

RODRIGO AUGUSTO ALVES

**TOPOLOGIA DOS SISTEMAS AUTÔNOMOS:
EVOLUÇÃO E PREDIÇÃO**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática do Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.
Orientadora: Profa. Cristina Duarte Murta

CURITIBA

2008



Universidade Federal do Paraná
Setor de Ciências Exatas
Programa de Pós-graduação em Informática

PARECER

Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática, do aluno Rodrigo Augusto Alves, avaliamos o trabalho intitulado, "*TOPOLOGIA DOS SISTEMAS AUTÔNOMOS: EVOLUÇÃO E PREDIÇÃO*", cuja defesa foi realizada no dia 30 de julho de 2008, às 14:00 horas, na sala de vídeo-conferência do CESEC do Setor de Tecnologia da Universidade Federal do Paraná. Após a avaliação, decidimos pela aprovação do candidato.

Curitiba, 30 de julho de 2008.

Prof^a. Dra. Cristina Duarte Murta
DINF/UFPR – Orientadora

Prof^a. Dra. Anelise Munaretto Fonseca
UTFPR – Membro Externo
Prof. Dr. Elias Procópio Duarte Jr.
DINF/UFPR – Membro Interno

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo CNPq. Agradeço ao grupo do Internet Research Laboratory da UCLA pela coleta e por tornar disponível os dados da topologia da Internet na Web.

Com admiração à minha esposa Larissa e aos meus pais, com profundo reconhecimento, pelo constante apoio e incentivo.

E, em especial, à Profa. Dra. Cristina Duarte Murta, que dedicou seu tempo a esclarecer dúvidas e contribuiu enormemente para o enriquecimento da pesquisa.

SUMÁRIO

LISTA DE FIGURAS	vi
LISTA DE TABELAS	vii
RESUMO	viii
ABSTRACT	ix
1 INTRODUÇÃO	1
1.1 Contexto	1
1.2 Objetivo da Pesquisa	2
1.3 Motivação	3
1.4 Organização do Trabalho	4
2 CONCEITOS E TRABALHOS RELACIONADOS	5
2.1 Trabalhos Relacionados	5
2.2 Base de Dados da Topologia	7
2.2.1 Coleta de Dados	7
2.2.2 A Base IRL	8
2.3 Métricas para Análise da Topologia	10
2.3.1 Grafo	10
2.3.2 Caminho Simples e Menor Caminho	10
2.3.3 Distância	11
2.3.4 Excentricidade e Métricas Derivadas	11
2.3.5 Distribuição dos Graus dos Vizinhos	12
2.3.6 Árvore nas Bordas	12
2.3.7 Coeficiente de Agrupamento	14
2.4 Modelagem dos Dados	15

2.4.1	Ferramenta Utilizada	16
2.4.2	Funções de Modelagem	17
2.4.3	Síntese do Capítulo	18
3	CARACTERIZAÇÃO DA EVOLUÇÃO DA TOPOLOGIA	20
3.1	Composição do Grafo da Topologia dos Sistemas Autônomos	20
3.2	Análise dos Graus dos Vértices	23
3.3	Distâncias entre Pares de Vértices	24
3.4	Caracterização da Excentricidade	29
3.5	Análise do <i>Status</i>	32
3.6	Árvore nas Bordas	33
3.7	Coefficiente de Agrupamento	37
3.8	Síntese do Capítulo	39
4	MODELOS DE EVOLUÇÃO E PREDIÇÃO PARA A TOPOLOGIA DA INTERNET	41
4.1	Composição dos Grafos e Graus dos Vértices	41
4.2	Distância entre Pares de Vértices	46
4.3	Status	50
4.4	Árvore nas Bordas	51
4.5	Coefficiente de Agrupamento	56
4.6	Síntese do Capítulo	58
5	CONCLUSÃO	59
	BIBLIOGRAFIA	64

LISTA DE FIGURAS

2.1	Grafo ilustrativo de G	13
2.2	Grafo ilustrativo G' , após a deleção na primeira iteração dos vértices com grau igual a 1 (Esquerda). Grafo ilustrativo G'' , após a deleção na segunda iteração dos vértices com grau igual a 1 (Direita).	13
2.3	Árvores nas bordas encontradas no grafo G' , com raiz em h e d	14
2.4	Grafo G exemplificando uma rede com coeficiente de agrupamento igual a 1.	14
3.1	Crescimento do número de vértices e arestas no período de 2004 a 2007.	21
3.2	Evolução da relação entre o número de vértices e de arestas no período de janeiro de 2004 a julho de 2007.	22
3.3	Evolução da média e da mediana dos graus dos vértices.	24
3.4	CDF do grau dos vértices.	25
3.5	CCDF do grau dos vértices.	26
3.6	Evolução de índices estatísticos das distâncias.	27
3.7	Frequência absoluta da distância entre pares de vértices no período.	28
3.8	Frequência relativa da distância entre pares de vértices no período.	28
3.9	Evolução da ocorrência de distâncias 3 e 4.	29
3.10	Evolução da ocorrência de distâncias 2, 5 e 6.	30
3.11	Evolução da excentricidade no período analisado.	31
3.12	Evolução das conexões em função da excentricidade.	32
3.13	Evolução do grau médio dos vértices por excentricidade no período.	32
3.14	Distribuição cumulativa dos status observados nas topologias no período analisado.	33
3.15	Estatísticas do Status para o período analisado.	34
3.16	Fração de nodos em árvores e no subgrafo 2 -core no período analisado.	35
3.17	Tamanho das árvores observadas no período (escala logarítmica).	36
3.18	Tamanho das árvores observadas nas topologias (escala linear).	37

3.19	Evolução na quantidade de nodos na maior árvore da rede no período. . . .	38
3.20	Evolução do número de árvores com menor número de nodos da rede no período	39
3.21	Evolução do coeficiente de agrupamento da rede no período.	40
4.1	Crescimento absoluto da rede no período de 2004 a 2010.	42
4.2	Padrão de crescimento da rede no período de 2004 a 2010.	44
4.3	Crescimento relativo de vértices e arestas no período de 2004 a 2010. . . .	45
4.4	Média e mediana do grau dos vértices no período de 2004 a 2010.	46
4.5	Aproximação da Evolução das Distâncias.	48
4.6	Evolução das distâncias 3 e 4.	50
4.7	Evolução das distâncias 2, 5 e 6.	51
4.8	Modelos para estatísticas do status dos vértices da rede no período de 2004 a 2010.	52
4.9	Frequência de nodos em árvores e de nodos do subgrafo <i>2-core</i> entre o período de janeiro de 2004 a 2010.	53
4.10	Exemplo de uma árvore fictícia na borda (acima). Exemplo da árvore fictícia (acima) na adição de uma aresta (abaixo).	54
4.11	Evolução na quantidade de nodos na maior árvore da rede entre o período de janeiro de 2004 a 2010.	55
4.12	Evolução na quantidade árvore com menor número de nodos da rede entre o período de janeiro de 2004 a 2010.	56
4.13	Evolução do coeficiente de agrupamento da rede entre o período de janeiro de 2004 a 2010.	58

LISTA DE TABELAS

3.1	Número de vértices e arestas em cada amostra da topologia.	21
3.2	Estatísticas dos graus dos vértices para as amostras da topologia.	23
3.3	Evolução de índices estatísticos das distâncias no período.	25
3.4	Evolução da excentricidade nas topologias analisadas.	30
3.5	Raio, diâmetro, centro e periferia nas topologias analisadas.	31
4.1	Composição dos grafos e extrapolação dos valores	43
4.2	Estatísticas dos graus dos vértices	47
4.3	Estimativas para estatísticas das distâncias	49

RESUMO

O conhecimento obtido na caracterização da topologia da Internet pode ser utilizado em várias pesquisas, testes e simulações de novos protocolos, serviços e aplicações. A caracterização da evolução da topologia em passado recente permite realizar previsões para este sistema em futuro próximo. Este trabalho apresenta uma caracterização extensiva da topologia real da Internet no período entre 2004 e 2007. Várias instâncias da topologia, coletadas em dias específicos deste período, são analisadas e a evolução da rede é modelada. A análise baseia-se em métricas de distância em grafos e suas derivadas, graus dos nodos, caracterização das bordas do grafo e métricas de conectividade como o coeficiente de agrupamento. Os modelos gerados para representar a evolução da rede em cada métrica são extrapolados para mostrar as previsões de crescimento da topologia no período de 2008 a 2010. Os resultados indicam que as métricas de distância são invariantes e características da topologia. Leis empíricas são propostas para enunciar aspectos característicos da topologia. A associação das métricas de distância e a distribuição dos graus pode servir de base para a descrição de regras empíricas sobre o crescimento e comportamento da Internet.

ABSTRACT

The knowledge obtained in the characterization of the Internet topology can be used in many ways, for instance, in tests and simulations of new protocols, services and applications. The characterization of topology evolution in recent past can be used to build forecast models for this system in near future. This dissertation presents an extensive characterization of the Internet topology based on topology data collected between 2004 and 2007. Several instances of the topology, collected in specific days during this period, are analyzed and the evolution of the network is modeled. The analysis is based in graph metrics, including distance-based metrics, node degrees, characterization of the edges of the graph, and connectivity metrics such as the coefficient of clustering. A mathematical model is built for every metric based on fitting. The models represent the evolution of the topology in relation to the chosen metrics. An extrapolation process is applied to these models to predict the topology growth and evolution in the period from 2008 to 2010. The results indicate that the distance metrics are invariant characteristics of the topology. Empirical laws are proposed to spell out some properties of the topology. The composition of distance metrics and distribution of vertex degrees can be a basis for an empirical description of rules on growth and behavior of the Internet topology.

CAPÍTULO 1

INTRODUÇÃO

1.1 Contexto

Redes complexas são sistemas físicos, biológicos ou sociais caracterizados por um amplo conjunto de entidades bem definidas que interagem dinamicamente entre si. Muitos sistemas podem ser representados na forma de redes complexas, por exemplo, as redes de distribuição de energia e água, redes de transporte e telecomunicações, redes sociais, além da Internet e da Web. A representação, o estudo e a modelagem das diversas redes complexas compõem uma área recentemente denominada ciência das redes, que tem como foco atual a caracterização da estrutura dessas redes.

A caracterização de redes complexas tem três objetivos. O primeiro objetivo é o de descrever suas propriedades estruturais, por exemplo, propriedades relacionadas a topologia, conectividade, distâncias, quantidade de árvores e distribuição dos graus dos nodos. Novas métricas podem ser propostas para caracterizar sua estrutura. A caracterização é focalizada nos aspectos topológicos das redes e não inclui, nesta fase, os aspectos funcionais como capacidade das conexões, hierarquia de nodos e latência no transporte dos dados.

O segundo objetivo é criar modelos que ajudem a entender o significado e o impacto dessas propriedades na estrutura e no funcionamento das redes. A produção de modelos sintéticos de redes que apresentem as mesmas características estruturais das redes reais é de grande valor para seu estudo, por meio de análise experimental e simulação.

O terceiro objetivo é prever o comportamento dessas redes com base nas propriedades estruturais medidas e nos seus aspectos funcionais. Há um relacionamento intrínseco entre a estrutura, a função e as propriedades das redes.

Um dos grandes desafios relacionados ao estudo das propriedades estruturais da Internet é o seu mapeamento topológico. A topologia da Internet pode ser descrita em

dois níveis: nível dos roteadores e nível dos sistemas autônomos. Um sistema autônomo é uma sub-rede que responde por um domínio da Internet e está sob um único domínio administrativo. A Internet é composta por dezenas de milhares de Sistemas Autônomos (ASes) [11] e as interconexões dinâmicas entre eles.

As redes complexas como a Internet podem ser representadas por grafos, que são compostos por nodos e arestas. Os nodos representam unidades do sistema e as arestas representam as interações entre os nodos. No caso da Internet, os ASes são representados pelos nodos e as conexões entre os nodos são representadas pelas arestas.

Não há um mecanismo na Internet para coletar e manter informações sobre sua topologia. A própria rede não oferece informação sobre a topologia. A topologia necessita, de fato, ser descoberta. Essa descoberta tem sido feita pela utilização de várias ferramentas de captura de dados de roteamento, que contêm informações sobre os ASes e os caminhos entre eles, e também a partir de informações obtidas em várias bases de dados. Esses métodos utilizam ferramentas e protocolos destinados a outras funções, gerando assim diversos problemas, dentre os quais se destacam incompletude dos dados da topologia e dificuldade de validação da topologia obtida. Além disso, métodos distintos obtêm topologias parcialmente diferentes, gerando discussões sobre a validade dos métodos e das topologias geradas.

No entanto, acredita-se que a topologia real não seja muito distinta da topologia descoberta. Há um constante aprimoramento dos métodos, e a cada dia a topologia descoberta se torna mais completa e próxima da topologia real. Uma das topologias mais completas e atualizadas é a elaborada pelo laboratório *Internet Research Laboratory* da *University of California* em Los Angeles, que será utilizada neste trabalho. Uma descrição completa dessa topologia, das formas de coleta de dados e problemas correlatos é apresentada na Seção 2.2.

1.2 Objetivo da Pesquisa

O objetivo deste trabalho é caracterizar e modelar a evolução da topologia da Internet no nível dos sistemas autônomos. Esse objetivo engloba a análise do crescimento e da

evolução da topologia por meio de métricas que a descrevam e o estudo da evolução da topologia ao longo dos últimos anos, uma vez que há dados disponíveis que permitem essa análise. Pode-se assim modelar o crescimento da rede e extrapolar a evolução caracterizada para tentar prever a evolução futura, isto é, tentar prever o crescimento e o desenvolvimento da rede no futuro.

A topologia da Internet pode ser descrita em dois níveis, no nível dos roteadores e no nível dos Sistemas Autônomos. Este trabalho apresenta uma caracterização da evolução da topologia da Internet, no nível dos sistemas autônomos, no período de 2004 a 2007. Várias métricas que caracterizam a topologia são analisadas e sua evolução nesse período é apresentada e discutida. A partir dos resultados da análise da evolução, foram construídos modelos que permitem fazer previsões sobre a evolução dessas métricas no período de 2008 a 2010.

1.3 Motivação

A topologia é a principal característica estrutural de uma rede, e compõe a estrutura a partir da qual todas as operações e funções nela se realizam. A caracterização e análise da topologia constituem o primeiro passo no estudo de uma rede complexa.

A Internet é uma rede dinâmica e em expansão, e sua topologia evolui ao longo do tempo. ASes são constantemente criados enquanto outros são desativados. Conexões entre ASes mudam freqüentemente em função de diversos aspectos relacionados à manutenção das redes, acordos comerciais, falhas nos enlaces, criação de novos SAs, extinção de SAs, dentre outros.

A evolução constante dos sistemas requer atividade contínua de caracterização para revelar seus padrões e identificar características passageiras e perenes ao longo do seu ciclo de vida. A Internet é um sistema que tem evoluído rapidamente, com alterações significativas em várias de suas medidas e características, incluindo sua topologia. O conhecimento obtido na análise da evolução da topologia da Internet é útil em várias áreas, por exemplo, no projeto de novas redes, no planejamento de expansão da rede atual, na avaliação de propostas de novos protocolos e aplicações, e no estudo das inter-

relações entre a evolução da Internet e aspectos econômicos e sociais do mundo atual.

O estudo sistemático da evolução da topologia da Internet contribui em diversos contextos. Em primeiro lugar, a partir de dados da evolução é possível fazer extrapolações e formular cenários para a topologia no futuro. A observação de comportamentos e características esperadas e não esperadas pode auxiliar a prever comportamentos indesejados no futuro, identificando falhas e restrições. Este estudo auxilia também a criação de novos modelos de geradores de topologias sintéticas, que se assemelhem mais à topologia real da Internet. Um outro aspecto de interesse é a geração de grafos que seriam pequenas amostras da Internet, mantendo suas propriedades estruturais para testes. A caracterização da topologia é também de interesse dos estudiosos da área de protocolos de roteamento, para o estudo de modificações e novas propostas.

1.4 Organização do Trabalho

Esta dissertação está organizada em cinco capítulos. O Capítulo 2 discute os conceitos e os trabalhos relacionados. A Seção 2.2 apresenta a descrição das fontes de dados e a caracterização básica das topologias reais. A Seção 2.3 apresenta discussões sobre as métricas utilizadas para a análise da rede, como distância, excentricidade, raio, diâmetro e outras. A Seção 2.4 discute métodos de aproximação utilizados para a extrapolação das métricas, assim como a apresentação da ferramenta utilizada para a geração das equações de modelagem. O Capítulo 3 discute resultados iniciais da análise da topologia de acordo com as métricas apresentadas na Seção 2.3. O Capítulo 4 apresenta resultados relacionados à aplicação das equações modeladas aos resultados obtidos no Capítulo 3. No Capítulo 5 são apresentadas as conclusões obtidas a partir do estudo.

CAPÍTULO 2

CONCEITOS E TRABALHOS RELACIONADOS

2.1 Trabalhos Relacionados

A importância e a complexidade da estrutura de roteamento da Internet têm aumentado o interesse no entendimento da sua topologia e de seu desempenho ponto a ponto. Grande parte dos estudos anteriores constituiu na construção de um grafo da topologia atualizado e mais próximo da realidade da rede, com a utilização de várias ferramentas para captura de rotas, permitindo coletas de quantidades de dados cada vez maiores. A utilização de vários métodos de captura de informações sobre o roteamento é uma das abordagens mais utilizadas atualmente.

Alguns estudos analisaram as relações lógicas entre os nodos da rede, criando assim uma estrutura de relacionamento entre os ASes. Grande parte dessas relações lógicas é apresentada na forma de hierarquia, e a relação é inferida a partir de informações pertinentes à comunicação. Assim, é possível descobrir e validar relações comerciais entre os nodos. Por exemplo, a partir de dados da tabela *Border Gateway Protocol* [26](BGP) publicados pelo projeto *RouteViews* [28], as relações entre os nodos foram inferidas [7]. Naquele estudo as informações obtidas foram posteriormente validadas com dados internos da AT&T, apresentando um fator de correção acima de 90% para os dados inferidos. Relações hierárquicas também são inferidas a partir das informações obtidas do protocolo BGP em [32], no qual os autores utilizaram uma medida para classificação dos nodos, dividindo assim o grafo em cinco diferentes classes hierárquicas.

Um grande conjunto de métricas pode ser utilizado para caracterizar a topologia dos sistemas autônomos da Internet, assim como a topologia de várias redes complexas [17, 5, 2, 20]. Não há um consenso na comunidade sobre quais são as métricas mais importantes e representativas da topologia. As métricas mais analisadas referem-se à distribuição dos graus dos vértices do grafo, as distâncias entre os vértices e métricas derivadas, e métricas

de conectividade.

Vários trabalhos contribuíram para revelar medidas e características da topologia dos sistemas autônomos da Internet [23, 4, 2, 20, 24]. Um exemplo da importância da caracterização é a descoberta da grande variabilidade na distribuição dos graus dos vértices da rede [6], resultado que teve forte influência em várias pesquisas [31, 22, 17, 20, 3, 25].

O trabalho que mais se relaciona ao presente estudo é o trabalho apresentado em [19]. Os autores da pesquisa apresentam uma caracterização detalhada da topologia da Internet em nível de sistemas autônomos. A partir de dados da topologia obtidos em bases de dados de roteamento BGP, coletados no período de novembro de 1997 a maio de 2000, o estudo avaliou o resultado de várias métricas que apresentam a possibilidade de caracterizar a rede, de forma evolucionar, durante um período de cerca de dois anos e meio. É feita a identificação de métricas com característica invariante durante a evolução da rede. Também são apresentadas algumas leis que foram obtidas a partir da análise dos resultados das métricas utilizadas. Tais leis foram denominadas empíricas e se mostraram válidas durante a evolução da rede no período analisado.

Nesse trabalho argumentamos que métricas relacionadas à distância são fundamentalmente mais importantes do que a métrica mais estudada e discutida no momento: a distribuição dos graus dos vértices [6, 20].

Para realizar esse estudo, utilizamos uma base de dados da topologia real da Internet. A base de dados escolhida é discutida em [34] e é uma base bem completa da topologia da Internet, obtida a partir de várias fontes e atualizada diariamente, contendo dados a partir de 2004. Além de calcular as diversas métricas para a topologia atual, estudamos, também, a evolução da topologia quanto a essas métricas, no período entre 2004 e 2007.

Resultados obtidos indicam que a rede é compacta e demonstra ser densamente conectada. A topologia apresenta as características “livre de escala”, que indica grande variabilidade nos graus dos vértices, e “mundo pequeno”, que significa que são observadas distâncias muito pequenas entre pares de vértices. A distribuição das distâncias, o raio e o diâmetro são métricas invariantes que capturam propriedades fundamentais da topologia. A rede apresenta árvores nas bordas cuja altura é, no máximo, igual a 2.

2.2 Base de Dados da Topologia

A base de dados da topologia da Internet utilizada neste trabalho é denominada IRL. As duas próximas seções discutem, nesta ordem, os aspectos da coleta de dados da topologia da rede, abordando os problemas e os resultados alcançados, e a descrição da base IRL e suas componentes.

2.2.1 Coleta de Dados

Várias pesquisas têm sido realizadas com o objetivo de capturar a maior quantidade de dados pertinentes à topologia da Internet em nível de sistemas autônomos [4, 13, 19, 20, 34], utilizando diferentes métodos, tecnologias e ferramentas para obtenção de dados topológicos.

Dentre os métodos mais comumente utilizados estão a observação de tabelas de roteamento e de mensagens *update* do protocolo BGP (*Border Gateway Protocol*), a utilização de ferramentas de captura de rotas ativas por mensagens de controle do protocolo TCP, feito pela ferramenta *traceroute* [1] e o uso de bases de dados da topologia cujas informações são inseridas e mantidas manualmente como na base *Internet Routing Registry* IRR [14].

Após a coleta das informações, estas são inseridas em bases de dados para que possam ser reutilizadas em futuras análises. As fontes de dados da topologia da Internet mais conhecidas e utilizadas atualmente são *RouteViews* [28], IRR, servidores *Looking Glasses* [1] e WHOIS [14].

Devido aos diferentes métodos utilizados para a captura de dados em cada uma das bases topológicas, a diferença de informações de uma para a outra é significativa [34, 4, 7, 12]. Sendo assim é interessante fazer a união dos dados das várias bases, considerando que todas as bases coletam dados de uma única topologia. Esta união revela um crescimento significativo tanto na descoberta de novas arestas entre os ASes quanto no aparecimento de novos ASes, em relação à topologia observada em uma única base. Esta abordagem vem sendo utilizada por vários pesquisadores.

A base de dados escolhida para esse trabalho foi a base elaborada pelo laboratório *Internet Research Laboratory* da *University of California, Los Angeles*, base essa denominada neste trabalho como base IRL [13]. A composição dessa base é descrita na seção a seguir.

2.2.2 A Base IRL

O principal objetivo da base IRL é capturar a topologia da Internet no nível dos ASes, utilizando bases de dados já existentes, para criar uma topologia mais completa e de fácil uso. A base IRL foi criada visando quatro características principais:

- **Completeness:** a topologia deve ser coletada a partir da maior quantidade possível de interdomínios, utilizando o estado da arte em métodos de captura de dados topológicos, com o objetivo de gerar a base mais completa possível;
- **Anotabilidade:** informações devem ser incluídas para auxiliar o uso da topologia, como a adição de informações sobre os tipos de nodos e arestas, assim como em qual base tais informações foram obtidas, facilitando a utilização dos dados;
- **Acessibilidade:** o acesso fácil à base pelos usuários é essencial para uma boa aceitação e utilização; a base pode ser facilmente acessada utilizando o protocolo HTTP, facilitando assim a automatização da análise;
- **Atualização:** a base é atualizada diariamente, mantendo informações disponíveis para utilização a partir de janeiro de 2004.

A seguir são descritas as bases de dados e ferramentas de captura utilizadas para compor a base IRL.

BGP Trace Collector é um sistema de medição que realiza a comunicação com redes comerciais de provedores de acesso (ISPs) utilizando seções BGP. O coletor apenas recebe mensagens BGP, não realizando nenhuma forma de aviso de rotas. Tipicamente a tabela de roteamento de um coletor possui mais de cem mil entidades, uma para cada faixa de prefixo alcançável pelo AS. Assim, o coletor possui uma “visão” da topologia da Internet.

Quanto maior a comunicação com os sistemas autônomos, maior e mais completa será a topologia capturada. Os dois maiores e mais importantes projetos de medição utilizados são o *Route Views* e o RIPE RIS [27], que tornaram públicas as informações de roteamento BGP. Essas duas bases juntas possuem cerca de 288 visões distintas da rede.

Route Servers [26] são roteadores acessíveis publicamente por algumas redes de provedores de acesso, com o objetivo de ajudar a solucionar alguns problemas da rede. Como o acesso a tais servidores é público, usuários podem fazer a utilização dos servidores por meio de `telnet` e facilmente obter informações sobre o roteamento utilizando o comando `show ip bgp`, que exibe a tabela de roteamento interna do servidor.

Looking Glasses [1] provê uma interface por meio da Web que permite a execução de um conjunto limitado de comandos. Tal interface permite aos usuários a verificação de rotas para prefixos particulares, mesmo não permitindo que o usuário tenha acesso a toda a tabela de roteamento do servidor. Ainda assim alguns *Looking Glasses* permitem a execução do comando `show ip bgp summary`, gerando informações que possibilitam a observação de seus vizinhos.

O principal propósito do IRR é o de permitir que operadores possam coordenar políticas globais de configuração. A base formada pelo IRR é operada manualmente por voluntários, podendo portanto conter informações incorretas, incompletas ou até mesmo desatualizadas. A porção RIPE da base IRR descreve a Internet européia e é usada com frequência pelos provedores de acesso europeus para a filtragem de anúncios de rotas, sendo que muitos IXP (*Internet Exchange Points* [15]) requerem que seus operadores registrem as informações de roteamento no RIPE. Dessa maneira, a porção RIPE é considerada a porção mais confiável de toda a base IRR.

Conforme os dados obtidos nos estudos [13, 23] foi que a base IRL contém cerca de 40% mais arestas e 5% mais ASes no grafo da topologia produzida em comparação ao grafo gerado pelos dados do *Route View* BGP, comparando grafos capturados em 24 de outubro de 2004.

2.3 Métricas para Análise da Topologia

A topologia da Internet em nível de sistemas autônomos (ASes) pode ser representada utilizando grafos não direcionados, nos quais os vértices representam ASes e as arestas representam as ligações lógicas entre os sistemas autônomos. A modelagem da topologia é feita de forma natural por meio de grafos, tornando possível a utilização de métricas de caracterização de grafos para a caracterização de redes complexas como a Internet. A seguir são descritas várias métricas utilizadas para análise do grafo da topologia da rede.

2.3.1 Grafo

A estrutura básica para a representação da topologia das redes complexas é a estrutura denominada de grafo [16]. Um grafo ou um grafo não direcionado G é definido como um par ordenado $G = (V, E)$ o qual satisfaz as seguintes condições:

- V é um conjunto de elementos não ordenados, denominados de nodos ou arestas.
- E é um conjunto de pares distintos de elementos que pertencem a V , tais pares são denominados arestas, *links*, conexões ou ligações.

2.3.2 Caminho Simples e Menor Caminho

As definições de caminho simples e de menor caminho em um grafo são necessárias para definir a distância entre dois pontos em um grafo. Um caminho simples de um vértice u a um vértice v , de tamanho n , em um grafo G , é uma seqüência ordenada e distinta de $n+1$ vértices $u = v_0, v_1, \dots, v_n = v$ que satisfaz $(v_i, v_{i+1}) \in E(G)$ para todo $i = 0, 1, \dots, n-1$, onde $E(G)$ é o conjunto de arestas do grafo G . O valor de n representa o número de arestas entre os vértices.

O menor caminho de u a v , em um grafo, é o caminho simples no qual o número n de arestas é mínimo, dentre todos os caminhos simples possíveis iniciados em u e terminados em v .

2.3.3 Distância

A distância é uma das métricas mais importantes relacionadas a grafos. A distância $d(u, v)$ entre dois nodos u e v é o número n de arestas do menor caminho simples entre os vértices u e v .

A partir da distância, várias outras métricas importantes são definidas, como, excentricidade, raio e diâmetro, descritos a seguir.

2.3.4 Excentricidade e Métricas Derivadas

A excentricidade de um nodo u é a maior distância entre u e v , considerando todos os pares de vértices $u, v \in V$ e $u \neq v$, no qual V é o conjunto de vértices do grafo. Com base na excentricidade de todos os nodos, é possível calcular o raio e o diâmetro do grafo. O raio R de um grafo G é a menor excentricidade encontrada para todos os nodos de G . O diâmetro D de um grafo G é a maior excentricidade para todos os nodos de G . A medida representativa, tanto para o raio quanto para o diâmetro, é bem intuitiva e fácil de ser observada.

Pode-se definir duas métricas a partir do raio e do diâmetro da rede, são elas: o conjunto do centro e o conjunto da periferia de um grafo. O conjunto do centro de um grafo é formado por todos os nodos que apresentam excentricidade igual ao raio da rede. Desta forma, o conjunto da periferia é formado por todos os nodos que apresentam excentricidade igual ao diâmetro da rede. Todos os nodos que não pertencem a nenhum desses conjuntos são referidos como nodos que estão entre o centro e a periferia. Como os nodos mais centrais do grafo apresentam menor excentricidade em relação a todos os outros, tais vértices estão em média mais próximos dos outros vértices do grafo. Já os nodos periféricos possuem um caminho mínimo mais longo, para outros nodos periféricos, que percorre o grafo de uma extremidade a outra.

2.3.5 Distribuição dos Graus dos Vizinhos

A distribuição dos graus é uma das maneiras de apresentar a coleção de graus e considerar a frequência dos graus dos nodos do grafo. Seja n_k , o número de vértices de grau k , a distribuição do grau de G pode ser representada como $\langle n_0, n_1, \dots, n_t \rangle$ onde t denota o maior grau em G . Supondo que a distribuição do grau $\langle n_0, n_1, \dots, n_t \rangle$ de um grafo G satisfaça a condição de que n_k é proporcional a $k^{-\beta}$, para algum número $\beta > 1$ fixo, então é dito que a distribuição dos graus em G pode ser representada por uma função na forma de uma lei de potência (*Power Law*) com expoente β .

2.3.6 Árvore nas Bordas

Para definir a métrica de árvore nas bordas primeiro definimos um subgrafo acíclico. Um subgrafo $G = (V, E)$ é chamado acíclico quando G possui n vértices e $n - 1$ arestas e para todo nodo $u, v \in V$ existe apenas um único caminho simples ligando u a v . Grafos acíclicos são também conhecidos como árvores.

Uma árvore na borda $K = (E_k, V_k)$ com raiz em r , de um grafo $G = (E, V)$, é uma árvore pertencente a uma floresta F formada pela extração iterativa de nodos $w \in V$ e arestas $(v, w) : v, w \in E$, no qual o nodo retirado w de G possui grau igual a 1. Tais árvores também incluem qualquer nodo $r \in V$ adjacente a uma ou mais raízes p de tais árvores pertencentes a F . As árvores são compostas pelos nodos retirados. O grafo G' restante após tal retirada iterativa é chamado de grafo *2-core*, pois possui apenas nodos com grau maior ou igual a 2.

A idéia mais intuitiva sobre as árvores nas bordas é a de que tal grafo $K = (E_k, V_k)$ é formado por um subconjunto acíclico maximal de E e V tal que toda aresta $(u, v) \in E_k$ se $(u, v) \in E$.

A Figura 2.1 ilustra um exemplo de um grafo $G = (E, V)$. Retiraram-se todos os nodos da rede G que contém grau igual a 1. No caso desse exemplo foram extraídos os nodos a, c, e e f na primeira iteração e os nodos b e g na segunda iteração, gerando o subgrafo G' e G'' apresentados nessa ordem na Figura 2.2.

O passo final na composição das árvores nas bordas é o da inclusão da raiz das árvores

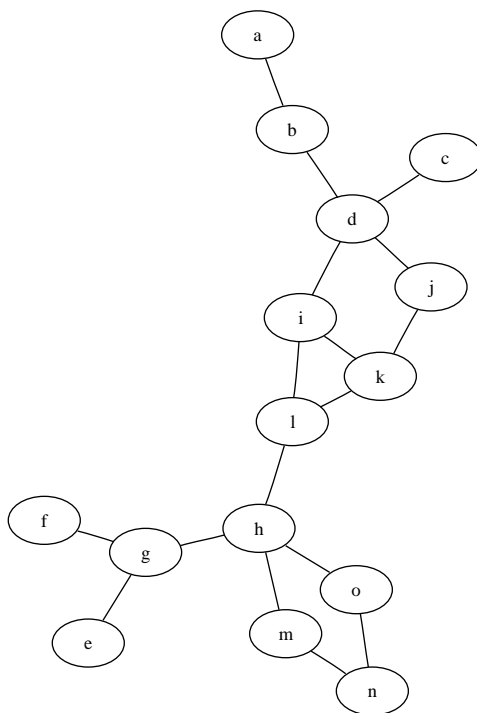


Figura 2.1: Grafo ilustrativo de G .

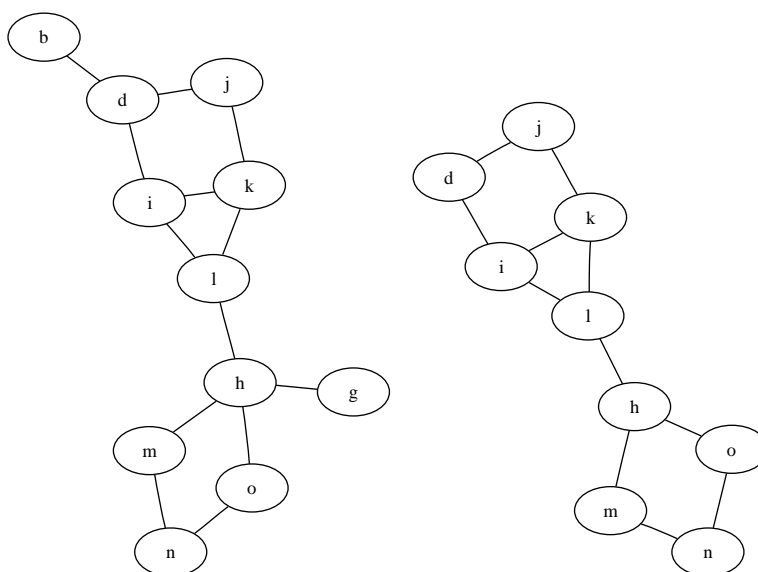


Figura 2.2: Grafo ilustrativo G' , após a deleção na primeira iteração dos vértices com grau igual a 1 (Esquerda). Grafo ilustrativo G'' , após a deleção na segunda iteração dos vértices com grau igual a 1 (Direita).

as quais são incidentes a um ou mais nodos na floresta. As árvores nas bordas obtidas podem ser observadas na Figura 2.3.

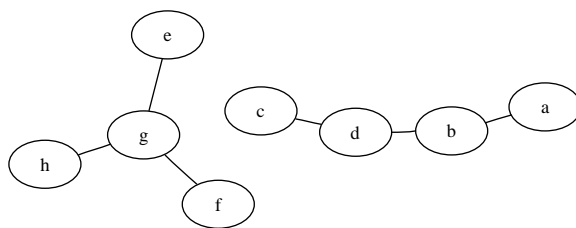


Figura 2.3: Árvores nas bordas encontradas no grafo G' , com raiz em h e d .

2.3.7 Coeficiente de Agrupamento

O coeficiente de agrupamento de um grafo G quantifica quão próximo $G = (E, V)$ está de ser um grafo completo. O coeficiente de agrupamento é calculado em relação a cada nodo do grafo e simboliza o quanto um nodo e seus vizinhos estão próximos de formarem uma clique.

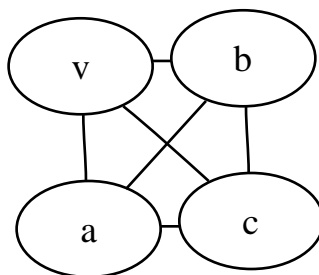


Figura 2.4: Grafo G exemplificando uma rede com coeficiente de agrupamento igual a 1.

O coeficiente de agrupamento C_v para um vértice $v \in V$ é calculado com base em duas funções, $\lambda(v)$ e $\tau(v)$. A função $\lambda(v)$ representa o número de triângulos formados com $v \in V$ para um grafo não direcionado G . Já a função $\tau(v)$ é igual à quantidade de subgrafos que possuem duas arestas e três vértices, no qual o nodo v é um dos três vértices e v é incidente aos outros dois nodos restantes.

Sendo assim, o coeficiente de agrupamento de um nodo i é igual a:

$$C_i = \frac{\lambda_G(i)}{\tau_G(i)}$$

Tendo o cálculo de C_v para todo $v \in V$ de G , o coeficiente de agrupamento do grafo G é definido como a média aritmética da soma de cada C_v , para todo $v \in V$, pelo número

total de nodos. Logo, o coeficiente de agrupamento \bar{C} de G é dado por:

$$\bar{C} = \frac{1}{|V|} \sum_{\forall v \in V} C_v$$

A Figura 2.4 ilustra um exemplo do cálculo do coeficiente de agrupamento para o grafo apresentado. Assumindo um grafo completo que possui 4 vértices nomeados a, b, c e v , $\lambda(v)$ é igual ao número de triângulos formados em v , nesse caso $\lambda(v) = 3$, que são formados pelos vértices (b, v, c) , (b, v, a) e (a, v, c) . O número de subgrafos com 2 arestas e 3 vértices sendo v incidente a ambos os nodos é igual a $\tau(v)$ para este exemplo $\tau(v) = 3$, são formados pelos nodos (b, v, a) , (b, v, c) e (a, v, c) . Dessa maneira $C_v = 3/3$, e de maneira análoga $C_a = 3/3$, $C_b = 3/3$ e $C_c = 3/3$. O coeficiente de agrupamento da rede é definido pela razão entre a soma de todos os coeficientes de agrupamento para cada nodo do grafo pelo número de vértices. Logo

$$C = \frac{3/3 + 3/3 + 3/3 + 3/3}{4} = 1$$

o valor resultante da métrica indica que a rede possui relação de 1 para 1 de formar um grafo completo.

2.4 Modelagem dos Dados

A teoria da aproximação [30] tem como objetivo o estudo de métodos matemáticos capazes de aproximar funções complexas, funções estas com relações e interações entre as incógnitas pouco previsíveis, em funções mais simples com poucas e conhecidas incógnitas, possibilitando o estudo de conjuntos de dados de uma maneira mais qualitativa e mensurável.

Na área de redes, a computação de modelos ou aproximações em relação ao conjunto de dados obtidos a partir da topologia real capturada é extremamente importante para a previsão e análise de comportamento da rede. Tais informações podem ser utilizadas no campo de redes complexas, para a criação de simuladores e a possibilidade da caracter-

ização da rede.

Tendo como fato que neste trabalho estamos interessados principalmente na caracterização e previsão do comportamento da rede, os conjuntos de dados a serem modelados na computação representam métricas para caracterização do sistema, provendo informações sobre o possível comportamento da topologia para os anos seguintes. A utilização de métricas de caracterização também possibilita a identificação de métricas invariantes que possam ser verificadas durante o estudo. Essas métricas invariantes são de grande importância para o campo de caracterização de redes complexas.

2.4.1 Ferramenta Utilizada

O estudo de equações matemáticas para modelar os conjuntos de dados obtidos a partir da captura da topologia da Internet foi feito com a utilização de um programa especializado na geração de equações. A ferramenta utilizada foi o aplicativo *Curve Fitting Tool* da seção de "Toolboxes" do programa MATLAB 7.10.246 (R14) *Service Pack 3*. Esse programa possui a capacidade de gerar e aproximar retas, curvas e outros tipo de equações a partir de um conjunto de dados apresentados pelo usuário. Neste trabalho, os dados utilizados representam as métricas de caracterização da topologia.

Os resultados obtidos pelo programa a partir das entradas de dados obtidas da topologia são equações que visam modelar resultados dos valores reais encontrados. Cada equação gerada foi calculada utilizando diferentes métodos de interpolação e regressão, que serão descritos posteriormente. Para cada equação obtida foi também calculado o intervalo de confiança, com não menos que 99% de confiança.

O intervalo de confiança está ligado diretamente aos coeficientes da equação gerada. Para cada modelo são gerados dois coeficientes distintos A e B . A substituição dos coeficientes na equação do modelo escolhido produz duas equações distintas, equação EqA e EqB , que são equações com os coeficientes A e B , nessa ordem. A área formada entre as equações EqA e EqB é chamada de intervalo de confiança com $k\%$ de confiança. Tal área representa o intervalo no qual há a probabilidade de $(100 - k)\%$ de uma observação não pertencer à área formada pelas curvas modeladas, considerando o menor e o maior coefi-

ciente. Nesta dissertação, todos os conjuntos de confiança foram gerados com confiança não menor que 99%, o que indica que um valor observado pode estar fora deste intervalo com probabilidade máxima de 1%.

2.4.2 Funções de Modelagem

Para gerar os modelos para as métricas características da rede foram utilizadas quatro diferentes equações, com o objetivo de obter o melhor modelo possível do conjunto de dados, o que possibilita uma previsão mais precisa dos valores das métricas para períodos futuros.

Os quatro diferentes tipos de equações foram:

- Modelagem por Polinômio - Esse modelo tem como resultado um polinômio da seguinte forma:

$$a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \dots + a_0 = 0, \quad a_n \neq 0$$

sendo que n é chamado de grau do polinômio e os coeficientes a_k são números reais.

Algumas vantagens são obtidas na utilização de equações polinomiais como, por exemplo, a propriedade de fechamento das quatro operações fundamentais do cálculo sobre o conjunto dos polinômios, facilitando assim a análise e a manipulação das equações obtidas e tendo em vista seu comportamento bem definido. Outra característica é a possibilidade da existência de várias inflexões na curva de aproximação, o que traz vantagem de uma melhor aproximação dos dados obtidos.

- Modelagem por Potência - A modelagem por potência tem como resultado uma equação de potência da seguinte forma:

$$ax^b + c, \quad a \neq 0$$

O valor a representa a multiplicidade da incógnita x elevada à potência b da equação,

sendo o valor c relacionado ao deslocamento da curva de potência no eixo das ordenadas.

- Modelagem Racional Quadrática - Esse modelo tem como resultado a divisão de dois polinômios de segundo grau com coeficientes reais. O termo racional na equação aproximada vem da sua semelhança com o conjunto de números racionais $Q = \{\frac{a}{b} \mid a \in Z, b \in Z^*\}$. O conjunto de números racionais é representado pelos números que são obtidos pela divisão de a por b sendo que $a \in Z$ e $b \in Z^*$. Sendo assim, a equação racional quadrática tem a forma $\frac{a}{b}$ tal que

$$a = P_1x^2 + P_2x + P_3$$

$$b = x^2 + q_1x + q_2$$

na qual P_1, P_2, P_3, q_1 e q_2 são coeficientes reais e a divisão de $\frac{a}{b}$ é um número racional.

- Modelagem Logarítmica - A modelagem por curva logarítmica foi escolhida nesse trabalho, tendo em vista o decrescimento característico de certas distâncias entre pares de vértices na rede. Uma curva logarítmica é representada pela seguinte equação:

$$k \log_b(x) + a$$

na qual os valores reais k e b determinam a curvatura da aproximação e o valor a determina o deslocamento da curva aproximada no eixo das ordenadas.

2.4.3 Síntese do Capítulo

Estudos da grande variabilidade do grau dos vértices, assim como trabalhos baseados nas métricas de distância e suas derivadas, vêm apresentando indicações de várias tendências, que possibilitam a caracterização da rede no campo de redes complexas. Para que a análise e caracterização fosse possível foi utilizada a base IRL, que possui características relevantes em relação as outras bases topológicas disponíveis. Para a caracterização da

topologia, foram escolhidos um conjunto de métricas e modelos os quais os resultados serão apresentados posteriormente e comparados com estudos relacionados para a validação das métricas e atualização de previsões realizadas sobre o comportamento da rede durante sua evolução.

CAPÍTULO 3

CARACTERIZAÇÃO DA EVOLUÇÃO DA TOPOLOGIA

Este capítulo apresenta uma caracterização da evolução da topologia dos sistemas autônomos da Internet no período de 2004 a 2007. Conforme citado anteriormente, a base de dados dessa topologia foi coletada e tratada pelo *Internet Research Laboratory* da UCLA [13]. Esta base é denominada IRL neste texto. Foram coletadas duas amostras da topologia por ano, uma em janeiro e outra em julho, sempre no dia 30 dos referidos meses. Assim, a análise apresentada é baseada em oito amostras, coletadas semestralmente.

A topologia da rede é representada na forma de um grafo $G = (V, E)$, em que V representa o conjunto de vértices e E o conjunto de arestas. Cada vértice representa um sistema autônomo e as arestas representam os caminhos lógicos ou rotas entre estes sistemas. Os resultados para as métricas analisadas são baseados em medições nesse grafo.

As métricas analisadas caracterizam a composição geral do grafo, as distâncias e suas derivadas da excentricidade, raio e diâmetro, métricas de status e de agrupamento. Estas métricas representam as características mais peculiares do grafo da topologia, além de serem de fácil entendimento e permitirem um mapeamento intuitivo e imediato em relação às funções da rede.

3.1 Composição do Grafo da Topologia dos Sistemas Autônomos

Na análise de um grafo, a primeira tarefa é verificar o número de componentes do grafo. É intuitivo pensar que o grafo dos sistemas autônomos é composto por um único componente gigante. A análise do número de componentes de cada amostra da base IRL no período de 2004 a 2007 revelou a existência de quatro componentes desconexas, sendo um componente gigante e três vértices isolados. Analisamos apenas o maior componente para obter todos os resultados apresentados neste trabalho.

A próxima análise refere-se à composição geral do grafo em termos do número de

vértices e arestas. A Tabela 3.1 apresenta estes números para as oito amostras da rede, coletadas no período de janeiro de 2004 a julho de 2007. Observa-se um crescimento contínuo e inequívoco da topologia nos últimos quatro anos.

Data	Vértices	Arestas
01/2004	16.573	39.143
07/2004	18.783	62.305
01/2005	20.830	76.976
07/2005	23.062	92.910
01/2006	24.971	106.141
07/2006	27.071	121.045
01/2007	29.167	136.888
07/2007	31.328	156.871

Tabela 3.1: Número de vértices e arestas em cada amostra da topologia.

A evolução do número de vértices e arestas pode também ser observada na Figura 3.1. Os cálculos indicam um aumento anual de cerca de 3.300 sistemas autônomos e 22.000 arestas, totalizando um crescimento de 189% no número de vértices e 400% na quantidade de arestas durante o período de observação.

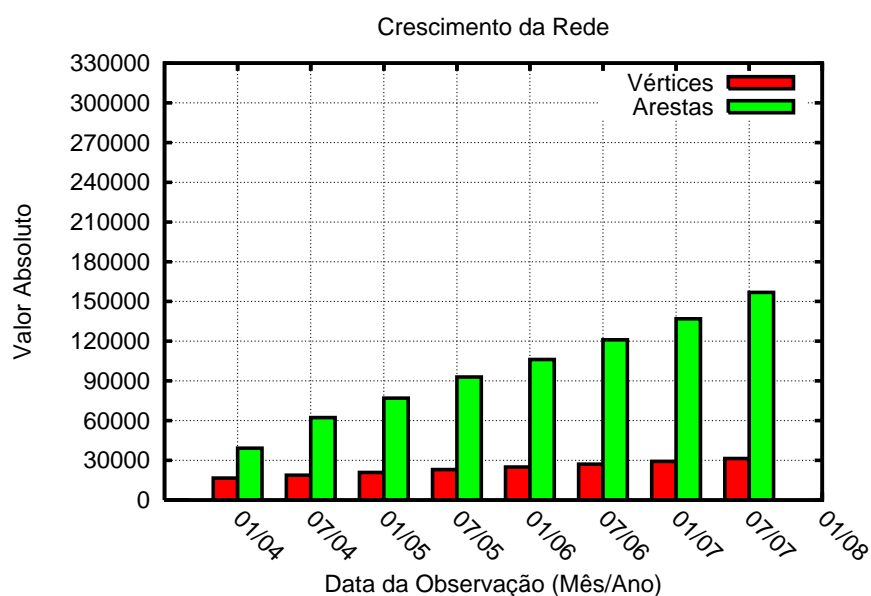


Figura 3.1: Crescimento do número de vértices e arestas no período de 2004 a 2007.

O gráfico da Figura 3.2 apresenta a evolução da relação entre a quantidade de vértices e arestas nas amostras do grafo da topologia capturadas entre janeiro de 2004 e julho

de 2007. Ambos os eixos do grafo estão na escala logarítmica. Cada ponto do gráfico representa uma amostra da rede, com seu respectivo número de nodos no eixo x e o número de arestas no eixo y . Aparentemente a evolução da relação entre o número de vértices e o número de arestas é linear, em especial nas amostras mais recentes. Esta relação é modelada e mais explorada no Capítulo 4.

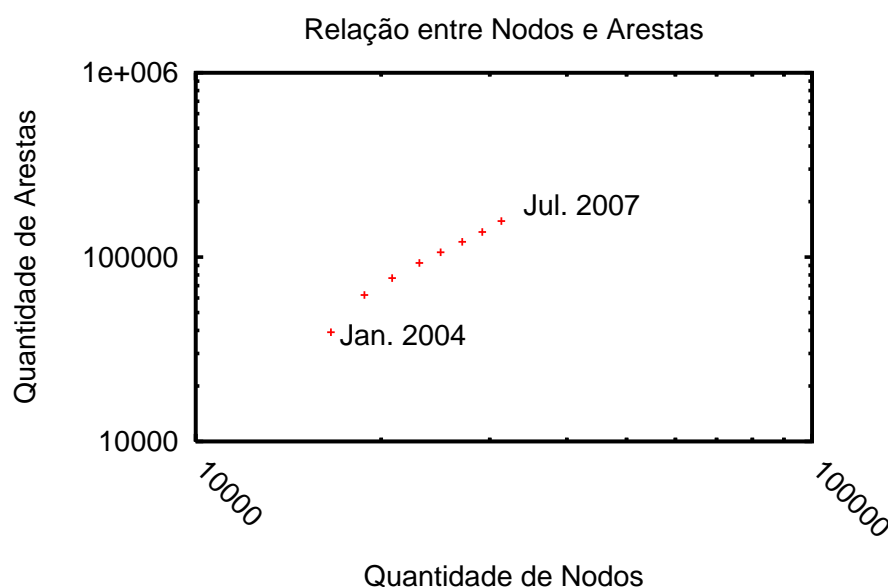


Figura 3.2: Evolução da relação entre o número de vértices e de arestas no período de janeiro de 2004 a julho de 2007.

Continuando a análise da composição do grafo, optamos por avaliar a composição da borda do grafo, utilizando um algoritmo que será descrito posteriormente na sub sessão 3.6. Esta análise tem o objetivo de identificar se há muitas ramificações na borda em forma de árvores que “pendem” da estrutura principal, qual é a altura destas árvores e o número de nodos e arestas envolvidos neste entorno. Estas árvores na borda compõem os subgrafos acíclicos encontrados na topologia.

Para fazer esta medição, implementamos um algoritmo para retirar todos os vértices de grau 1 e suas respectivas arestas. A seguir, o mesmo procedimento foi repetido até que nenhum nodo de grau 1 fosse encontrado. Em todas as amostras da topologia, observamos que foram necessárias apenas duas iterações para eliminar todos os subgrafos acíclicos da rede. Portanto, as árvores presentes na borda têm altura máxima igual a 2.

A caracterização geral da topologia dos sistemas autônomos mostra que a rede contém

um grande componente conectado, e este componente contém árvores nas bordas de altura máxima igual a dois. Além disso, a rede cresceu de forma consistente nos últimos quatro anos. Neste período o número de sistemas autônomos praticamente dobrou, enquanto o número de arestas foi multiplicado por quatro, aproximadamente. Esta relação indica que a rede está ficando cada vez mais conectada, e que a conectividade aumenta de forma mais acelerada do que seu tamanho.

3.2 Análise dos Graus dos Vértices

A análise dos graus dos vértices é uma caracterização essencial da topologia. Alguns índices estatísticos dos graus dos vértices para todas as amostras são apresentados na Tabela 3.2, a saber, a mediana, a média, o maior grau observado e o coeficiente de variação do grau dos vértices (COV). O coeficiente de variação é definido como a razão entre o desvio padrão e a média, e é uma medida adimensional da variabilidade do conjunto.

Data	Mediana	Média	Maior	COV
01/2004	2	4,72	2.388	4,72
07/2004	2	6,63	2.917	6,37
01/2005	2	6,36	3.205	6,22
07/2005	2	8,05	3.505	6,07
01/2006	3	8,50	3.710	6,02
07/2006	3	8,94	3.908	5,96
01/2007	3	9,38	4.122	5,90
07/2007	3	10,01	4.989	5,99

Tabela 3.2: Estatísticas dos graus dos vértices para as amostras da topologia.

Observa-se que a média, a mediana e o maior grau cresceram de forma consistente no período de observação. O COV se manteve em torno do valor 6, exceto na primeira amostra. O gráfico da Figura 3.3 apresenta a evolução da média e da mediana dos graus dos vértices. No período da observação, a mediana apresentou uma alteração discreta, de 2 para 3. O grau médio aumentou de 5 para próximo de 10, o que é uma alteração significativa. A pequena variação da mediana indica um pequeno aumento no grau dos nodos menos conectados, enquanto o aumento da média pode ser explicado pelo maior crescimento dos graus dos nodos mais conectados. O aumento do grau médio é citado como

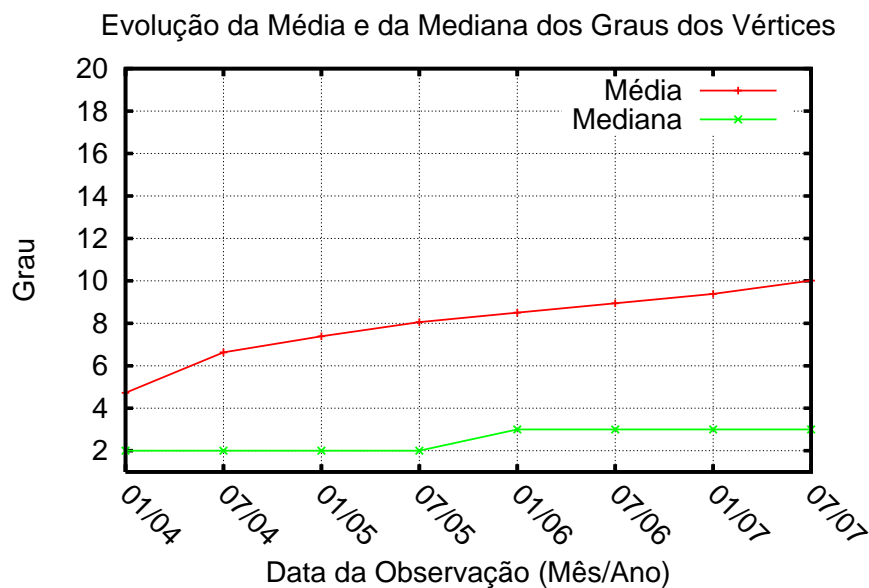


Figura 3.3: Evolução da média e da mediana dos graus dos vértices.

uma característica da evolução de várias redes complexas [18]. Observa-se também que a variabilidade dos graus não aumentou no período, mesmo com o crescimento expressivo da rede e do grau médio. Tendências como o aumento do maior grau dos vértices e da pouca variabilidade no coeficiente de variação do grau podem ser observadas na Figura 3.4 que apresenta a CDF do grau dos vértices e na Figura 3.5 que ilustra a CCDF durante o período da observação.

3.3 Distâncias entre Pares de Vértices

Muitos aspectos do desempenho de protocolos de roteamento e de aplicações dependem da distribuição da distância entre pares de vértices. Em particular, distâncias pequenas e pouco variáveis são desejáveis em algoritmos de roteamento. Porém, distâncias curtas tornam a rede mais vulnerável a ataques baseados em propagação de agentes. A distância caracteriza também o efeito denominado “mundo pequeno”, indicando que os vértices estão conectados por caminhos bem curtos.

A distância entre um par de vértices em um grafo é definida como o caminho mais curto entre os dois vértices, e é expressado pelo número de arestas entre os vértices. Para analisar o conjunto de distâncias foi implementado o algoritmo de Dijkstra, que encontra

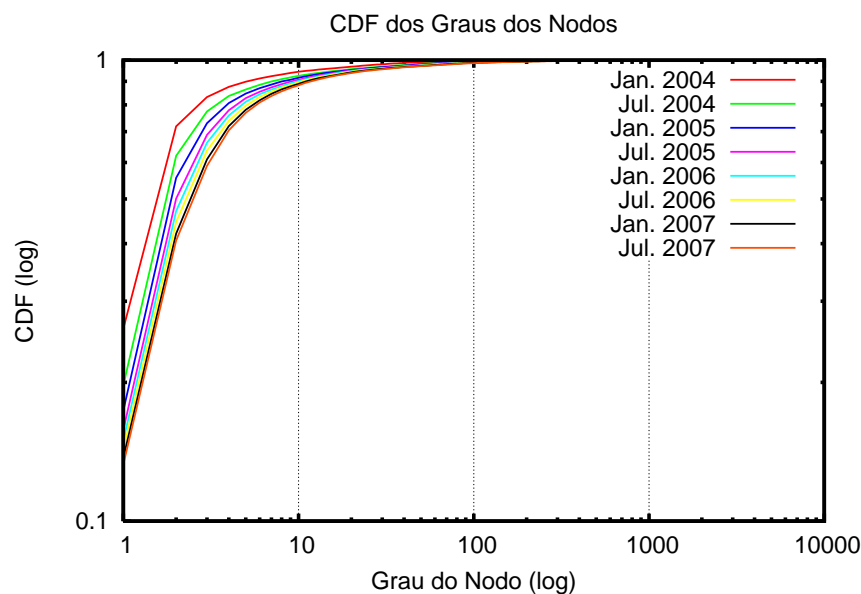


Figura 3.4: CDF do grau dos vértices.

a menor distância entre dois vértices quaisquer de um grafo. Em um grafo não dirigido com n vértices há $n(n-1)/2$ pares de vértices e suas respectivas distâncias.

A Tabela 3.3 apresenta estatísticas da distância entre pares de vértices. Observa-se que os vértices estão a distâncias muito pequenas. A maior distância observada é de apenas 10 arestas, um número muito pequeno considerando que as redes contêm milhares de vértices e arestas. A distância média está em torno de 3. A variabilidade, expressa pelo COV, é baixa. Resultados similares para a distância média e a maior distância são reportados na literatura [8].

Data	Média	Maior	COV
01/2004	3,708	10	0,241
07/2004	3,561	9	0,230
01/2005	3,531	9	0,230
07/2005	3,492	9	0,224
01/2006	3,480	9	0,221
07/2006	3,471	9	0,218
01/2007	3,461	10	0,215
07/2007	3,431	9	0,215

Tabela 3.3: Evolução de índices estatísticos das distâncias no período.

O gráfico da Figura 3.6 apresenta a evolução de índices estatísticos da distância no

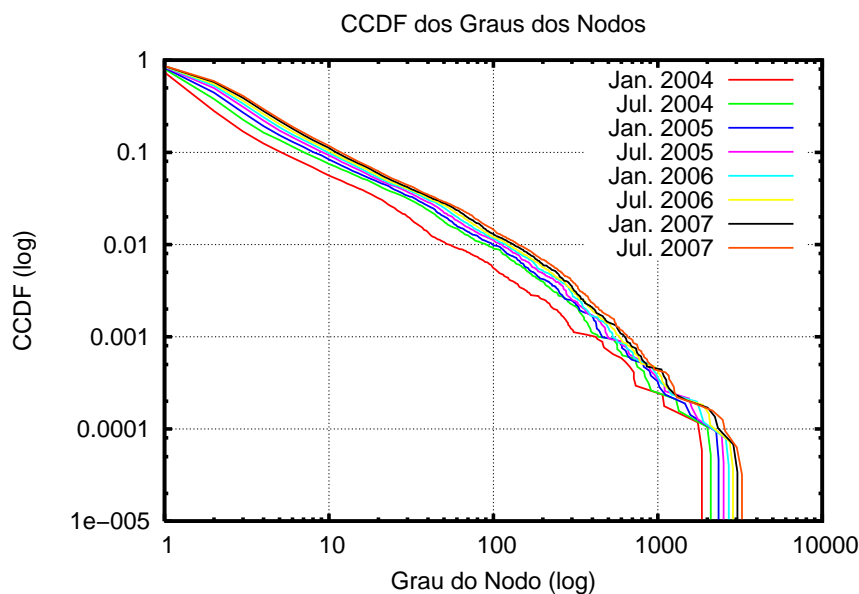


Figura 3.5: CCDF do grau dos vértices.

período de observação. Observa-se que a mediana e o terceiro quartil se mantiveram praticamente constantes durante o período. Observa-se também uma diminuição pequena porém consistente da distância média entre 2004 e 2007, de 3,708 para 3,431. Essa tendência foi observada também em estudos que analisaram topologias nos períodos entre 1997 e 2000 [19, 18] e entre 1999 e 2002 [3]. Apesar do crescimento do número de vértices, há um crescimento ainda maior no número de conexões entre eles ao longo do tempo, o que contribui para diminuir a distância média. A rede fica mais conectada, mais compacta e mais densa.

As Figuras 3.7 e 3.8 apresentam, respectivamente, as frequências absolutas e relativas das distâncias para a topologia IRL no período de 2004 a 2007. Na Figura 3.7 observamos uma evolução bem característica das distâncias no período, com crescimento bastante nítido da distância 3. Na Figura 3.8 observamos que houve uma mudança importante em relação às distâncias. No início do período de observação predominava a distância 4 (janeiro de 2004). A partir da segunda amostra há predomínio da distância 3, seguida das distâncias 4, 2 e 5, nessa ordem. Observa-se um aumento progressivo na percentagem de distâncias 3 no conjunto de distâncias ao longo do período.

A evolução da ocorrência das distâncias mais frequentes é analisada nas Figuras 3.9

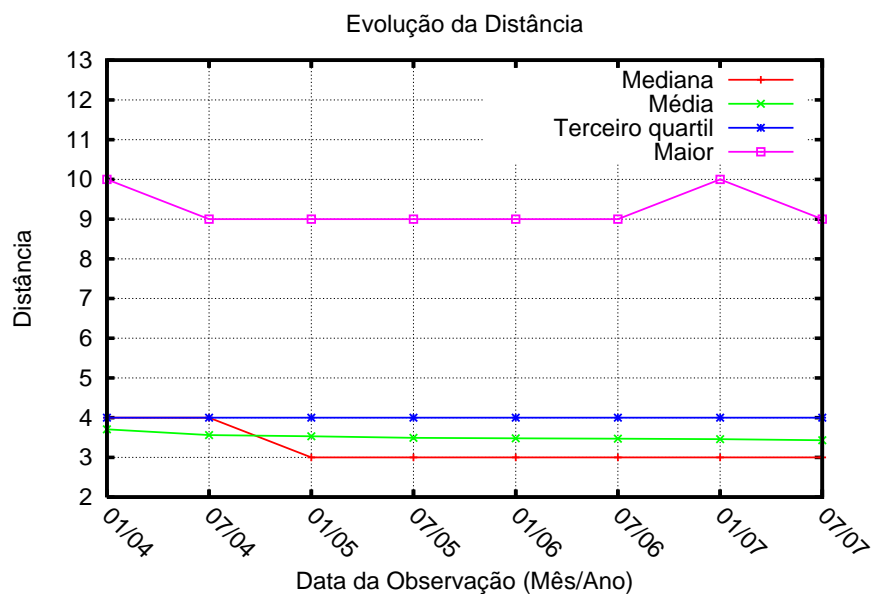


Figura 3.6: Evolução de índices estatísticos das distâncias.

e 3.10. O gráfico da Figura 3.9 ilustra a evolução das frequências das distâncias 3 e 4 no período observado. A porcentagem de pares de vértices que estão à distância 3 entre si aumentou de cerca de 37% para 48%, enquanto a fração de distâncias 4 diminuiu, entre 2004 e 2007, de 40% para 38%.

O gráfico da Figura 3.10 apresenta a evolução da frequência das distâncias 2, 5 e 6 no período. Enquanto a frequência da distância 2 permaneceu razoavelmente estável, as ocorrências de distâncias 5 e 6 sofreram uma queda perceptível. A análise conjunta da evolução das distâncias indica que a rede, mesmo tendo crescido significativamente durante o período, está se compactando em torno da distância de 3 arestas.

Esses resultados indicam que a distância é uma métrica característica da rede, apresentando um comportamento bastante regular e invariante. Não obstante o crescimento da rede em termos do número de vértices, a distância média entre os vértices tem decrescido de forma lenta e consistente. Pode-se concluir que, enquanto a rede cresce com o passar do tempo, ela se torna cada vez mais conexa e compacta. Isso ocorre devido ao crescimento mais acelerado do número de conexões lógicas entre os sistemas autônomos, em relação ao crescimento da quantidade de sistemas autônomos. O crescimento do número médio de arestas ao longo do período tem como consequência a diminuição da distância entre os

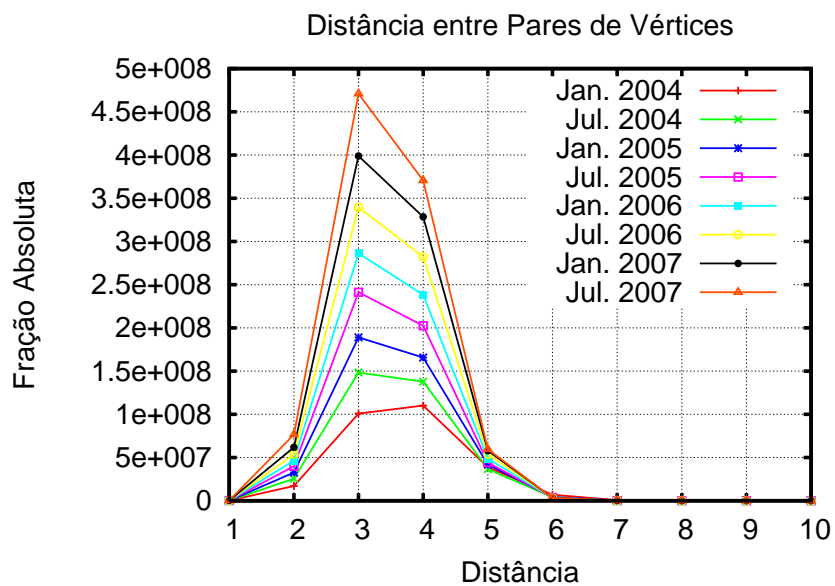


Figura 3.7: Frequência absoluta da distância entre pares de vértices no período.

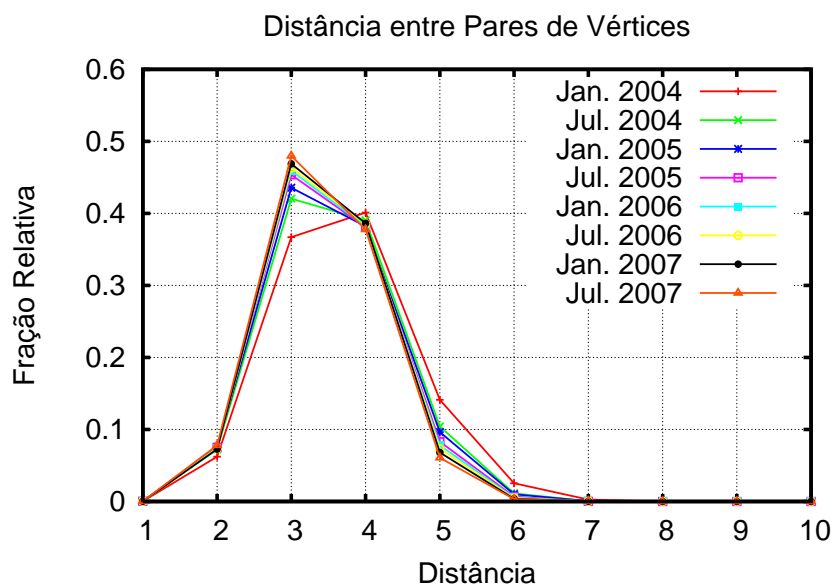


Figura 3.8: Frequência relativa da distância entre pares de vértices no período.

vértices. Estes resultados indicam que a distância é uma métrica bastante característica da topologia da rede e fundamental em seu estudo.

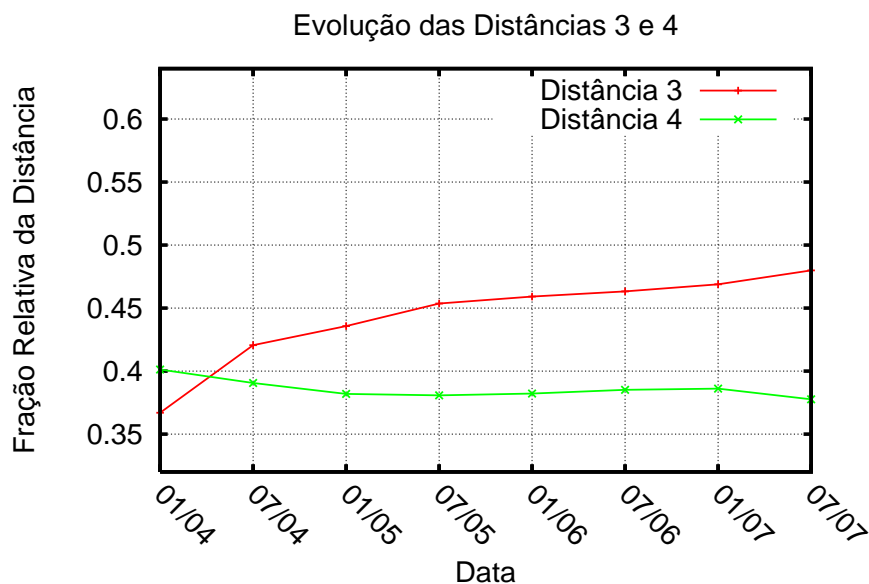


Figura 3.9: Evolução da ocorrência de distâncias 3 e 4.

3.4 Caracterização da Excentricidade

A partir da medida de distância, é possível definir a excentricidade. A excentricidade 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 e indica quão centrais ou periféricos são os vértices.

A Tabela 3.4 apresenta estatísticas relativas à excentricidade nas topologias analisadas. A excentricidade média está entre 6 e 7, indicando que, em média, os vértices mais longínquos estão à distância 6 ou 7. A excentricidade máxima está entre 9 e 10. Esses resultados indicam que a rede é bastante compacta e que os caminhos mais longos são, de fato, caminhos curtos.

Métricas de centralidade relativas a um grafo são definidas a partir da excentricidade de seus vértices. 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 nodo v é centro se $e(v) = R(G)$. O centro de G é o conjunto de vértices de menor excentricidade. Um nodo v é periférico se $e(v) = D(G)$. De forma análoga ao centro de G , a periferia de G é o conjunto de vértices do grafo com a maior excentricidade.

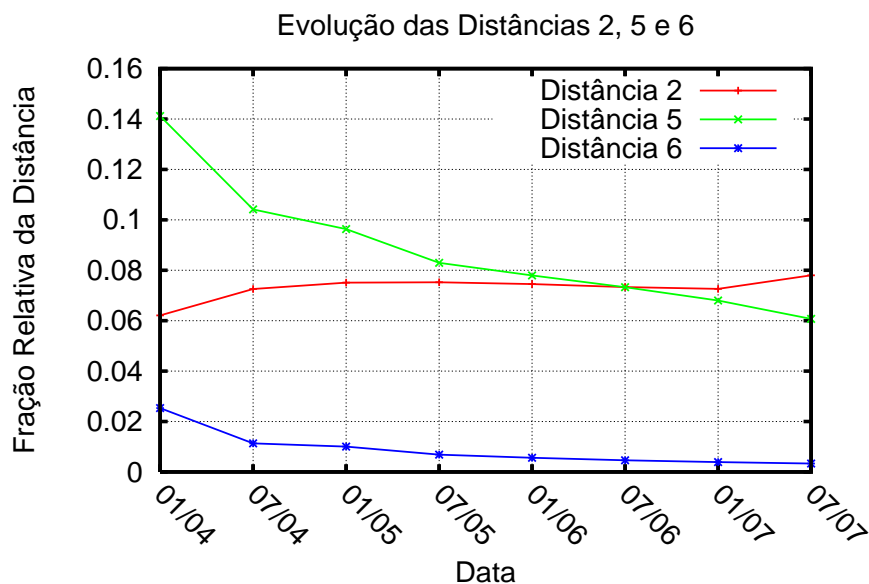


Figura 3.10: Evolução da ocorrência de distâncias 2, 5 e 6.

Data	Mínima	Moda	Média	Máxima
01/2004	5	7	7.135	10
07/2004	5	6	6.346	9
01/2005	5	6	6.360	9
07/2005	5	6	6.263	9
01/2006	5	6	6.199	9
07/2006	5	6	6.214	9
01/2007	5	7	6.981	10
07/2007	5	7	6.760	9

Tabela 3.4: Evolução da excentricidade nas topologias analisadas.

Os resultados dessas métricas para a topologia analisada estão na Tabela 3.5. A distância máxima entre os vértices do centro do grafo e os demais vértices é representada pelo valor 5 do raio, constante em todas as amostras da topologia. O diâmetro varia entre 9 e 10, e indica a maior distância entre vértices periféricos. A tabela também apresenta a quantidade de vértices no centro e na periferia de cada grafo. Observamos que o centro e a periferia são compostos por uma pequena fração dos vértices do grafo. Grande parte dos nodos, mais de 90%, está entre o centro e a periferia. Esses resultados indicam que as métricas, como raio e diâmetro, são também invariantes e características da rede.

As Figuras 3.11 e 3.12 apresentam gráficos que exibem a evolução da excentricidade na topologia IRL no decorrer dos últimos anos. Na Figura 3.11 observa-se a predominância

Data	Raio	Diâmetro	Centro	Periferia
01/2004	5	10	1	13
07/2004	5	9	173	10
01/2005	5	9	128	15
07/2005	5	9	542	8
01/2006	5	9	1059	2
07/2006	5	9	895	3
01/2007	5	10	6	3
07/2007	5	9	16	67

Tabela 3.5: Raio, diâmetro, centro e periferia nas topologias analisadas.

das excentricidades 6 e 7 sobre os valores 5 e 8 ao longo do período.

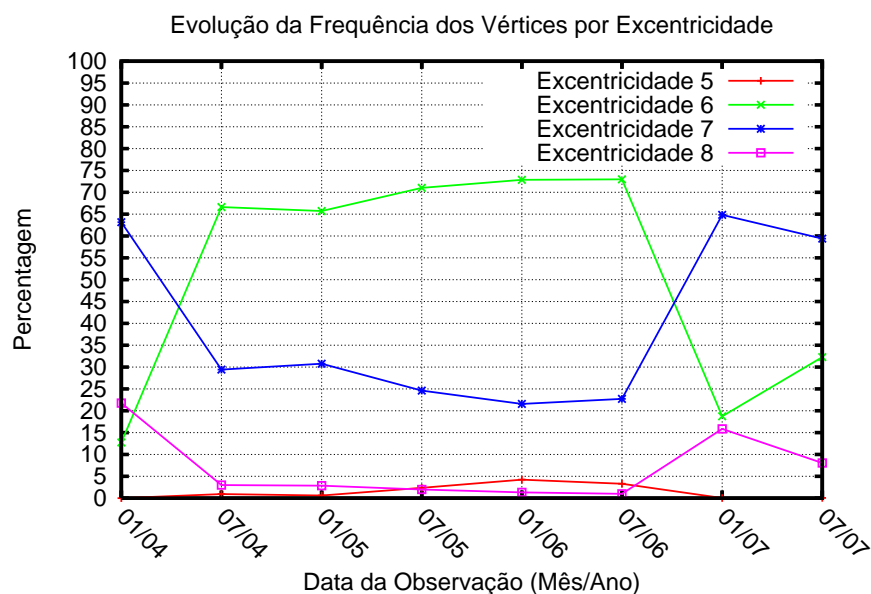


Figura 3.11: Evolução da excentricidade no período analisado.

O gráfico da Figura 3.12 ilustra a distribuição das conexões entre vértices, de acordo com sua excentricidade. Esse gráfico mostra que quase 50% das arestas ligam vértices de excentricidades 5 e 6 (5 – 6). Em 2007, houve aumento significativo das conexões entre vértices de excentricidades 6 e 7 (6 – 6, 6 – 7 e 7 – 7).

A Figura 3.13 exibe o grau médio dos vértices em função de sua excentricidade, para todas as topologias analisadas no período. O eixo y está em escala logarítmica. Observamos que os vértices mais centrais apresentam maiores graus em comparação com os vértices periféricos, o grau médio diminui à medida que o vértice se afasta do centro.

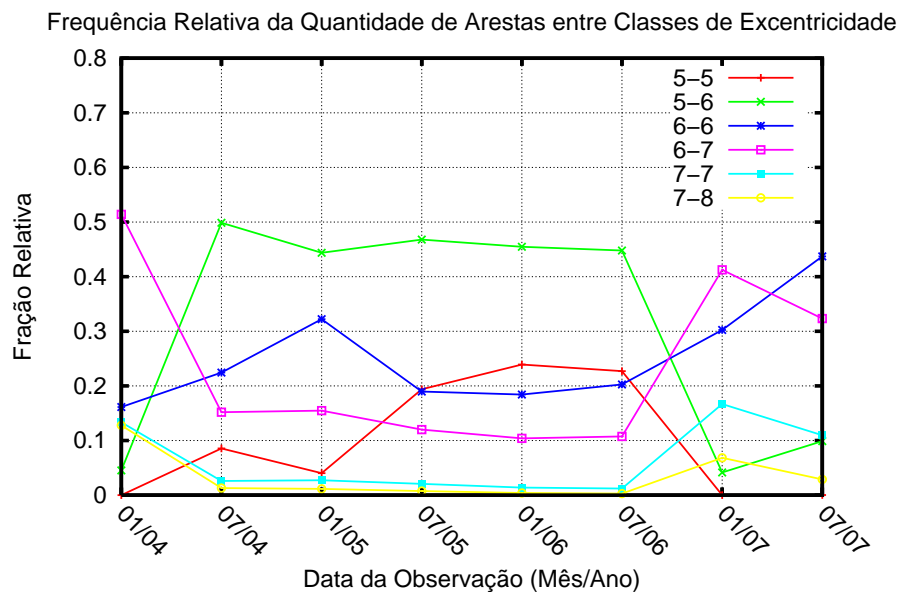


Figura 3.12: Evolução das conexões em função da excentricidade.

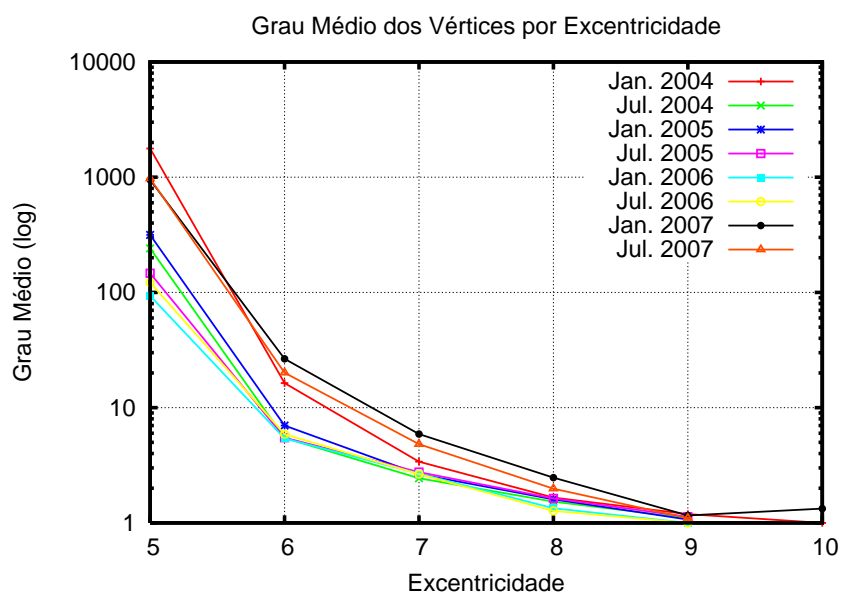


Figura 3.13: Evolução do grau médio dos vértices por excentricidade no período.

3.5 Análise do *Status*

Nessa seção apresentamos resultados relacionados à métrica status, definida para os nodos da rede, e sua relação com as medidas de centro e periferia. O status $s(v)$ de um nodo $v \in V$ em um grafo $G = (V, E)$ é igual à soma das distâncias de v a u para todo $u \in V$. O status é, portanto, a soma das distâncias de um nodo a todos os demais nodos da rede.

A mediana $M(G)$ de um grafo G é o conjunto de nodos com status mínimo, isto é, o menor valor para o status. O status total de um grafo é a soma dos status de todos os seus vértices.

A Figura 3.14 apresenta a função de distribuição cumulativa dos status dos nodos para todos os grafos das topologias analisadas. Observa-se que o status aumentou regularmente ao longo do período, refletindo o aumento do número de vértices e de conexões na rede.

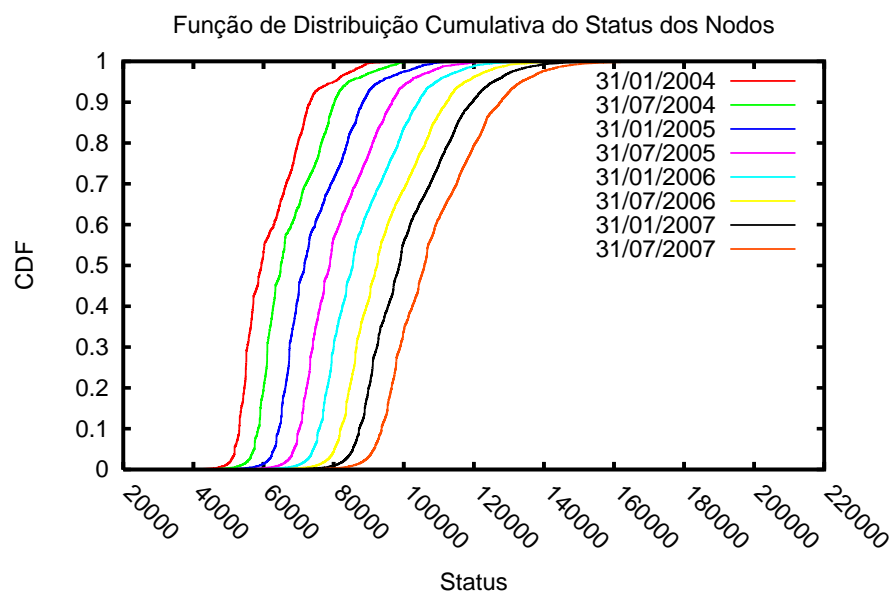


Figura 3.14: Distribuição cumulativa dos status observados nas topologias no período analisado.

A Figura 3.15 ilustra a evolução das métricas: menor status, maior status e status mais comum (moda) no período. O aumento destes índices está diretamente relacionado ao aumento na quantidade dos nodos da rede, pois há um aumento do número de distâncias a serem somadas para o cálculo do status. Sendo assim, a soma de todas as distâncias também sofre um acréscimo.

3.6 Árvore nas Bordas

Árvores nas bordas da topologia são subgrafos acíclicos conectados nas bordas do grafo. Estas árvores são características da topologia apresentada. Para caracterizar e quantificar estes subgrafos na topologia, o primeiro passo foi analisar a altura das árvores. O caminho

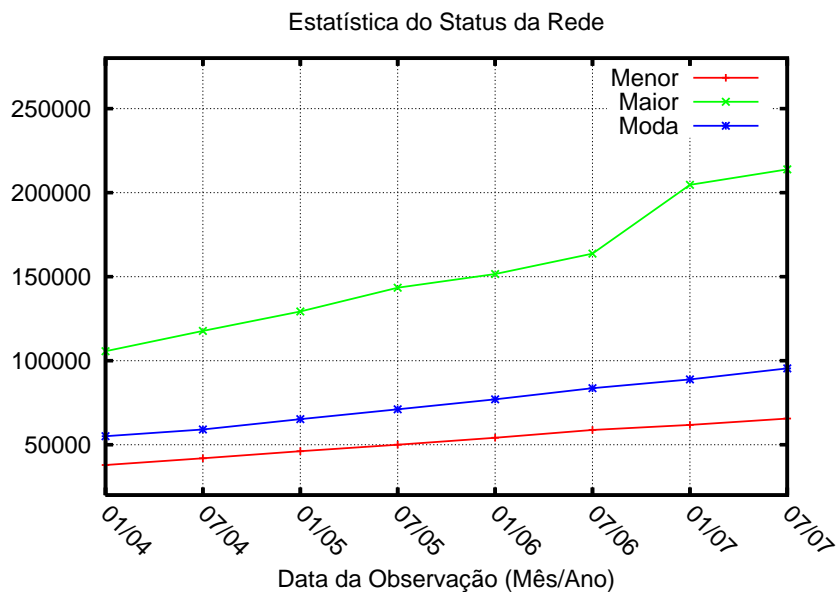


Figura 3.15: Estatísticas do Status para o período analisado.

foi obter o subgrafo 2 -core da rede. Um subgrafo k -core de um grafo G é o subgrafo maximal no qual todos os seus vértices possuem grau igual ou maior a k . Sendo assim, no subgrafo 2 -core, todos os vértice possuem grau no mínimo igual a 2. Para obter o subgrafo 2 -core da rede capturada, foi utilizado o seguinte algoritmo:

```

1: Enquanto(existem vértices com grau 1){
2:     retire vértices de grau 1 e sua aresta que se conecta ao grafo;
3: }

```

Observe que a retirada de um nodo de grau um e sua aresta altera o grau de outros nodos da rede. Por exemplo, um nodo de grau dois em um subgrafo acíclico pode se tornar um nodo de grau um. Por isso, o algoritmo foi executado até que nenhum nodo da rede possuísse grau menor que 2. O número de execuções da segunda linha, do algoritmo apresentado, é igual à altura da maior árvore encontrada no grafo. Para todo o período do estudo, de janeiro de 2004 a julho de 2007, foi observado que apenas 2 iterações foram necessárias para eliminar todos os subgrafos acíclicos da rede, mostrando, assim, que as árvores presentes têm altura máxima igual a 2. A literatura reporta resultados muito próximos aos valores apresentados neste trabalho. Por exemplo, em [19], análise similar

indicou árvores de altura máxima igual a 3, e menos de 10% das árvores com altura maior ou igual a 2.

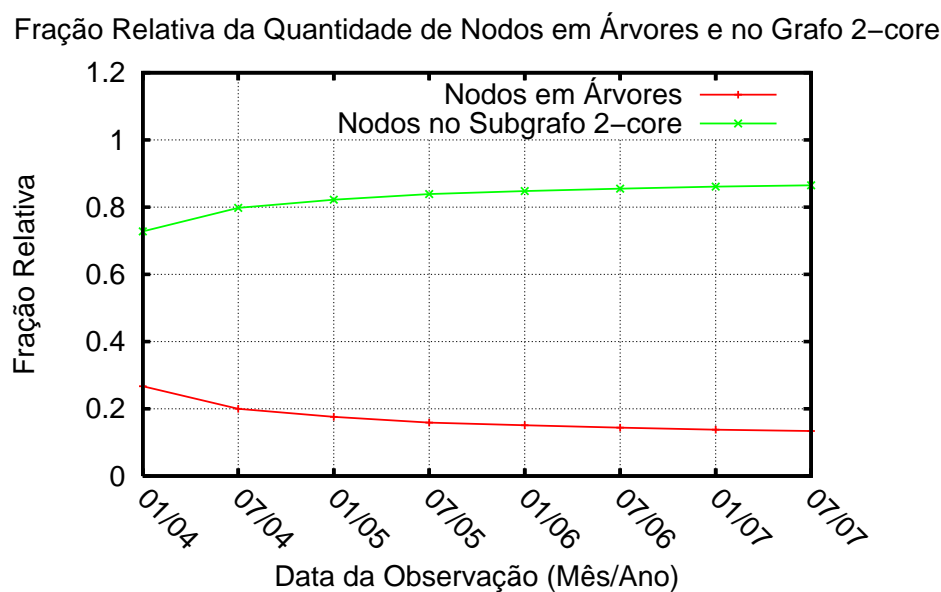


Figura 3.16: Fração de nodos em árvores e no subgrafo *2-core* no período analisado.

A Figura 3.16 apresenta a fração de nodos pertencentes a árvores e de nodos no subgrafo *2-core*, em relação à quantidade total de nodos na rede. Observa-se um crescimento na frequência de nodos do grafo *2-core* e proporcionalmente, um decréscimo na quantidade de nodos pertencentes a árvores que foram retirados utilizando o algoritmo descrito anteriormente.

A quantidade de nodos envolvidos nestas árvores foi também analisada. A Figura 3.17 apresenta o número de árvores e a quantidade de nodos pertencentes a estas árvores, para as topologias capturadas. A escala do gráfico é logarítmica. Observa-se que há poucas árvores com centenas de nodos e muitas árvores com poucos nodos. A maior árvore encontrada durante o período de estudo continha 335 vértices, em janeiro de 2004. O mesmo gráfico é reexibido na Figura 3.18, em escala linear, para facilitar a observação. Identificamos grande quantidade de árvores com apenas dois nodos, que é a menor árvore encontrada no grafo. O número destas árvores cresceu ao longo dos anos. Como as árvores têm altura máxima igual a 2, elas são, portanto, compostas por uma raiz, conectada ao grafo, e possuem dezenas a centenas de filhos, que são nodos de grau 1.

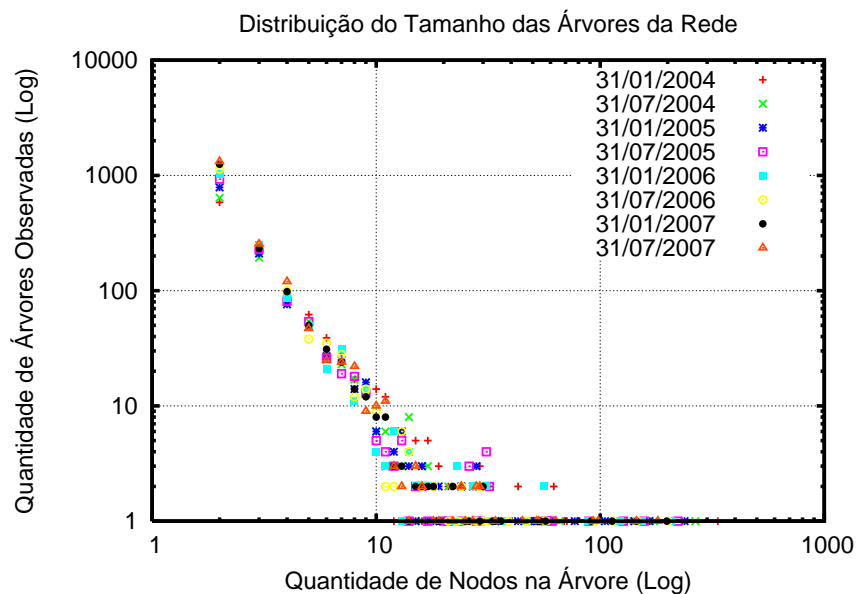


Figura 3.17: Tamanho das árvores observadas no período (escala logarítmica).

A evolução da quantidade de nodos em cada árvore observada também foi analisada. Observa-se que o número máximo de nodos presentes nas árvores das bordas, durante o período analisado, tendeu a diminuir com o passar do tempo. Durante o mesmo período, a quantidade de árvores com 2 nodos, que é a menor árvore encontrada na rede, tendeu a aumentar. Estes dados são apresentados nas Figuras 3.19 e 3.20. A Figura 3.19 apresenta o número de nodos da maior árvore encontrada na topologia, para todos os grafos analisados. Observa-se diminuição do número de nodos na maior árvore durante o período de observação, de 335 nodos em janeiro de 2004, para 181 nodos em julho 2007. Isso pode ser explicado pelo aumento no número de conexões, que aumenta o grau médio dos nodos e diminui a possibilidade de ter nodos de grau 1.

Por outro lado, conforme mostra a Figura 3.20, o número de árvores com dois nodos aumentou no mesmo período, de 584 árvores observadas em janeiro de 2004 para 1320 em julho de 2007. Este resultado indica maior distribuição das árvores pelos nodos da rede.

Uma das hipóteses sobre o decréscimo da quantidade máxima de nodos observados na rede e sobre o crescimento da quantidade de árvores de apenas 2 vértices é que a razão do crescimento do número de arestas é maior que o crescimento da quantidade de nodos da rede. Logo, a adição de uma aresta ligando um nodo de uma árvore ocasiona a diminuição

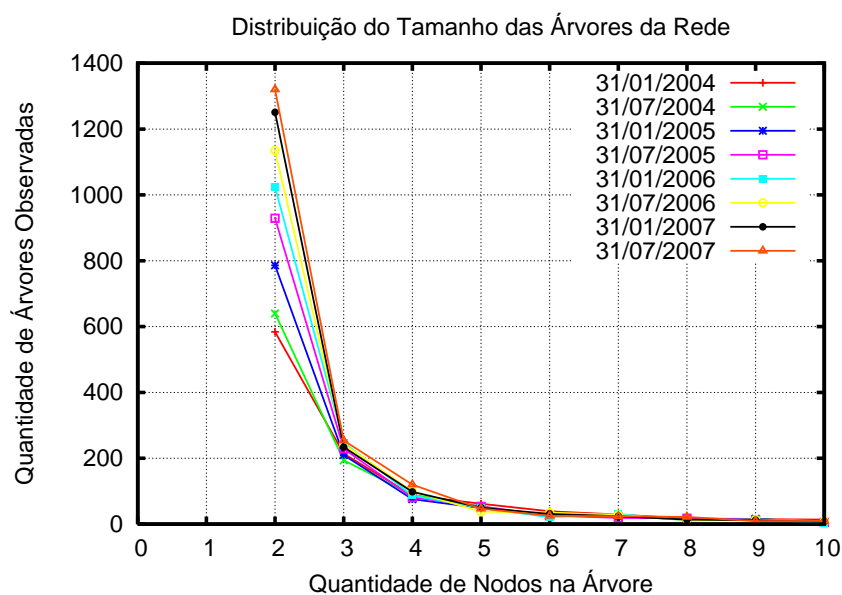


Figura 3.18: Tamanho das árvores observadas nas topologias (escala linear).

de seu tamanho, conforme a definição de árvore nas bordas.

3.7 Coeficiente de Agrupamento

Devido ao grande crescimento da rede, na quantidade de nodos e arestas é de grande importância verificar a influência de tal crescimento na conectividade dos vértices e na expansão da rede. Uma das métricas que tem como objetivo caracterizar tal propriedade estrutural é conhecida como o coeficiente de agrupamento. O coeficiente de agrupamento de um nodo i é dado como a taxa do número de conexões entre os vizinhos de nodo i e o número de conexões se i e tais vizinhos fossem totalmente conectados. Sendo assim, o coeficiente de agrupamento de um grafo é expresso pela média da soma dos coeficientes de agrupamento de cada nodo em relação a todos os nodos da rede. O valor do coeficiente de agrupamento de um grafo varia seu valor entre 0 a 1, o qual quantifica a conectividade dos nodos entre si. Um grafo com valor de coeficiente de agrupamento igual a 0, é uma rede a qual não apresenta nenhuma conectividade entre os nodos, sendo assim não existe nenhuma aresta interligando os vértices. De maneira análoga, um grafo com coeficiente igual a 1 é uma rede da forma completa, no qual existe um caminho do tamanho 1 de qualquer vértice a todos os outros. O coeficiente de agrupamento foi introduzido em [33],

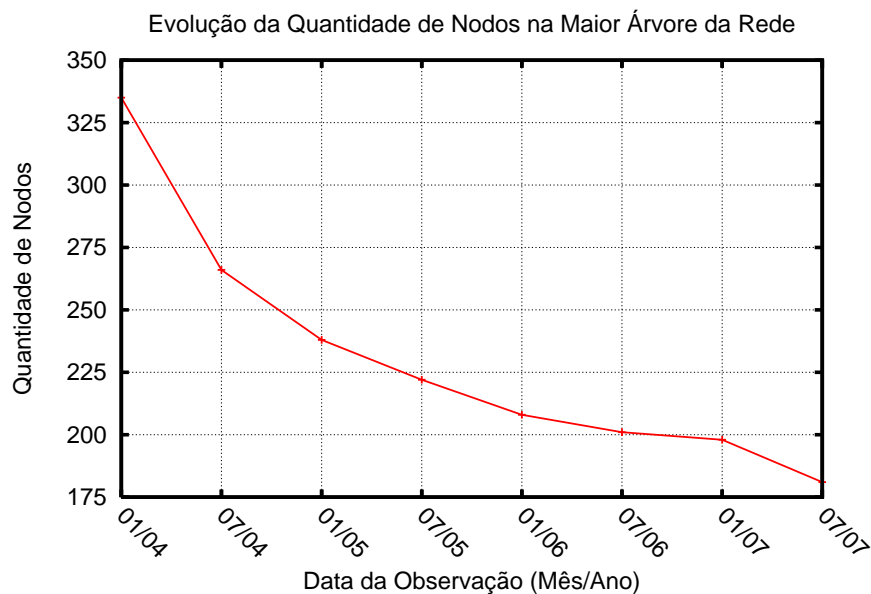


Figura 3.19: Evolução na quantidade de nodos na maior árvore da rede no período.

que teve como contexto principal a análise de redes sociais, como por exemplo: as redes caracterizadas como "mundo pequeno". Uma das principais características de tais redes é a presença de grupos de nodos fortemente conectados, o que leva a curta distâncias entre os nodos.

A Figura 3.21 apresenta os resultados para os coeficientes de agrupamento obtidos nas topologias analisadas durante o período de janeiro de 2004 a julho de 2007. É possível verificar que, com o passar do tempo, a rede apresenta a tendência de aumentar o agrupamento entre seus nodos. Isso indica que, cada vez mais, a distância entre os vértices tende a diminuir e em contrapartida, a quantidade de triângulos, cliques e conexões tende a aumentar. Este resultado era esperado, tendo em vista que a razão do crescimento da quantidade de arestas é maior que o crescimento da quantidade de vértices. Os resultados indicam também que houve um crescimento maior do coeficiente no período entre 2004 e 2005, e posteriormente o crescimento foi mais lento.

O resultado do coeficiente de agrupamento é importante para o presente estudo, pois demonstra de fato que a rede, com o crescimento no número de vértices e arestas, torna-se cada vez mais compacta. Desse modo, o crescimento atual observado na Internet, em relação aos sistemas autônomos, contribui muito mais para diminuir as distâncias entre

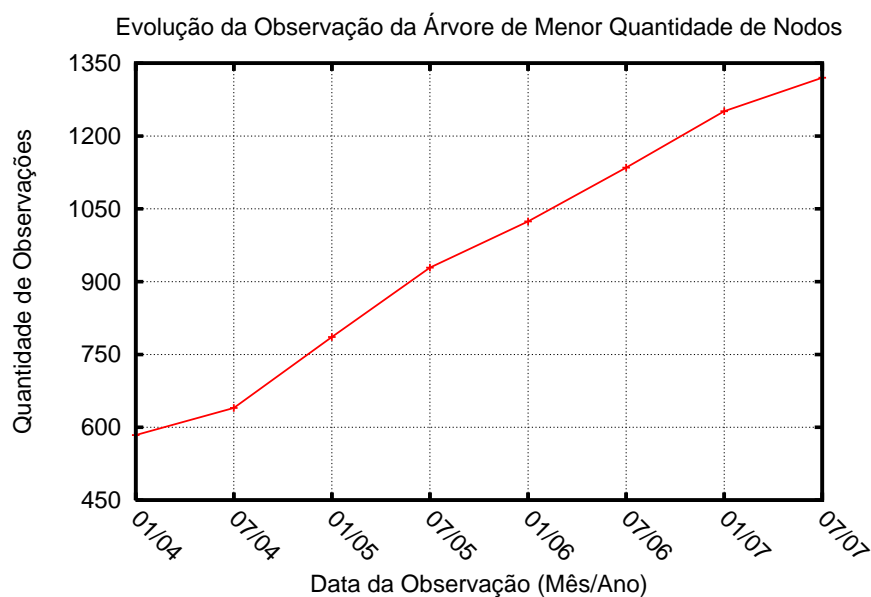


Figura 3.20: Evolução do número de árvores com menor número de nodos da rede no período

os nodos, do que para criar uma rede esparsa e com caminhos longos entre os vértices e árvores nas bordas.

3.8 Síntese do Capítulo

Com base nas métricas apresentadas, foi possível analisar e caracterizar a rede sob vários aspectos, observando tendências e relações evolutivas na adição de novos nodos e arestas. O próximo capítulo apresenta modelos matemáticos gerados a partir dos dados obtidos pela aplicação das métricas características na rede.

Os modelos a serem apresentados têm como objetivos a obtenção de várias equações que quantifiquem tendências evolutivas, possibilitando assim, a identificação e enumeração de leis empíricas sobre a rede durante a sua evolução.

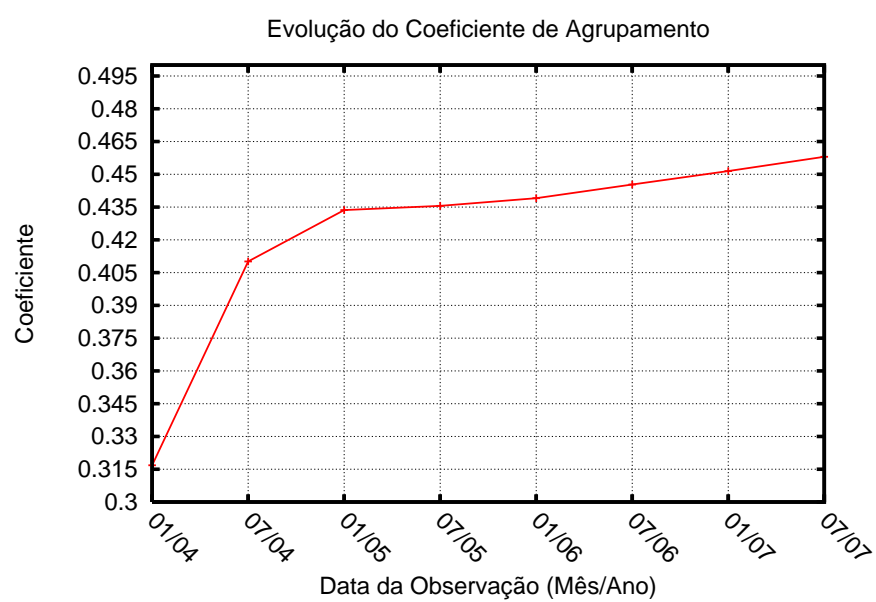


Figura 3.21: Evolução do coeficiente de agrupamento da rede no período.

CAPÍTULO 4

MODELOS DE EVOLUÇÃO E PREDIÇÃO PARA A TOPOLOGIA DA INTERNET

Neste capítulo propomos modelos baseados nos dados, resultados e métricas apresentados no Capítulo 3. Esses modelos nos permitem fazer extrapolações e inferir medidas futuras da rede. Os modelos foram obtidos utilizando o *Curve Fitting Toolbox* do MATLAB. A ferramenta utiliza várias técnicas de ajuste de curvas e extrapolação. A qualidade dos modelos gerados foi medida pelo intervalo de confiança da medição. Todos os resultados apresentam intervalos de confiança não inferiores a 99%. Cada métrica foi modelada por duas equações que delimitam os valores com base no intervalo de confiança. Para a geração dos gráficos foram utilizados os valores da média aritmética dos parâmetros das curvas que delimitam o conjunto de confiança, isto é, para duas curvas EqA e EqB que delimitam um conjunto de confiança, a curva exibida no gráfico correspondente utiliza a média aritmética dos valores numéricos de cada termo das curvas.

A partir dos modelos obtidos para as métricas, são enunciadas algumas leis de crescimento da Internet.

4.1 Composição dos Grafos e Graus dos Vértices

A primeira modelagem refere-se ao crescimento da topologia quanto ao número de vértices e arestas. Para esta modelagem foram gerados dois polinômios de primeiro grau, um para cada métrica. Com o resultado obtido, foi realizada a verificação da possível utilização de tais equações para a extrapolação das medidas estudadas. O objetivo foi utilizar os dois polinômios criados para extrapolar resultados para o período de 2008 a 2010. A equação obtida para a métrica número de vértices é $14.553,28 + 2.093,29 * x$ com conjunto de confiança igual a $(14.340 + 2.051 * x; 14.760 + 2.135 * x)$. Para a medida da quantidade de arestas do grafo, a equação obtida é $27.117,96 + 15.981,53 * x$ com conjunto

de confiança igual a $(19.130 + 14.400 * x; 35.510 + 17.560 * x)$. As equações apresentadas na Figura 4.1 foram obtidas calculando-se a média dos coeficientes que delimitam o conjunto de confiança.

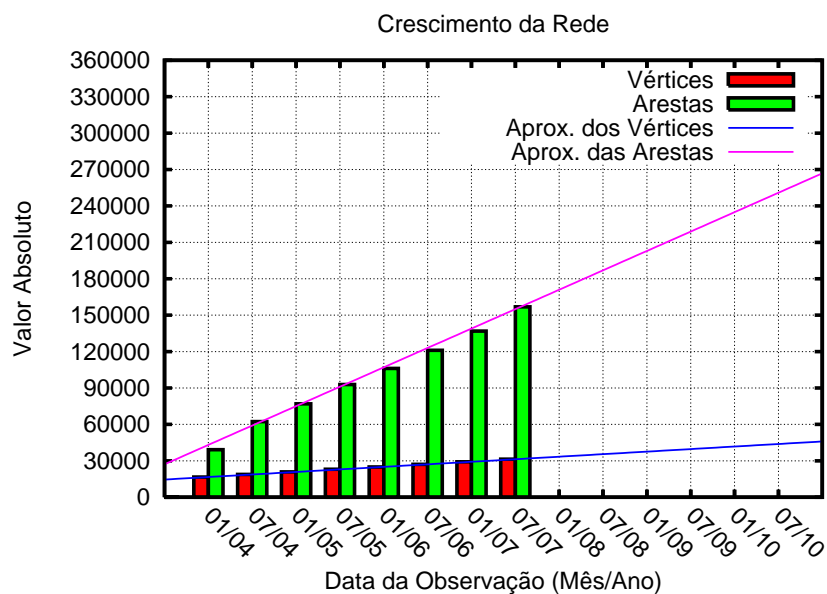


Figura 4.1: Crescimento absoluto da rede no período de 2004 a 2010.

A Figura 4.1 apresenta as curvas modeladas e os valores reais obtidos para a quantidade de nodos e arestas no grafo. É possível observar que o crescimento tanto do número de arestas quanto da quantidade de vértices segue as equações das retas encontradas, o que indica tendência de crescimento linear dessas métricas com o passar do tempo. Essas duas medidas apresentaram ser lineares durante todo o período de observação, de janeiro de 2004 a julho de 2007. Isso pode ser verificado pelo conjunto de confiança obtido. Para ambos os casos, a diferença entre as curvas delimitadoras do conjunto de confiança, igual a $420 + 84 * x$ para os vértices e igual a $16.380 + 3.160 * x$ para as arestas. Quanto menor a diferença entre as curvas delimitadoras de um conjunto de confiança, maior a regularidade do conjunto de dados obtidos, assim como uma previsão mais precisa dos resultados extrapolados.

Os valores numéricos fornecidos pelas equações podem ser observados na Tabela 4.1. Os dados indicam que em 2010 a Internet terá crescido 260% em relação à primeira medida em 2004, quanto ao número de vértices, e 640% em relação ao número de arestas.

Data	Vértices	Arestas
01/2004	16573	39143
07/2004	18783	62305
01/2005	20830	76976
07/2005	23062	92910
01/2006	24971	106141
07/2006	27071	121045
01/2007	29167	136888
07/2007	31328	156871
01/2008	33392	170951
07/2008	35486	186933
01/2009	37579	202914
07/2009	39672	218896
01/2010	41766	234877
07/2010	43859	250859

Tabela 4.1: Composição dos grafos e extrapolação dos valores

Quanto ao crescimento da rede, os resultados indicam um padrão bastante diferente do observado em [19]. Leis empíricas enunciadas naquele trabalho indicam crescimento anual de 45% para o número de ASes, e 53% para o número de conexões entre ASes (arestas).

O padrão atual revela uma taxa de crescimento decrescente, conforme mostra a Figura 4.2. Cada ponto representa o crescimento da rede em relação ao ponto (semestre) anterior. A rede continua crescendo, mas a taxas cada vez menores, o que evidencia uma nova fase em seu ciclo de vida. As equações que modelam a taxa de crescimento semestral da rede são, quanto ao número de vértices, $0,1604 - 0,04292 * \log(x)$ e quanto ao número de arestas $18,01 * e^{-1,91*x} + 0,2259 * e^{-0,069*x}$.

Tendo em vista os resultados quanto ao crescimento da topologia, declaramos duas leis empíricas sobre o crescimento do número de nodos e de arestas, baseado na observação da rede durante o período de quatro anos:

Lei Empírica 1 (*Crescimento dos ASes*): *O crescimento da quantidade de Sistemas Autonomos é subexponencial da forma $\Theta(e^{-kx} + e^{-px})$, na qual k e p são constantes.*

O crescimento não é linear e nem apresenta ser constante com o passar do tempo, como

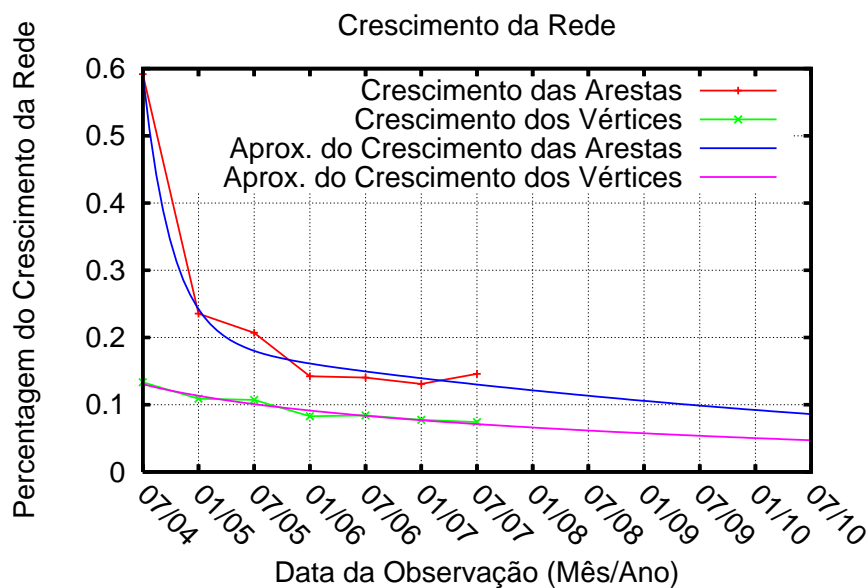


Figura 4.2: Padrão de crescimento da rede no período de 2004 a 2010.

proposto em [19]. Nesse trabalho obtivemos modelos subexponenciais para o crescimento de SAs da rede.

Lei Empírica 2 (*Crescimento das Conexões*): *O crescimento da quantidade de conexões que ligam os ASes na rede é da ordem logarítmica da forma $\Theta(\log(x))$.*

O crescimento das conexões, assim como o dos ASes, não é linear em relação a quantidade de arestas da rede no passado como apresentado em [19].

Uma preocupação em relação à expansão da rede é relativa ao esgotamento do espaço de endereçamento dos ASes. Cada AS recebe um número único, e o número máximo de endereços que o espaço de endereçamento do protocolo IPv4 permite é de $2^{16} = 65.536$. Segundo o modelo proposto, o número de ASes chegará a esse limite entre janeiro e julho de 2015, como mostra a Figura 4.3. A equação desta curva é $84650,94 + \frac{x}{0,130511}$. Este fato é enunciado na Lei Empírica 3. A previsão feita por [19] indicava que o esgotamento dos endereços se daria em janeiro de 2006, o que não ocorreu.

Lei Empírica 3 (*Esgotamento de Identificadores dos ASes*): *Os números válidos para identificadores dos ASes se esgotarão no período entre janeiro e julho de 2015.*

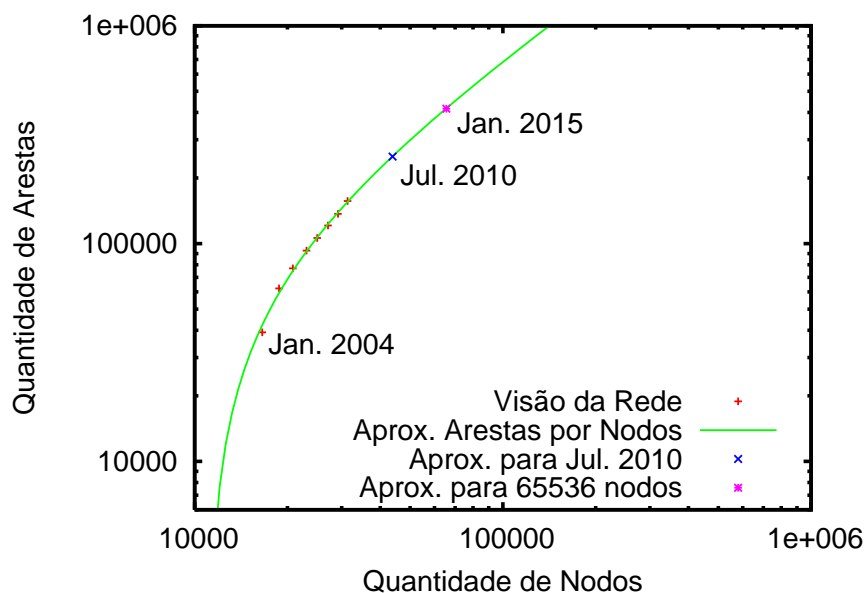


Figura 4.3: Crescimento relativo de vértices e arestas no período de 2004 a 2010.

Se o crescimento da quantidade de ASes continuar seguindo o mesmo padrão identificado na modelagem dos dados, o número possível de identificadores para os ASes se esgotará no período de janeiro a julho de 2015.

As estatísticas dos graus dos vértices foram modeladas. O gráfico da Figura 4.4 apresenta os dados coletados da topologia real, bem como os modelos gerados para as métricas média e mediana dos graus dos vértices, para o período de 2008 a julho de 2010. Os modelos são equações polinomiais de primeiro grau. A equação polinomial obtida para a média é $4,962 + 0,665 * x$, com conjunto de confiança igual a $(3,526 + 0,3811 * x; 6,398 + 0,9497 * x)$. Em relação à modelagem da mediana, a equação gerada pela ferramenta utilizada foi igual a $1,643 + 0,1905 * x$ com conjunto de confiança igual a $(0,829 + 0,02931 * x; 2,457 + 0,3516 * x)$. Para a métrica da mediana, o valor obtido na equação de modelagem foi arredondado, levando em consideração que a mediana é representada por um valor inteiro. O arredondamento de y foi realizado da seguinte forma: para valores da parte fracionária de y maiores ou iguais a 0,5, o valor resultante de y é igual a $\lceil y \rceil$. Caso os valores da parte fracionária de y sejam estritamente menores que 0,5, o valor obtido para y é igual a $\lfloor y \rfloor$.

A Tabela 4.1 apresenta os resultados numéricos produzidos pelos modelos para a me-

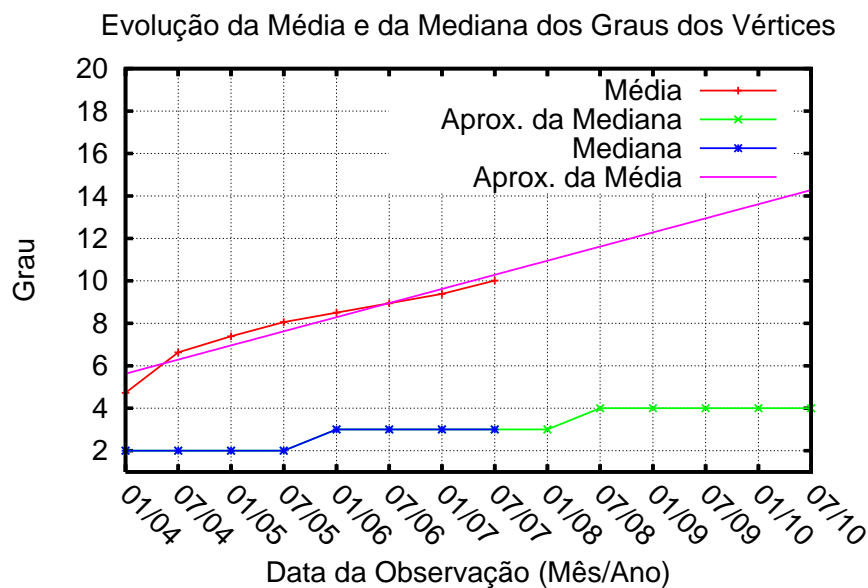


Figura 4.4: Média e mediana do grau dos vértices no período de 2004 a 2010.

diana e a média dos graus dos vértices. Observamos que a tendência de aumento da conectividade da rede pelo aumento expressivo do grau médio dos vértices no período. Observamos também que o aumento do grau ocorre de forma genérica para todos os vértices, pois há aumento de grau não apenas para os vértices mais conectados, como podemos observar nos números do maior grau, mas também nos vértices menos conectados, o que produz aumento da mediana.

A métrica maior grau observado, apresentada na Tabela 4.1, foi modelada por um polinômio de primeiro grau igual a $4,665 + 0,702 * x$ apresentando um conjunto de confiança $(3,241 + 0,42 * x; 6,089 + 0,984 * x)$. A partir das equações obtidas pode-se observar que tanto a média quanto a mediana apresentam comportamento linear, e tendem ao crescimento com o passar do tempo.

4.2 Distância entre Pares de Vértices

A métrica distância entre pares de vértices foi modelada com funções para a distância média, a maior distância e o coeficiente de variação das distâncias. A Figura 4.5 apresenta os dados reais até 2007 e as curvas geradas pelas equações obtidas para os modelos. A distância média entre os vértices diminuiu de 3,708, valor observado em janeiro de 2004,

Data	Mediana	Média	Maior	COV
01/2004	2	4,72	2.388	4,72
07/2004	2	6,63	2.917	6,37
01/2005	2	6,36	3.205	6,22
07/2005	2	8,05	3.505	6,07
01/2006	3	8,50	3.710	6,02
07/2006	3	8,94	3.908	5,96
01/2007	3	9,38	4.122	5,90
07/2007	3	10,01	4.989	5,99
01/2008	3	10,98	5.015	
07/2008	4	11,68	5.331	
01/2009	4	12,38	5.647	
07/2009	4	13,08	5.963	
01/2010	4	13,79	6.279	
07/2010	4	14,49	6.595	

Tabela 4.2: Estatísticas dos graus dos vértices

para 3,431 em julho de 2007. A equação para esta métrica é um polinômio de primeiro grau igual a $3,658 - 0,03132 * x$ com conjunto de confiança igual a $(3,528 - 0,05696 * x; 3,787 - 0,005683 * x)$. O modelo indica que diminuição da distância média com o passar do tempo. Extrapolando os resultados, estimamos que a distância média, terá valor próximo a 3,219 em julho de 2010. Os valores numéricos são apresentados na Tabela 4.2.

Observa-se também que coeficiente de variação das distâncias calculadas entre pares de vértices diminuiu de 0,241 em janeiro de 2004 para 0,215 em julho de 2007. A continuar esta tendência, o valor desta métrica será 0,207 em julho de 2010. A equação obtida para o coeficiente de variação é $0,241 - 0,01262 * \log(x)$, um polinômio de primeiro grau, de confiança $(0,2362 - 0,01586 * \log(x); 0,2458 - 0,009376 * \log(x))$.

Para modelar a maior distância observada, escolhemos a mediana das maiores distâncias, uma vez que a métrica maior distância manteve-se praticamente invariante no período de janeiro de 2004 a julho de 2007. Nesse caso o valor mediano obtido foi igual a 9, usado para extrapolar a métrica até julho de 2010, tendo em vista a grande estabilidade da maior distância encontrada em cada análise da topologia.

Os gráficos das Figuras 4.6 e 4.7 apresentam as funções modeladas para as distâncias

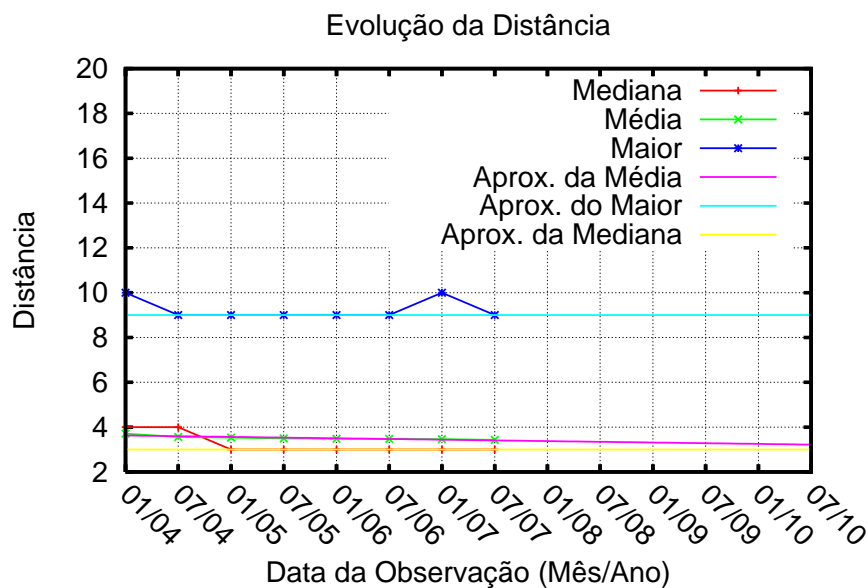


Figura 4.5: Aproximação da Evolução das Distâncias.

2 a 6. Estas funções representam a evolução da ocorrência das respectivas distâncias no conjunto de distâncias da topologia. Para as distâncias 2 e 4 foram gerados polinômios de primeiro grau. Para a modelagem da distância 3 foi gerada uma equação exponencial e para as distâncias 5 e 6 foram geradas equações logarítmicas. As equações obtidas para as distâncias 1 a 6 podem ser verificadas na lista a seguir.

Distância 1: $y = 0$

Distância 2: $y = 0,06728 + 0,001257 * x$, com conjunto de confiança igual a $(0,05614 - 0,0009495 * x; 0,07842 + 0,003464 * x)$.

Distância 3: $y = 0,4276 * \exp(0,01395 * x) - 0,2154 * \exp(-1,176 * x)$ com conjunto de confiança igual a $(0,3929 * \exp(0,001565 * x) - 0,4242 * \exp(-2,414 * x); 0,4622 * \exp(0,02633 * x) - 0,006672 * \exp(0,06141 * x))$.

Distância 4: $y = 0,3951 - 0,002101 * x$ com conjunto de confiança igual a $(0,3786 - 0,005375 * x; 0,4117 + 0,001174 * x)$.

Distância 5: $y = 0,1362 - 0,03627 * \log(x)$ com conjunto de confiança igual a $(0,1246 - 0,04408 * \log(x); 0,1477 - 0,02847 * \log(x))$.

Data	Média	Maior	COV
01/2004	3,708	10	0,241
07/2004	3,561	9	0,230
01/2005	3,531	9	0,230
07/2005	3,492	9	0,224
01/2006	3,480	9	0,221
07/2006	3,471	9	0,218
01/2007	3,461	10	0,215
07/2007	3,431	9	0,215
01/2008	3,375	9	0,213
07/2008	3,344	9	0,211
01/2009	3,313	9	0,210
07/2009	3,281	9	0,209
01/2010	3,250	9	0,208
07/2010	3,219	9	0,207

Tabela 4.3: Estimativas para estatísticas das distâncias

Distância 6: $y = 0,02188 - 0,0092803 * \log(x)$ com conjunto de confiança igual a $(0,01489 - 0,01453 * \log(x); 0,02887 - 0,005076 * \log(x))$.

Os modelos indicam que a frequência das distâncias 2 e 3 aumenta com o passar do tempo, sendo o aumento da frequência da distância 3 mais significativo do que o aumento da frequência da distância 2. Observando os valores gerados para a distância 3, verifica-se que em julho de 2010, cerca de 52% das distâncias calculadas serão iguais a 3. As demais distâncias ocorrerão na proporção de 36% para a distância 4, 4% para a distância 5 e 8% para a distância 2. A equação gerada para a distância 6 é uma curva logarítmica decrescente com o passar do tempo. Portanto, tende a números negativos. Assim, foi considerado que a partir de janeiro de 2009 a frequência da distância 6 atinge números muito próximos ou iguais a zero.

Considerando os resultados obtidos para a distância, enunciamos uma lei empírica sobre a distância média observada entre os ASes.

Lei Empírica 4 (*Distância entre Vértices*): *A distância média entre pares de vértices tende para o valor 3.*

Este resultado deve-se ao fato de que o crescimento do número de conexões na topologia é significativamente maior do que o crescimento do número de vértices (ASes). A

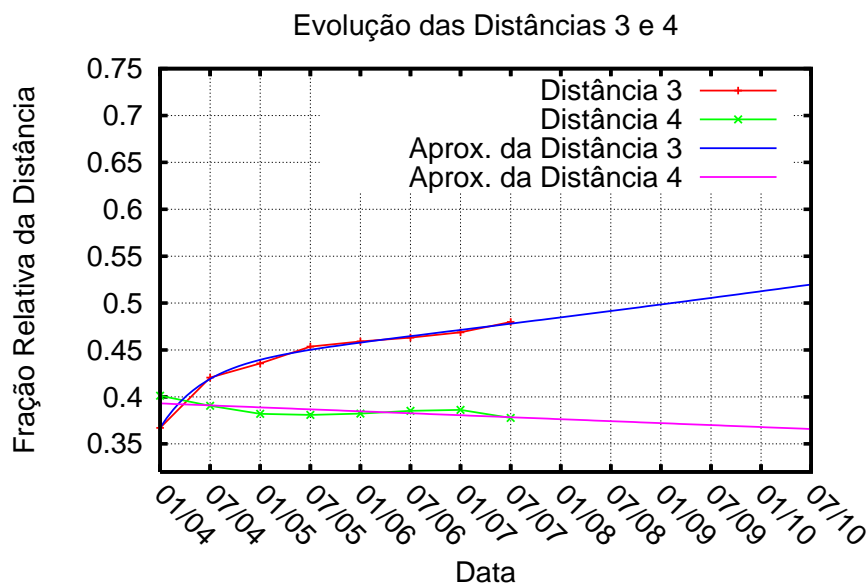


Figura 4.6: Evolução das distâncias 3 e 4.

maior proporção de conexões em relação ao número de vértices contribui para aumentar a conectividade da rede, aumentando o grau médio e, portanto, diminuindo a distância média entre os vértices. Caminhos alternativos são criados. No período analisado, a maior distância se manteve praticamente constante apesar da rede ter crescido cerca de 260%. O raio e o diâmetro da rede são constantes, resultados que concordam com o trabalho [19]. A distância média decresce lenta e consistentemente, o que induz a lei formulada.

4.3 Status

Conforme definido no Capítulo 3, o status $s(v)$ de um nodo $v \in V$ em um grafo $G = (V, E)$ é igual à soma das distâncias de v a u para todo $u \in V$, e o status do grafo G é a soma dos status de todos os seus vértices. A caracterização do status apresentada no Capítulo 3, indicou que a métrica cresce à medida que a rede evolui. Embora a distância entre os vértices tenha decrescido com o decorrer do tempo, o status do grafo da topologia tem aumentando devido ao aumento do número de vértices.

A Figura 4.8 apresenta a extrapolação das métricas relacionadas ao menor, ao maior status e o status mais observado na rede (moda). As três medidas são crescentes em relação à evolução da rede, ou seja, tanto o menor, maior status e o status mais observado

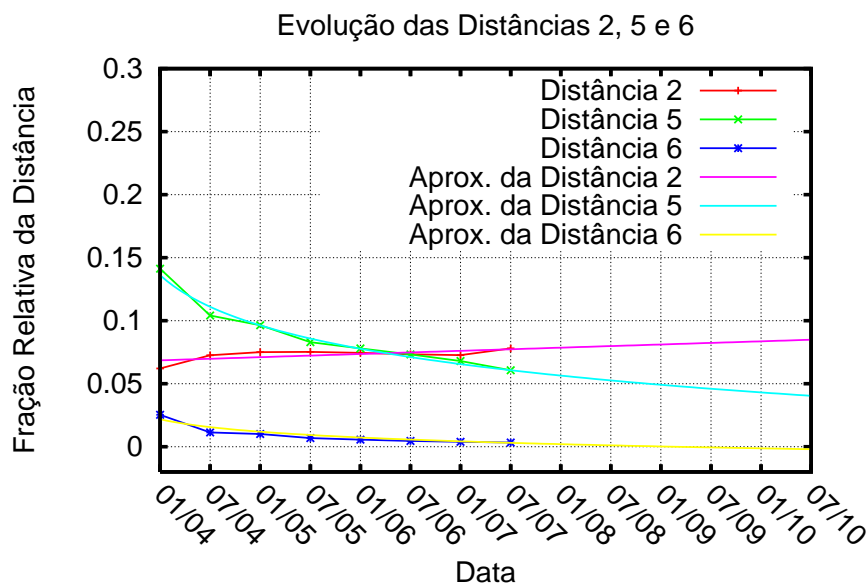


Figura 4.7: Evolução das distâncias 2, 5 e 6.

possuem a tendência de aumentar conforme a rede cresce. Isso ocorre principalmente devido ao próprio crescimento da rede. Com adição de novos nodos, o status tende a aumentar devido à adição de mais distâncias calculadas. Em contrapartida a adição de arestas pode reduzir o status, devido à criação de caminhos menores entre os vértices da rede. A adição de um novo vértice tem maior peso pois para cada vértice adicionado à rede, há mais $n - 1$ distâncias entre o novo vértice e os demais vértices a serem adicionadas ao status.

Para o maior status, a curva encontrada foi uma equação exponencial igual a $93.900 * \exp(0,1033 * x)$ com conjunto de confiança igual a $(78.990 * \exp(0,07655 * x); 108.800 * \exp(0,13 * x))$. Da mesma maneira, para os dados de menor status e status mais observado, foi gerada uma equação de primeiro grau igual a $3.979 * x + 34.140$, para o menor status, e $5.863 * x + 48.030$, para o status mais observado, com conjunto de confiança igual a $(3.749 * x + 32.970; 4.209 * x + 35.300)$ e $(5.467 * x + 46.030; 6.259 * x + 50.030)$, respectivamente.

4.4 Árvore nas Bordas

Para caracterizar e prever a evolução das bordas da topologia, modelamos a frequência relativa da quantidade de nodos em árvores e de nodos do subgrafo *2-core*.

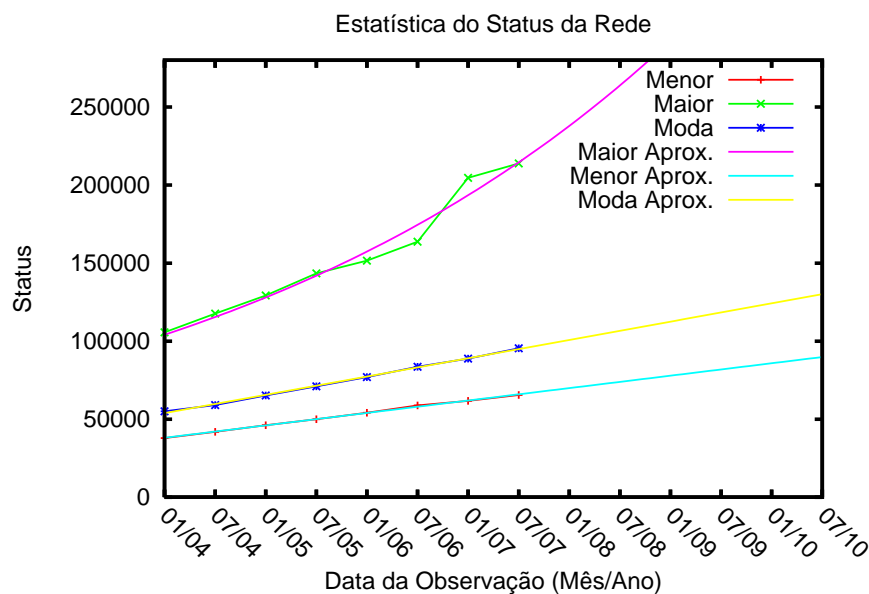


Figura 4.8: Modelos para estatísticas do status dos vértices da rede no período de 2004 a 2010.

A equação obtida para a modelagem da frequência de nodos em árvores foi igual a $0,2663 * x^{-0,3465}$, com conjunto de confiança igual a $(0,2616 * x^{-0,3907}; 0,2616 * x^{-0,2907})$. Para a modelagem da frequência relativa da quantidade de nodos no subgrafo 2 -core foi gerada uma equação de potência igual a $0,7452 * x^{0,07759}$, com conjunto de confiança igual a $(0,7125 * x^{0,04883}; 0,7779 * x^{0,1063})$.

A Figura 4.9 apresenta a frequência relativa dos nodos em árvores e tamanho do subgrafo 2 -core, em relação à quantidade total de nodos na rede, com as suas respectivas retas geradas. Observamos que há cada vez mais nodos que pertencem a ciclos na rede e cada vez menos nodos em subgrafos acíclicos. Este fato também é resultado do aumento de conectividade da rede.

Os resultados desta análise permitem enunciar as seguintes leis empíricas:

Lei Empírica 5 (*Profundidade das Árvores*): *A altura das árvores nas bordas é invariante em relação ao crescimento da rede, com valor igual a 2.*

Assim como no trabalho [19], a lei empírica 5 enuncia que a altura dos subgrafos acíclicos, ou árvores, encontradas na borda da topologia, é invariante e igual a 2. Para a quantidade de árvores no grafo, a rede apresenta a tendência de queda subexponencial, fato esse que leva ao enunciado da lei empírica 6.

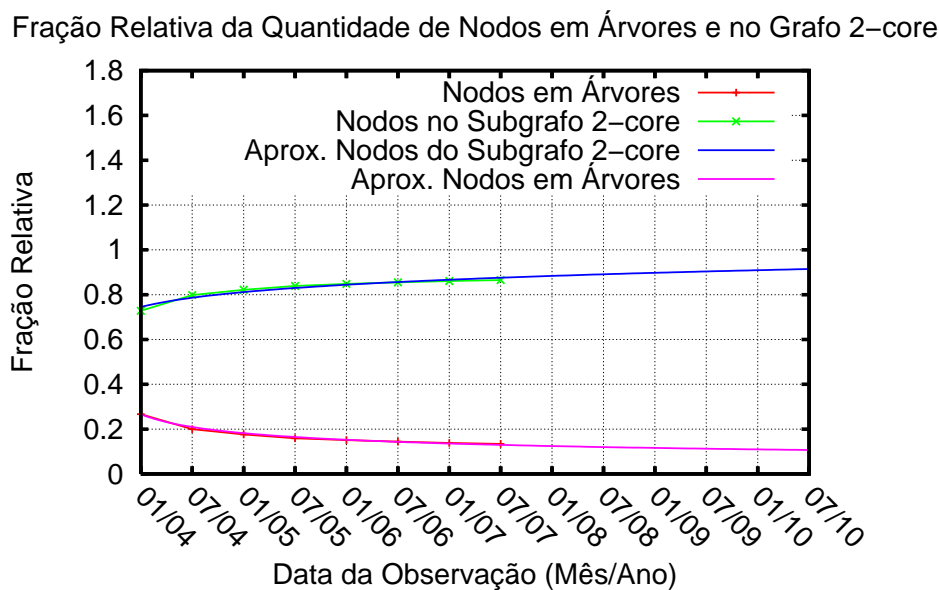


Figura 4.9: Frequência de nodos em árvores e de nodos do subgrafo 2-core entre o período de janeiro de 2004 a 2010.

Lei Empírica 6 (*Crescimento dos Subgrafos Acíclicos*): *A evolução na quantidade de subgrafos acíclicos apresenta tendência de queda subexponencial $\Theta(x^{-k})$.*

Para a lei empírica 6, o resultado obtido não foi similar ao resultado apresentado no estudo [19], que enunciou um crescimento na quantidade de árvores de 32% a cada ano.

Como descrito no capítulo anterior na Seção 3.6, as árvores nas bordas apresentaram uma característica evolutiva relevante durante o período da observação da rede, de janeiro de 2004 a julho de 2007. A caracterização e estudo das árvores nas bordas pode ser de grande importância devido a sua evolução e relação direta com o crescimento e formação da topologia da rede. Tais informações podem ser utilizadas no projeto de geradores de topologias sintéticas de redes complexas como a Internet, que gerem topologias mais realistas.

Ainda em relação ao estudo da borda da topologia, observamos que as árvores nas bordas tem número bastante variável de filhos, que pode chegar a centenas. Avaliamos o número de nodos que compõem estas árvores e observamos um decréscimo, ao longo do tempo, na quantidade total de nodos na árvore com maior número de nodos e o crescimento na quantidade de árvores de dois nodos. A Figura 4.10 ilustra um exemplo

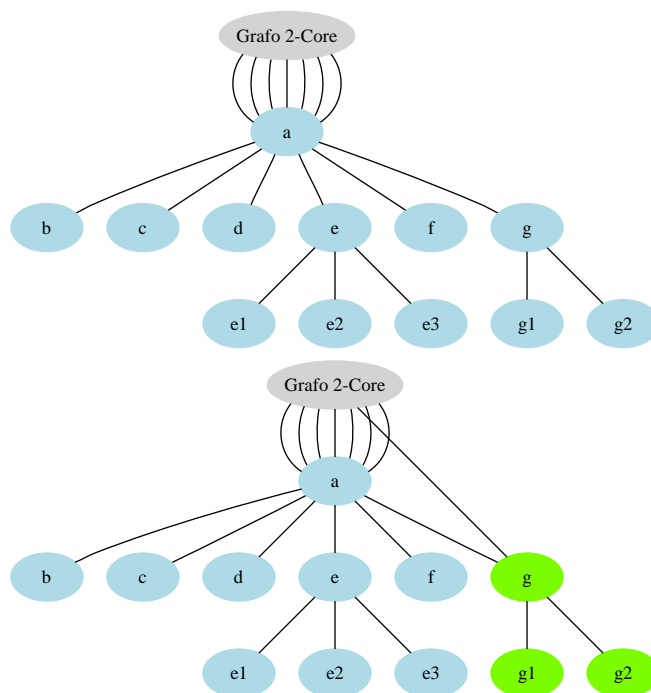


Figura 4.10: Exemplo de uma árvore fictícia na borda (acima). Exemplo da árvore fictícia (acima) na adição de uma aresta (abaixo).

da alteração de uma árvore fictícia na borda, durante a adição de uma aresta, conectando um nodo da árvore ao grafo *2-core*. O grafo acima mostra uma árvore na borda, com altura igual a 2, com 12 nodos e que possui raiz no nodo *a*. Tal árvore está diretamente conectada ao grafo *2-core* a partir do vértice *a*. No caso da adição de uma aresta em tal árvore, representada pela Figura 4.10 abaixo, conectando o nodo *g* da árvore ao grafo *2-core*, fazendo com que a árvore de 12 nodos seja dividida em duas árvores com raízes em *a* e *g* com 9 e 3 nodos respectivamente, representadas pelos nodos em azul e verde, respectivamente. A Figura 4.11 apresenta resultados para a evolução da quantidade de nodos na maior árvore apresentando a tendência de diminuir com o tempo. Esta função é modelada por uma curva do tipo subexponencial igual a $330,6 * x^{-0,2841}$ com conjunto de confiança igual a $(315,2 * x^{-0,3208}; 346,1 * x^{-0,2474})$.

Tendo em vista estes resultados, estima-se que a quantidade de nodos na maior árvore encontrada decresce com o passar do tempo. A hipótese mais concreta para explicar este resultado é a de que o maior crescimento em relação ao número de arestas contribui para a criação de ciclos na rede, e diminui a quantidade de vértices das árvores nas bordas. A

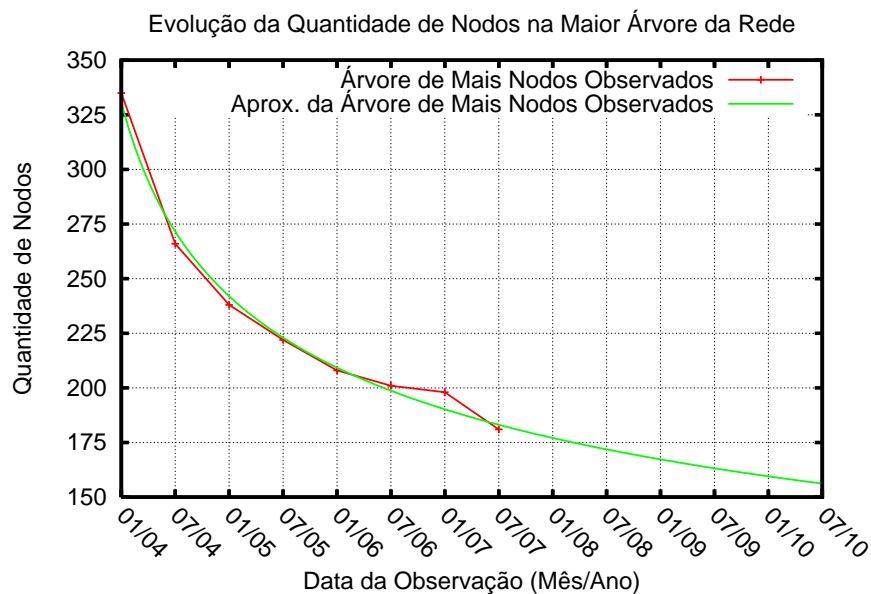


Figura 4.11: Evolução na quantidade de nodos na maior árvore da rede entre o período de janeiro de 2004 a 2010.

árvore com maior quantidade de nodos prevista para julho de 2010 deverá ter cerca de 156 nodos, sendo que, em julho de 2007, a maior árvore continha 181 nodos.

Paralelamente ao decréscimo na quantidade de vértices nas maiores árvores nas bordas, existe um crescimento na quantidade de árvores de apenas dois nodos. É possível observar na Figura 4.12 tal crescimento, modelado por uma curva da forma exponencial igual a $496.2 * x^{0.4635}$ com conjunto de confiança igual a $(388, 3 * x^{0.3355}; 604, 1 * x^{0.5914})$.

É importante ressaltar que ambas as aproximações revelaram equações de mesma ordem assintótica, igual a $(k_1 x^{k_2}) = \Theta(x^k)$, em que x representa a data da captura dos dados e k_1 , k_2 e k representam constantes numéricas. Mesmo sendo os valores de k_1 e k_2 diferentes nas equações modeladas para as métricas de maior quantidade de nodos observados em uma árvore e maior quantidade de árvores observadas, é possível que o fenômeno relacionado ao crescimento na quantidade de árvores mais observadas possua uma relação direta com o decréscimo no número de vértices observados na maior árvore. Sendo assim, há a possibilidade de que o decréscimo observado contribui para o aumento na quantidade de árvores de menor quantidade de nodos e no tamanho de outras árvores.

Desta maneira enunciamos mais duas leis empíricas relacionadas as árvores nas bordas.

Lei Empírica 7 (*Diminuição de Nodos na Maior Árvore*): A diminuição na

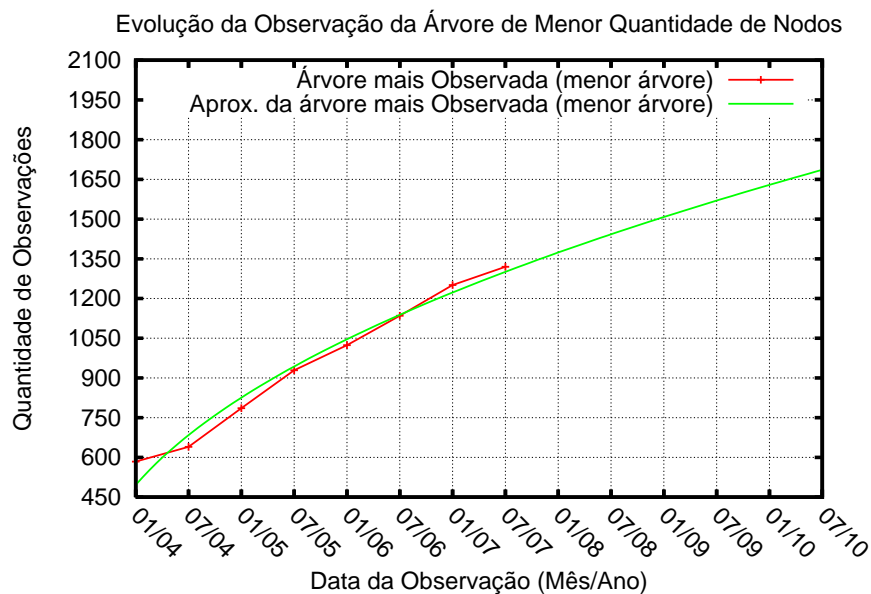


Figura 4.12: Evolução na quantidade árvore com menor número de nós da rede entre o período de janeiro de 2004 a 2010.

quantidade de vértices da maior árvore é de ordem subexponencial da forma $\Theta(x^{-k})$.

Lei Empírica 8 (*Aumento na Quantidade de Menor Árvores*): *O crescimento no número de árvores de 2 nós (menor árvore), é da ordem exponencial da forma $\Theta(x^k)$.*

4.5 Coeficiente de Agrupamento

O coeficiente de agrupamento é uma medida que indica o quanto a rede está próxima de formar um grafo completo. Por isso, é uma medida bastante relevante da topologia, que indica diretamente a densidade das conexões dos Sistemas Autônomos uns com os outros. Considerando a grande adição na quantidade de vértices e arestas com a evolução da rede, assim como os modelos gerados para as mesmas métricas, é de grande importância medir não só o crescimento quantitativo da rede obtida, mas também sua influência nas características e funções da rede. É intuitivo pensar que, com a evolução da Internet e a adição de ASes e conexões, a rede tende a se expandir, no sentido de que as distâncias entre os Sistemas Autônomos aumentem, pois os novos SAa ficariam cada vez mais distantes

dos outros. Mas, da mesma maneira que novos nodos são adicionados na rede a uma determinada taxa λ , conexões entre eles são criadas em uma taxa maior que λ . Assim, podemos também concluir que a rede pode estar se compactando, em relação as distâncias médias, mais do que está se expandindo pela adição de novos vértices. Isso só pode ser comprovado pelo uso de uma métrica que tem a capacidade de capturar a quantificação do agrupamento dos nodos.

Os resultados obtidos e apresentados no Capítulo 3, sobre a métrica do coeficiente de agrupamento, indicaram que, de fato, a rede está se compactando ao evoluir. Foi possível observar que, com a evolução da rede, o coeficiente de agrupamento se mostrou crescente durante todo o período da observação de 2004 a 2007. Para extrapolar tal métrica, foi gerada uma curva exponencial igual a $0,4181 * \exp(0,01095 * x) - 0,697 * \exp(-1,883 * x)$, com um conjunto de confiança igual a $(0,3977 * \exp(0,002943 * x) - 1,431 * \exp(-3,06 * x); 0,4385 * \exp(0,01894 * x) + 0,03813 * \exp(-0,7286 * x))$.

A Figura 4.13 apresenta os resultados do coeficiente de agrupamento. São apresentados os resultados obtidos da análise das topologias reais, bem como a curva modelada para a métrica. Observa-se que o coeficiente de agrupamento cresceu bastante no período de 2004 a 2005, e continua apresentando crescimento contínuo, porém não tão acentuado. Assim, podemos prever que a rede se tornará cada vez mais conectada e compacta.

Esse capítulo apresentou extrapolações dos modelos gerados para as métricas escolhidas para descrever a topologia da Internet. As extrapolações permitiram avaliar a evolução da topologia da Internet no período entre 2008 e 2010. Com base nos modelos gerados, foi possível traçar um perfil da evolução da rede e enumerar leis empíricas sobre suas características. Os resultados indicam que a rede está se compactando com o passar do tempo, as distâncias estão ficando cada vez menores e as estruturas acíclicas presentes nas bordas da rede estão diminuindo. Tais resultados e previsões são úteis em várias áreas de estudo sobre a Internet, tais como protocolos de roteamento, simulações, geradores de topologias sintéticas, segurança e proteção contra ataques.

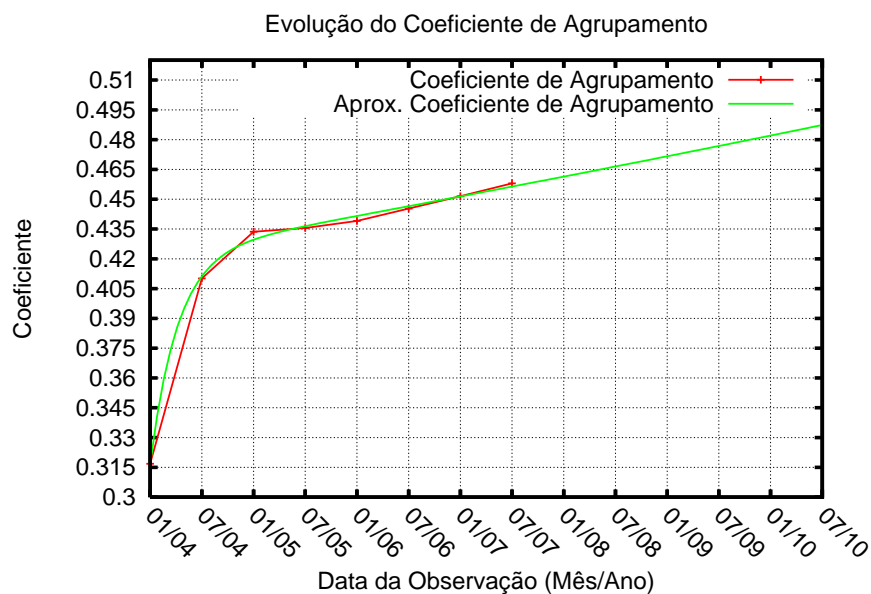


Figura 4.13: Evolução do coeficiente de agrupamento da rede entre o período de janeiro de 2004 a 2010.

4.6 Síntese do Capítulo

Utilizando os modelos gerados a partir da análise da rede, foi possível observar e validar tendências evolutivas da topologia estudada. Foram identificadas oito leis empíricas, sendo quatro delas, sobre o crescimento das arestas e vértices, esgotamento dos identificadores dos ASes e sobre a distância média observada, comparando com leis apresentadas em trabalhos relacionados. Observou-se que tais leis empíricas mostram que a rede apresenta uma nova fase no seu crescimento, levando em conta os diferentes valores obtidos em relação aos trabalhos relacionados. Outra contribuição do estudo é a da apresentação das equações modeladas para as métricas analisadas, acrescentando informações sobre a suas ordens de crescimento ou decrescimento.

CAPÍTULO 5

CONCLUSÃO

Este trabalho apresentou uma análise extensiva da evolução da topologia da Internet no nível dos sistemas autônomos, baseada em dados da topologia coletados no período de janeiro de 2004 a julho de 2007. Esta análise contemplou métricas reconhecidamente importantes na caracterização da topologia da Internet. A análise de dados em um período de quatro anos permitiu observar e modelar o crescimento, o comportamento da rede e os efeitos de sua expansão.

Este trabalho atualiza estudos anteriores sobre a topologia da Internet, encontrados na literatura para dados coletados até 2002, e mostra que alguns padrões de evolução se mantêm e outros foram alterados. Os padrões de evolução que continuam sendo observados são a diminuição da distância média entre pares de vértices e o crescimento mais acelerado no número de arestas em relação ao número de vértices, com conseqüente aumento do grau médio. A tendência alterada referem-se à taxa de crescimento da rede, que atualmente é menor do que as taxas observadas no passado, indicando que a evolução da rede iniciou uma nova etapa. O número de nodos de grau 1 e as árvores nas bordas tendem a diminuir, em mais uma evidência de que a rede está cada vez mais compacta e conectada.

Os resultados e previsões podem ser úteis em várias áreas de estudo sobre a Internet, tais como protocolos de roteamento, simulações, geradores de topologias sintéticas, segurança e proteção contra ataques em protocolos de roteamento.

Os resultados iniciais deste trabalho foram publicados no Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2008) [21]. Diante dos resultados obtidos no trabalho, pode ser objeto de uma futura pesquisa a utilização de métricas características da rede para coloração dos nodos na inferência de uma gramática de grafos [29, 9, 10] que tem como objetivo descrever o crescimento da rede. Com essa

informação será possível gerar topologias sintéticas da rede e ao mesmo tempo caracterizar a rede de forma qualitativa.

BIBLIOGRAFIA

- [1] Traceroute. traceroute. disponível em: www.traceroute.org acesso em: 2008.
- [2] David Alderson, Lun Li, Walter Willinger, e John C. Doyle. Understanding Internet Topology: principles, models, and validation. *IEEE/ACM Trans. Netw.*, 13(6):1205–2118, Piscataway, NJ, USA, IEEE Press, 2005.
- [3] Tain Bu e D. Towsley. On Distinguishing Between Internet Power Law Topology Generators. 2:638–647, Dept. of Comput. Sci., Massachusetts Univ. Amherst, MA, USA, IEEE INFOCOM, 2002.
- [4] Hyunseok Chang, Ramesh Govindan, Sugih Jamin, Scott J. Shenker, e Walter Willinger. Towards Capturing Representative AS-Level Internet Topologies. *SIGMETRICS Perform. Eval. Rev.*, 30(1):280–281, New York, NY, USA, ACM, 2002.
- [5] S. N. Dorogovtsev e J. F. F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Inc., New York, NY, USA, 2003.
- [6] M. Faloutsos, P. Faloutsos, e C. Faloutsos. On Power-law Relationships of the Internet Topology. *ACM SIGCOMM*, páginas 251–262, New York, NY, USA, ACM Press, 1999.
- [7] L. Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Trans. Netw.*, 9(6):733–745, Piscataway, NJ, USA, IEEE Press, 2001.
- [8] Ramesh Govindan e Anoop Reddy. An Analysis of Internet Inter-Domain Topology and Route Stability. *Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution*, páginas 850–857, Washington, DC, USA. IEEE INFOCOM, 1997.

- [9] H-J Kreowski Grzegorz Rozenberg H. Ehrig, G. Engels, editor. *Handbook of Graph Grammars and Computing by Graph Transformation: Applications, Languages and Tools*, volume 2. World Scientific Publishing Company, 1 edition, Outubro, 1999.
- [10] U.Montanari Grzegorz Rozenberg H. Ehrig, H-J Kreowski, editor. *Handbook of Graph Grammars and Computing by Graph Transformations: Concurrency, Parallelism, and Distribution*, volume 3. World Scientific Publishing Company, 1 edition, Outubro, 1999.
- [11] J. Hawkinson e T. Bates. Guidelines for Creation, Selection, and Registration of an Autonomous System (AS). United States, IETF RFC, 1996.
- [12] Siganos-G. Faloutsos M. Krishnamurthy S. (2007). He, Y. A Systematic Framework for Unearthing the Missing Links: Measurements and Impact. In USENIX/SIGCOMM NSDI 2007., 2007.
- [13] TOPOLOGY, Internet. Internet Topology Collection. Disponível em: <http://irl.cs.ucla.edu/topology> Acesso em: 2006.
- [14] SERVICES, Merit IRR. Internet Routing Registry. Disponível em: <http://irr.net> Acesso em 2008.
- [15] Exchanges, internet / exchanges points / peering points. bgp: The border gateway protocol advanced internet routing resources. disponível em: <http://bgp4.as/internet-exchanges> acesso em 2008.
- [16] U.S.R. Murty J.A. Bondy. *Graph Theory (Graduate Texts in Mathematics)*. Springer, 1 edition, Agosto, 2008.
- [17] Rosenberg A. L. Jaiswal, S. e D. Towsley. Comparing the Structure of Power-law Graphs and the Internet AS Graph. páginas 294–303, Washington, DC, USA, IEEE ICNP, 2004.
- [18] Kleinberg J. Leskovec, J. e C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):2, New York, NY, USA, 2007.

- [19] D. Magoni e J. J. Pansiot. Analysis of the Autonomous System Network Topology. *ACM SIGCOMM Comput. Commun. Rev.*, 31(3):26-37., 31(3):26–37, New York, NY, USA, ACM Press, 2001.
- [20] Krioukov D. Fomenkov M. Dimitropoulos X. k c claffy Mahadevan, P. e A. Vahdat. The Internet AS-Level Topology: Three Data Sources and One Definitive Metric. *SIGCOMM Comput. Commun. Rev.*, 36(1):17–26, ACM Press, 2006.
- [21] Alves R. A. Murta, C. D. Análise da Topologia dos Sistemas Autônomos: Evolução e Predição. páginas 37–50, Rio de Janeiro, RJ, Brasil, SBRC, 2008.
- [22] C. R. Palmer e J. G. Steffan. Generating Network Topologies that Obey Power Laws. *GOBLECOM*, 1(1):434–438, San Francisco, CA, USA, 2000.
- [23] Pennock D. M. Park, S.-T. e C. L. Giles. Comparing Static and Dynamic Measurements and Models of the Internet’s AS Topology. *IEEE INFOCOM*, 2004.
- [24] R. Pastor-Satorras e A. Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press. Cambridge University Press, 2004.
- [25] Govindan R. Jamin S. Quian Chen, Hyunseok Chang. The Origin of Power Laws in Internet Topologies Revisited. 2:608–617, Dept. of Electr. Eng. and Comput. Sci. Michigan Univ. Ann Arbor, MI, USA, *IEEE INFOCOM*, 2002.
- [26] Y. Rekhter e T. Li. A Border Gateway Protocol 4 (BGP-4). Relatório técnico, IETF - RFC 1771, 1995.
- [27] RIS, Project. Routing Information Service. Disponível em: <http://www.ris.ripe.net> Acesso em: 2008.
- [28] PAGE, Route Views Projects. University of Oregon Route Views Projects. Disponível em: <http://antc.uoregon.edu/route-views> Acesso em: 2008.

- [29] Grzegorz Rozenberg, editor. *Handbook of Graph Grammars and Computing by Graph Transformation*, volume 1. World Scientific Publishing Company, 1 edition, Janeiro, 1997.
- [30] Karl-Georg Steffens. *The history of approximation theory: from Euler to Bernstein*. Boston: Birkhäuser, c2006. Includes bibliographical references (p.[201]-216) and index.
- [31] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, e W. Willinger. Network topology generators: Degree-based vs Structural. *ACM SIGCOMM Comput. Commun. Rev.*, 32(4):147-159. *SIGCOMM - National Academy of Science, volume 99*, 2002.
- [32] Tian Bu and Lixin Gao and Don Towsley. On Characterizing BGP Routing Table Growth. *Comput. Networks*, 45(1):45-54, New York, NY, USA, Elsevier North-Holland, Inc., 2004.
- [33] D. J. Watts e S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440-442, Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA. djw24@columbia.edu, Julho de 1998.
- [34] Beichuan Zhang, Raymond A. Liu, Daniel Massey, e Lixia Zhang. Collecting the Internet AS-level Topology. *ACM Computer Communication Review. ACM Computer Communication Review*, 35(1):53-61, 2005.