



ANÁLISE DA ESTABILIDADE DO RANQUEAMENTO DE GRAU EM REDES  
LIVRE DE ESCALA

Fernando Fraga Machado

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Miguel Elias Mitre Campista

Rio de Janeiro  
Janeiro de 2016

ANÁLISE DA ESTABILIDADE DO RANQUEAMENTO DE GRAU EM REDES  
LIVRE DE ESCALA

Fernando Fraga Machado

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO  
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE  
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE  
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A  
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA  
ELÉTRICA.

Examinada por:

---

Prof. Miguel Elias Mitre Campista, D.Sc.

---

Prof. Artur Ziviani, Dr.

---

Prof. Daniel Ratton Figueiredo, Ph.D.

---

Prof. Luís Henrique Maciel Kosmowski Costa, Dr.

RIO DE JANEIRO, RJ – BRASIL  
JANEIRO DE 2016

Machado, Fernando Fraga

Análise da Estabilidade do Ranqueamento de Grau em Redes Livre de Escala/Fernando Fraga Machado. – Rio de Janeiro: UFRJ/COPPE, 2016.

XIII, 40 p.: il.; 29, 7cm.

Orientador: Miguel Elias Mitre Campista

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2016.

Referências Bibliográficas: p. 38 – 40.

1. Estabilidade do Ranqueamento de Grau. 2. Evolução de Redes Livre de Escala. 3. Redes Complexas. I. Campista, Miguel Elias Mitre. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*À minha família. Ao meu pai  
Fernando (in memoriam).*

# Agradecimentos

Agradeço aos meus pais por tudo, inclusive por sempre apoiar meus estudos. Agradeço também à minha esposa pelo carinho e ao meu filho pela alegria que trouxe para a minha vida.

Agradeço à Petrobras e à minha equipe de trabalho pelo suporte que me foi dado.

Agradeço ao meu orientador e amigo Miguel Elias Mitre Campista pelo incansável incentivo e pelo apoio.

Agradeço a todos do Grupo de Teleinformática e Automação pelo apoio e pelo companheirismo. Agradeço também aos professores do Programa de Engenharia Elétrica pela grande dedicação, e em especial ao professor Daniel Ratton Figueiredo cujos ensinamentos e dicas foram fundamentais para este trabalho.

Agradeço aos professores Artur Ziviani e Luís Henrique Maciel Kosmowski Costa pela participação na banca examinadora.

Agradeço a todos os funcionários e colaboradores da COPPE/UFRJ que contribuíram de forma indireta com esse trabalho.

Por fim, agradeço aos órgãos CNPq, CAPES e FAPERJ por financiar parcialmente este trabalho.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## ANÁLISE DA ESTABILIDADE DO RANQUEAMENTO DE GRAU EM REDES LIVRE DE ESCALA

Fernando Fraga Machado

Janeiro/2016

Orientador: Miguel Elias Mitre Campista

Programa: Engenharia Elétrica

Muitas redes possuem propriedades que tendem a se manter ao longo do tempo, mesmo havendo grande variação na quantidade de vértices e arestas. Uma dessas propriedades é a estabilidade do ranqueamento dos principais vértices que pode ser avaliada através dos dados históricos da rede. Neste trabalho, técnicas anteriormente desenvolvidas são utilizadas como base para a análise da estabilidade temporal do ranqueamento de grau em redes livre de escala. Adicionalmente, uma análise empírica é proposta para avaliar a influência de fatores externos à estrutura das redes. O objetivo é compreender como os fatores internos e externos afetam a dinâmica temporal do ranqueamento, de modo a permitir prever o nível de estabilidade se mantidas as mesmas condições. Os resultados mostram que a distribuição de grau livre de escala da rede garante maior estabilidade nas posições do topo do ranqueamento e que as variações ocorridas nessas posições são predominantemente causadas por fatores externos. A estabilidade das demais posições do ranqueamento também é afetada por um ruído formado pelas variações regulares de grau da rede.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

## ANALYSIS OF THE DEGREE RANKING STABILITY IN SCALE-FREE NETWORKS

Fernando Fraga Machado

January/2016

Advisor: Miguel Elias Mitre Campista

Department: Electrical Engineering

Many networks have properties that remain invariant over time, even when many vertices and edges are added and removed. One of such properties is the ranking stability of the most important vertices that can be obtained from network historical data. In this work, we analyze the degree ranking stability over time of scale-free networks based on techniques from prior works. Additionally, we propose an empirical analysis to evaluate the influence of external factors to the network structure. The purpose is to understand how the external and internal factors affect the temporal dynamics of the degree ranking, in order to predict the stability level when the same conditions are maintained. Results show that the scale-free degree distribution in the network make the top-ranked positions more stable and the variations occurred in these positions are predominantly caused by external factors. The stability of remaining positions are also affected by a noise formed by common degree variations in the network.

# Sumário

<b>Lista de Figuras</b>	<b>x</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>Lista de Abreviaturas</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Trabalhos Relacionados</b>	<b>4</b>
<b>3 Conjuntos de Dados</b>	<b>9</b>
3.1 Rede de Sistemas Autônomos . . . . .	10
3.2 Rede de Citações . . . . .	10
<b>4 Metodologia da Análise de Estabilidade</b>	<b>15</b>
<b>5 Resultados</b>	<b>18</b>
5.1 Avaliação da estabilidade do ranqueamento de grau da rede de Siste- mas Autônomos . . . . .	18
5.1.1 Análise da quantidade e da estabilidade temporal dos vértices superestáveis . . . . .	18
5.1.2 Análise do impacto do ruído no ranqueamento . . . . .	19
5.1.3 Análise do impacto de fatores pontuais no ranqueamento . . . .	23
5.1.4 Avaliação da previsão de estabilidade do ranqueamento . . . .	25
5.2 Avaliação da estabilidade do ranqueamento de grau da rede de Citações	26
5.2.1 Análise da quantidade e da estabilidade temporal dos vértices superestáveis . . . . .	26
5.2.2 Análise do impacto do ruído no ranqueamento . . . . .	28
5.2.3 Análise do impacto de fatores pontuais no ranqueamento . . . .	31
5.2.4 Previsão de estabilidade do ranqueamento . . . . .	33
<b>6 Conclusões e Trabalhos Futuros</b>	<b>35</b>



# Lista de Figuras

2.1	Valor calculado de $m_c$ em relação ao expoente $\gamma$ para cinco tamanhos de rede ( $N$ ) diferentes. . . . .	6
3.1	Caracterização do conjunto de dados da rede de Sistemas Autônomos ao longo do período em análise. . . . .	11
3.2	Variações nas dez primeiras posições do ranqueamento de grau na rede de Sistemas Autônomos. Cada linha de cor diferente representa um vértice diferente da rede. . . . .	12
3.3	Dez maiores graus na rede de Citações em intervalos de dez anos. . .	12
3.4	Caracterização do conjunto de dados da rede de Citações ao longo do período em análise. . . . .	13
3.5	Variações nas dez primeiras posições do ranqueamento de grau na rede de Citações. Cada linha de cor diferente representa um vértice diferente da rede. . . . .	14
4.1	Metodologia de análise da estabilidade de ranqueamento. A metodologia é formada pela análise de três aspectos: os vértices superestáveis da rede, o ruído e os fatores pontuais. . . . .	17
5.1	Estabilidade média por posição de ranqueamento de grau na rede de Sistemas Autônomos. O tempo médio de estabilidade é calculado ao dividir o tempo total do período pela quantidade de variações ocorridas em cada posição do ranqueamento no mesmo período. . . .	20
5.2	Características de variação do grau relativo na rede de Sistemas Autônomos. . . . .	21
5.3	Adições e remoções de arestas vs. amplitude de ruído $B$ . . . . .	22
5.4	Evolução do grau relativo ( $x_i$ ) dos principais sistemas autônomos. . .	23
5.5	Evolução do grau dos principais sistemas autônomos. As marcações destacam as ocorrências dos fatores pontuais. . . . .	25

5.6	Estabilidade média por posição de ranqueamento de grau na rede de Citações, calculada separadamente em quatro períodos. O tempo médio de estabilidade é calculado ao dividir o tempo total do período pela quantidade de variações ocorridas em cada posição no mesmo período. . . . .	28
5.7	Dispersão suavizada das variações de grau relativo ( $\sigma_{\Delta x}$ ) da rede de Citações no período de 1970 a 2009. . . . .	29
5.8	Características de variação do grau relativo sob diferentes configurações do conjunto de dados Citações para o período de 1970 a 2009. . . . .	30
5.9	Adições e remoções de arestas vs. amplitude de ruído $B$ . . . . .	31
5.10	Evolução do grau relativo ( $x_i$ ) dos principais artigos da rede. . . . .	32
5.11	Evolução do grau absoluto dos principais artigos da rede. As marcações indicam os momentos de laureamento com prêmio Nobel. . . . .	32
5.12	Evolução do grau dos dez artigos com maior quantidade de citações no final do período em análise. . . . .	34

# Lista de Tabelas

3.1	Características das redes utilizadas neste trabalho: rede de “Sistemas Autônomos” e rede de “Citações”. . . . .	9
5.1	Propriedades da rede e quantidade prevista de vértices superestáveis em cinco diferentes instantes no tempo. . . . .	19
5.2	Amplitude do ruído e variações em cada ano na rede de Sistemas Autônomos. . . . .	22
5.3	ASN e grau dos 5 sistemas autônomos de maior grau ao final do período analisado e após 1 e 2 anos. . . . .	26
5.4	Propriedades da rede de Citações e quantidade prevista de vértices superestáveis em diferentes instantes no tempo. . . . .	27
5.5	Amplitude do ruído e variações a cada período de vinte anos na rede de Citações. . . . .	31

# Lista de Abreviaturas

APS	<i>American Physical Society</i> , p. 10
ASN	<i>Autonomous System Number</i> , p. 24, 25
BGP	<i>Border Gateway Protocol</i> , p. 10
CAIDA	<i>Center for Applied Internet Data Analysis</i> , p. 10
DOI	<i>Digital Object Identifier</i> , p. 30
HITS	<i>Hyperlink-Induced Topic Search</i> , p. 4
IPv4	<i>Internet Protocol version 4</i> , p. 24
MLE	<i>Maximum Likelihood Estimation</i> , p. 15

# Capítulo 1

## Introdução

O estudo empírico de redes reais vem identificando propriedades que se repetem em redes de origens distintas. Uma das propriedades comuns a diversas redes reais, como redes tecnológicas e redes sociais, é a distribuição de grau com cauda pesada, na qual há a ocorrência de vértices com grau muito superior à média [1]. Dessas redes, as que possuem distribuição de grau próxima a uma lei de potência, ou seja, seguindo o modelo  $P(k) \sim k^{-\gamma}$  com  $\gamma > 1$  constante, onde  $P(k)$  denota a fração de vértices com grau  $k$ , são chamadas de redes livre de escala [2]. Muitas redes apresentam essa propriedade, tais como a rede formada por *links* de páginas *web* [3], a rede de vizinhança de sistemas autônomos da Internet [4] e até mesmo a rede formada por citações de artigos científicos [5]. Essa diversidade em termos de aplicabilidade explica o atual interesse científico na área de redes complexas [6].

Um importante aspecto de qualquer rede é o ranqueamento de seus vértices segundo algum critério de importância. Por exemplo, o PageRank é um algoritmo de ranqueamento inicialmente proposto pelo Google para ranquear os vértices da rede de páginas *web* em função da estrutura da rede [7]. Outra forma mais simples de ranqueamento é a contagem de grau, usada por exemplo ao ranquear usuários do Twitter por quantidade de seguidores. Nas redes livres de escala os vértices com grau muito superior à média da rede, também chamados de *hubs*, são essenciais para a rede. Por estarem conectados a muitos vértices, os *hubs* são responsáveis por reduzir drasticamente as distâncias na rede [8] e por manter a rede conectada [9]. Por outro lado, também são os principais atores na propagação de epidemias [10].

Frequentemente redes reais sofrem modificações ao longo do tempo com a adição e remoção de vértices e arestas. Surge assim o interesse na dinâmica do ranqueamento ao longo do tempo, com a evolução da rede. Em particular, o estudo da estabilidade das primeiras posições do ranqueamento, que se refere à propriedade dos principais vértices da rede manterem suas posições no ranqueamento inalteradas ao longo do tempo. São muitos os trabalhos que buscam modelar a evolução das redes, mas até onde se sabe, poucos trabalhos avaliam o impacto da evolução no

ranqueamento da rede.

Muitas das características estruturais das redes tendem a se manter ao longo do tempo, mesmo havendo grande variação no tamanho da rede. Uma dessas características é a distribuição de grau livre de escala, que muitas redes reais mantêm com pouca variação. A partir da análise de redes livre de escala, Ghoshal e Barabási [11] demonstram a existência de vértices com diferenciada estabilidade de ranqueamento de PageRank, chamados de superestáveis, e ainda que é possível prever a quantidade aproximada desses vértices a partir de um grafo instantâneo da rede. Após a observação de redes reais que evoluem no tempo, Ghoshal e Barabási concluem que a mesma técnica pode ser aplicada no ranqueamento de grau dessas redes. Já em outro trabalho, Blumm *et al.* provam que o nível de estabilidade geral do ranqueamento está relacionado ao ruído do sistema cuja amplitude absoluta pode ser estimada a partir do histórico das variações da característica analisada, que pode ser o grau [12]. Porém, esses trabalhos não se aprofundam no cálculo do tempo de estabilidade por posição do ranqueamento e também não analisam o impacto de fatores externos à estrutura da rede, aqui chamados de pontuais. São exemplos de fatores pontuais os eventos que aumentam a aptidão de um ou vários vértices em atrair arestas. Tanto o tempo quanto os fatores pontuais têm influência sobre o ranqueamento da rede, mas não foram ainda modelados dada a complexidade.

Esta dissertação de mestrado propõe uma metodologia de análise da estabilidade do ranqueamento de grau considerando o impacto do tempo e dos fatores pontuais em redes livre de escala. As redes analisadas neste trabalho são a rede de sistemas autônomos da Internet e a rede de citações de artigos da coleção Physical Review. A metodologia é baseada em três etapas: (i) na análise da existência de vértices superestáveis considerando o tempo, (ii) na análise do impacto do ruído no nível de estabilidade do ranqueamento ao longo do tempo e (iii) na análise dos fatores pontuais que afetam o ranqueamento. Parte da metodologia é baseada nas técnicas dos trabalhos relacionados [11, 12] e parte é baseada na análise empírica dos dados das redes livre de escala escolhidas.

Os resultados mostram que, devido às estruturas das redes livre de escala analisadas, parte das posições do topo do ranqueamento de grau apresenta elevada estabilidade temporal quando comparada às demais posições do ranqueamento. Além disso, demonstram que os fatores pontuais afetam de forma importante a estabilidade do ranqueamento ao alterar a aptidão dos vértices em atrair conexões. Ao final, os resultados obtidos permitem realizar previsões sobre o nível futuro de estabilidade do ranqueamento das duas redes. No caso da rede de sistemas autônomos é possível ainda verificar o acerto das previsões através da análise de fotografias futuras da rede.

O restante desta dissertação de mestrado está organizada da seguinte forma. O

Capítulo 2 resume os trabalhos relacionados. O Capítulo 3 detalha os conjuntos de dados utilizados. O Capítulo 4 descreve a metodologia empregada na análise da estabilidade do ranqueamento de grau. O Capítulo 5 apresenta os resultados encontrados. O Capítulo 6 conclui este trabalho e apresenta as direções futuras.

## Capítulo 2

# Trabalhos Relacionados

O ranqueamento dos vértices de uma rede pode ser realizado em função de diferentes métricas de centralidade, que visam capturar a importância de um vértice na rede. Nessa direção, duas métricas bastante utilizadas são o grau e o PageRank [13]. Apesar do grau inferir a centralidade baseada apenas no número de adjacências de um dado vértice, ele é frequentemente utilizado pela simplicidade e por exigir apenas conhecimento local. Já o PageRank é uma métrica que foi proposta pelo Google com o objetivo de calcular a importância de páginas *web* a partir da estrutura de links na rede *World Wide Web* de forma recursiva [7]. O valor de PageRank de um vértice é influenciado não apenas pelo seu grau de entrada, mas também pelo valor de PageRank e do grau de saída dos vértices que apontam para ele. Com isso, a importância de um vértice depende também da importância dos vértices que apontam para ele.

A estabilidade do ranqueamento de PageRank foi inicialmente analisada por Ng *et al.* [14], utilizando um conjunto de dados de citações de artigos da área de ciência da computação. Os autores analisaram as variações no ranqueamento de PageRank ao inserir perturbações no conjunto de dados. Por exemplo, foi comparado o ranqueamento do conjunto de dados completo com o ranqueamento do mesmo conjunto após remoção de 30% dos dados. Dentre as perturbações simuladas nesse trabalho, não há simulações relacionadas ao tempo ou ao crescimento da rede. No mesmo trabalho também foi analisada a estabilidade do ranqueamento do HITS (*Hyperlink-Induced Topic Search*), um outro algoritmo desenvolvido para ranqueamento de páginas *web* [15]. Os resultados mostraram maior estabilidade do PageRank em relação ao HITS.

Os dois trabalhos apresentados a seguir são utilizados como base na análise da estabilidade do ranqueamento de grau realizada nesta dissertação.

A análise da estabilidade de ranqueamento em redes livres de escala foi inicialmente proposta no trabalho de Ghoshal e Barabási [11]. Nesse trabalho os autores afirmam que, em redes livre de escala, alguns vértices podem ser considerados

superestáveis, ou seja, alguns vértices podem manter suas respectivas posições no ranqueamento por muito tempo durante o qual a rede evolui. A premissa se baseia na capacidade desses vértices manterem as suas posições no ranqueamento de PageRank. Através de simulações numéricas, os autores atestam que, em determinadas redes, as perturbações que alteram a relação de vizinhança entre os vértices sem haver alteração de grau, não alteram a ordem dos vértices no topo do ranqueamento. Isso ocorre quando a diferença entre o PageRank de um vértice do topo do ranqueamento (vértice chamado de superestável) e do próximo colocado no ranqueamento é maior do que a variação de PageRank do vértice. A cauda pesada da distribuição de grau pode levar à ocorrência de vértices com grau muito maior do que os demais, que manterão alto PageRank mesmo que as arestas de entrada originem em vértices com baixo PageRank. Isto é, a grande diferença de grau entre os vértices de maior grau garante a estabilidade no ranqueamento desses vértices. Em contrapartida, redes com distribuição de grau exponencial não possuem vértices superestáveis por não apresentarem cauda pesada.

A partir desta premissa, Ghoshal e Barabási propõem uma formulação matemática com o objetivo de calcular a quantidade de vértices superestáveis existentes na rede. Segundo essa proposta, se o expoente ( $\gamma$ ) da lei de potência que melhor representa a distribuição de grau da rede livre de escala for entre  $2 \leq \gamma < 3$ , essa rede contém pelo menos um vértice superestável. Para  $\gamma \geq 3$ , a rede só apresentará vértice superestável se a quantidade total de vértices da rede ( $N$ ) for maior que

$$N_c(\gamma) \approx (\alpha(\gamma - 1)/(\gamma - 3)^{1/2})^{2(\gamma-1)}, \quad (2.1)$$

onde  $\alpha$  é o fator de amortecimento da equação de PageRank, tipicamente igual a 0,85. Quando  $N > N_c(\gamma)$ , a quantidade de vértices superestáveis ( $m_c$ ) é de aproximadamente

$$m_{c\pm} = N^{1/(2\gamma-1)} \left[ \alpha(\gamma - 1) \left( \frac{(\gamma - 1)(\gamma - 2)}{\gamma} \right)^{1/2} \times \left( \pm \frac{1 - N^{(3-\gamma)/(\gamma-1)} \Gamma(1 - (\gamma - 1)^{-1})}{\gamma - 3} \right)^{1/2} \right]^{-(\gamma-1)/(\gamma-1/2)} \quad (2.2)$$

onde  $\Gamma(z)$  é a função de Euler-Gama.

A Figura 2.1 mostra o resultado da Equação 2.2 para diferentes valores de  $\gamma$  e  $N$ . Esses resultados são aproximadamente iguais aos resultados verificados em simulações numéricas realizadas pelos autores, onde grafos de redes sintéticas e grafos que representam redes reais foram perturbados de modo a alterar as conexões sem modificar o grau de cada vértice. Tais resultados confirmam que as quantidades de

vértices superestáveis verificadas nas redes livre de escala são pequenas em relação ao total de vértices da rede, mas em geral aumentam com o tamanho da rede. Como pode ser observado na Figura 2.1, uma rede com  $10^7$  vértices apresenta menos de dez vértices superestáveis.

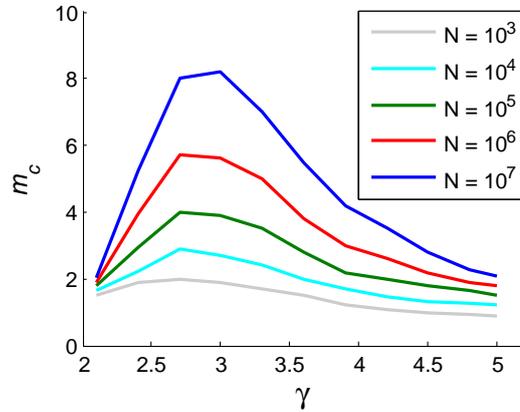


Figura 2.1: Valor calculado de  $m_c$  em relação ao expoente  $\gamma$  para cinco tamanhos de rede ( $N$ ) diferentes.

Adicionalmente, Ghoshal e Barabási realizam uma análise empírica do comportamento do ranqueamento de grau de redes reais ao longo do tempo. São analisados os conjuntos de dados contendo a evolução temporal de três redes reais, onde as quantidades previstas de vértices superestáveis são calculadas a partir de uma fotografia estática de cada rede. A análise da evolução do ranqueamento de grau, feita individualmente nas três redes, aponta que a mesma quantidade aproximada de vértices superestáveis manteve seu ranqueamento de grau por um longo tempo enquanto que os demais vértices apresentaram maior instabilidade. Isto é, as primeiras  $m_c$  posições do ranqueamento de grau se mantiveram estáveis ou apresentaram pouca variação ao longo do tempo, enquanto que as demais posições do ranqueamento apresentaram muito mais variações no mesmo período. Vale ressaltar que essa avaliação não se baseia em prova matemática, apenas na observação empírica do ranqueamento de grau ao longo do tempo.

Em outro trabalho, Blumm *et al.* [12] analisam a dinâmica do ranqueamento em sistemas complexos, considerando seis diferentes conjuntos de dados de sistemas reais que não representam necessariamente redes. Cada sistema é formado por um conjunto de itens e existe uma característica que é medida individualmente ao longo do tempo para formar o ranqueamento dos itens. Por exemplo, no sistema de citações de artigos é considerada a quantidade de citações recebidas anualmente por cada artigo, que por sua vez é um item do sistema. Repare que esse estudo considera o número de citações que cada artigo recebe em cada ano, sem acumular diferentes anos.

Em um sistema complexo, considere  $a_i(t)$  a quantidade absoluta de uma característica de interesse do item  $i$  em um instante de tempo  $t$ , e  $s(t)$  a soma das quantidades absolutas de todos os itens no mesmo instante. Por conseguinte, o valor relativo da característica do item  $i$  pode ser calculado como  $x_i(t) = a_i(t)/s(t)$ , onde  $\sum_i x_i(t) = 1$ . A partir da análise empírica dos seis conjuntos de dados, Blumm *et al.* constataram que a dispersão das variações de  $x$  a cada passo de tempo é proporcional a  $x^\beta$ , ou seja

$$\sigma_{\Delta x|x} = \sqrt{\frac{1}{R} \times \sum_{\forall i,t|x_i(t)=x}^R (x_i(t+1) - x_i(t))^2} \sim x^\beta, \quad (2.3)$$

onde  $R$  é a quantidade de ocorrências de  $x$ .

Diante da constatação dada pela Equação 2.3, os autores formularam uma representação matemática da dinâmica do sistema, onde o valor de  $x_i$  é influenciado não só pela aptidão individual do item em relação à qualidade medida mas também pelo ruído presente no sistema. Segundo essa proposta, o valor futuro de  $x_i$  é determinado pela equação

$$\dot{x}_i = f(x_i) + g(x_i)\xi_i(t) - \phi(t)x_i, \quad (2.4)$$

onde  $f(x_i)$  representa a aptidão do item em conjunto com outros mecanismos que definem o provável valor de  $\dot{x}_i$ . O termo  $g(x_i)\xi_i(t)$  representa o ruído, onde  $g(x_i)$  modela a sua amplitude e  $\xi_i(t)$  é uma variável aleatória Gaussiana de média zero. O termo  $\phi(t)x_i$  garante que  $\sum_i x_i(t) \equiv 1$ . Devido ao resultado da 2.3, os autores assumem que a amplitude do ruído é definida como

$$g(x_i) = Bx_i^\beta, \quad (2.5)$$

onde os valores do coeficiente  $B$  e do expoente  $\beta$  são obtidos a partir de  $\sigma_{\Delta x|x}$ .

Ao analisar os resultados extraídos dos seis conjuntos de dados, os autores concluem que o valor do coeficiente  $B$  está diretamente relacionado à estabilidade do ranqueamento do sistema. Ou seja, quando o coeficiente  $B$  é maior do que a diferença  $|x_i - x_j|$ , sendo que  $x_i$  e  $x_j$  são os mais próximos em valor dentre os itens do sistema, o ranqueamento sofre variações causadas pelo ruído. Dentre os conjuntos de dados avaliados, os sistemas considerados estáveis apresentam valor de  $B$  na ordem de  $10^{-3}$  enquanto que os sistemas instáveis possuem  $B$  na ordem de  $10^{-2}$  ou  $10^{-1}$ . Já o valor do expoente  $\beta$  é menor que 1 nos seis sistemas, indicando que a amplitude do ruído é proporcionalmente menor nos itens mais próximos do topo do ranqueamento.

Diferente dos trabalhos anteriores, esta dissertação de mestrado analisa a estabi-

lidade do ranqueamento de grau acrescentando o cálculo do tempo de permanência em cada posição do ranqueamento e a análise do impacto de fatores externos à estrutura da rede.

# Capítulo 3

## Conjuntos de Dados

Os resultados deste trabalho são obtidos a partir de dois conjuntos de dados: um que descreve a vizinhança entre sistemas autônomos da Internet (chamado de agora em diante de “Sistemas Autônomos”) e outro que retrata as citações a artigos da coleção de revistas Physical Review (chamado de agora em diante de “Citações”). Neste trabalho, os conjuntos de dados são modelados como grafos, sendo que as redes resultantes compartilham algumas propriedades, notadamente a distribuição de grau livre de escala [4, 5, 16]. Por outro lado, também possuem características estruturais distintas, como o fato das arestas serem direcionadas na rede de Citações e não-direcionadas na rede de Sistemas Autônomos. A Tabela 3.1 resume as principais características de cada uma das redes.

Tabela 3.1: Características das redes utilizadas neste trabalho: rede de “Sistemas Autônomos” e rede de “Citações”.

Rede	Vértices	Arestas	Tipo de arestas	Variações no tempo
<b>Sistemas Autônomos</b>	Sistemas autônomos	Conexões lógicas	Não-direcionadas	Ocorrem adições e remoções de vértices e arestas
<b>Citações</b>	Artigos	Citações	Direcionadas	Ocorrem adições de vértices e arestas

Uma característica importante dos grafos utilizados é que eles evoluem no tempo uma vez que as redes sofrem variações nas quantidades de vértices e arestas. Deste modo, a dinâmica do ranqueamento é analisada a partir de fotografias periódicas do grafo de cada rede.

Nas seções a seguir são detalhadas as características de cada rede.

## 3.1 Rede de Sistemas Autônomos

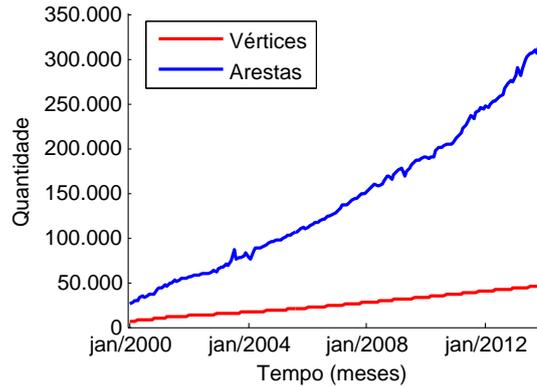
Os conjuntos de dados contendo as listas mensais de conexões entre sistemas autônomos da Internet foram obtidos através da página *web* do CAIDA (*Center for Applied Internet Data Analysis*). Os dados disponibilizados pela CAIDA foram inferidos a partir das tabelas de roteamento divulgadas na Internet por meio do protocolo BGP (*Border Gateway Protocol*). O grafo completo da rede em um determinado mês é determinado através da coleta das tabelas do BGP a partir de diversas posições geográficas, durante intervalos de oito horas ao longo dos cinco primeiros dias do mês [17].

Neste trabalho, as análises estão baseadas nos dados referentes ao período de janeiro de 2000 a novembro de 2013, por ser o maior período sem interrupção disponível na página da CAIDA. A análise do ruído, baseada na técnica proposta por Blumm *et al.* [12], neste trabalho considera os dados em intervalos de um mês, portanto uma lacuna de tempo maior do que um mês no meio do período inviabilizaria essa análise. Entre o início e o fim desse período, a rede de Sistemas Autônomos cresceu mais de 5 vezes em quantidade de vértices e mais de 10 vezes em número de arestas (Figura 3.1(a)). Apesar disso, a distribuição de grau da rede manteve sua característica livre de escala com pouca variação, conforme ilustrado na Figura 3.1(b) que contém a distribuição de grau da rede no início, meio e fim do período em análise neste trabalho.

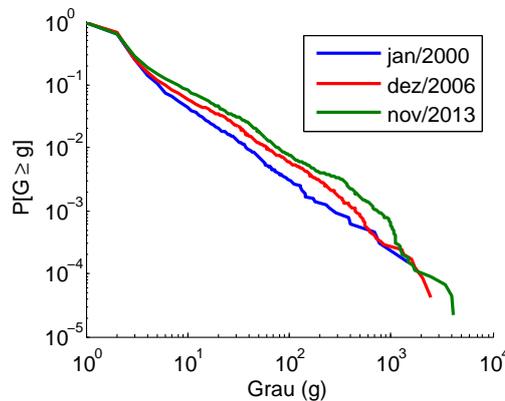
A Figura 3.2 apresenta a evolução do ranqueamento dos dez maiores graus da rede a cada mês. Cada linha com cor diferente representa um sistema autônomo diferente. Note que no período em análise, 29 sistemas autônomos diferentes ocuparam as 10 primeiras posições do ranqueamento de grau ao fim do período, demonstrando certa instabilidade geral mesmo no topo do ranqueamento. Em particular, repare que o primeiro no ranqueamento em 2000 se manteve na mesma posição por nove anos consecutivos, mas ao fim do período ocupou a sexta posição. Isso indica que fatores pontuais influenciam a aptidão dos sistemas autônomos e conseqüentemente a habilidade em manter posições no ranqueamento. Por outro lado, observa-se que o tempo que os sistemas autônomos permanecem na mesma posição é maior nas primeiras posições do ranqueamento.

## 3.2 Rede de Citações

O conjunto de dados de citações da APS (*American Physical Society*) é fornecido gratuitamente através de solicitação e contém as citações feitas entre os artigos da coleção *Physical Review* desde 1893 até 2013 [18]. Os arquivos disponibilizados possuem a lista de citações, onde cada artigo citante e citado é representado por um



(a) Evolução temporal da quantidade de vértices e arestas.



(b) Distribuição de grau da rede em três instantes no tempo.

Figura 3.1: Caracterização do conjunto de dados da rede de Sistemas Autônomos ao longo do período em análise.

identificador, além dos dados cadastrais de cada artigo. Neste trabalho, a data de surgimento de cada citação é determinada pela data de publicação do artigo citante.

O mesmo conjunto de dados de citações foi utilizado nos trabalhos relacionados [11] e [12], porém de formas diferentes. No trabalho de Ghoshal e Barabási [11], o conjunto de dados foi modelado como uma rede, considerando tanto os artigos citantes quanto os artigos citados no cálculo do tamanho da rede. Além disso, a quantidade de citações é contabilizada desde o surgimento de cada artigo até o instante em questão. Já no trabalho de Blumm *et al.* [12], apenas a quantidade anual bruta de citações recebidas por cada artigo foi utilizada, sem considerar a relação entre artigo citado e o que citou.

Neste trabalho, a estabilidade do ranqueamento de grau de entrada da rede de Citações é avaliada, ou seja, o ranqueamento da quantidade de citações recebidas por cada artigo desde o surgimento do artigo até a data alvo do ranqueamento. Nesse sentido, é analisada a evolução do ranqueamento a partir de 1930, quando a rede atinge um tamanho capaz de formar um ranqueamento de grau com o topo

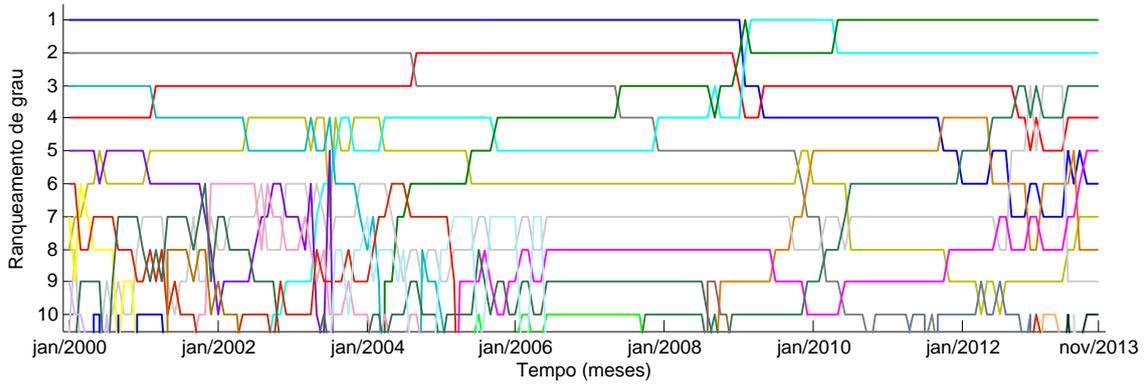


Figura 3.2: Variações nas dez primeiras posições do ranqueamento de grau na rede de Sistemas Autônomos. Cada linha de cor diferente representa um vértice diferente da rede.

definido. Como mostrado na Figura 3.3, no ano de 1910 o maior grau da rede é três apenas e há seis artigos com essa quantidade de citações. Em 1920 o maior grau da rede aumenta para cinco mas existem também cinco artigos com essa quantidade. Esses empates causam indefinição no topo do ranqueamento de grau, dificultando a análise da dinâmica do ranqueamento nesses períodos. Já em 1930 o maior grau é de vinte e seis e as quatro primeiras posições do ranqueamento não apresentam empate. Em 1940 e 1950 não são mais observados empates nas dez primeiras posições do ranqueamento de grau.

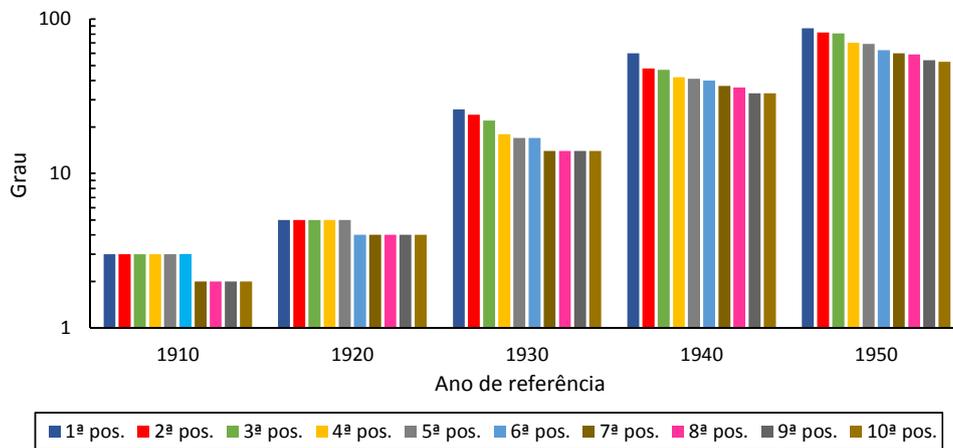
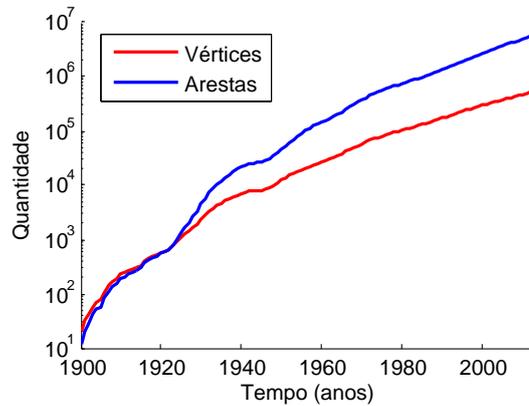


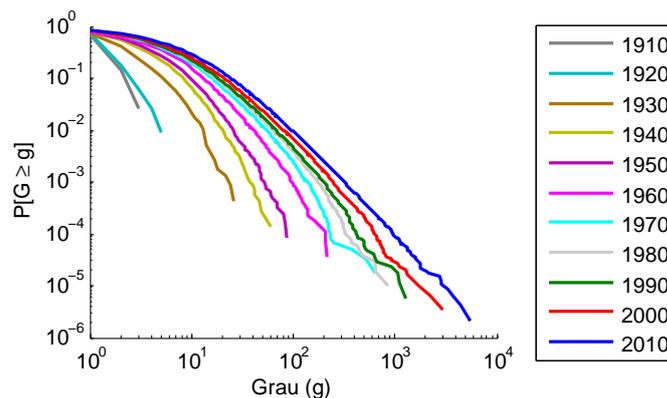
Figura 3.3: Dez maiores graus na rede de Citações em intervalos de dez anos.

A Figura 3.4(a) mostra a evolução da quantidade de vértices e de arestas na rede de Citações desde 1900 até 2013. Note que o eixo vertical está em escala logarítmica, pois a rede cresce várias ordens de grandeza nesse período. Ao contrário da rede de Sistemas Autônomos, a distribuição de grau da rede de Citações apresenta significativa variação ao longo do período em análise, conforme ilustrado na Figura 3.4(b). A rede só desenvolve uma distribuição próxima a uma lei de potência na cauda em

torno de 1940 [11], e a partir daí essa cauda se torna cada vez mais pesada com o passar do tempo. Em outras palavras, há um aumento gradual da ocorrência de vértices com grau muito maior que o grau médio. Esse resultado está de acordo com a constatação anterior de ocorrência de muitos empates nas primeiras posições do ranqueamento de grau nos primeiros anos de coleta de dados.



(a) Evolução temporal da quantidade de vértices e arestas.



(b) Distribuição de grau da rede em diferentes instantes no tempo.

Figura 3.4: Caracterização do conjunto de dados da rede de Citações ao longo do período em análise.

A Figura 3.5 apresenta a evolução do ranqueamento dos dez maiores graus da rede de Citações a cada ano. Cada linha com cor diferente representa um artigo diferente. Note que as primeiras posições alternam períodos de maior e de menor estabilidade. Por exemplo, entre 1940 e 1950 há menos variações nessas posições do que na década seguinte. Já a partir de 1980 há um aumento de estabilidade por maior período. Assim como foi observado no ranqueamento da rede de Sistemas Autônomos, as primeiras posições do ranqueamento de Citações também apresentam maiores períodos de estabilidade do que as posições seguintes.

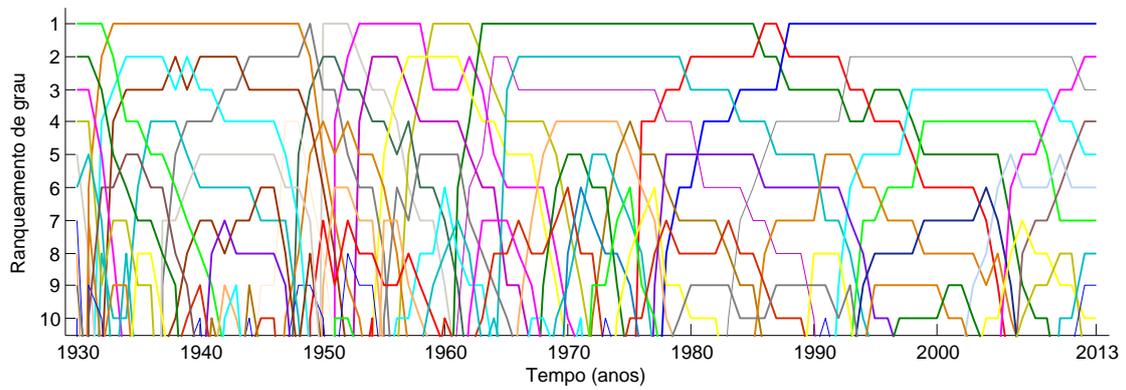


Figura 3.5: Variações nas dez primeiras posições do ranqueamento de grau na rede de Citações. Cada linha de cor diferente representa um vértice diferente da rede.

# Capítulo 4

## Metodologia da Análise de Estabilidade

Os fatores que influenciam a estabilidade do ranqueamento de grau são analisados neste trabalho sob três diferentes aspectos: dos vértices superestáveis da rede, do ruído e dos fatores pontuais. Para isso, a metodologia de análise descrita nesta seção é utilizada, tendo como entrada os dados da rede (grafo) e, no caso da análise dos fatores pontuais, dados adicionais de fatos históricos. O objetivo final da análise é compreender como cada característica ou fator afeta o ranqueamento, de modo que seja possível também prever a dinâmica do ranqueamento. A seguir, cada um dos aspectos analisados é apresentado.

1. **Análise da quantidade e da estabilidade temporal dos vértices superestáveis:** A quantidade prevista de vértices superestáveis ( $m_c$ ) no período em análise é calculada a partir da equação proposta por Ghoshal e Barabási [11]. Os dados de entrada dessa equação são a quantidade total de vértices ( $N$ ) e o expoente da lei de potência aproximada da distribuição de grau ( $\gamma$ ) de uma fotografia da rede. Como essas características da rede podem sofrer alteração ao longo do tempo, uma única fotografia da rede, como usada por Ghoshal e Barabási, não é suficiente para avaliar a dinâmica da rede. Portanto, este trabalho considera diferentes fotografias correspondentes a diferentes instantes de tempo. Para estimar  $\gamma$  a partir da cauda da distribuição de grau, foi escolhida a técnica de estimativa de máxima verossimilhança (*Maximum Likelihood Estimation* - MLE). Em seguida, para avaliar a relação da estabilidade com o tempo, também não tratada por Ghoshal e Barabási, foi introduzido neste trabalho o cálculo do tempo médio de estabilidade do ranqueamento, que é a média de tempo em que um mesmo vértice se manteve ininterruptamente em cada posição do ranqueamento. Por fim, o tempo médio de estabilidade dos vértices apontados como superestáveis é comparado com o tempo médio de

estabilidade dos demais vértices.

2. **Análise do impacto do ruído no ranqueamento:** Conforme proposto por Blumm *et al.* [12], o nível de estabilidade geral do ranqueamento de um sistema complexo está relacionado à amplitude do ruído, que é calculada a partir do histórico de variações de grau da rede ao longo do período em análise. Para reproduzir a técnica, mediu-se as variações de grau proporcionais ( $\Delta x|x$ ) e em seguida extraiu-se a amplitude do ruído ( $B$ ) a partir da dispersão suavizada de  $\Delta x|x$ , conforme proposto em [12]. Os programas desenvolvidos para realizar esses cálculos foram avaliados através da análise do conjunto de dados de citações da coleção Physical Review [18]. A estratégia de refazer os cálculos para o mesmo conjunto de dados analisado no artigo original permitiu comparar os resultados e validar o programa. Neste trabalho, além de calcular o valor de  $B$  para todo o período analisado, o mesmo cálculo é feito para períodos menores. A ideia é verificar o quanto que a estabilidade da rede muda ao longo do tempo. Para isso, diferente dos trabalhos anteriores, os resultados são comparados com as variações observadas no ranqueamento de grau e nas quantidades de adições e remoções de arestas no mesmo período. Ao final, é possível avaliar se a variação de  $B$  tem correlação com as variações estruturais observadas empiricamente na rede.

3. **Análise do impacto de fatores pontuais no ranqueamento:** Fatores pontuais, que influenciam as variações de grau de um grupo de vértices ou de um único vértice, também afetam a estabilidade do ranqueamento da rede. São alguns exemplos de fatores pontuais investigados nesse trabalho:

- questões econômicas;
- estratégias de negócio;
- intervenções de órgãos reguladores no mercado;
- limitações tecnológicas;
- conflitos mundiais;
- premiações importantes.

Como as origens dos fatores pontuais são muito diversas e suas ocorrências pouco documentadas, a análise de todos os fatores pontuais que afetam as redes do porte da Internet e da coleção de periódicos Physical Review torna-se complexa. Neste trabalho, porém, são levantados os fatores pontuais relacionados aos principais vértices do ranqueamento de grau, isto é, relacionados aos vértices que ocupam as primeiras posições do ranqueamento ao longo do período em análise. Após a identificação dos fatores pontuais, o impacto de cada

um no ranqueamento de grau é avaliado de forma empírica, através da análise da evolução histórica do grau desses vértices. Essa análise, não realizada em trabalhos anteriores, permite avaliar como os fatores pontuais podem alterar a estabilidade de ranqueamento dos vértices do topo, já que o ruído pode não ser o único responsável por provocar as mudanças de posição no ranqueamento.

A Figura 4.1 resume as etapas da metodologia empregada na análise de estabilidade do ranqueamento, destacando a contribuição deste trabalho em cada uma das etapas.

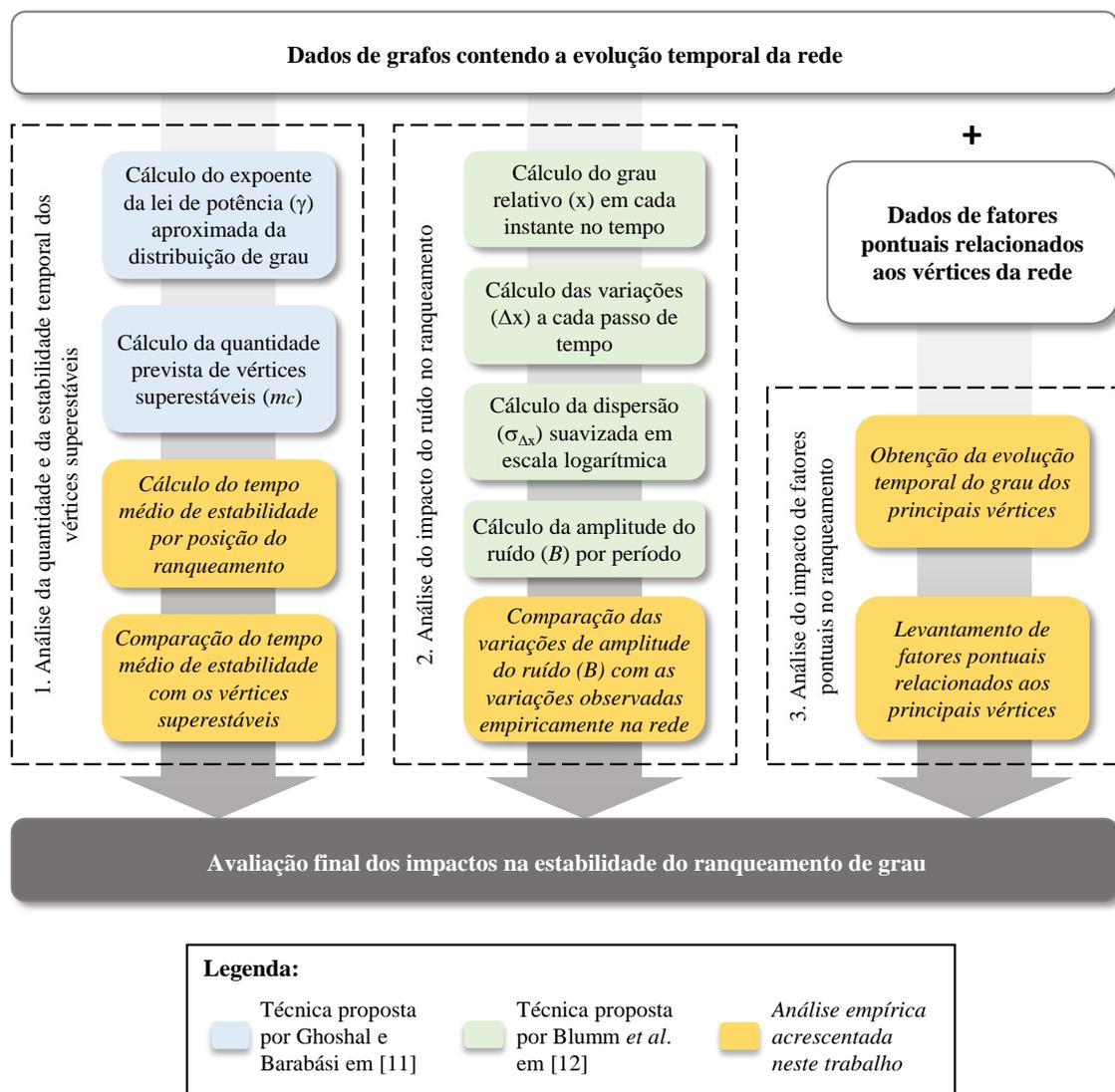


Figura 4.1: Metodologia de análise da estabilidade de ranqueamento. A metodologia é formada pela análise de três aspectos: os vértices superestáveis da rede, o ruído e os fatores pontuais.

# Capítulo 5

## Resultados

São apresentados a seguir os resultados das análises propostas no Capítulo 4 para a rede de Sistemas Autônomos e a rede de Citações.

### 5.1 Avaliação da estabilidade do ranqueamento de grau da rede de Sistemas Autônomos

Conforme explicado no Capítulo 3, as análises da rede de Sistemas Autônomos são baseadas nos dados do período de janeiro de 2000 a novembro de 2013, a partir de fotografias mensais da rede disponíveis no conjunto de dados. Ao final desta seção é realizada uma avaliação da previsão feita a partir dos resultados, aproveitando o fato de estarem disponíveis também os dados de alguns meses dos anos de 2014 e 2015.

#### 5.1.1 Análise da quantidade e da estabilidade temporal dos vértices superestáveis

A quantidade prevista de vértices superestáveis ( $m_c$ ) na rede de Sistemas Autônomos, calculada através da equação proposta por Ghoshal e Barabási [11], é de aproximadamente dois vértices em todo o período analisado. Isso quer dizer que as duas primeiras posições no ranqueamento de grau da rede de Sistemas Autônomos devem apresentar estabilidade muito maior do que as demais posições do ranqueamento. Na Tabela 5.1 estão os valores de  $N$  (quantidade total de vértices) e  $\gamma$  (expoente da lei de potência aproximada da distribuição de grau) utilizados no cálculo de  $m_c$ , além do grau médio e dos três maiores graus, referentes ao estado da rede no oitavo mês dos anos de 2000, 2003, 2006, 2009 e 2013. É possível constatar que os três maiores graus são muito maiores do que o grau médio em todos os casos. Em 2009 os três maiores graus estavam muito próximos, sugerindo uma baixa esta-

bilidade no ranqueamento entre eles (fato que pode ser confirmado na Figura 3.2). Também é observado um aumento da densidade de arestas na rede, o que significa que a rede de Sistemas Autônomos está cada vez mais conectada, aumentando a disponibilidade de sistemas autônomos para trânsito na Internet.

Tabela 5.1: Propriedades da rede e quantidade prevista de vértices superestáveis em cinco diferentes instantes no tempo.

Rede	Tamanho $N$ da rede	$\gamma$ apro- ximado	$m_c$ previsto	Grau médio	Maior grau	2° maior grau	3° maior grau
ago/2000	8.284	2,4	2	4,2	1852	996	814
ago/2003	15.821	2,3	2	4,8	2446	1808	1667
ago/2006	23.102	2,3	2	5,2	2407	2029	1740
ago/2009	32.265	2,3	2	5,7	2487	2350	2241
ago/2013	45.067	2,2	2	6,8	4042	3815	3217

A Figura 5.1 mostra o tempo médio de estabilidade de cada posição do ranqueamento de grau ao longo de todo o período analisado. As duas primeiras posições do ranqueamento apresentam tempo médio de estabilidade bem superior às demais posições, confirmando o resultado obtido através da equação proposta por Ghoshal e Barabási. Entretanto, ao contrário do que sugere o nome superestável, as duas primeiras posições do ranqueamento não são estáveis por todo o período analisado. Considerando o período total analisado de 167 meses, o tempo médio de estabilidade foi de aproximadamente 42 meses (ou 25% do tempo total) para a primeira posição e 28 meses (ou 17% do tempo total) para a segunda posição. O tempo médio de estabilidade da terceira posição foi de aproximadamente 12 meses (ou 7% do tempo total), sendo menos da metade da estabilidade observada na segunda posição. Apesar da estabilidade das duas primeiras posições do ranqueamento se destacarem, vale ressaltar que o tempo médio apresenta uma tendência contínua de queda à medida que se observa posições mais distantes do topo do ranqueamento.

### 5.1.2 Análise do impacto do ruído no ranqueamento

O nível de estabilidade geral do ranqueamento da rede é avaliado a partir da técnica desenvolvida por Blumm *et al.* [12]. Para isso, foram calculadas as variações entre meses subsequentes dos valores relativos de grau ( $\Delta x$ ) e em seguida a dispersão dessas variações ( $\sigma_{\Delta x}$ ).

As Figuras 5.2(a) e 5.2(b) apresentam, respectivamente, os gráficos de  $\Delta x$  e  $\sigma_{\Delta x}$  considerando todo o período em análise. Na Figura 5.2(a), as cores mais quentes indicam maior densidade de pontos. É possível observar uma assimetria na distribuição dos pontos de cor azul (pontos de menor densidade) em torno de  $\Delta x = 0$ ,

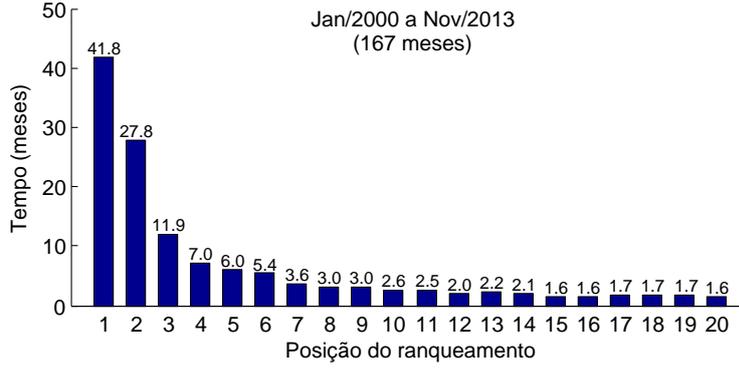
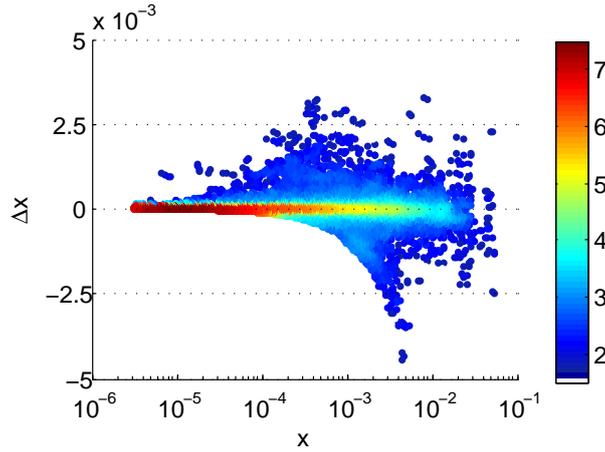


Figura 5.1: Estabilidade média por posição de ranqueamento de grau na rede de Sistemas Autônomos. O tempo médio de estabilidade é calculado ao dividir o tempo total do período pela quantidade de variações ocorridas em cada posição do ranqueamento no mesmo período.

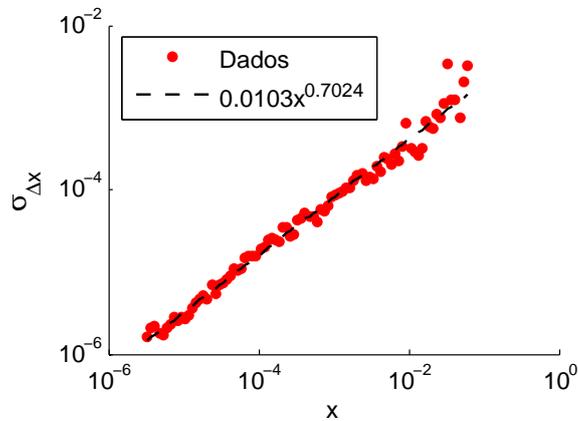
sendo que os pontos relativos aos valores mais baixos  $x$  estão mais concentrados acima de  $\Delta x = 0$ , enquanto que os pontos relativos aos valores mais altos de  $x$  estão mais concentrados abaixo de  $\Delta x = 0$ . Isso significa que há uma tendência dos sistemas autônomos de menor grau relativo subir no ranqueamento e os de maior grau relativo cair no ranqueamento. Contudo, essa tendência não é tão frequente pois os pontos com cores mais quentes (pontos de maior densidade) não apresentam assimetria em torno de  $\Delta x = 0$ .

A linha tracejada na Figura 5.2(b) representa a linha de tendência (isto é, regressão linear) dos valores de  $\sigma_{\Delta x}$ , a partir da qual é estimada a amplitude do ruído ( $B$ ) presente na rede de Sistemas Autônomos. Nessa avaliação, considerando todo o período em análise,  $B = 0,0103 \sim 1 \times 10^{-2}$ . Esse resultado classifica essa rede em uma situação intermediária na classificação de estabilidade proposta por Blumm *et al.* Isto é, uma parcela das posições próximas ao topo do ranqueamento apresentam estabilidade enquanto que as demais posições apresentam instabilidade. Essa avaliação obtida através do valor estimado de  $B$  está coerente com a avaliação do tempo médio de estabilidade por posição do ranqueamento mostrada na Figura 5.1, uma vez que as primeiras posições no ranqueamento têm um tempo médio de estabilidade bem superior aos demais.

Uma questão interessante é a relação entre  $B$  e as variações ocorridas na rede. Tal relação foi abordada calculando o valor de  $B$  em períodos menores e comparando com as variações detectadas na rede no mesmo período. A Tabela 5.2 mostra, para cada período de um ano, o valor estimado da amplitude de ruído  $B$ , as quantidades proporcionais de arestas adicionadas e removidas na rede, e as quantidades de alterações nas duas primeiras posições do ranqueamento de grau. Para facilitar a visualização, foi acrescentada uma gradação de cor onde o menor valor de cada coluna é marcado de branco e o maior valor de vermelho.



(a) Variações do grau relativo ( $\Delta x$ ).



(b) Dispersão suavizada das variações de grau relativo ( $\sigma_{\Delta x}$ ).

Figura 5.2: Características de variação do grau relativo na rede de Sistemas Autônomos.

A Tabela 5.2 mostra uma certa correlação entre o valor de  $B$  e a quantidade de adições e remoções de arestas. Por exemplo, no ano de 2000, o valor do ruído  $B$  é elevado, assim como a porcentagem de adições e remoções. Já em 2006, o valor de  $B$  é menor, assim como as porcentagens de adições e remoções de arestas. Essa correlação é mais bem visualizada no gráfico da Figura 5.3. Note que nos primeiros anos o valor de  $B$  é mais elevado, assim como as porcentagens de adição e remoção de arestas. Nos anos seguintes, de 2004 até 2012, há uma redução na porcentagem de adição e remoção, seguido pelo valor  $B$ , em parte por causa do crescimento da rede. Por outro lado, não há uma correlação aparente do valor de  $B$  com as variações ocorridas nas duas primeiras posições do ranqueamento de grau. Por exemplo, apesar do ano de 2009 ter sido o ano com o maior número de alterações nas primeiras posições, o valor de  $B$  não foi dos mais elevados. Isso leva a crer que o ruído tem pouca ou nenhuma influência nas trocas das primeiras posições

Tabela 5.2: Amplitude do ruído e variações em cada ano na rede de Sistemas Autônomos.

Período	Coefficiente $B$ do ruído	% média de arestas adicionadas por mês	% média de arestas removidas por mês	Alterações na 1ª posição do ranqueamento	Alterações na 2ª posição do ranqueamento
2000	0,017	11,4%	6,8%		
2001	0,009	9,5%	7,5%		
2002	0,006	7,2%	6,3%		
2003	0,020	8,3%	6,0%		
2004	0,008	6,5%	5,2%		1
2005	0,004	5,4%	4,3%		
2006	0,002	5,1%	3,9%		
2007	0,003	5,2%	3,9%		
2008	0,004	5,9%	4,9%		
2009	0,006	5,3%	4,4%	2	3
2010	0,003	4,6%	4,0%	1	1
2011	0,003	5,0%	3,6%		
2012	0,004	5,2%	4,3%		
2013	0,005	6,0%	4,8%		

do ranqueamento dessa rede.

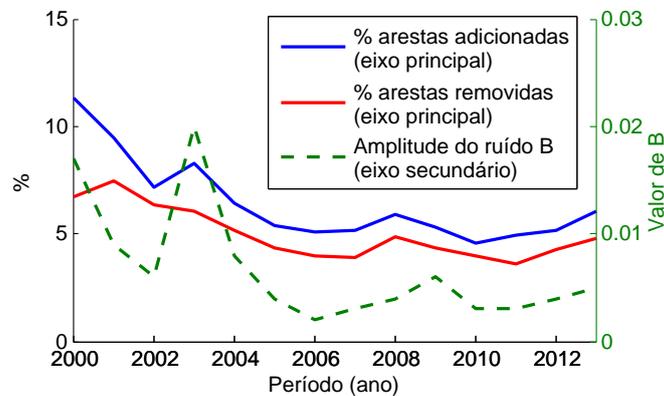


Figura 5.3: Adições e remoções de arestas vs. amplitude de ruído  $B$ .

Outra forma de avaliar o impacto do ruído no topo do ranqueamento é observar a evolução do grau relativo ( $x_i(t)$ ) dos vértices de maior grau da rede. Na Figura 5.4, é possível identificar um ruído de maior amplitude ao longo dos anos de 2000 e de 2003, causando variações rápidas no grau relativo dos vértices. Essa figura contém o grau relativo de todos os sistemas autônomos que ocuparam a primeira e a segunda posições do ranqueamento de grau no período em análise, portanto são mostradas todas as trocas ocorridas nas duas primeiras posições do ranqueamento. Note que

entre 2000 e 2008 a diferença de grau relativo dos dois primeiros colocados é grande, maior do que os picos de variação causados pelo ruído. A partir de 2009, a diferença entre os dois primeiros colocados se mantém pequena e passam a ocorrer algumas trocas na primeira posição. Essa condição de proximidade entre o grau da primeira e da segunda posições é inesperada em uma distribuição de grau livre de escala, indicando que a evolução de grau desses vértices sofre influência de fatores pontuais.

Os picos de amplitude de ruído observados nos anos de 2000 e 2003 estão coerentes com os picos do valor estimado de  $B$  mostrados na Figura 5.3. A Figura 5.4 também permite avaliar que de forma geral o grau relativo dos principais vértices sofre redução ao longo de quase todo o período analisado, devido ao crescimento da rede em quantidade total de arestas.

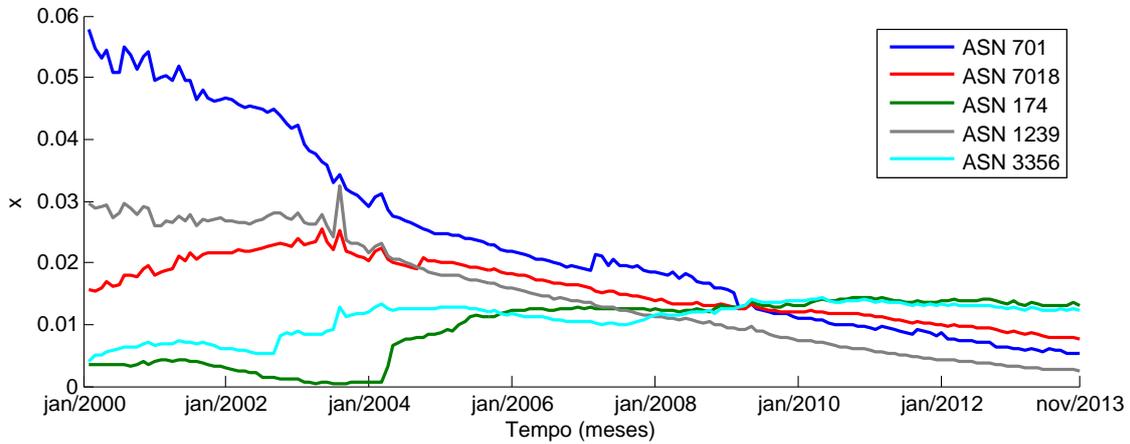


Figura 5.4: Evolução do grau relativo ( $x_i$ ) dos principais sistemas autônomos.

### 5.1.3 Análise do impacto de fatores pontuais no ranqueamento

Conforme a premissa adotada por Blumm *et al.* [12], o valor de  $x_i$  é determinado não apenas pelo ruído presente no sistema, mas também pela aptidão individual do item  $i$ . Analogamente, a estabilidade de ranqueamento de cada sistema autônomo depende também da sua capacidade de atrair conexões. Se um sistema autônomo aumentar sua capacidade de atrair conexões em relação aos demais, ele tende a subir de posição no ranqueamento. Se reduzir essa capacidade, tende a perder posição. Apesar de prever a influência da aptidão individual de  $i$  no cálculo de  $x_i$ , Blumm *et al.* não definiram uma forma de calcular a aptidão.

Os resultados até aqui indicam que, devido às características da rede, as duas primeiras posições do ranqueamento de grau possuem elevada estabilidade e o nível de ruído do sistema não é determinante para causar as variações observadas nessas

posições. Resta, portanto, avaliar se há fatores pontuais responsáveis por alterar a aptidão dos principais sistemas autônomos de modo a influenciar as variações no topo do ranqueamento.

No período em análise, as duas primeiras posições do ranqueamento de grau foram ocupadas pelos sistemas autônomos de ASN (*Autonomous System Number*) 174, 701, 1239, 3356 e 7018. Esses ASNs pertencem a companhias privadas e, portanto, a expansão das redes desses sistemas autônomos depende fundamentalmente da estratégia de crescimento e da capacidade de investimento de cada companhia. A seguir são listados os fatos relevantes a respeito dessas companhias divulgados publicamente dentro do período analisado:

- A) Em 2002 a companhia detentora do ASN 701 (MCI/WorldCom) entrou em concordata [19]. A companhia continuou passando por dificuldades nos anos seguintes, culminando na venda da companhia em 2006 para uma concorrente (Verizon) [20]. Nesse período, o grau do ASN 701 encerrou a tendência anterior de crescimento e começou a recuar. Houve uma redução da aptidão desse sistema autônomo.
- B) Em 2004 ocorreu a expansão da empresa detentora do ASN 174 (Cogent) para a Europa e a taxa de crescimento do seu grau aumentou significativamente [21]. Nesse caso ocorreu um aumento da aptidão do sistema autônomo.
- C) Entre 2004 e 2005 houve a fusão da companhia detentora do ASN 1239 (Sprint) com uma companhia operadora de telefonia móvel (Nextel). Essa fusão foi considerada problemática, dentre outros motivos, por causa da disparidade das tecnologias das duas redes [22]. A concentração de esforços na integração das redes de telefonia móvel nos anos seguintes à fusão causou redução de investimentos na rede do sistema autônomo, levando à redução da sua aptidão.
- D) Entre 2004 e 2007 a companhia detentora do ASN 3356 (Level 3) adquiriu diversas empresas concorrentes [23]. A partir de 2007 a taxa de crescimento do grau aumentou significativamente como resultado da estratégia de expansão, levando ao aumento da aptidão desse sistema autônomo.

Os órgãos reguladores governamentais têm poder de limitar a atuação das maiores companhias do mercado para estimular a competição, como ocorreu com o sistema Bell na década de 1940. Contudo, não foi identificado impacto desse tipo no período em análise. Outro fato relevante é o esgotamento recente do endereçamento IPv4 (*Internet Protocol version 4*) mas não há impactos evidentes desse tipo nas redes dos cinco principais sistemas autônomos no período em análise.

A Figura 5.5 mostra a evolução de grau dos principais sistemas autônomos com a indicação da atuação dos fatores pontuais. É possível observar que o fator determinante para a definição das duas primeiras posições do ranqueamento de grau é a taxa de crescimento de grau de cada sistema autônomo, que está associada à aptidão individual do sistema autônomo. Conforme analisado na seção anterior, o ruído apresentou amplitude insuficiente para causar troca nas duas primeiras posições do ranqueamento, com exceção de um pequeno período entre 2009 e 2010 onde o grau dos principais sistemas autônomos esteve muito próximo e o ruído pode ter influenciado momentaneamente o ranqueamento entre eles.

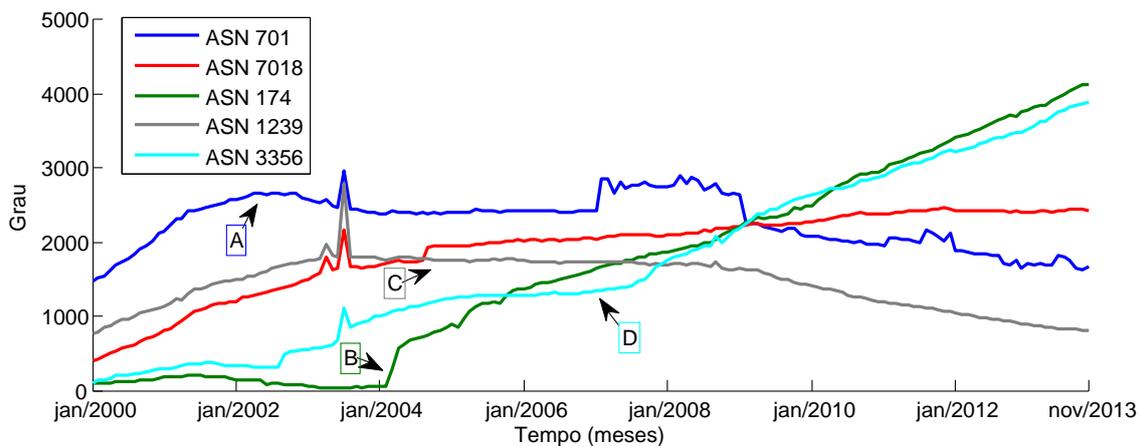


Figura 5.5: Evolução do grau dos principais sistemas autônomos. As marcações destacam as ocorrências dos fatores pontuais.

#### 5.1.4 Avaliação da previsão de estabilidade do ranqueamento

Diante dos resultados obtidos, é esperado que as duas primeiras posições do ranqueamento de grau da rede de Sistemas Autônomos continuem com alta estabilidade em relação às demais posições. Essa estabilidade será mantida até que ocorram alterações nas aptidões dos principais vértices da rede de forma a causar uma troca nessas posições. Considerando o histórico, a primeira posição do ranqueamento tende a permanecer estável por cerca de 42 meses, ou seja, por mais que três anos. Seguindo a mesma linha de raciocínio, a segunda posição tende a se manter estável por cerca de 28 meses ou mais que dois anos.

Dados mais recentes mostram que, de fato, os mesmos sistemas autônomos continuam a ocupar as duas primeiras posições do ranqueamento de grau, mesmo após dois anos desde o fim do período analisado neste trabalho. A Tabela 5.3 mostra o ASN e o grau dos cinco sistemas autônomos de maior grau no fim do período analisado e nos dois anos seguintes, com os dois ASNs de maior grau destacados em vermelho. A terceira posição do ranqueamento também foi ocupada pelo mesmo

sistema autônomo nos dois anos seguintes, porém a quarta e quinta posições apresentaram variações conforme esperado. Nota-se que o crescimento do grau dos três maiores sistemas autônomos se manteve regular nesse período, indicando que não houve mudança na aptidão desses sistemas autônomos, garantindo a estabilidade do ranqueamento.

Tabela 5.3: ASN e grau dos 5 sistemas autônomos de maior grau ao final do período analisado e após 1 e 2 anos.

Ranqueamento de grau	Nov/2013		Nov/2014		Nov/2015	
	ASN	Grau	ASN	Grau	ASN	Grau
1°	174	4137	174	4306	174	4765
2°	3356	3897	3356	3990	3356	4284
3°	6939	3408	6939	3598	6939	4270
4°	7018	2433	3549	3573	3549	3560
5°	4323	1710	7018	2340	24482	2694

## 5.2 Avaliação da estabilidade do ranqueamento de grau da rede de Citações

A evolução da rede de Citações é analisada considerando intervalos de tempo de um ano, desta forma são utilizadas as fotografias da rede ao final de cada ano. Ao final desta seção é realizada uma previsão a partir dos resultados obtidos.

### 5.2.1 Análise da quantidade e da estabilidade temporal dos vértices superestáveis

Conforme exposto na Figura 3.4(b), a distribuição de grau da rede de Citações evolui de forma a aumentar gradualmente o peso da cauda, sendo assim o expoente ( $\gamma$ ) da lei de potência aproximada diminui com o passar dos anos. Isso significa que há um aumento da concentração de arestas nos vértices de maior grau da rede. Além disso, a partir de 1930 a quantidade de vértices e arestas cresce de forma quase exponencial. Diante dessas variações expressivas, a quantidade de vértices superestáveis ( $m_c$ ) calculada através da Equação 2.2 não é a mesma em todo o período avaliado. A Tabela 5.4 mostra as propriedades da rede de Citações e a quantidade de vértices superestáveis prevista no primeiro ano de cada década a partir de 1930. Note que no início do período o valor de  $m_c$  é zero e em 2010 chega a cinco. O grau médio da rede também aumenta mais de cinco vezes ao longo do período, evidenciando um aumento da densidade de arestas na rede. Observe que a

diferença de grau entre o vértice de maior grau e o segundo colocado em 1930 é de apenas dois graus, e a diferença entre os próximos colocados até o quinto varia entre um e dois graus. Nesse ano a distribuição de grau não se aproxima de uma lei de potência e, portanto, a quantidade prevista de vértices superestáveis é zero. Já nos anos de referência seguintes foi possível estimar os respectivos expoentes ( $\gamma$ ) da lei de potência a partir da cauda da distribuição de grau de cada ano, sendo que o valor estimado de  $\gamma$  varia de 6,8 (em 1940) até 3,2 (em 2010). Aqui vale lembrar que, de acordo com a Figura 2.1 do Capítulo 2, a quantidade de vértices superestáveis aumenta a medida que  $\gamma$  se aproxima da faixa entre 2,5 e 3,0.

Tabela 5.4: Propriedades da rede de Citações e quantidade prevista de vértices superestáveis em diferentes instantes no tempo.

Rede	Tamanho $N$ da rede	$\gamma$ apro- ximado	$m_c$ pre- visto	Grau médio	Maior grau	2° maior grau	3° maior grau	4° maior grau	5° maior grau
1930	2.297	-	0	1,9	26	24	22	18	17
1940	7.082	6,8	0	3,1	60	48	47	42	41
1950	12.271	5,6	1	3,8	87	82	81	70	69
1960	27.275	4,3	1	5,3	214	208	207	181	158
1970	58.060	3,9	2	6,5	648	518	372	246	232
1980	103.350	3,6	3	7,2	871	646	641	476	447
1990	173.679	3,4	3	7,7	1300	1099	1065	915	656
2000	297.701	3,3	4	8,8	2937	2211	1705	1463	1336
2010	469.526	3,2	5	10,7	5573	4411	3697	3340	2857

Diante da grande variação da quantidade prevista de vértices superestáveis, o tempo médio de estabilidade do ranqueamento mostrado na Figura 5.6 está dividido em quatro períodos de vinte anos cada. Podemos observar que no período de 1930 a 1950 e também no período de 1950 a 1970 o tempo médio de estabilidade da primeira posição do ranqueamento foi de cinco anos (ou 25% do tempo total), tempo maior do que o dobro da segunda posição, que por sua vez apresentou tempo médio de estabilidade próximo ao das demais posições. Note que o cálculo de  $m_c$  para os anos de 1930 e 1940 previu não haver vértice superestável, porém previu a ocorrência de um vértice superestável em 1950 e em 1960. O período de 1970 a 1990 apresentou tempo médio de estabilidade decrescente da primeira posição até aproximadamente a sexta posição, havendo relativo destaque para os tempos médios da primeira (33% do tempo total), segunda (25% do tempo total) e terceira (20% do tempo total) posições. A previsão de vértices superestáveis para esse período foi de dois a três vértices. Por último, no período de 1990 a 2010 ocorre um aumento significativo do tempo médio de estabilidade das duas primeiras posições, sendo que a primeira posição se manteve totalmente estável (tempo de estabilidade foi de 100% do tempo

total) e a segunda apresentou apenas uma variação (tempo médio de estabilidade foi de 50% do total). A previsão para esse período foi de três a quatro vértices superestáveis, porém a terceira e quarta posições não apresentaram aumento de tempo médio de estabilidade em relação ao período anterior. Assim como a análise feita para a rede de Sistemas Autônomos, os resultados aqui mostram que os vértices apontados como superestáveis pela Equação 2.2 em geral apresentam maior estabilidade, e também que o tempo médio de estabilidade apresenta tendência contínua de queda ao se afastar do topo do ranqueamento.

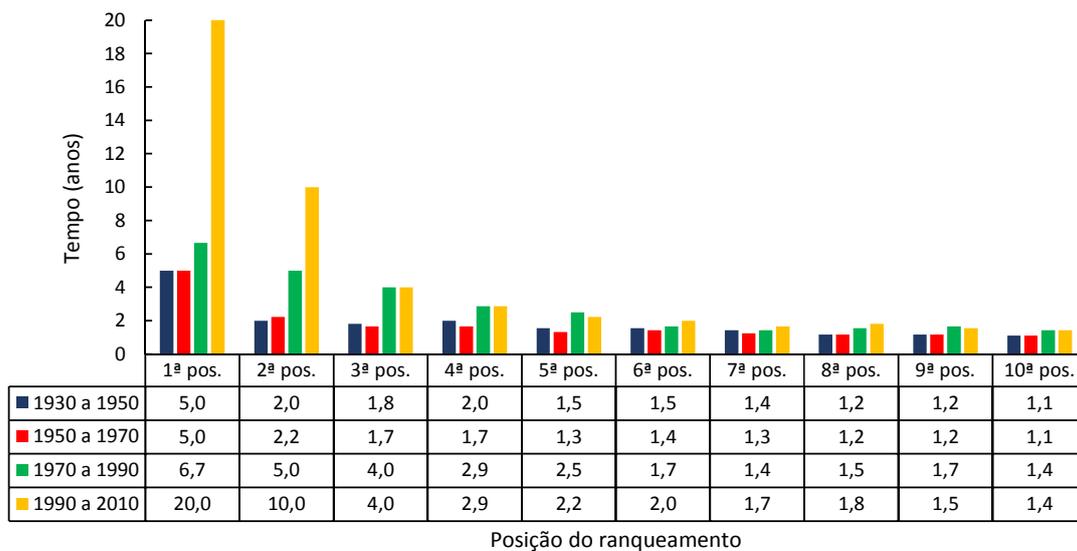


Figura 5.6: Estabilidade média por posição de ranqueamento de grau na rede de Citações, calculada separadamente em quatro períodos. O tempo médio de estabilidade é calculado ao dividir o tempo total do período pela quantidade de variações ocorridas em cada posição no mesmo período.

## 5.2.2 Análise do impacto do ruído no ranqueamento

Em [12] Blumm *et al.* analisaram a dinâmica do ranqueamento de citações de artigos da coleção Physical Review, contabilizando apenas a quantidade de citações recebidas por cada artigo por ano, sem acumular para o ano seguinte. Ao contrário de Blumm *et al.*, neste trabalho, o conjunto de Citações é analisado como sendo uma rede, que aumenta continuamente de tamanho a cada nova citação. Portanto, para formar o ranqueamento em um ano qualquer são consideradas as citações desde a origem em 1893 até o ano em questão.

Blum *et al.* avaliaram que o sistema de citações apresentou ranqueamento instável no período de 1970 a 2009, sendo que o valor estimado de  $B$  foi de  $5 \times 10^{-2}$ . Levando em conta o mesmo período de 1970 a 2009, porém considerando que as citações são conexões em uma rede que nunca são desfeitas, o valor encontrado de

$B$  é de  $1,8 \times 10^{-3}$  (Figura 5.7). A princípio, esse resultado indica que o ranqueamento se torna mais estável na configuração de rede. Por outro lado, a Figura 5.8(a) mostra que as variações de grau relativo ( $\Delta x$ ) da rede de Citações são assimétricas em torno do eixo  $\Delta x = 0$  e, portanto, os vértices de maior grau relativo tendem a perder posição no ranqueamento. Isso significa que, apesar da baixa amplitude de ruído ( $B$ ), o ranqueamento dessa rede é instável. Repare que as variações de grau relativo do sistema de citações no mesmo período de 1970 a 2009 (Figura 5.8(b)) também apresenta assimetria em torno do eixo  $\Delta x = 0$ . Porém a amplitude das variações é cerca de dez vezes maior. Essa diferença de amplitude explica a discrepância entre o valor de  $B$  calculado por Blumm *et al.* e o valor calculado para a rede neste trabalho.

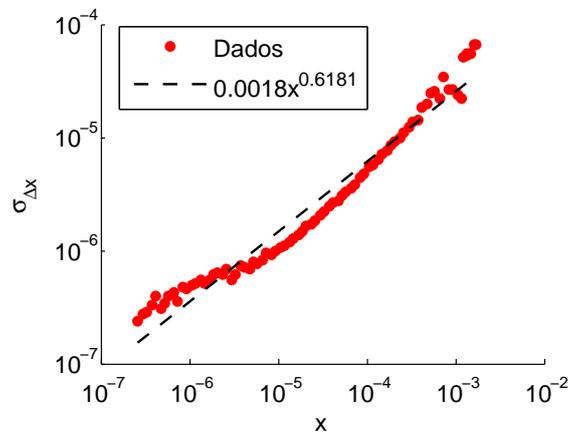
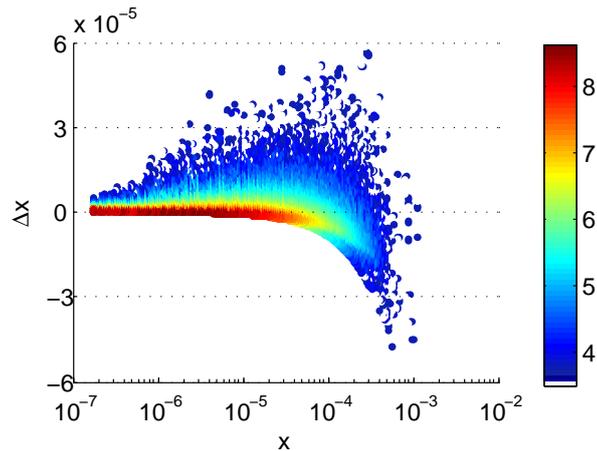
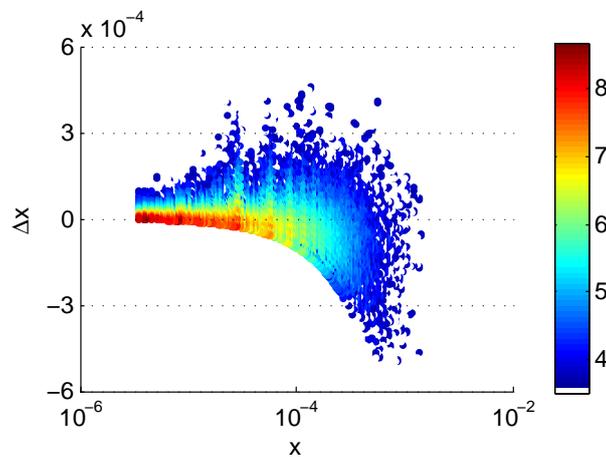


Figura 5.7: Dispersão suavizada das variações de grau relativo ( $\sigma_{\Delta x}$ ) da rede de Citações no período de 1970 a 2009.

Considerando agora apenas a rede de Citações, na Tabela 5.5 estão os valores de  $B$  calculados para cada intervalo de dez anos a partir de 1930. A relação entre a amplitude de ruído  $B$  e a quantidade relativa de adições de arestas na rede é ainda mais evidente do que na rede de Sistemas Autônomos (Figura 5.9). Uma explicação é o fato de haver apenas adições de arestas na rede de Citações enquanto que na rede de Sistemas Autônomos também ocorrem remoções, trazendo maior complexidade para a análise. Já a avaliação do impacto de  $B$  no topo do ranqueamento nesse caso precisa ser feita com maior atenção a outros fatores. Há um pico no valor de  $B$  no período de 1930 a 1940, que coincide com a maior quantidade de alterações observadas nas três primeiras posições do ranqueamento. Em 1930, a rede ainda não apresentava cauda pesada na distribuição de grau e então é possível que as trocas de posição no topo do ranqueamento neste período tenham sido influenciadas pelo ruído. Em contrapartida, nos períodos de menor valor de  $B$ , que são de 1940 a 1950 e de 1980 a 1990, também houve uma quantidade alta de variações nas três primeiras posições do ranqueamento. Isso indica que outros fatores que não somente



(a) Variações do grau relativo ( $\Delta x$ ) sob a configuração de rede, onde são contabilizadas todas as citações recebidas até o ano de referência.



(b) Variações do grau relativo ( $\Delta x$ ) ao contabilizar apenas as citações recebidas no espaço de um ano.

Figura 5.8: Características de variação do grau relativo sob diferentes configurações do conjunto de dados Citações para o período de 1970 a 2009.

o ruído foram responsáveis por causar instabilidade nessas posições.

A Figura 5.10 mostra a evolução do grau relativo dos vértices que atingiram a primeira posição do ranqueamento, onde os artigos referentes a esses vértices estão identificados pelo sufixo DOI (*Digital Object Identifier*) determinado pela APS. O ruído, da forma que foi definido por Blumm *et al.*, causa variações aleatórias a cada passo de tempo, que no caso dessa análise é de um ano. Portanto, as variações da primeira posição do ranqueamento não foram causadas pelo ruído da rede, mas sim pela mudança no grau de inclinação das curvas de grau relativo. De maneira geral, as curvas possuem uma fase de ascensão seguida de declínio, indicando que a aptidão dos vértices sofre grande variação com o tempo. É interessante notar que o maior grau relativo dessa rede apresenta uma evolução (redução gradual) muito parecida com a observada na Figura 5.4, onde o maior grau relativo ao final do período é

Tabela 5.5: Amplitude do ruído e variações a cada período de vinte anos na rede de Citações.

Período	Coefficiente $B$ do ruído	% média de arestas adicionadas por ano	Alterações na 1ª posição do ranqueamento	Alterações na 2ª posição do ranqueamento	Alterações na 3ª posição do ranqueamento
1930 a 1940	0,014	19,8%	1	6	6
1940 a 1950	0,001	7,2%	2	3	4
1950 a 1960	0,003	13,0%	2	4	6
1960 a 1970	0,002	10,0%	1	4	5
1970 a 1980	0,002	7,5%			1
1980 a 1990	0,001	5,8%	2	3	3
1990 a 2000	0,002	6,9%		1	3
2000 a 2010	0,002	6,8%			1

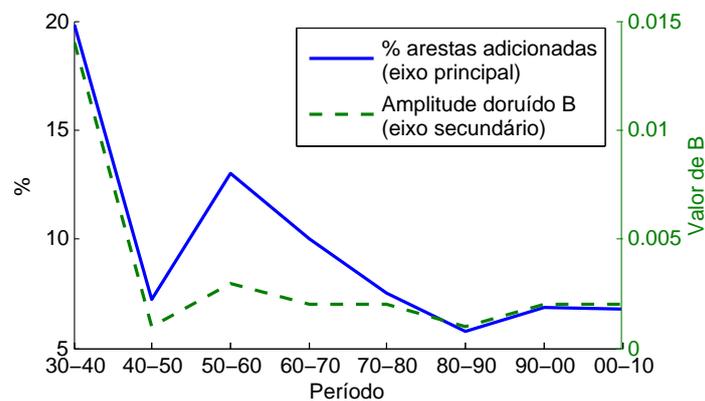


Figura 5.9: Adições e remoções de arestas vs. amplitude de ruído  $B$ .

cerca de três vezes menor do que o maior grau relativo no início do período. Essa variação está relacionada com o crescimento acelerado das redes, que causa redução no grau relativo de todos os vértices. Porém, mesmo com essa redução, o nível de estabilidade no topo do ranqueamento de grau da rede de Citações aumenta com o tempo.

### 5.2.3 Análise do impacto de fatores pontuais no ranqueamento

A Figura 5.11 mostra a evolução do grau dos vértices que chegaram à primeira posição do ranqueamento. Repare que o eixo vertical está em escala logarítmica para facilitar a visualização. Os principais vértices da rede de Citações possuem um perfil de evolução de grau onde ocorre um alto crescimento logo após sua entrada

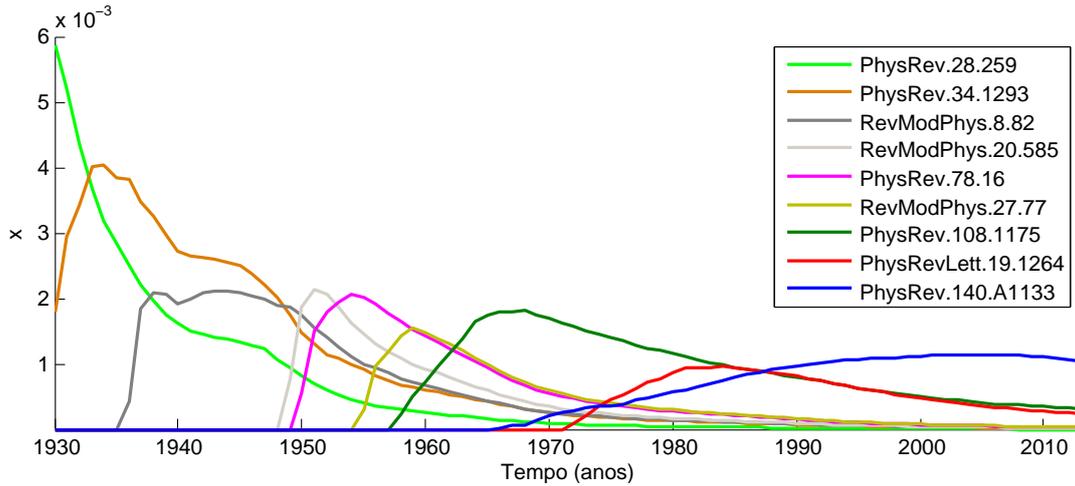


Figura 5.10: Evolução do grau relativo ( $x_i$ ) dos principais artigos da rede.

na rede e em seguida esse crescimento perde força de forma gradual e se aproxima de zero. Ao contrário da rede de Sistemas Autônomos, o grau de cada vértice na rede de Citações não reduz. Portanto, sua permanência no topo do ranqueamento depende fundamentalmente da intensidade e duração desse período de crescimento do grau, que está relacionado com a aptidão do artigo em atrair citações.

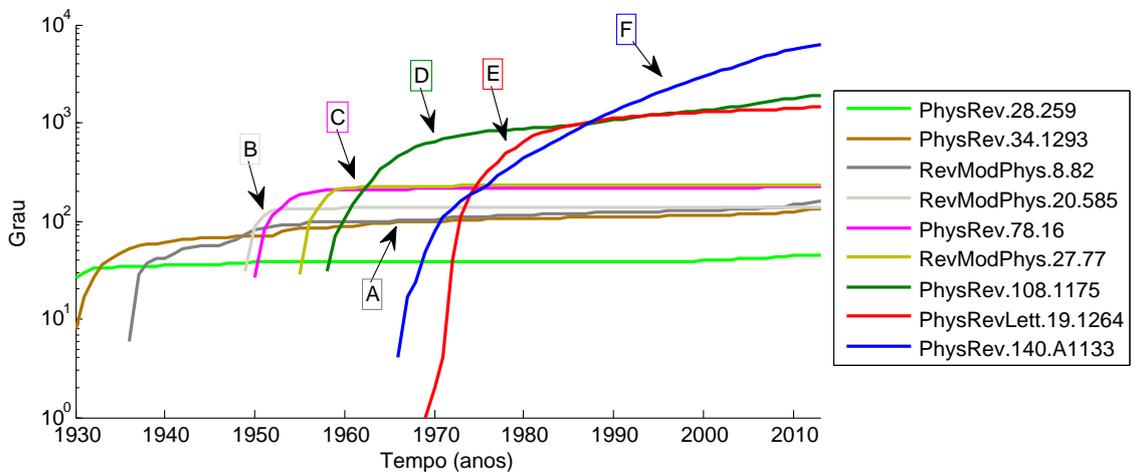


Figura 5.11: Evolução do grau absoluto dos principais artigos da rede. As marcações indicam os momentos de laureamento com prêmio Nobel.

Conforme detalhado abaixo, a maior parte desses artigos está relacionada a vencedores do prêmio Nobel de Física e de Química:

- A) Artigo RevModPhys.8.82: O autor principal do artigo “Nuclear Physics A. Stationary States of Nuclei” (Hans Albrecht Bethe) foi laureado com o Nobel de Física em 1967 [24].

- B) Artigo RevModPhys.20.585: O autor do artigo “Table of Isotopes” (Glenn Theodore Seaborg) foi laureado com o Nobel de química em 1951 [25].
- C) Artigo PhysRev.78.16: A autora do artigo “Nuclear Configurations in the Spin-Orbit Coupling Model. I. Empirical Evidence” (Maria Goeppert Mayer) foi laureada com o Nobel de física em 1963 [26].
- D) Artigo PhysRev.108.1175: Os autores do artigo “Theory of Superconductivity” (John Bardeen, Leon Neil Cooper e John Robert Schrieffer) foram laureados com o Nobel de física em 1972 [27].
- E) Artigo PhysRevLett.19.1264: O autor do artigo “A Model of Leptons” (Steven Weinberg) foi laureado com o Nobel de física em 1979 [28].
- F) Artigo PhysRev.140.A1133: O autor principal do artigo “Self-Consistent Equations Including Exchange and Correlation Effects” (Walter Kohn) foi laureado com o Nobel de química em 1998 [29].

Portanto, o fator determinante para o alto crescimento do grau é a grande notoriedade do trabalho dentro da comunidade de pesquisa, o que aumenta a aptidão do artigo em receber novas citações.

Devido ao aumento gradual da cauda pesada na distribuição de grau da rede de Citações, a estabilidade das primeiras posições do ranqueamento tende a aumentar com o tempo. Esse aumento de estabilidade pode ser observado na Figura 3.5 do Capítulo 3. A exceção é o período de 1930 a 1950, que apresentou estabilidade maior do que os anos seguintes. Nesse caso, o aumento momentâneo de estabilidade foi causado pela Segunda Guerra Mundial, um fator externo à estrutura da rede que reduziu drasticamente a publicação de novos artigos [30]. Esse “freio” temporário no crescimento da rede nos anos próximos à Segunda Guerra, que durou de 1939 a 1945, também pode ser observado na Figura 3.4(a) do Capítulo 3.

### 5.2.4 Previsão de estabilidade do ranqueamento

Ao final do período disponível para análise, em 2013, o tamanho ( $N$ ) da rede de Citações é de 527.494 vértices e o expoente da lei de potência aproximada da distribuição de grau ( $\gamma$ ) é igual a 3,1. A quantidade de vértices superestáveis ( $m_c$ ) calculada a partir da Equação 2.2 é de aproximadamente 5. Mantidos o crescimento da rede e a mesma cauda pesada na distribuição de grau da rede, a tendência é que essa quantidade de vértices superestáveis seja mantida ou aumente. Porém, conforme avaliação feita a partir do tempo médio de estabilidade mostrado na Figura 5.6, a estabilidade diminui conforme se distancia do topo do ranqueamento, mesmo entre as posições superestáveis.

Na seção anterior foi visto que a aptidão de um artigo está diretamente relacionada com a sua notoriedade na comunidade de pesquisa. Contudo, é difícil de mensurar essa notoriedade. Por outro lado, o perfil da evolução individual de grau na rede de Citações, observado na Figura 5.11, pode ser levado em conta para prever as alterações no ranqueamento. A Figura 5.12 contém a evolução do grau, no período de 2000 a 2013, dos dez vértices de maior grau em 2013. Considerando a inclinação das curvas, é possível prever uma troca de posição entre o primeiro e o segundo colocados do ranqueamento nos próximos dois ou três anos. Note que a última alteração na primeira posição ocorreu há vinte e cinco anos. Isso indica que, após a inversão de posição prevista para 2015 ou 2016, a primeira posição deverá se manter estável por outro longo período. A terceira, a quarta e a quinta posições do ranqueamento sofreram alteração recente, entre 2011 e 2013. Considerando a tendência de manter os tempos médios de estabilidade de ranqueamento da última década, calculados na Seção 5.2.1, essas posições devem manter a estabilidade por mais cerca de três anos.

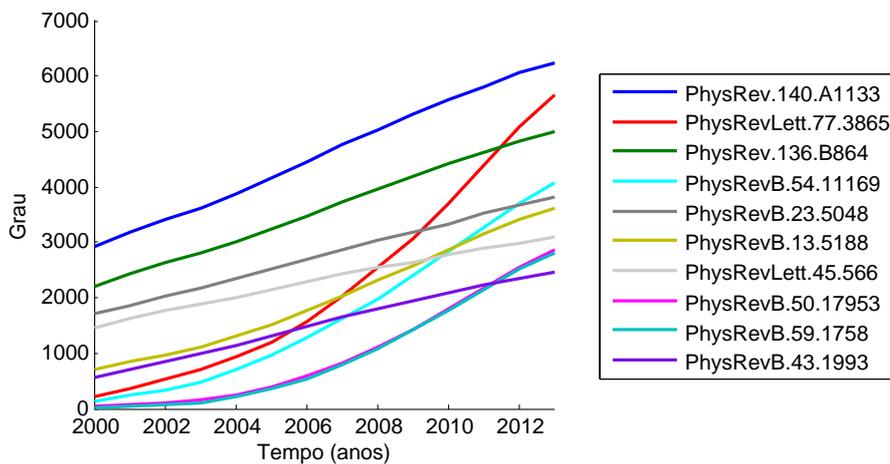


Figura 5.12: Evolução do grau dos dez artigos com maior quantidade de citações no final do período em análise.

As previsões feitas para a rede de Citações só poderão ser avaliadas em uma ocasião futura, quando for disponibilizado o conjunto de dados da rede referente aos anos posteriores a 2013.

# Capítulo 6

## Conclusões e Trabalhos Futuros

As duas redes analisadas possuem uma distribuição de grau com cauda pesada que se manteve ou apresentou aumento ao longo dos períodos analisados. Com isso, o ranqueamento por grau dos vértices leva a grandes diferenças de grau no topo do ranqueamento. Consequentemente, as trocas de posição próximas ao topo do ranqueamento são mais raras do que nas posições mais distantes do topo, onde as diferenças de grau entre vértices de grau comparável são menores.

As variações de grau regulares em torno de um valor central, que podem ser comparadas a um ruído, também causam relativa instabilidade no ranqueamento de grau das redes. Porém, como analisado neste trabalho, a estabilidade de algumas posições no topo do ranqueamento não é afetada pelo ruído, dado que a sua amplitude é muito menor do que a amplitude da distância entre os vértices de maior grau em redes livre de escala.

Na rede de Sistemas Autônomos, as duas primeiras posições do ranqueamento de grau apresentam tempo médio de estabilidade muito maior do que as demais posições. O resultado desse cálculo de tempo médio de estabilidade, introduzido nesse trabalho, está de acordo com o resultado da Equação 2.2. A análise da evolução de grau relativo dos principais vértices dessa rede mostra que, na maior parte do tempo, a distância de grau relativo entre os três primeiros colocados no ranqueamento de grau é maior do que os picos de variação causados pelo ruído.

De forma a complementar essa avaliação, neste trabalho foram identificados os fatores pontuais responsáveis por alterar individualmente a aptidão dos principais Sistemas Autônomos de atrair conexões. A análise empírica da evolução de grau desses vértices e dos dados dos fatores pontuais permitem concluir que as alterações de aptidão dos principais vértices, em conjunto, determinaram as trocas de posição ocorridas na primeira e na segunda posições do ranqueamento. Diante dessa avaliação, é previsto que a estabilidade seja mantida nas duas primeiras posições do ranqueamento de grau até que fatores pontuais afetem a aptidão dos principais vértices de forma a levar ao cruzamento das suas curvas de grau. É possível levar em

conta também o tempo médio de estabilidade dessas posições nos anos anteriores. Dados um período posterior ao período inicialmente analisado indicam que após dois anos a estabilidade das duas primeiras posições do ranqueamento foi mantida.

A quantidade de vértices superestáveis previstos na rede de Citações variou de nenhum até cinco vértices dentro do período analisado. Essa variação se deve ao crescimento da rede e também ao aumento da cauda pesada da distribuição de grau da rede. Como reflexo dessa mudança na estrutura da rede, o tempo médio de estabilidade calculado neste trabalho para as primeiras posições do ranqueamento também aumentou com o tempo.

A evolução do grau relativo dos principais vértices da rede de Citações mostra que as alterações da primeira posição do ranqueamento foram causadas pela mudança de aptidão dos principais vértices da rede em atrair novas citações, e não pelo ruído da rede. A análise empírica realizada neste trabalho aponta que o fator pontual determinante para um vértice atingir a primeira posição do ranqueamento de grau é conseguir grande notoriedade na comunidade de pesquisa. Essa avaliação é confirmada pelo fato de que a maioria dos vértices que ocuparam a primeira posição do ranqueamento foram laureados com o prêmio Nobel.

Na rede de Citações é observada uma tendência geral de redução gradual da aptidão dos principais vértices com o passar do tempo, levando à estagnação do grau do vértice e consequente perda de posições no ranqueamento de grau. A partir dessa observação, é possível estimar a evolução da curva de crescimento do grau dos principais vértices e prever o tempo de estabilidade das primeiras cinco posições do ranqueamento. O tempo médio de estabilidade de ranqueamento calculado para a década anterior também pode ser considerado como tendência para estimar tempo de estabilidade previsto para os principais vértices.

Os resultados das análises da rede de Sistemas Autônomos e da rede de Citações revelam que os fatores aqui chamados de pontuais devem ser levados em conta na análise de estabilidade do ranqueamento de grau, principalmente dos principais vértices da rede, algo que não foi considerado em trabalhos anteriores. Esses fatores foram determinantes nas variações das primeiras posições do ranqueamento. Apesar da dificuldade de se criar um modelo matemático para esses fatores, foi possível realizar importantes avaliações empíricas a partir de informações públicas relacionadas aos vértices das redes.

Como trabalho futuro, será conduzido um estudo com o intuito de modelar a aptidão em função do tempo (termo  $f(x_i)$  da Equação 2.4) a partir do histórico da evolução do grau dos principais vértices. Para isso será necessário também aprofundar a compreensão de como os fatores pontuais afetam a aptidão em cada rede.

Outro objetivo é desenvolver um método que permita prever o tempo médio de estabilidade das posições superestáveis do ranqueamento de grau, que se aplique a

qualquer rede livre de escala. Esse método será desenvolvido a partir da Equação 2.2, que se mostrou eficiente para prever a quantidade de posições do ranqueamento de grau que apresentam alta estabilidade.

# Referências Bibliográficas

- [1] FIGUEIREDO, D. R. “Introdução a Redes Complexas”. In: de Souza, A. F., Meira Jr., W. (Eds.), *Atualizações em Informática*, PUC-Rio, cap. 7, pp. 303–358, Rio de Janeiro, 2011.
- [2] BARABÁSI, A.-L., ALBERT, R. “Emergence of Scaling in Random Networks”, *Science*, v. 286, n. 5439, pp. 509–512, out. 1999.
- [3] ALBERT, R., JEONG, H., BARABÁSI, A.-L. “Diameter of the world wide web”, *Nature Communications*, v. 2, n. 6749, pp. 130–131, set. 1999.
- [4] SIGANOS, G., FALOUTSOS, M., FALOUTSOS, P., et al. “Power Laws and the AS-level Internet Topology”, *IEEE/ACM Transactions on Networking*, v. 11, n. 4, pp. 514–524, ago. 2003.
- [5] REDNER, S. “How popular is your paper? An empirical study of the citation distribution”, *The European Physical Journal B - Condensed Matter and Complex Systems*, v. 4, n. 2, pp. 131–134, jul. 1998.
- [6] BARABÁSI, A.-L. “Scale-Free Networks: A Decade and Beyond”, *Science*, v. 352, n. 5939, pp. 412–413, jul. 2009.
- [7] PAGE, L., BRIN, S., MOTWANI, R., et al. *The PageRank Citation Ranking: Bringing Order to the Web*. Relatório Técnico 1999-66, Stanford InfoLab, 1999.
- [8] COHEN, R., HAVLIN, S. “Scale-Free Networks Are Ultrasmall”, *Physical Review Letters*, v. 90, n. 5, pp. 058701, fev. 2003.
- [9] COHEN, R., EREZ, K., BEN AVRAHAM, D., et al. “Resilience of the Internet to Random Breakdowns”, *Physical Review Letters*, v. 85, n. 21, pp. 4626–4628, nov. 2000.
- [10] PASTOR-SATORRAS, R., VESPIGNANI, A. “Epidemic Spreading in Scale-Free Networks”, *Phys. Rev. Lett.*, v. 86, pp. 3200–3203, Apr 2001.

- [11] GHOSHAL, G., BARABÁSI, A.-L. “Ranking stability and super-stable nodes in complex networks”, *Nature Communications*, v. 2, n. 394, pp. 1–7, jul. 2011.
- [12] BLUMM, N., GHOSHAL, G., FORRÓ, Z., et al. “Dynamics of Ranking Processes in Complex Systems”, *Physical Review Letters*, v. 109, n. 12, pp. 128701, 2012.
- [13] BOLDI, P., VIGNA, S. “Axioms for centrality”, *Internet Mathematics*, v. 10, n. 3-4, pp. 222–262, abr. 2014.
- [14] NG, A. Y., ZHENG, A. X., JORDAN, M. I. “Stable Algorithms for Link Analysis”. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '01, pp. 258–266, New York, NY, USA, 2001. ACM.
- [15] GIBSON, D., KLEINBERG, J., RAGHAVAN, P. “Inferring Web Communities from Link Topology”. In: *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia : Links, Objects, Time and Space—structure in Hypermedia Systems*, HYPERTEXT '98, pp. 225–234, New York, NY, USA, 1998. ACM.
- [16] AHMAD, M. Z., GUHA, R. “Studying the Effects of Internet Exchange Points on Internet Topology”, *J Inform Tech Softw Eng*, v. 2, pp. 114, dez. 2012.
- [17] CAIDA. “The CAIDA AS Relationships Dataset”. Disponível em: <http://www.caida.org/data/as-relationships/>. Acessado em Novembro de 2015.
- [18] APS. “APS Data Sets for Research”. Solicitado em: <http://journals.aps.org/datasets>, 2015.
- [19] CNN. “WorldCom files largest bankruptcy ever”. Disponível em: [http://money.cnn.com/2002/07/19/news/worldcom\\_bankruptcy/](http://money.cnn.com/2002/07/19/news/worldcom_bankruptcy/), jul. 2002. Acessado em Novembro de 2015.
- [20] VERIZON. “Verizon Communications 2007 Annual Report”. Disponível em: [https://www.verizon.com/investor/app\\_resources/interactiveannual/2007/note08.html](https://www.verizon.com/investor/app_resources/interactiveannual/2007/note08.html), 2007. Acessado em Novembro de 2015.
- [21] COGENT. “Cogent History”. Disponível em: <http://www.cogentco.com/en/about-cogent/history>. Acessado em Novembro de 2015.

- [22] FORBES. “Was Sprint Buying Nextel One Of The Worst Acquisitions Ever At \$35b?” Disponível em: <http://www.forbes.com/sites/quora/2012/11/29/was-sprint-buying-nextel-one-of-the-worst-acquisitions-ever-at-35b/>, nov. 2012. Acessado em Novembro de 2015.
- [23] WIKIPEDIA. “Level 3 — Wikipedia, The Free Encyclopedia”. Disponível em: [https://en.wikipedia.org/wiki/Level\\_3\\_Communications](https://en.wikipedia.org/wiki/Level_3_Communications). Acessado em Novembro de 2015.
- [24] NOBELPRIZE.ORG. “The Nobel Prize in Physics 1967”. Disponível em: [http://www.nobelprize.org/nobel\\_prizes/physics/laureates/1967/](http://www.nobelprize.org/nobel_prizes/physics/laureates/1967/), . Acessado em Dezembro de 2015.
- [25] NOBELPRIZE.ORG. “The Nobel Prize in Chemistry 1951”. Disponível em: [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1951/](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1951/), . Acessado em Dezembro de 2015.
- [26] NOBELPRIZE.ORG. “The Nobel Prize in Physics 1963”. Disponível em: [http://www.nobelprize.org/nobel\\_prizes/physics/laureates/1963/](http://www.nobelprize.org/nobel_prizes/physics/laureates/1963/), . Acessado em Dezembro de 2015.
- [27] NOBELPRIZE.ORG. “The Nobel Prize in Physics 1972”. Disponível em: [http://www.nobelprize.org/nobel\\_prizes/physics/laureates/1972/](http://www.nobelprize.org/nobel_prizes/physics/laureates/1972/), . Acessado em Dezembro de 2015.
- [28] NOBELPRIZE.ORG. “The Nobel Prize in Physics 1979”. Disponível em: [http://www.nobelprize.org/nobel\\_prizes/physics/laureates/1979/](http://www.nobelprize.org/nobel_prizes/physics/laureates/1979/), . Acessado em Dezembro de 2015.
- [29] APS. “Focus: Nobel Focus: Chemistry by Computer”. Disponível em: <http://physics.aps.org/story/v2/st19>, out. 1998. Acessado em Dezembro de 2015.
- [30] REDNER, S. “Citation Statistics from 110 Years of Physical Review”, *Physics Today*, v. 58, n. 6, pp. 49–54, jun. 2005.