade do gerador NIT quanto as medidas de distância média, excentricidade média, raio, diâmetro, grau médio e coeficientes de agrupamento global.

Vértices	Menor	Maior	Média	COV
NIT 1/4	3,71	3,74	3,7272	0,0018
NIT 6/4	3,57	3,60	3,587	0,0021
NIT 1/5	3,51	3,54	3,5297	0,0023
NIT 6/5	3,49	3,51	3,4999	0,0020
NIT 1/6	3,46	3,49	3,4788	0,0024
NIT 6/6	3,45	3,48	3,4671	0,0024
NIT 1/7	3,44	3,47	3,4544	0,0023
NIT 6/7	3,43	3,45	3,4411	0,0017

Tabela 5.6: Estatística das distâncias médias nas topologias NIT.

Nas Tabelas 5.6 e 5.7 são descritos os dados estatísticos da variabilidade das distância média e da excentricidade média. Pode-se notar que o desvio padrão e o coeficiente de variação (COV) são próximos de zero, demonstrando que o algoritmo é pouco variável quanto a essas medidas.

Nos gráficos da Figura 5.9 são apresentados os diagramas de *Box and Whisker* para distância média (esquerda) e excentricidades médias (direita) de cada topologia analisada. Nota-se que a variabilidade da distância média é bem pequena, caracterizando também a baixa variabilidade. Embora a excentricidade média apresente variabilidade maior, o

Vértices	Menor	Média	Maior	COV
NIT 1/4	6,10	7,27	6,51	0,047
NIT 6/4	6,12	6,69	6,32	0,021
NIT 1/5	6,00	6,94	6,30	0,032
NIT 6/5	6,03	6,68	6,22	0,020
NIT 1/6	6,04	6,65	6,23	0,023
NIT 6/6	6,07	6,55	6,24	0,019
NIT 1/7	5,93	6,80	6,21	0,027
NIT 6/7	5,98	6,54	6,18	0,020

Tabela 5.7: Estatísticas das excentricidades médias nas topologias NIT.

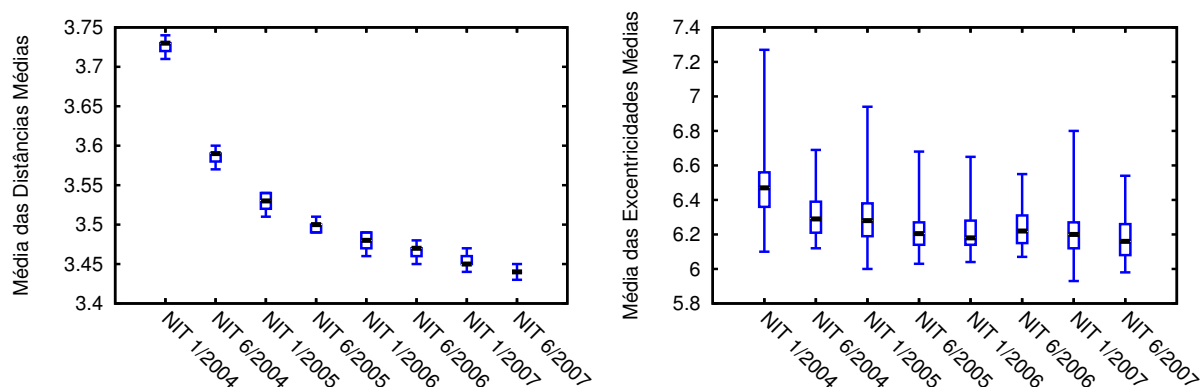


Figura 5.9: Variabilidade da distância média (direita) e da excentricidade média (esquerda) nas topologias NIT.

maior e menor valores não chegam a 1 ponto de diferença em relação a mediana (Figura 5.9 - direita).

Vértices	Menor	Média	Maior	COV
NIT 1/4	4	5	4,94	0,0480
NIT 6/4	5	5	5,00	0
NIT 1/5	4	5	4,91	0,0583
NIT 6/5	5	5	5,00	0
NIT 1/6	5	5	5,00	0
NIT 6/6	5	5	5,00	0
NIT 1/7	4	5	4,96	0,0395
NIT 6/7	5	5	5,00	0

Tabela 5.8: Estatísticas dos raios nas topologias NIT.

Nas Tabelas 5.8 e 5.9 são exibidos os dados estatísticos da variabilidade do raio e do diâmetro. Estes valores são importantes pois definem a característica de mundo pequeno (*small world*) existente na topologia dos sistemas autônomos da Internet [43, 24].

Pode-se notar que o desvio padrão e o COV são próximos de zero, ou zero em alguns

Vértices	Menor	Média	Maior	COV
NIT 1/4	8	10	9,02	0,05
NIT 6/4	8	9	8,92	0,03
NIT 1/5	8	10	8,93	0,04
NIT 6/5	8	10	8,85	0,05
NIT 1/6	8	10	8,98	0,03
NIT 6/6	8	9	8,90	0,03
NIT 1/7	8	10	8,89	0,05
NIT 6/7	8	9	8,81	0,04

Tabela 5.9: Estatísticas dos diâmetros nas topologias NIT.

casos, e demonstram que os valores de raio e de diâmetro são poucos variáveis, tendo limites entre 4 e 5 para o raio e 8 e 10 para o diâmetro.

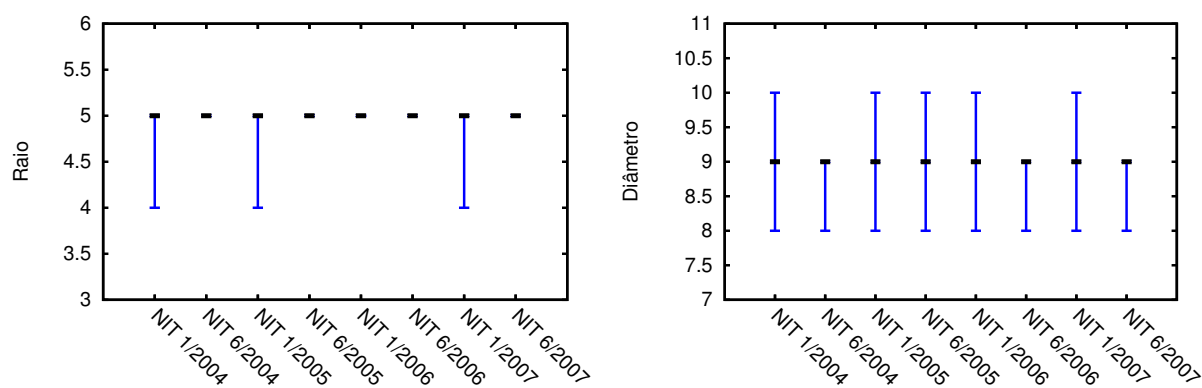


Figura 5.10: Variabilidade do Raio (esquerda) e do diâmetro (direita) nas topologias NIT.

Nos gráficos da Figura 5.10 são mostradas a variabilidade do raio (esquerda) e do diâmetro (direita) dos conjuntos de topologias analisados, os quais apresentam mediana do raio igual a 5 e mediana do diâmetro igual a 9. Percebe-se também baixa variabilidade do raio e do diâmetro e, especificamente, o diâmetro apresentando maior variabilidade do que o raio.

A seguir é analisada a variabilidade do grau médio e do coeficiente de agrupamento global (CC_g) das topologias do NIT. Conforme a Tabela 5.10, o grau médio das topologias geradas pelo NIT apresentou uma variabilidade muito próxima a zero em todas as análises.

Considerando que a equação utilizada na geração da distribuição dos graus, descrita na seção 4.1, fornece sempre a mesma distribuição dos graus para um determinado número de vértices, a baixa variabilidade acontece exatamente pela diferença do número final de

arestas das topologias, originada a partir das conexões aleatórias do algoritmo NIT.

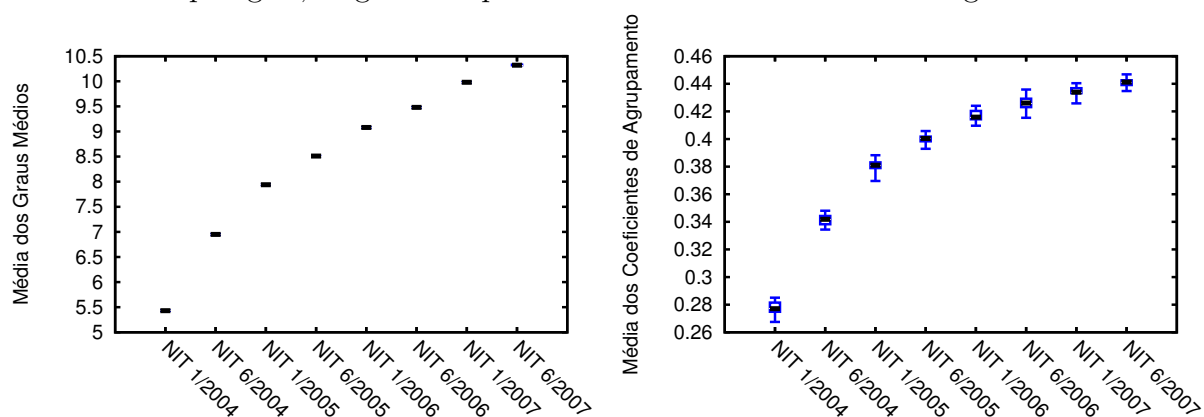


Figura 5.11: Variabilidade do grau médio (esquerda) e coeficiente de agrupamento (direita) nas topologias NIT.

Para exemplificar essa variabilidade, que também é apresentada na Figura 5.10 (esquerda), tendo que todas as topologias possuem exatamente o mesmo número de vértices (n), e que o número de arestas (m) pode ter uma pequena variação de acordo com a semente empregada nas conexões aleatórias, ao final temos o desvio padrão e de COV do grau médio gm apresentando valores baixos, pois $gm = 2m/n$.

De acordo com a Tabela 5.11 e a Figura 5.10 (direita), conforme o esperado, o CC_m exibiu baixa variabilidade demonstrando que as comparações realizadas na seção 5.4 têm poucas diferenças quando é alterada a semente utilizada.

5.5 Síntese

Nesse capítulo foram apresentadas as análises topológicas relativas as métricas descritas na seção 3.2, bem como a comparação dos resultados entre as topologias Inet, IRL e

Topologia	Menor	Média	Maior	COV
NIT 1/4	5,43	5,43	5,43	4,90706e-16
NIT 6/4	6,95	6,95	6,95	8,94568e-16
NIT 1/5	7,94	7,94	7,94	2,23723e-15
NIT 6/5	8,51	8,51	8,51	8,3495e-16
NIT 1/6	9,08	9,08	9,08	1,56507e-15
NIT 6/6	9,48	9,48	9,48	1,12428e-15
NIT 1/7	9,98	9,98	9,98	1,24594e-15
NIT 6/7	10,32	10,33	10,32	0,000430081

Tabela 5.10: Estatísticas do grau médio nas topologias NIT.

Vértices	Menor	Média	Maior	COV
NIT 1/4	0,27	0,29	0,28	0,014
NIT 6/4	0,33	0,35	0,34	0,010
NIT 1/5	0,37	0,39	0,38	0,010
NIT 6/5	0,39	0,41	0,40	0,008
NIT 1/6	0,41	0,42	0,42	0,009
NIT 6/6	0,42	0,44	0,43	0,011
NIT 1/7	0,43	0,44	0,43	0,008
NIT 6/7	0,43	0,45	0,44	0,007

Tabela 5.11: Estatísticas do coeficiente de agrupamento global nas topologias NIT.

do gerador NIT, o qual é proposto neste trabalho. Adicionalmente, avaliou-se aspectos relacionados a variabilidade do gerador proposto.

Em relação a composição dos grafos e a distribuição dos graus, o gerador NIT demonstrou maior semelhança com as topologias IRL, conforme descrito na seção 5.1. Os resultados das métricas de distâncias e derivadas exibiram maior similaridade com as topologias IRL, principalmente quanto aos valores de raio e diâmetro (seção 5.2), empregados na definição da característica de *small world* característica da topologia da Internet. Esses melhores resultados de modelagem também foram obtidos para o coeficiente de agrupamento a distribuição das cliques, conforme analisado na seção 5.3.

De acordo com resultados exibidos na seção 5.4, a variabilidade do gerador NIT apresentou baixa para todas as métricas analisadas. Essa análise da variabilidade não foi realizada para a distribuição das cliques por esse problema pertencer a classe NP-Completo, ou seja, ainda não foi encontrado um algoritmo com complexidade de tempo polinomial que viabilizasse a análise em todas as 800 topologias geradas. Para finalizar, maiores detalhes referentes a estes resultados, algumas conclusões e trabalhos futuros são temas abordados no próximo capítulo.

CAPÍTULO 6

CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho abordou o problema da análise e caracterização da topologia da Internet. Nesse contexto, foram estudados aspectos históricos dessa área de pesquisa, foram analisados diversos geradores de topologias sintéticas da Internet, bem como as principais métricas para caracterização da topologia. Posteriormente, foi realizada uma caracterização da topologia atual da Internet, baseada nessas métricas. A seguir, propusemos um novo gerador de topologias dos SAs da Internet, denominado NIT. Por fim, apresentamos e analisamos os resultados desse gerador e fizemos uma comparação com um dos mais utilizados geradores de topologias, o Inet, e com a topologia real da Internet, obtida a partir das bases IRL.

A partir da caracterização realizada em [19], que revelou grande variabilidade na distribuição dos graus dos vértices da topologia, os geradores de topologias da Internet são construídos tendo como estrutura fundamental a distribuição dos graus dos vértices. No entanto, a observação de que um grande número de topologias pode ser construído a partir de uma mesma distribuição dos graus [25] torna a abordagem baseada em graus insuficiente para modelar a topologia da Internet. Adicionalmente, a topologia da Internet está em constante evolução, requerendo caracterização contínua e cada vez mais completa. Assim, acreditamos que a associação de métricas baseadas em distância, conectividade e graus, podem constituir um novo paradigma para construção de geradores de topologias sintéticas que representam com maior exatidão as características da Internet.

Nesse trabalho foi proposto e apresentado o NIT, um novo gerador de topologias sintéticas similares à topologia da Internet, com o objetivo foi melhorar a modelagem da Internet quanto as métricas analisadas. O NIT foi implementado a partir do código fonte do Inet 3.0. Portanto, é também baseado na distribuição dos graus dos vértices. As topologias sintéticas geradas pelo NIT e pelo Inet foram comparadas com a topologia extraída

da Internet e registrada nas bases IRL. De modo geral, os resultados demonstram que o NIT constrói topologias mais semelhantes às da Internet do que o Inet [27]. Os resultados indicam que o Inet não modela bem as métricas de distâncias e derivadas, nem reproduz a distribuição das cliques e o coeficiente de agrupamento, nos valores apresentados pela topologia atual da Internet. Possivelmente, o Inet que é um gerador fundamentado principalmente na distribuição dos graus dos vértices e a contínua evolução da topologia da Internet, são fatores que contribuem para esses resultados do Inet 3.0.

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. Este trabalho pretende contribuir para a geração de topologias sintéticas mais semelhantes à da Internet de acordo com as principais métricas analisadas na atualidade. Desenvolver um algoritmo que contemple simultaneamente as principais métricas de caracterização topológica da Internet é um desafio bastante complexo, pois pequenas alterações no algoritmo interferem nos resultados de várias métricas. As métricas são inter-relacionadas, porém as funções que regem estas relações não são conhecidas. No estado atual de desenvolvimento desta pesquisa, as topologias são caracterizadas por métricas independentes, sem comparação ou associação com outras métricas, e geralmente baseando-se na distribuição dos graus dos vértices.

Conforme apresentado neste trabalho e em [27], os resultados do gerador NIT são satisfatórios quando comparados com a topologia real, para todas as métricas analisadas. Os resultados do NIT são também superiores aos do Inet 3.0, principalmente em relação às topologias mais recentes. Embora métricas tais como a distribuição das cliques não tenham sido modeladas de modo idêntico, as alterações efetuadas no gerador Inet tornam as topologias do NIT mais realísticas em relação à Internet.

Uma possibilidade de trabalho futuro é direcionada à construção de geradores de topologias baseados em métricas de conectividade, tais como a distribuição das cliques e coeficientes de agrupamento. O estudo e a descoberta das inter-relações entre as métricas auxiliaria bastante a construção de geradores de topologias sintéticas. Outra direção

importante é a geração de topologias reduzidas. Os geradores de topologias são projetados para produzir topologias do tamanho da Internet atual, com dezenas de milhares de vértices. A geração de topologias similares porém de tamanho reduzido, por exemplo, com cerca de 10% ou 1% do tamanho da topologia real seria extremamente útil para testes de protocolos e serviços em simuladores e ambientes de laboratório.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] C. Alaettinoglu, C. Villamizar, E. Gerich, D. Kessens, D. Meyer, T. Bates, D. Karrenberg, e M. Terpstra. Routing Policy Specification Language (RPSL). Relatório Técnico 2622, Internet Engineering Task Force (IETF), 1999.
- [2] Réka Albert e Barabasi Albert-László Barabási. Topology of evolving networks: local events and universality. *Physical Review Letters*, 85:5234, 2000.
- [3] D. Alderson, L. Li, W. Willinger, e J. C. Doyle. Understanding Internet Topology: principles, models, and validation. *IEEE/ACM Transactions on Networking*, páginas 1205–1218, 2005.
- [4] David A. Bader e Kamesh Madduri. GTgraph: A Synthetic Graph Generator Suite. Relatório técnico, College of Computing - Georgia Institute of Technology, Atlanta, 2006.
- [5] Albert-László Barabási e Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [6] Coen Bron e Joep Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM*, 16(9):575–577, 1973.
- [7] Tian Bu e Donald F. Towsley. On Distinguishing between Internet Power Law Topology Generators. *INFOCOM 2002: Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, páginas 638–647, 2002.
- [8] J. M. Carlson e J. Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E (Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics)*, páginas 1412–1427, 1999.
- [9] Deepayan Chakrabarti e Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Computing Survey*, 38(1):2, 2006.

- [10] Deepayan Chakrabarti, Yiping Zhan, e Christos Faloutsos. R-MAT: A Recursive Model for Graph Mining. *SIAM International Conference on Data Mining*, 2004.
- [11] Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shenker, e Walter Willinger. Towards capturing representative AS-level Internet topologies. *SIGMETRICS '02: Proceedings of the 2002 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, páginas 280–281, 2002.
- [12] Hyunseok Chang e Walter Willinger. Difficulties Measuring the Internet's AS-Level Ecosystem. Invited Paper. *INFOCOM 2006: Information Sciences and Systems*, páginas 1479–1483, 2006.
- [13] Qian Chen, Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shenker, e Walter Willinger. The Origin of Power Laws in Internet Topologies Revisited. *IEEE INFOCOM 2002*, 2001.
- [14] National Research Council Committee on Network Science for Future Army Applications. *Network Science*. The National Academies Press, 2006.
- [15] DIMES. Distributed Internet Measurements and Simulations. Site: <http://www.netdimes.org/new> (acessado em 2007), 2006.
- [16] Matthew B. Doar. A Better Model for Generating Test Networks. *Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity*, páginas 86–93, 1996.
- [17] P. Erdős e A. Rényi. On Random Graphs. *I Publicationes Mathematicae*, páginas 290–297, 1959.
- [18] Alex Fabrikant, Elias Koutsoupias, e Christos H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the Internet. *Lecture Notes In Computer Science*, páginas 110–122, 2002.
- [19] Michalis Faloutsos, Petros Faloutsos, e Christos Faloutsos. On power-law relationships of the Internet topology. *SIGCOMM '99: Proceedings of the conference on*

- Applications, technologies, architectures, and protocols for computer communication*, páginas 251–262, 1999.
- [20] Ulysses Doria Filho. *Introdução à Bioestatística*. ELSEVIER, 1999.
- [21] Internet Research Laboratory. Site: <http://irl.cs.ucla.edu/topology> (acessado em 2007), 2006.
- [22] C. Jin, Q. Chen, e S. Jamin. Inet: Internet Topology Generator. Relatório Técnico 443-00, Department of EECS, University of Michigan, 2000.
- [23] Yihua He Georgos Siganos Michalis Faloutsos Srikanth Krishnamurthy. A Systematic Framework for Unearthing the Missing Links: Measurements and Impact. *USENIX/SIGCOMM NSDI 2007*, 2007.
- [24] Jure Leskovec, Jon Kleinberg, e Christos Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. *The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD*, 2005.
- [25] Lun Li, David Alderson, Walter Willinger, e John Doyle. A first-principles approach to understanding the internet’s router-level topology. *SIGCOMM Comput. Commun. Rev.*, páginas 3–14, 2004.
- [26] Yan Li, Jun-Hong Cui, Dario Maggiorini, e Michalis Faloutsos. Characterizing and Modelling Clustering Features AS-Level Internet Topology. *INFOCOM 2008: The 27th Conference on Computer Communications. IEEE*, páginas 271–275, 2008.
- [27] Joylan N. Maciel e Cristina D. Murta. Um Novo Gerador de Topologias da Internet. *SBRC 2008: 26º Simpósio Brasileiro de Redes de Computadores*, páginas 51–63, 2008.
- [28] Damien Magoni e Jean Jacques Pansiot. Analysis of the autonomous system network topology. *ACM SIGCOMM Comput. Commun. Rev.*, páginas 26–37, 2001.

- [29] Priya Mahadevan, Dmitri Krioukov, Kevin Fall, e Amin Vahdat. Systematic topology analysis and generation using degree correlations. *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, páginas 135–146, 2006.
- [30] Priya Mahadevan, Dmitri Krioukov, Marina Fomenkov, e Bradley Huffaker. The Internet AS-level Topology: Three Data Sources and One Definitive Metric. *SIGCOMM Comput. Commun. Rev.*, páginas 17–26, 2006.
- [31] Alberto Medina, Anukool Lakhina, Ibrahim Matta, e John Byers. BRITE: Universal Topology Generation from a User’s Perspective. Relatório Técnico 2001-003, Boston University, 1 de 2001.
- [32] Alberto Medina, Anukool Lakhina, Ibrahim Matta, e John W. Byers. BRITE: An Approach to Universal Topology Generation. *IEEE MASCOTS*, páginas 346–361, 2001.
- [33] Alberto Medina, Ibrahim Matta, e John Byers. On the origin of power laws in Internet topologies. *SIGCOMM Comput. Commun. Rev.*, páginas 18–28, 2000.
- [34] Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, e Lixia Zhang. In Search of the elusive Ground Truth: The Internet’s AS-level Connectivity Structure. *Proceedings of ACM SIGMETRICS’08*, 2008.
- [35] Ricardo V. Oliveira, Beichuan Zhang, e Lixia Zhang. Observing the Evolution of Internet AS Topology. *SIGCOMM Comput. Commun. Rev.*, páginas 313–324, 2007.
- [36] Seung-Taek Park, David M. Pennock, e C. Lee Giles. Comparing Static and Dynamic Measurements and Models of the Internet’s AS Topology. *INFOCOM 2004: The 23rd Conference of the IEEE Communications Society - IEEE*, páginas 1616–1627, 2004.
- [37] R. Pastor-Satorras e A. Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach. Book Review*. Cambridge University Press, 2004.

- [38] Danny Raz e Rami Cohen. The Internet Dark Matter - on the Missing Links in the AS Connectivity Map. *INFOCOM 2006: 25th IEEE International Conference on Computer Communications*, páginas 1–12, 2006.
- [39] Réseaux IP Européens. Site: <http://www.ripe.net> (acessado em 2006), 2006.
- [40] The Route Views Project. Site: <http://www.routeviews.org> (acessado em 2006), 2006.
- [41] Hongsuda Tangmunarunkit, Ramesh Govindan, Sugih Jamin, Scott Shenker, e Walter Willinger. Network topology generators: degree-based vs. structural. *ACM SIGCOMM'02*, páginas 147–159, 2002.
- [42] CAIDA Macroscopic Topology Project Team. CAIDA Skitter Topology Traces, 2006. Site: <http://www.caida.org/tools/measurement/skitter> (acessado em 2007).
- [43] D. J. Watts e S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, páginas 440–442, 1998.
- [44] Bernard M. Waxman. Routing of Multipoint Connections. *IEEE Journal on Selected Areas in Communications*, páginas 1617–1622, 1988.
- [45] Clustering coefficient. Site: http://en.wikipedia.org/wiki/Clustering_coefficient (acessado em 2008), 2007.
- [46] Walter Willinger, Ramesh Govindan, Sugih Jamin, Vern Paxson, e Scott Shenker. Scaling Phenomena in the Internet: Critically examining Criticality. *National Academy of Science, volume 99*, 2002.
- [47] Jared Winick e Sugih Jamin. Inet-3.0: Internet Topology Generator. Relatório Técnico 456-02, EECS, University of Michigan, 2002.
- [48] Ellen W. Zegura, Kenneth L. Calvert, e Michael J. Donahoo. A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking*, páginas 770–783, 1997.

- [49] Beichuan Zhang, Raymond A. Liu, Daniel Massey, e Lixia Zhang. Collecting the Internet AS-level Topology. *ACM SIGCOMM Computer Communication Review (CCR)*, special issue on *Internet Vital Statistics*, páginas 53–61, 2005.

JOYLAN NUNES MACIEL

UM NOVO GERADOR DE TOPOLOGIAS DA INTERNET

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientadora: Profa. Dra. Cristina Duarte Murta.

CURITIBA

2008