
UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

MAGDA CARDOSO MANTOVANI

FÍSICA ESTATÍSTICA APLICADA A
PROCESSOS ELEITORAIS

Maringá, Setembro de 2013.

UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

MAGDA CARDOSO MANTOVANI

FÍSICA ESTATÍSTICA APLICADA A
PROCESSOS ELEITORAIS

*Tese de Doutorado apresentada ao Programa de Pós-Graduação em Física
da Universidade Estadual de Maringá como parte dos requisitos
para a obtenção do título de doutora em Física.*

Orientador: Prof. Dr. Renio dos Santos Mendes

Área de Concentração: Física Estatística

Maringá, Setembro de 2013.

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

M293f Mantovani, Magda Cardoso
Física estatística aplicada a processos
eleitorais / Magda Cardoso Mantovani. -- Maringá,
2013.
116 f. : il. (algumas color.), figs.

Orientador: Prof. Dr. Renio dos Santos Mendes.
Tese (doutorado) - Universidade Estadual de
Maringá, Departamento de Física, Programa de Pós-
Graduação em Física, 2013.

1. Física estatística. 2. Sistemas complexos -
Física estatística. 3. Sistemas sociais - Física
estatística. 4. Processos eleitorais - Física
estatística. 5. Análise de dados - Física
estatística. 6. Distribuições de probabilidade -
Física estatística. 7. Escala alométrica. 8.
Eleições. I. Mendes, Renio dos Santos, orient. II.
Universidade Estadual de Maringá. Departamento de
Física. Programa de Pós-Graduação em Física. III.
Título.

CDD 21.ed. 530.13

AMMA-001231

UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

Banca Examinadora:

Prof. Dr. Renio dos Santos Mendes - Universidade Estadual de Maringá

Prof. Dr. Luciano Rodrigues da Silva - Universidade Federal do Rio Grande do
Norte

Prof. Dr. Gabriel Alves Mendes - Universidade Federal do Ceará

Prof. Dr. Luis Carlos Malacarne - Universidade Estadual de Maringá

Prof. Dr. Haroldo Valentin Ribeiro - Universidade Estadual de Maringá

Prof. Dr. Rodolfo Teixeira de Souza - Universidade Tecnológica Federal do Paraná
(suplente)

Prof. Dr. Ervin Kaminski Lenzi - Universidade Estadual de Maringá (suplente)

Maringá, Setembro de 2013.

Resumo

Usando conceitos e ferramentas da Física Estatística, neste trabalho são analisados vários aspectos de processos eleitorais. Primeiramente, é apresentada uma revisão de investigações empregando dados de eleições de vários países: distribuições de votos, modelos usando redes para reproduzir distribuições de votos e análises empregando taxa de comparecimento às urnas (correlações espaciais, análise entrópica e processos fraudulentos). Em seguida, são apresentadas as investigações desta tese. A primeira foi direcionada à escolha de candidatos para eleições (processo de candidatura). Encontrou-se uma correspondência média não linear (lei de potência) entre o número de candidatos (a prefeito e a vereador) e o tamanho do eleitorado em cada unidade eleitoral (número de eleitores), em que as distribuições de probabilidade das flutuações são aproximadamente lognormais (processo multiplicativo). Em adição a essa investigação, é apresentada uma análise estatística da ocupação de posições políticas em todas as cidades brasileiras considerando os números de candidatos (a prefeito e a vereador) e o número de membros de partidos políticos. Identificou-se que essas relações são também bem descritas por leis de potência (alometrias) médias, subjetivadas por ruídos multiplicativos com desvios padrões aproximadamente constantes. Os resultados indicam que a ocupação de posições políticas é caracterizada por uma hierarquia: posições mais influentes correspondem a menores valores para o expoente alométrico.

Palavras-chave: física estatística, sistemas complexos, sistemas sociais, processos eleitorais, análise de dados, distribuições de probabilidade, escala alométrica.

Abstract

By using concepts and tools of statistical physics, various aspects of electoral processes were analyzed. First, investigations employing elections data from several countries were reviewed. Afterwards, the focus was directed to the main research findings of this thesis. The first one concerns to the selection of candidates for elections (candidature process) in several countries: distributions of votes, models using networks to reproduce distributions of votes and analyzes employing turnout rate (spatial correlations, entropy analysis and fraudulent processes). We found an average nonlinear (power law) behavior between the number of candidates (for mayor and councilor) and the size of the electorate in each electoral unit (number of voters), where the probability distributions of the fluctuations are approximately lognormal (multiplicative process). The second one was obtained from a statistical analysis of the occupation of political posts in all Brazilian cities by considering the numbers of candidates (for mayor and councilor) and the number of members of political parties. As in the previous case, average power laws (allometries) subject to multiplicative noises with standard deviations being approximately constant were identified. The results indicate that the occupation of political posts (mayor, councilor and party member) is characterized by a hierarchy: most influential positions correspond to the smaller allometric exponents.

Key words: statistical physics, complex systems, social systems, electoral processes, data analysis, probability distributions, allometric scaling.

Sumário

Introdução	4
1 Distribuição de votos em eleições	10
1.1 Distribuição de votos em eleições proporcionais	11
1.2 Ajuste via lei de Zipf	13
1.3 Distribuição de votos em eleições pluralistas	17
2 Simulação de distribuição de votos em redes	26
2.1 Modelo de Sznajd na rede Barabási-Albert	26
2.2 Redes fractal e de Barabási-Albert	28
2.3 Propriedades de mundo pequeno em simulações de eleições	34
3 Escala e universalidade em eleições proporcionais	38
3.1 Relevância dos partidos	39
3.2 Procura por classes de universalidade	40
3.3 Similaridade entre distribuições	45

4	Comparecimento às urnas	47
4.1	Decaimento logarítmico da taxa de correlação	47
4.2	Entropia do envolvimento cívico do eleitorado	53
4.3	Deteção de fraude em eleições	58
5	Escolha de candidatos para eleições	63
5.1	Apresentação dos dados	64
5.2	Análise dos dados	65
5.3	Processo multiplicativo	71
5.4	Modelo de rede	72
5.5	Discussões	77
6	Ocupação de posições políticas	79
6.1	Apresentação dos dados	80
6.2	Análise dos dados	80
6.3	Processo Multiplicativo	87
6.4	Regra de ocupação das posições políticas	87
6.5	Discussões	90
	Conclusões	92
A	Sistemas eleitorais	96
A.1	Sistema proporcional	96
A.2	Sistema majoritário	98

B Física Estatística e sistemas complexos	99
B.1 Leis de potências	99
B.2 Distribuição gaussiana e o teorema do limite central	100
B.3 Modelo de Sznajd	102
B.4 Redes complexas	103
B.4.1 Distribuição de grau	104
B.4.2 Grafos aleatórios e mundo pequeno	104
B.4.3 Redes livres de escala	106

Introdução

A população mundial, assim como o número de cidades, em geral cresce continuamente e conseqüentemente os desafios sociais aumentam. Além disso, as interações sociais são reforçadas devido à alta mobilidade das pessoas e à grande facilidade de comunicação. Com esses fatores em evidência, compreender os diversos aspectos de nossa sociedade é uma necessidade também crescente. Nos últimos anos, muitas pesquisas têm sido realizadas utilizando conceitos e ferramentas de Mecânica Estatística na tentativa de entender fenômenos coletivos emergentes das interações de indivíduos como unidades elementares em estruturas sociais, isto é, com o intuito de aumentar a compreensão dos sistemas sociais. Esses estudos têm-se mostrado frutíferos para descrever sistemas sociais e nesse contexto tem-se usado o termo *sócio-física*.

Uma dificuldade conceitual surge imediatamente ao tentar aproximar a dinâmica social empregando o ponto de vista da física estatística. Tradicionalmente, os componentes elementares dos sistemas estudados, os átomos e moléculas, são objetos relativamente simples, cujo comportamento é bem conhecido. Por outro lado, os seres humanos são exatamente o oposto de tais entidades simples: o comportamento detalhado de cada um deles é o resultado complexo de muitos processos fisiológicos

e psicológicos. Ainda, não é possível saber a dinâmica exata de um único indivíduo e nem a forma detalhada de como ele interage com os outros. Portanto, qualquer modelagem de agentes sociais, inevitavelmente, envolve uma grande simplificação do problema real. Na maioria das situações e com boa aproximação, propriedades qualitativas (e até mesmo quantitativas) de fenômenos de grande escala podem não depender de vários detalhes do processo. Características de níveis mais elevados como simetrias, dimensionalidades e leis de conservação são relevantes para o comportamento global. Com esses conceitos em mente, torna-se mais fácil a abordagem de modelos para alguns sistemas sociais. Para dar consistência a um modelo, um passo crucial é a comparação com dados empíricos, tornando possível verificar se o modelo “microscópico” é uma boa abordagem para os dados reais.

Padrões em muitos sistemas sociais vêm sendo cada vez mais identificados usando física estatística, como por exemplo, em religião [1, 2], economia [3, 4], propagação de notícias [5, 6], popularidade [7, 8, 9, 10], dinâmicas de opinião [11, 12, 13, 14] e política [15, 16]. Nesse sentido, também têm surgido trabalhos revisando resultados que utilizam física estatística no estudo de sistemas sociais [17, 18, 19].

A vida cotidiana apresenta muitas situações nas quais há necessidade de um grupo chegar a decisões compartilhadas, tornando uma posição mais forte e ampliando seu impacto na sociedade. Nesse contexto, a física estatística também é aplicada em vários estudos de dinâmicas de opinião visando a definir os estados de opinião de uma população e os processos elementares que determinam as transições entre esses estados. Muitas vezes, as pessoas são confrontadas com um número limitado de escolhas em uma questão específica, como por exemplo: direita ou esquerda, Windows ou

Linux, comprar ou vender, candidato 1 ou 2, etc. Se opiniões podem ser representadas por números, o desafio é encontrar um conjunto adequado de regras matemáticas para descrever os mecanismos responsáveis por suas evoluções e mudanças. A primeira dinâmica de opinião criada por um físico foi um modelo proposto por Weidlich [20], em 1971. Mais tarde, o modelo de Ising fez sua primeira aparição na dinâmica de opinião nos trabalhos de Galam [21, 22]. O acoplamento spin-spin representa a interação de pares entre os indivíduos e o análogo do campo magnético geralmente representa o efeito externo. Entre os modelos de dinâmica de opinião que têm recebido maior atenção estão o modelo do eleitor [23, 24], a regra da maioria [25], os modelos baseados em teoria de impacto social (o impacto de um grupo social em um indivíduo depende do número de indivíduos no grupo) [26, 27, 28] e o modelo de Sznajd (o impacto exercido por um grupo social em um indivíduo cresce com o tamanho do grupo) [11].

Uma grande contribuição na abordagem física para a dinâmica de opinião está centrada em aspectos quantitativos do fenômeno de formação de consenso, ou seja, começando pela análise de dados reais, além de abordar as questões meramente qualitativas de quando e como as pessoas concordam ou discordam. A disponibilidade atual de grandes bases de dados e computadores capazes de lidar com eles contribuem imensamente com as investigações.

Entre os fenômenos de dinâmica de opinião, eleições constituem dinâmicas de importância fundamental para uma sociedade, já que seus resultados influenciam diretamente em muitos aspectos da vida das pessoas. Ademais, atualmente encontra-se disponível ao público uma grande quantidade de dados de eleições, em formato

eletrônico. Diante dessas considerações, várias investigações das dinâmicas eleitorais empregando dados têm sido realizadas.

Um histórico, muito breve, de resultados já encontrados até o momento inicia-se com Costa filho *et al.* [29, 30], argumentando que o número de candidatos que recebem uma dada fração de votos, em eleições proporcionais brasileiras, tem distribuição seguindo uma lei de potência com expoente -1 e que esse comportamento pode ser modelado como um processo multiplicativo; em seguida, os mesmos dados foram melhor ajustados por Lira *et al.* [31] usando uma generalização da lei de Zipf. Para a distribuição de votos em eleições pluralistas brasileiras, Araripe *et al.* [32] sugerem distribuições e observam que os votos tendem a ser polarizados entre os dois primeiros colocados; Araujo *et al.* [33] constroem um modelo considerando um mecanismo de “votação tática” obtendo esse comportamento de polarização através do impacto dos resultados de pesquisas. Com o objetivo de simular o comportamento de lei de potência encontrado por Costa Filho *et al.*, Bernardes *et al.* [34] construíram um modelo que utiliza uma generalização da dinâmica de opinião de Sznajd [11] na rede de Barabási-Albert. González *et al.* [35] também abordam um modelo usando a dinâmica de Sznajd aplicada à rede de Barabási-Albert e a uma rede determinística. Travieso e Costa [37] construíram uma dinâmica de opinião e a aplicaram a cinco modelos de redes. Analisando as eleições proporcionais, e agora considerando a “força” dos partidos, Fortunato e Castellano [38] apontam que a distribuição do número de votos recebidos por candidatos para três países é ajustada por uma lognormal; em contraste, Araripe e Costa [39] indicam que os resultados para eleições brasileiras são diferentes. Dando continuidade às investigações de For-

tunato e Castellano, Chatterjee *et al.* [40] usam dados de quinze países na busca por classes de universalidade das distribuições de votos, pontuando que um conjunto de cinco países tem o comportamento lognormal. Dados como taxa de comparecimento às urnas também foram usados para analisar dinâmicas eleitorais. Borghesi e Bouchaud [41] indicam que a correlação da taxa de comparecimento às urnas decai logaritmicamente com a distância entre as cidades para dados de eleições francesas. Continuando essa investigação e considerando uma base de dados de onze países, a mesma investigação e o mesmo resultado são obtidos por Borghesi *et al.* [43]. Para investigar o estado de ordem do sistema levando em conta as proporções de abstenções, votos brancos e nulos e votos válidos, Borghesi *et al.* [44] realizaram uma análise entrópica. Considerando a taxa de comparecimento às urnas e a taxa de votos para o candidato vencedor, Klimek *et al.* [45] constroem um modelo para detectar mecanismos fraudulentos em eleições.

Nas várias investigações citadas acima, usando conceitos e técnicas de física estatística, um único ponto de vista é focado como consequência para os resultados de eleições: a escolha de candidatos no dia da eleição. No entanto, o eleitor tem sua opinião restrita a um conjunto de candidatos pré-estabelecido pelos partidos políticos. Tal fato aponta que o resultado de uma eleição é também influenciado pelo processo de candidatura, ou seja, pela formação do conjunto de candidatos. Um aspecto central desta tese foi direcionado a buscar regularidade no processo de formação do conjunto de candidatos para eleições, um aspecto que ainda não havia sido investigado. Analisando dados de eleições para prefeitos e vereadores, encontraram-se um comportamento médio lei de potência e um aparente crescimento do expoente

com o número de posições. Nesse contexto, foi também realizada a mesma análise para dados de filiados a partidos políticos [73], confirmando o resultado observado (comportamento médio lei de potência) e um crescimento do expoente com o número de posições. Para ambas as análises, a flutuação dos dados em torno do comportamento médio tem uma distribuição lognormal, indicando que as variáveis seguem um processo multiplicativo.

Motivado pela relevância das dinâmicas eleitorais, este trabalho contém um histórico, nos capítulos 1-4, dos resultados citados acima, focando os objetivos e resultados obtidos nas investigações, e a seguir, nos capítulos 5 e 6, um detalhamento das investigações que culminaram nesta tese. No primeiro capítulo, abordaram-se as investigações que analisam distribuições de votos em eleições proporcionais e pluralistas. No segundo capítulo, apresentam-se os resultados de modelagem via dinâmicas de opinião em redes para o comportamento encontrado no capítulo 1 para eleições proporcionais. No terceiro capítulo, é retomado o contexto de distribuição de votos em eleições proporcionais, porém levando em conta o peso dos partidos. No capítulo 4, são mostrados os resultados de dinâmicas que levam em conta a taxa de comparecimento às urnas. No capítulo 5, apresenta-se uma análise do crescimento do número de candidatos a prefeito e a vereador com o tamanho do eleitorado das cidades. No capítulo 6, é apresentada a mesma análise para o crescimento do número de filiados a partidos políticos, indicando que as regularidades encontradas (comportamento lei de potência) nos capítulos 5 e 6 sugerem uma regra relacionada ao crescimento das posições políticas (o expoente cresce com o número de posições). Encerrando esta tese há ainda as conclusões e dois apêndices.

Capítulo 1

Distribuição de votos em eleições

Resultados de eleições proporcionais no Brasil indicam que a distribuição do número de candidatos com uma dada fração de votos, para uma parte considerável dos dados, tem um comportamento hiperbólico, o qual pode ser relacionado a um processo multiplicativo. No entanto, uma generalização da lei de Zipf ajusta uma quantidade maior dos dados. Para eleições pluralistas, encontrou-se um comportamento aproximadamente exponencial, com uma tendência à polarização dos votos entre os dois primeiros colocados. Tal tendência mostrou-se consistente com um mecanismo de votação tática devido a resultados de pesquisas durante a campanha eleitoral.

1.1 Distribuição de votos em eleições proporcionais

Análises objetivando identificar padrões nas distribuições de votos em eleições no sistema proporcional (Apêndice A.1) vêm sendo feitas. Em particular, investigações referentes à eleições realizadas no Brasil nos anos de 1998 e 2002 indicaram um regime de lei de potência para uma parte considerável da distribuição dos dados de eleições proporcionais (deputados federais e estaduais) [29, 30]. Considerando os resultados de eleições em cada estado para a posição de deputado estadual e normalizando os votos de cada candidato pelo número total de eleitores do respectivo estado, o histograma do número de candidatos N recebendo uma fração de votos ν é mostrado na Figura 1.1 (a). Nesse caso, há um comportamento hiperbólico, isto é, $N(\nu)$ segue muito próximo a uma lei de potência

$$N(\nu) \propto \nu^{-\alpha} , \quad (1.1)$$

com $\alpha \approx 1$, por duas ordens de magnitude. Na Figura 1.1 (b), o mesmo procedimento foi realizado para a posição de deputado federal.

Um caminho possível para interpretar as leis de potência encontradas é olhar para $N(\nu)$ como um “grande processo” multiplicativo. Nesse sentido [29], associa-se a cada candidato fatores p_i 's, e cada p_i está intrinsecamente relacionado aos atributos do candidato que realiza o subprocesso i entre os eleitores. Além disso, assume-se

que os p_i 's são positivos e independentes, de modo que sua fração de voto deve ser

$$\nu \propto p_1 p_2 \dots p_n, \quad (1.2)$$

em que n é a quantidade de subprocessos. Para n suficientemente grande e pelo teorema do limite central, a distribuição de ν é aproximadamente lognormal (Apêndice B.2). Já que a lognormal é proporcional a $\frac{1}{\nu} \exp(-a(\ln \nu)^2 + b \ln \nu)$ e se seu expoente, em módulo, for muito pequeno, a lognormal aproxima-se de ν^{-1} , que é o padrão encontrado na Figura 1.1.

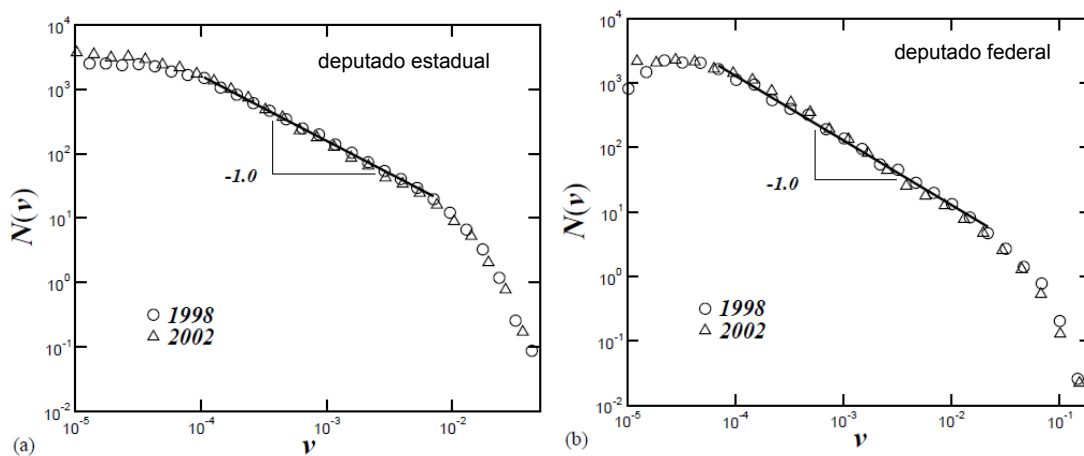


Figura 1.1: **Histogramas das frações de votos** - Os gráficos, em log-log, mostram as distribuições de votos para as eleições de 1998 (círculos) e de 2002 (triângulos). As linhas sólidas representam os ajustes para as regiões de escala (a) para deputados estaduais e (b) deputados federais. Adaptada de Costa Filho *et al.* [29, 30].

1.2 Ajuste via lei de Zipf

Observando os histogramas da Figura 1.1, é perceptível que uma parte considerável da distribuição dos dados não é ajustada pela lei de potência sugerida [29, 30]. Todavia, as curvas são aparentemente “bem comportadas”, surgindo a questão: será que um outro modelo não ajustaria melhor? Baseando-se nos resultados das eleições proporcionais do Brasil nos anos de 1998 e 2000, essa questão foi respondida para as posições de vereadores, deputados estaduais e federais. Mostrou-se [31] que as distribuições dos votos podem ser bem aproximadas por uma lei de Zipf generalizada, propostas dentro de uma ferramenta estatística não-extensiva [48].

Para dados das eleições brasileiras no ano 2000 e considerando os resultados para as 15 maiores cidades para a posição de vereador, 15 histogramas estão agrupados, por meio de médias, e mostrados em log-log na Figura 1.2.

A curva encontrada é muito próxima da lei de Zipf

$$N(\nu) = \frac{A}{(1 + C\nu)^\alpha}, \quad (1.3)$$

em que A é uma constante de normalização e C regula a transição entre o platô inicial e o regime de lei de potência caracterizada pelo expoente α . A lei acima pode ser relacionada ao formalismo termodinâmico não-extensivo [48]. Nesse formalismo, a função exponencial, que ocorre no peso estatístico de Boltzmann-Gibbs, é substituída por uma generalização da exponencial a um parâmetro q . O desvio de 1 desse parâmetro q está associado ao grau de não-extensividade. Por sua vez, a exponencial

q , $\exp_q(t)$ é a solução da equação

$$\frac{dW}{dt} = \lambda_q W^q, \quad (1.4)$$

com $\lambda_q = 1$ e $W(0) = 1$, isto é, $W(t) = \exp_q(t) = [1 + (1 - q)t]^{1/(1-q)}$. Note-se que o caso $q = 1$ corresponde à exponencial usual. Tomando $W = N$ e $t = \nu$ na equação (1.4), obtém-se a solução

$$N(\nu) = \frac{N(0)}{[1 + (q - 1)\lambda'_q \nu]^{1/(q-1)}}, \quad (1.5)$$

com $\lambda'_q = -\lambda_q N(0)^{q-1}$. O melhor ajuste de $N(\nu)$ para eleições de vereadores para a equação (1.5) é mostrado como uma linha sólida na Figura 1.2 e fornece, quando comparada à equação (1.3), $\alpha = 1/(q - 1) = 2,63$ ($q = 1,68$), um expoente consideravelmente maior do que um.

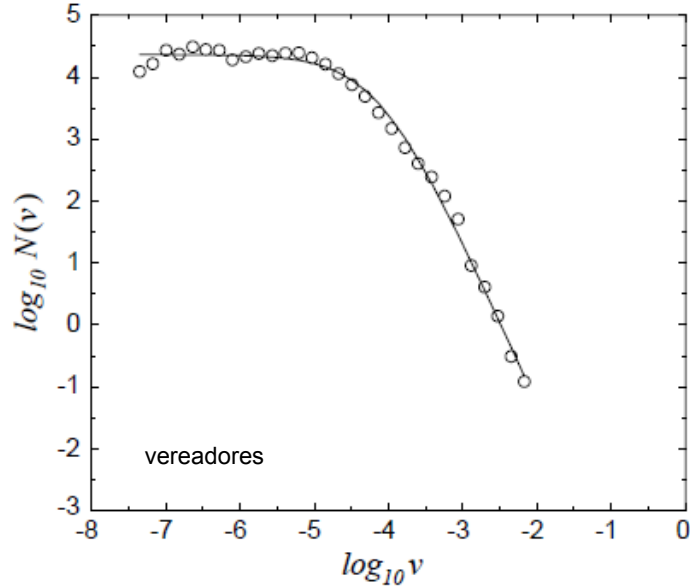


Figura 1.2: **Ajuste via lei de Zipf - vereadores** - O gráfico mostra a distribuição de votos para representantes locais (vereadores) em log-log. Os círculos são referentes aos dados dos resultados de eleições proporcionais brasileiras de 2000, considerando as 15 maiores capitais. A linha sólida é o melhor ajuste via lei de Zipf, equação (1.5), cujo expoente é $\alpha = 2,63$. Adaptada de Lyra *et al.* [31].

A Figura 1.3 mostra $N(\nu)$ para dados de eleições para deputados estaduais em (a) e federais em (b) para 15 estados, correspondendo às 15 maiores capitais do Brasil, como na análise para vereadores. Observa-se que depois do platô inicial, um regime de escala lei de potência emerge com expoente muito próximo de -1,0, em ambos os casos, mas pode-se observar que o regime de escala muda para frações grandes de votos e um rápido decaimento ocorre. Como visualizado na Figura 1.2, essa quebra de invariância de escala não é vista nos dados de eleições para vereadores.

É perceptível portanto, que os gráficos da Figura 1.3 parecem ser composições entre a exponencial usual ($q = 1$) e a exponencial q ($q \neq 1$). Assim, é razoável supor

que $N(\nu)$ pode ser a solução de uma equação que é a interpolação entre a equação (1.4) com $(q = 1)$ e a equação (1.4) com $(q \neq 1)$, isto é, considera-se que $N(\nu)$ satisfaz a equação

$$\frac{dN(\nu)}{d\nu} = -\lambda N - (\lambda_q - \lambda)N^q, \quad (1.6)$$

cuja solução é

$$N(\nu) = \frac{N(0)}{[1 - \lambda'_q/\lambda + (\lambda'_q/\lambda)e^{(q-1)\lambda\nu}]^{1/(q-1)}}, \quad (1.7)$$

com $\lambda'_q = \lambda - (\lambda - \lambda_q)N(0)^{q-1}$. Na Figura 1.3, a linha sólida corresponde ao melhor ajuste via equação (1.7). Nesse caso, tem-se $\lambda'_q = 12227,3$, $\lambda_q = 135,1$ e $\alpha = 1/(q-1) = 1,03$ ($q = 1,97$) para deputados estaduais e $\lambda'_q = 6774,4$, $\lambda_q = 33,1$ e $\alpha = 1/(q-1) = 1,07$ ($q = 1,93$) para deputados federais. Note que o expoente α está de acordo com o estimado anteriormente. Os processos de votação não-locais sugerem um comportamento de escala universal. No entanto, a passagem para o comportamento exponencial ocorre em diferentes frações (não universal) de votos. A diferença nos expoentes de escala que aparecem na lei de Zipf generalizada para representantes locais (vereadores) e não-locais (deputados estaduais e federais) sugere distintas dinâmicas subjacentes.

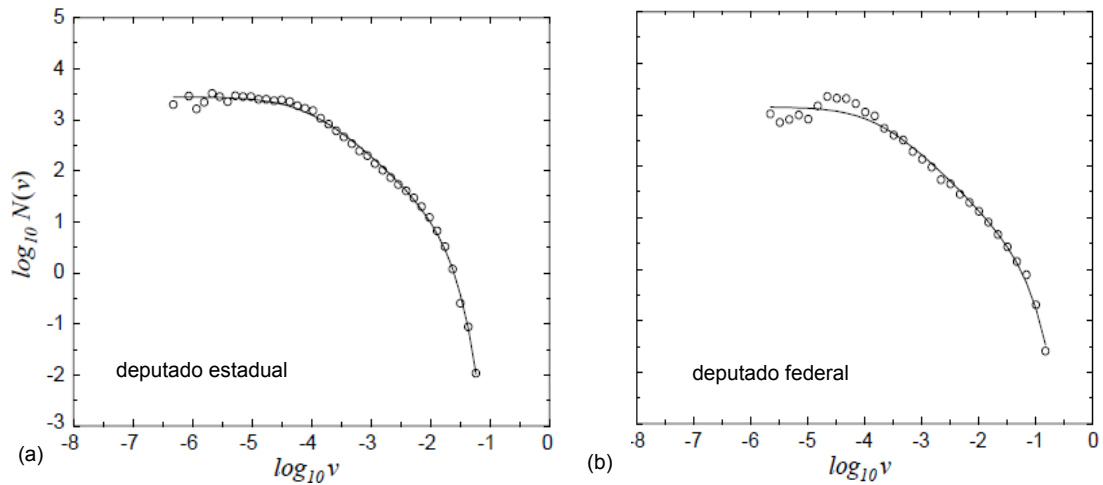


Figura 1.3: **Ajuste via lei de Zipf generalizada - deputados estaduais e federais** - A distribuição de votos para deputados estaduais e federais em log-log. Os círculos são referentes às eleições brasileiras no ano de 1998, considerando os estados com as 15 maiores capitais. A linha sólida é o melhor ajuste via lei de Zipf generalizada, equação (1.7). O expoente de escala da lei de potência é $\alpha = 1,03$ para (a) deputados estaduais e $\alpha = 1,07$ para (b) deputados federais. Adaptada de Lyra et al. [31].

1.3 Distribuição de votos em eleições pluralistas

Os estudos apresentados até o momento focavam resultados de eleições proporcionais, em que o número de vagas em cada cargo é sempre superior a um. A seguir, analisam-se eleições em que há somente uma vaga, como é o caso do executivo (presidente, governador e prefeito). Para esse tipo de eleição, um dos formatos mais comuns é o chamado sistema de votação pluralista, no qual o candidato vencedor deve apenas receber o maior número de votos. Esse sistema é aplicado em 43 dos 191 países das Nações Unidas, e o Brasil é a maior democracia desse grupo. Na Figura 1.4, encontra-se a distribuição da fração de votos ν para o vencedor à direita

e para o perdedor à esquerda se somente dois candidatos estão disputando. Por construção, essa distribuição é simétrica em relação à $\nu = 0,5$. O conjunto de dados completo foi ajustado pela distribuição de Laplace

$$P(\nu) \propto \exp\left(\frac{-|\nu - 0,5|}{\lambda}\right), \quad (1.8)$$

com o parâmetro $\lambda = 0,08$ [32].

Essa distribuição foi usada em uma generalização do modelo de fragmentação [46] para descrever distribuições de votos para prefeito. Qualitativamente, fragmentações ocorrem como consequência de sucessivas partições. No contexto de eleições, fragmentação pode ser entendida como um fracionamento do eleitorado. No modelo de fragmentação empregado em [32], começa-se dividindo o eleitorado em uma sequência de frações, ν_1 e $r_1 = 1 - \nu_1$, $r_2 = r_1 - \nu_2$ e assim sucessivamente. Diferente de [46], em que investigou-se o limite de um número infinito de frações, considerou-se aqui o processo para um número finito de frações, generalizando com tamanhos que podem ser reescritos como: $\nu_1 = x_1$, $\nu_2 = (1 - x_1)x_2$, $\nu_3 = (1 - x_1)(1 - x_2)x_3$, \dots , $\nu_n = (1 - x_1)\dots(1 - x_{n-1})x_n$, com $0 \leq x_i \leq 1$ sendo uma variável aleatória obedecendo a uma distribuição $\rho(x)$. Atribui-se a cada candidato i uma fração de votos ν_i para $i = 1, 2, \dots, n$. Isso é justificado quando assume-se que ν_i deve ser muito próximo da fração de eleitores decididos a votar no candidato i . Desse modo, é razoável adotar a distribuição dada pela equação (1.8) como uma primeira aproximação para $\rho(x)$. A fração máxima de votos entre os n candidatos mais votados é dada por

$$\nu_{max} = \max[\nu_1, \nu_2, \dots, \nu_{n-1}, \nu_n]. \quad (1.9)$$

Para a solução numérica do modelo, primeiramente geraram-se $n - 1$ números aleatórios distribuídos de acordo com (1.8). A partir disso, calculou-se todo o conjunto de frações ν_i e determinou-se a maior delas, ν_{max} . Repetiu-se o processo $N = 10^5$ vezes para produzir um histograma para ν_{max} para o caso de eleições com 3 ou 4 candidatos, como ilustra a Figura 1.5.

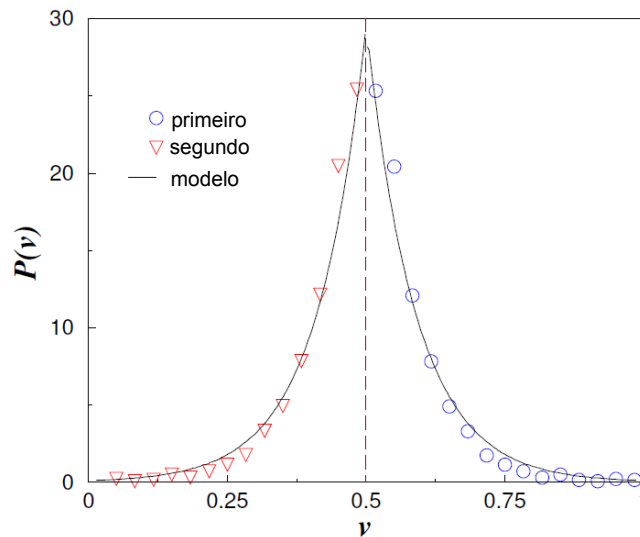


Figura 1.4: **Distribuição de votos para prefeito** - Histograma da fração de votos das eleições para prefeito nas cidades brasileiras em 2004. Os dados correspondem às eleições com somente dois candidatos e todas as coordenadas foram divididas por um fator de 10^3 . Os círculos mostram a frequência da fração de votos do vencedor e os triângulos representam os resultados para o perdedor. A linha contínua é simétrica em relação à $\nu = 0,5$ e representa o melhor ajuste dos dados pela distribuição de Laplace (1.8). Adaptada de Araripe *et al.* [32].

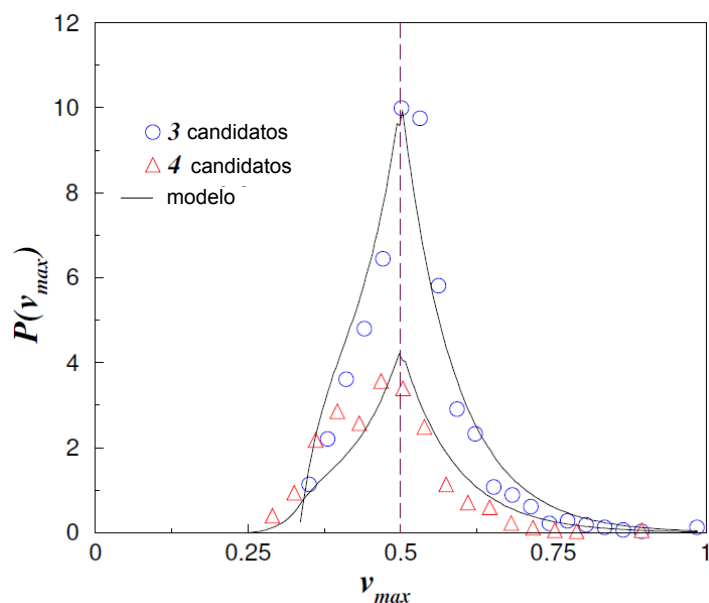


Figura 1.5: **Distribuição de votos dos candidatos mais votados** - Histograma da fração de votos para o candidato mais votado v_{max} em eleições com 3 (círculos) e 4 (triângulos) candidatos. As linhas sólidas são as previsões do modelo de fragmentação. Adaptada de Araripe *et al.* [32].

Observa-se que a Figura 1.6 indica que a seleção de um entre n candidatos pela população durante uma campanha eleitoral certamente não é um processo estático. A evolução temporal dessas enquetes ilustram bem a polarização dos votos entre o primeiro e o segundo lugar. O crescimento da popularidade de ambos está claramente ligada à perda de votos dos dois menos votados.

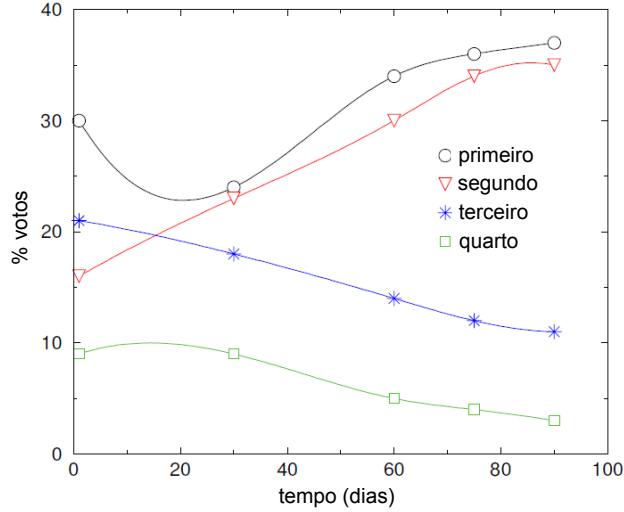


Figura 1.6: **Evolução de uma campanha eleitoral** - Evolução temporal da fração das intenções de votos para os 4 primeiros candidatos mais votados na eleição para prefeito de São Paulo em 2004 durante a campanha eleitoral. Essas enquetes foram realizadas pela agência brasileira IBOPE. O tempo em dias é contado a partir da data da primeira enquete, 28 de junho de 2004. A aproximação gradual entre o primeiro e o segundo candidatos mais votados mostra o fenômeno de polarização. Adaptada de Araripe *et al.* [32].

Em eleições pluralistas, diferentes mecanismos podem afetar a formação da opinião dos eleitores durante um processo eleitoral. Um tipo de mecanismo chamado de votação tática, caracterizado pela força relativa do retorno das pesquisas eleitorais e da interação entre os indivíduos da sociedade, é identificado em [33]. Um modelo foi introduzido para estudar os efeitos da votação tática em eleições pluralistas, o qual oferece uma possível explicação para a polarização final de votos sugerida na Figura 1.6.

Para descrever o modelo, foram considerados n candidatos identificados pelo índice $\sigma = 1, \dots, n$ e uma função balanço \mathcal{F} para quantificar o grau de indecisão

na sociedade devido à coexistência de diferentes opiniões, definida por

$$\mathcal{F} = -H \sum_{\langle i,j \rangle} \delta(\sigma_i, \sigma_j) + \alpha(\nu_f - \nu_s) \sum_i \delta(\sigma_i, f), \quad (1.10)$$

em que σ_i é o candidato para o qual o eleitor i vota e f é o índice do primeiro colocado nas pesquisas. A função $\delta(\sigma_i, \sigma_j)$ é 1 se i e j escolheram o mesmo candidato e 0 caso contrário. O coeficiente H mede a “força” de interação entre indivíduos conectados e α o impacto das pesquisas. Uma tendência natural na sociedade é minimizar esse grau de indecisão. Junto com o princípio de máxima entropia, isso leva à distribuição de probabilidade de cada estado $P \propto \exp(-\mathcal{F})$.

Na ausência de pesquisas ($\alpha = 0$), o modelo torna-se o modelo de Potts de n estados do magnetismo [49]. Para H acima de H_c , a opinião de uma maioria estável pode surgir, enquanto que abaixo desse valor a interação entre as pessoas não é suficiente para levar a uma maioria fixada. O sistema é então controlado pela competição entre o comprimento relativo do impacto das pesquisas α/H e seu grau de subcriticalidade $(H - H_c)/H_c$, dado pela razão

$$R = \frac{\alpha/H}{(H - H_c)/H_c}. \quad (1.11)$$

O numerador mede a tendência à votação tática, que quanto maior, mais forte é a resposta às pesquisas. A “força” da ligação psicológica entre indivíduos de uma sociedade é representada no modelo pelo grau de subcriticalidade. Para H acima H_c , no limite de $R = 0$, quando os eleitores decidem independentemente da opinião global, rapidamente uma clara maioria aparece. Com o crescimento de R , a tendência

para a o voto tático torna-se mais relevante. No limite de R grande, a fração de votos dos vencedores é um pouco acima de $1/n$, ou seja, não há evidência de maioria clara.

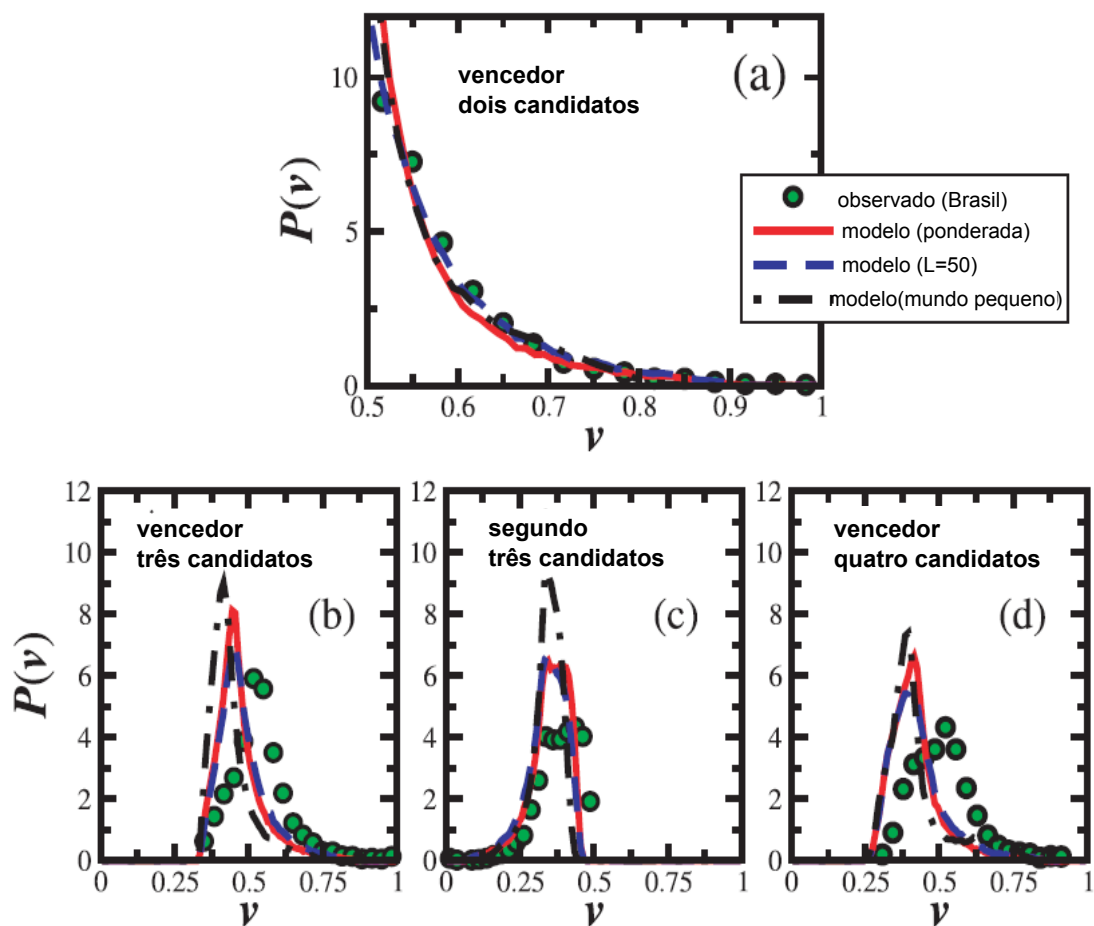


Figura 1.7: **Resultados de eleições brasileiras e o modelo de voto tático** - Distribuição dos votos dos vencedores (verde) para eleições com (a) dois, (b) três e (d) quatro candidatos. Devido à resiliência dos resultados das eleições de 2000, 2004 e 2008, as médias foram realizada sobre as três eleições para melhorar as estatísticas. As linhas vermelhas são as mesmas distribuições obtidas com o modelo para $R = 0,69$, considerando uma média ponderada sobre três sistemas de tamanhos 50^2 , 75^2 e 100^2 eleitores. As linhas azuis correspondem a uma malha quadrada com 50^2 eleitores. As linhas pontilhadas pretas são obtidas em uma rede com topologia de “mundos pequenos” com $R = 0,46$. A distribuição do segundo candidato para eleições com três candidatos é mostrada em (c). Todos os resultados são médias sobre 5^4 amostras. Adaptada de Nuno *et al.* [33].

Para uma primeira verificação do modelo, foram considerados dados de eleições para prefeito no Brasil. A Figura 1.7 mostra a distribuição de votos do vencedor em eleições para prefeito com dois candidatos. Esses resultados foram usados para caracterizar o eleitorado em termos da razão R da equação (5.1). Ajustando os resultados simulados para $n = 2$ com os dados empíricos, obtém-se $\alpha/H = 0,0069$ e $(H - H_c)/H_c = 0,01$, correspondendo a $R = 0,69$.

A Figura 1.8 (a) mostra, juntamente com dados do Brasil, resultados de eleições para prefeito nos Estados Unidos (2008) e no Canadá (2008). Independentemente do número de candidatos, é possível notar, Figura 1.8 (b), um máximo da fração de votos em torno de 0,5 com reduzidas margens de vitória.

Comparando a Figura 1.8 (a) com a Figura 1.8 (b), e recordando que o crescimento de R é obtido por meio de uma diminuição no grau de subcriticalidade ou um aumento do impacto dos resultados das pesquisas, observa-se que o impacto das eleições nos Estados Unidos e Canadá são muito semelhantes. No Brasil, contudo, o impacto revela-se ainda mais forte.

O modelo aplicado a uma malha quadrada, ou à rede com propriedades de “pequenos mundos” reproduzem as principais características: a pequena diferença e a polarização entre os dois primeiros colocados.

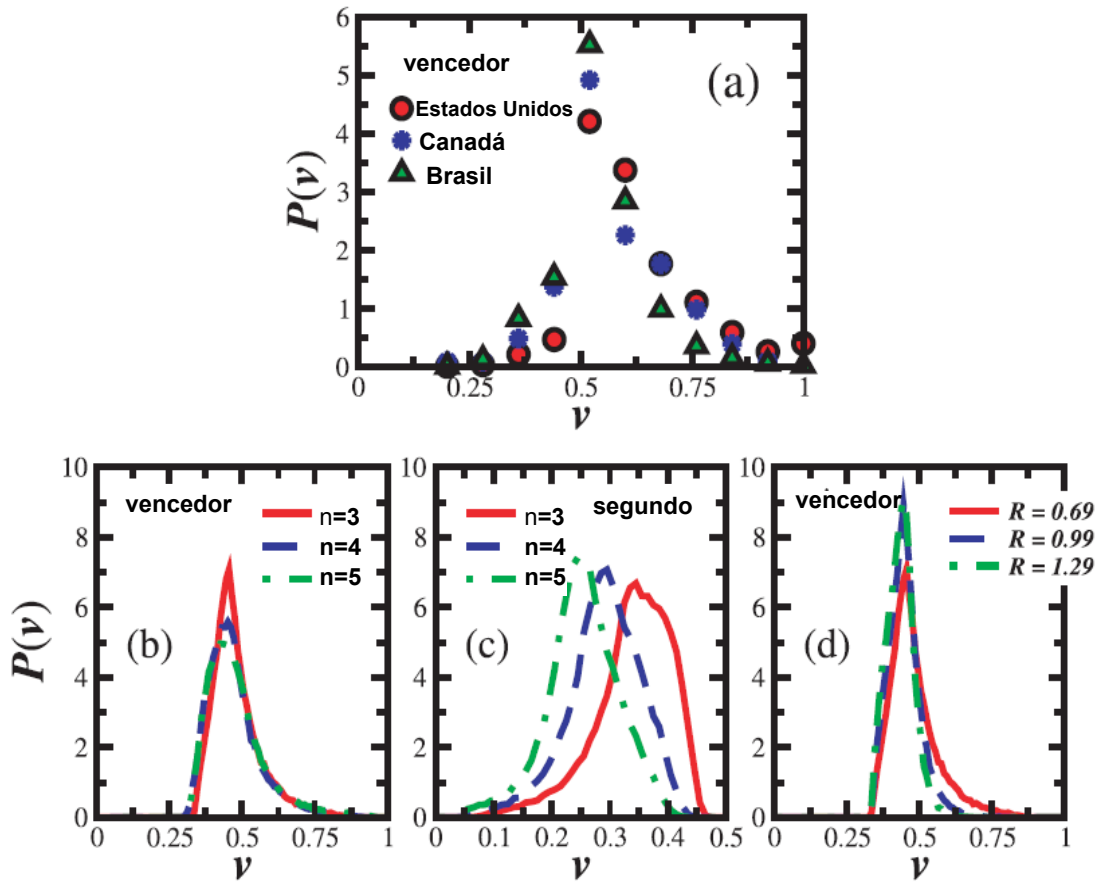


Figura 1.8: **Comparação entre diferentes sistemas** - (a) Distribuição da fração de votos do candidato vencedor em eleições para prefeito, nos Estados Unidos em 2008 (círculos vermelhos), Canadá em 2008 (estrelas azuis) e no Brasil foi tomada a média sobre os resultados de 2000, 2004 e 2008 (triângulos verdes). (b) Distribuição da fração de votos do vencedor para 3 (linha contínua vermelha), 4 (linha tracejada azul) e 5 (linha pontilhada verde) candidatos. (c) Distribuição de votos do segundo colocado. Distribuição de votos do vencedor relativas a diferentes valores do impacto das pesquisas, para $R = 0,69$ (linha contínua vermelha), $R = 0,99$ (linha tracejada azul) e $R = 1,29$ (linha pontilhada verde). Todos os resultados foram obtidos realizando a média sobre 5^4 amostras em uma malha quadrada com 2.500 eleitores. Adaptada de Nuno *et al.* [33].

Capítulo 2

Simulação de distribuição de votos em redes

Com o objetivo de simular o comportamento hiperbólico em distribuição de votos, alguns modelos foram construídos usando redes com diferentes topologias. A dinâmica de opinião de Sznajd foi aplicada à rede de Barabási-Albert e a uma outra com topologia fractal. Outro modelo simples de dinâmica de opinião foi aplicado a vários tipos de redes visando analisar o quão robusto se torna o modelo em relação à propriedade de “mundo pequeno” presente na rede.

2.1 Modelo de Sznajd na rede Barabási-Albert

No modelo de Sznajd, pequenos conjuntos de pessoas influenciam seus vizinhos próximos se e somente se todas as pessoas dentro do conjunto concordam (Apêndice B.3). Uma generalização desse modelo foi aplicada a duas redes diferentes: em uma

rede cúbica, que é uma versão tridimensional de [47], e na rede de Barabási (Apêndice B.4.3).

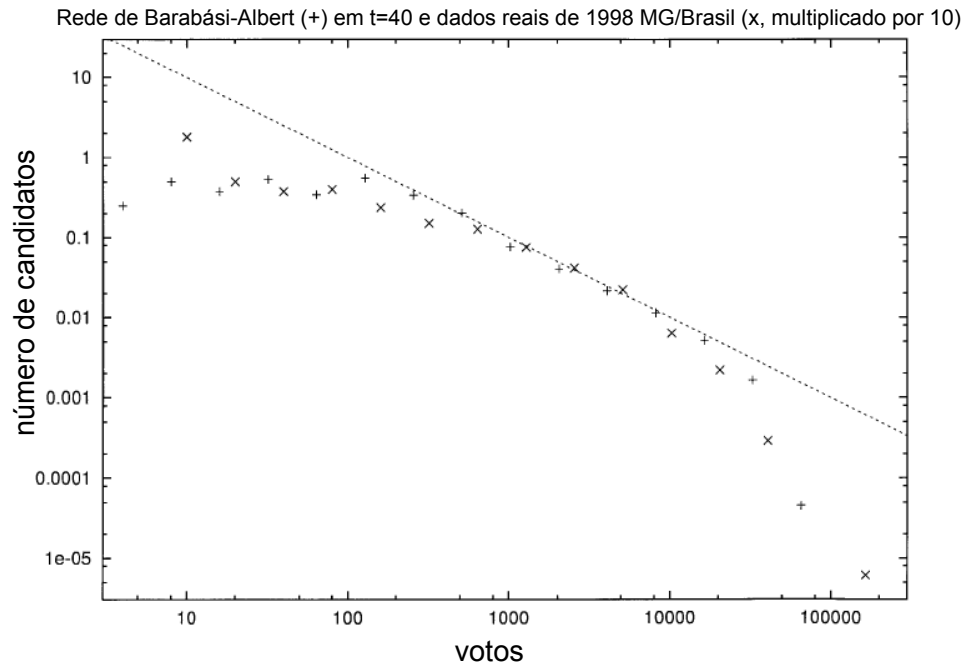


Figura 2.1: **Simulação em rede livre de escala** - Distribuição de $N(\nu)$ para meio milhão de nós na rede de Barabási-Albert, em que cada nó adicionado é ligado a 5 nós já existentes. Adaptada de Bernardes *et al.* [34].

Para aplicar o modelo na rede de Barabási-Albert (Apêndice B.4.3), a rede foi construída partindo de 6 nós e $m = 5$. Mil candidatos foram distribuídos aleatoriamente desconsiderando o número de conexões de um nó, e a partir daí começou a campanha. Em cada espaço de tempo, todos os nós são visitados. Para cada nó, tem-se o seguinte processo:

- Se um nó i já tem preferência por um candidato, escolhe-se um nó conectado j aleatoriamente. Se o nó i não tem candidato, segue-se para outro nó selecionado

aleatoriamente.

- Se o nó j tem o mesmo candidato que o nó i , eles tentam convencer todos os nós conectados a eles. A probabilidade de convencer outros, para cada um dos dois nós, é agora inversamente proporcional ao número de nós conectados a ele, significando que cada nó convence, em média, um outro nó em cada processo.
- Se o nó j não tem candidato, o nó i tenta convencê-lo com a mesma probabilidade descrita acima.
- Se o nó j tem um candidato diferente do nó i , passa-se para outro nó i . A Figura 2.1 mostra que, exceto para valores pequenos e grandes de ν , a lei hiperbólica concorda bem com o modelo para tempos intermediários como $t = 40$. Porém, aplicando o mesmo modelo em uma rede cúbica [34], apesar de também capturar esse comportamento, ele é bem menos robusto.

2.2 Redes fractal e de Barabási-Albert

Os resultados relatados sugerem que a dinâmica dada pelo modelo de formação de opinião de Sznajd modela bem sistemas eleitorais. Nesse sentido, ele foi aplicado em diferentes redes para simular resultados de eleições. Em [35], a simulação do modelo aplicado à rede de Barabási-Albert é comparada a uma simulação do mesmo modelo aplicado a uma rede dada por um grafo determinístico. Os resultados são ainda comparados com dados de eleições brasileiras e indianas.

Esse grafo determinístico cresce da seguinte maneira: em cada intervalo de tempo, cada aresta gera um vértice adicional, o qual é ligado a ambos os vértices finais da

aresta. Em $t = 0$, tem-se um triângulo de arestas conectando os três vértices; em $t = 1$, o grafo consiste de seis vértices conectados por nove arestas, e assim por diante. O número total de vértices em t interações é dado por $N_t = 3(3^t + 1)/2$ e o grafo descrito é chamado pseudo-fractal. A distribuição de grau (Apêndice B.4.1) assintótica para essa rede é $P \sim k^{-\gamma}$ com $\gamma = 1 + \ln 3 / \ln 2$. Apesar de não discutido em [35], outras redes de topologias parecidas, grafos determinísticos, poderiam ser usadas na simulação de distribuições de votos. Um exemplo é a rede de Apolônio [36], em que $N_t = 3 + (3^t - 1)/2$.

O modelo aplicado às redes pseudo-fractal e Barabási-Albert é o mesmo descrito na seção anterior, porém faz-se a seguinte consideração: se o nó j tem o mesmo candidato que o nó i , cada nó convence todos os nós conectados a ele com uma probabilidade:

$$p(k) = \frac{1}{[n(k)]^{1/\xi}}, \quad (2.1)$$

em que $n(k)$ é o número de nós conectados a i ou j e $\xi \geq 1$. O caso $\xi = 1$ corresponde ao empregado na seção anterior.

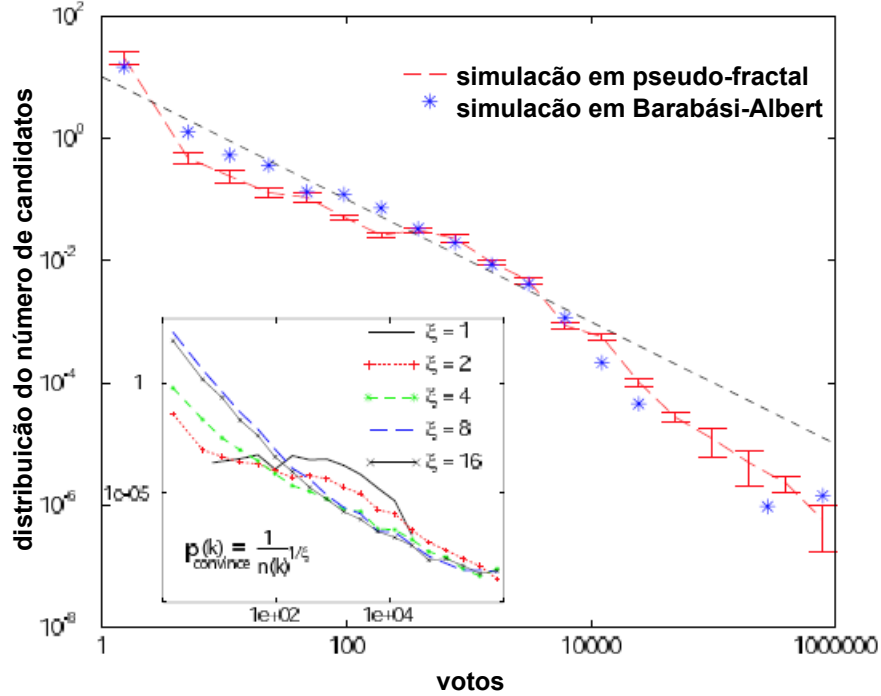


Figura 2.2: **Comparação dos modelos** - O gráfico da distribuição de votos de um processo eleitoral na rede pseudo-fractal comparado com a simulação em uma rede Barabási-Albert. Ambas as redes têm 797.163 sítios (eleitores) e 500 candidatos. Considera-se a média de 20 realizações em cada rede. A distribuição da pseudo-fractal é obtida depois de 15 interações do processo de convencimento e $\xi = 2$. A distribuição da Barabási-Albert é obtida depois de 83 interações. A inserção mostra os resultados da simulação na rede pseudo-fractal para $\xi = 1, 2, 4, 8$ e 16 , depois de 400, 20, 4, 2 e 1 interações, respectivamente. Veja que os resultados dependem da regra escolhida. Adaptada de Sousa *et al.* [35].

Na Figura 2.2, observa-se que as distribuições de votos obtidas via simulações empregando a rede pseudo-fractal e a rede Barabási-Albert concordam muito bem. No caso da rede-pseudo fractal, foi empregado $\xi = 2$ na equação (2.1). Para a rede de Barabási-Albert, foi usado o mesmo valor de [34], $\xi = 1$. Uma comparação entre resultados de eleições para dois estados brasileiros, São Paulo e Minas Gerais, para a posição de deputado estadual e o modelo na rede pseudo-fractal são mostrados na

Figura 2.3.

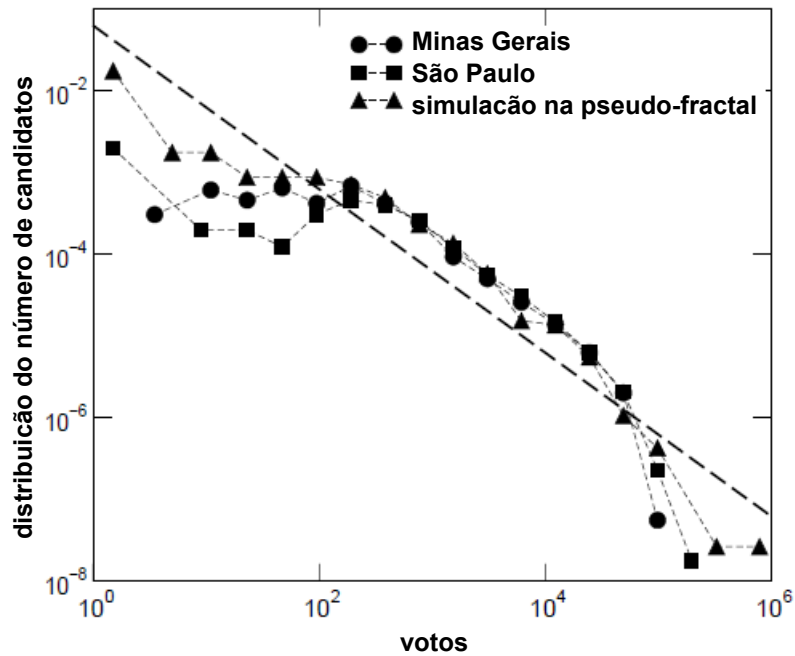


Figura 2.3: **Modelo pseudo-fractal e dados de Minas Gerais e São Paulo** - Resultado da simulação de um processo eleitoral depois de 20 iterações na rede pseudo-fractal de 21.523.362 nós (eleitores) e 1.144 candidatos, comparado com a distribuição de votos para o estado de Minas Gerais em 1998, com 11.815.183 eleitores e 819 candidatos, e o estado de São Paulo em 1998, com 23.321.034 eleitores e 1.260 candidatos. Ambos os eixos estão em escala logarítmica. A linha pontilhada é um ajuste com inclinação -1 . O tamanho do intervalo de votos cresce por um fator 2. A altura da distribuição a partir da rede pseudo-fractal é multiplicada por 10. Adaptada de Sousa *et al.* [35].

Como é possível perceber, o comportamento da distribuição de votos na simulação de eleições na rede pseudo-fractal tem boa concordância com os resultados observados em eleições reais. A distribuição de votos para a simulação e para os dados reais segue a lei hiperbólica, $N(\nu) \propto 1/\nu$, mais precisamente, para o número N de candidatos com ν votos, estendendo-se sobre duas ou três ordem de magnitude, com desvios

para pequenos e grandes números de votos.

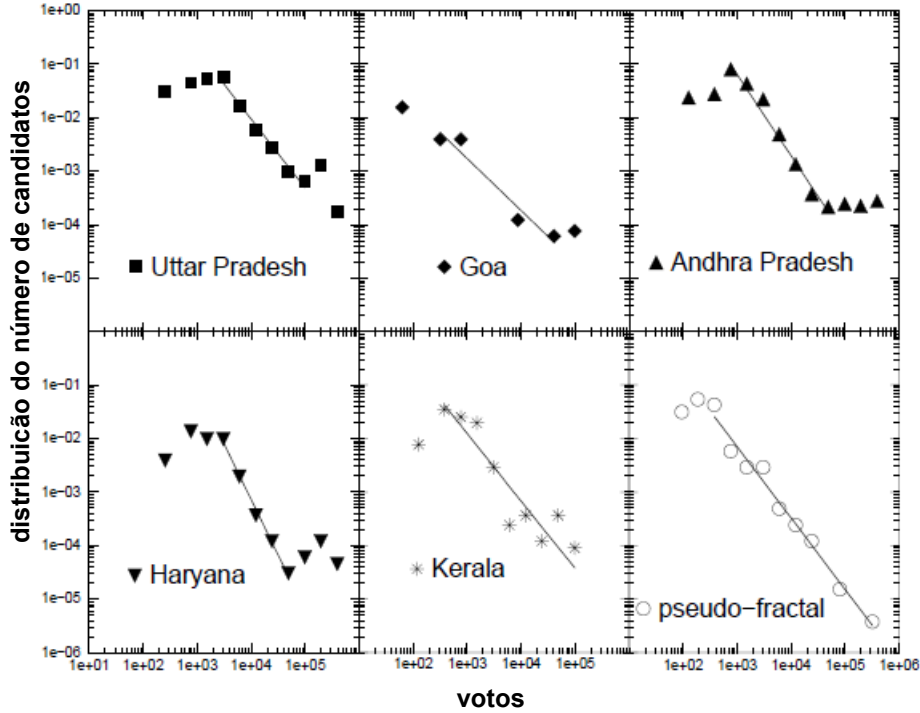


Figura 2.4: **Eleições em estados indianos** - Distribuição de votos para eleições estaduais na Índia em 1998. Para Uttar Pradesh (■), com 5.015.804 eleitores e 649 candidatos; Goa (◆), com 532.766 eleitores e 12 candidatos; Andhra Pradesh (▲), com 31.829.338 eleitores e 301 candidatos; Haryana (▼), com 7.516.884 eleitores e 84 candidatos; e Kerala (*) com 13.036.581 eleitores e 108 candidatos. Obteram-se resultados quantitativamente comparáveis com a simulação na rede pseudo-fractal com 12 gerações e 80 candidatos, depois de 5 interações. As linhas sólidas são os ajustes do dados, com inclinações: -1,32 (Uttar Pradesh), -0,97 (Goa), -1,51 (Andhra Pradesh), -2,06 (Haryana), -1,26 (Kerala) e -1,32 (pseudo-fractal). Adaptada de Sousa *et al.* [35].

Foram também analisadas as distribuições de votos em cada estado nas eleições indianas para a câmara parlamentar. A distribuição de votos para cada estado é a superposição de diferentes processos eleitorais para todos os constituintes do estado. A Figura 2.4 mostra o resultado final do processo eleitoral para estados

indianos, apontando perfis do comportamento hiperbólico ($1/\nu$), o que foi também observado em muitos estados brasileiros. A Figura 2.5 apresenta uma comparação da distribuição de votos na Índia, para o país todo, com a distribuição obtida via modelo pseudo-fractal. Observa-se que $N(\nu) \propto \nu^\alpha$ com $\alpha = -1,3$ (para o Brasil $\alpha = -1$).

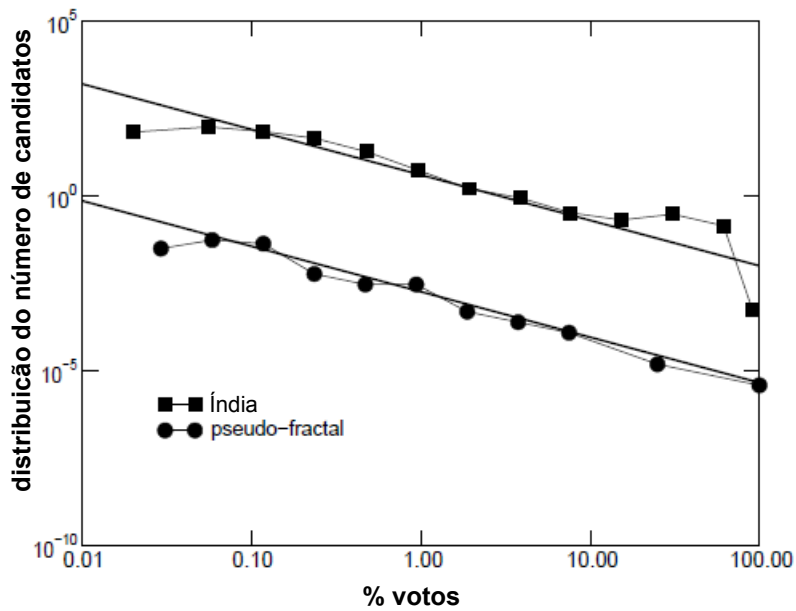


Figura 2.5: **Eleição na Índia e o modelo pseudo-fractal** - Distribuição de votos para a Índia em 1998 (■), comparado, com a simulação na rede pseudo-fractal (●) com 12 gerações e 80 candidatos, depois de 5 interações. A linha sólida é um ajuste com inclinação -1,3. Adaptada de Sousa *et al.* [35].

2.3 Propriedades de mundo pequeno em simulações de eleições

Evidencia-se que, existem muitas possibilidades para se criar modelos que simulam dinâmicas de eleições, dado que uma dinâmica de opinião pode ser aplicada a vários tipos de redes. Quatro modelos são abordados em [37] para simular o regime hiperbólico. Uma mesma dinâmica de opinião é aplicada a cinco tipos de redes.

Para a simulação do modelo de formação de opinião, adotaram-se as redes de Erdős-Rényi, de Barabási-Albert, uma generalização do modelo de Barabási-Albert (descritas no Apêndice B.4.2 e B.4.3), uma malha bidimensional e uma malha bidimensional com conexões aleatórias. A rede de Erdős-Rényi é caracterizada por uma distribuição de grau de Poisson e pela presença da propriedade de “mundo pequeno” (a distância média entre nós é pequena e cresce lentamente com o número de nós na rede). No modelo de Barabási-Albert, apesar de também possuir propriedades de “mundo pequeno”, a distribuição de grau segue uma lei de potência. É notável que na rede regular a adição de nós aleatórios conduz à propriedade de “mundo pequeno” de maneira controlada. A generalização do modelo de Barabási-Albert é caracterizada por uma distribuição de grau do tipo lei de potência, cujo expoente é ajustável.

No modelo de opinião, com N eleitores (nós), começa-se por distribuir C candidatos entre nós escolhidos aleatoriamente com probabilidade uniforme. Os eleitores restantes começam como indecisos. O processo, a seguir, é repetido subsequentemente SN vezes: escolhe-se aleatoriamente um eleitor i que já tenha um candidato

associado C_i ; para todos os vizinhos do eleitor i , se são indecisos, são associados ao candidato C_i ; caso contrário, eles mudam de opinião, com probabilidade p . Esse modelo de opinião é motivado pelo seguinte:

(i) eleitores indecisos são passivos, no sentido de que eles não espalham suas opiniões para outros eleitores;

(ii) eleitores indecisos são facilmente convencidos por eleitores que já têm opinião formada;

(iii) a flexibilidade de mudar de opinião, quantificada por um parâmetro p , é a mesma para todos os eleitores.

As simulações do modelo nas cinco redes consideradas geraram os seguintes resultados:

- O resultado para a rede de Erdős-Rényi exibe um regime de lei de potência para números intermediários de votos com expoente -1, um platô para pequenos números de votos e um corte para grandes números de votos.

- Para o modelo Barabási-Albert, embora tenha apresentado dois regimes de lei de potências, nenhum deles corresponde ao valor experimental -1. A cauda da curva segue uma lei de potência com inclinação -1,45.

- Na malha bidimensional simples, não existe sinal de um regime de lei de potência e um pico em torno de 1000 votos pode ser notado, em desacordo com a natureza livre de escala.

- O efeito da adição de ligações aleatórias na malha bidimensional simples mostrou que a adição de apenas um pequeno número de novas ligações é o suficiente para obter um resultado semelhante ao do modelo de Erdős-Rényi.

- Para verificar se o expoente da lei de potência da distribuição de candidatos com um dado número de votos é influenciada pela inclinação da distribuição de grau para redes com leis de potência, usou-se o modelo de Barabási-Albert generalizado (veja o último parágrafo do Apêndice B.4.3). No modelo Barabási-Albert original, o expoente da distribuição de grau $P \sim k^{-\gamma}$ é $\gamma = 3$. Para simulação do modelo de Barabási-Albert generalizado foram escolhidos dois conjuntos de parâmetros, para os quais os expoentes obtidos foram $\gamma = 2$ e $\gamma = 2,5$. Os resultados mostraram que a inclinação da lei de potência da distribuição de votos é pouco afetada pela inclinação da distribuição de grau. De fato, ajustando para uma lei de potência na região entre 100 e 200.000 votos, obtém-se para o Barabási-Albert original ($\gamma = 3$) um expoente -1,45; para o modelo generalizado ($\gamma = 2,5$) um expoente -1,44 e ($\gamma = 2$) um expoente -1,40.

Uma comparação do modelo na rede de Erdős-Rényi com resultados de eleições reais é mostrado, na Figura 2.6, a qual apresenta a distribuição do número de votos para eleições brasileiras em 1998 para deputado estadual no estado de São Paulo. De uma maneira geral, segundo [37], efeitos de “mundo pequeno” parecem ser de importância central para modelar o comportamento lei de potência dos dados reais. O modelo aplicado à rede de Erdős-Rényi foi o que apresentou o melhor ajuste aos dados reais. Para os modelos de Barabási-Albert e da malha sem a adição aleatória de trajetórias curtas os resultados são inconsistentes com os de eleições reais como a de São Paulo.

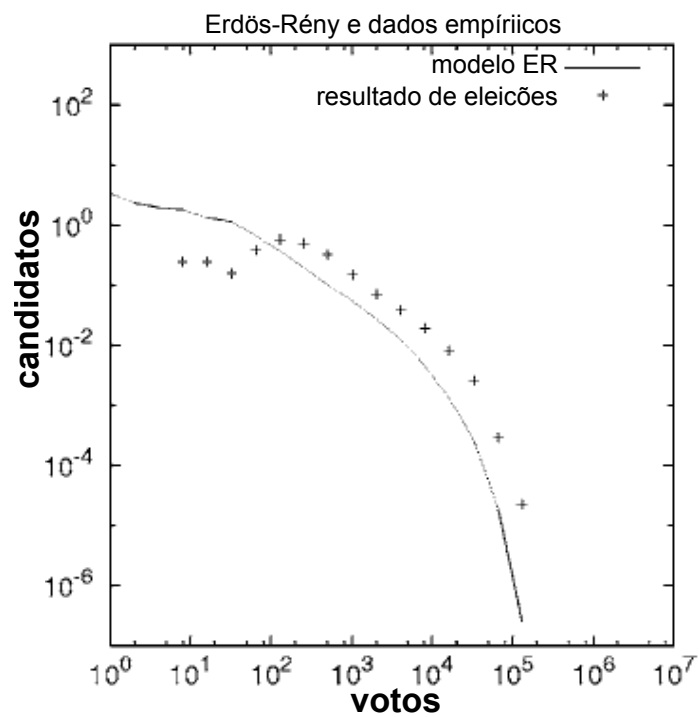


Figura 2.6: **Comparação do modelo Erdős-Rényi com resultados de eleições reais**
 - Dados referentes a eleições no estado de São Paulo em 1998, com 23.321.034 eleitores.
 Adaptada de Travieso *et al.* [37].

Capítulo 3

Escala e universalidade em eleições proporcionais

No tocante às eleições proporcionais, todos os resultados apresentados, até o momento, referem-se a distribuições do número de candidatos em função do número de votos. A seguir, serão abordados resultados que consideram, em eleições proporcionais, a relevância dos partidos. Nesse sentido, o número de votos é reescalado de acordo com a “força” do partido a que o candidato pertence. Inicialmente, são considerados três países nos quais a distribuição é aproximadamente lognormal, o que sugere uma universalidade. Entretanto, uma mesma análise, realizada posteriormente, mostra diferentes distribuições para resultados de eleições brasileiras e finlandesas. Com o intuito de investigar mais profundamente esses dois resultados foi usada uma base de dados de eleições de vários países em diferentes anos, levando em conta dois sistemas eleitorais diferentes: lista abertas e semiabertas.

3.1 Relevância dos partidos

Para as discussões que seguem sobre eleições proporcionais, considera-se:

- ν_i o número de votos do candidato i ;
- l_i a lista de candidatos do partido ao qual i pertence;
- Q_i o número de candidatos da lista l_i ;
- N_{l_i} o número total de votos recebidos pelos Q_i candidatos.

Para dados da Itália (1958; 1972; 1987), Polônia (2005) e Finlândia (2003), considerou-se que a distribuição de probabilidade do número de votos recebidos por candidato $P(\nu; Q, N)$ pode ser reescrita como uma função de uma única variável reescalada, $F(\nu Q/N)$.

Uma comparação entre as funções de escala F para os cinco conjuntos de dados mostra um resultado de universalidade ainda mais notável, a função $F(\nu Q/N)$ é bem reproduzida pela função lognormal [38]

$$F(\nu Q/N) = \frac{N}{\sqrt{2\pi} \sigma \nu Q} e^{-(\log(\nu Q/N) - \mu)^2 / 2\sigma^2}, \quad (3.1)$$

com $\mu = 0,54$ e $\sigma^2 = -2\mu = 1,08$.

No entanto, o caráter universal desse resultado foi contestado [39], visto que dados de eleições no Brasil e na Finlândia apontam outro comportamento. Usando o mesmo procedimento de [38], para três eleições brasileiras o melhor ajuste para o dados é a função exponencial

$$F(\nu/\nu_0) = \gamma e^{-\beta\nu/\nu_0}, \quad (3.2)$$

em que ν_0 é o número médio de votos dos candidatos de cada partido, $\gamma = 1,86$ e $\beta = 1,262$. No caso da Finlândia, a distribuição também é ajustada por uma lognormal como em [38], mas com parâmetros diferentes.

3.2 Procura por classes de universalidade

Na busca por uma escala que conduza à classes de universalidade das distribuições de probabilidade, três diferentes escalas foram usadas para investigar resultados de eleições para uma ampla base de dados [40]. Tal base de dados contém resultados de eleições parlamentares, com sistemas proporcionais, de 15 países: Itália (antes de 1992), Polônia, Finlândia, Dinamarca, Estônia, Suécia, Bélgica, Suíça, Eslovênia, República Checa, Grécia, Eslováquia, Holanda, Uruguai e Brasil, com listas abertas ou semiabertas (Apêndice A.1). Os resultados mostraram que a escala que leva em conta o desempenho dos candidatos dentro de seu respectivo partido e listas abertas são importantes para indicar classes de universalidade.

Na discussão a seguir, D^i é o distrito do candidato i , N_{D^i} o número total de votos no distrito D^i e N_T número total de votos em uma eleição no país inteiro. Além disso, usam-se as seguintes escalas:

- A escala de Castellano e Fortunato [38], na qual o número de votos ν_i de um candidato é dividido pelo número médio de votos $\nu_0 = N_{D^i}/Q_i$ de todos candidatos na lista do seu partido, chamada de escala FC.
- A escala de Costa Filho, Almeida, Andrade e Moreira (CAAM), que considera a fração de votos recebida por um candidato. Aqui, serão consideradas duas

normalizações:

- (a) a fração de votos no distrito, ν_i/N_{D^i} (denotada por CAAMd);
- (b) a fração do total de votos no país ν_i/N_T (denotada por CAAMn).

Analisando o desempenho dos candidatos em sistemas com listas abertas, para a escala FC, observa-se que a universalidade encontrada para Finlândia, Polônia e Itália em vários anos [38], se estende para a Dinamarca e Estônia, conforme pode ser visto na Figura 3.1. Como já visto, essa curva universal é bem representada por uma função lognormal. Outros países usando listas abertas são Eslovênia, Grécia, Suíça, Brasil e Uruguai, cujos resultados empregando a escala FC também são ilustrados na Figura 3.1. Enquanto existe um histórico persistente da distribuição a nível nacional, as curvas não seguem um padrão comum através dos países. Distinguem-se aqui duas classes de comportamento: Eslovênia, Grécia e Suíça são caracterizadas por um nítido pico em $\nu/\nu_0 = 1$ e suas caudas se assemelham bastante. Brasil e Uruguai apresentam um padrão monotônico, bem diferente das outras três curvas. A curva brasileira segue muito próxima do perfil da curva universal, da Figura 3.1 (f), sobre a cauda $\nu/\nu_0 > 1$. Conclui-se que sistemas de listas abertas não garantem distribuições idênticas, mas podem ser agrupados em classes de comportamentos semelhantes.

A Figura 3.2 mostra uma sequência de gráficos análoga à da Figura 3.1 usando a escala CAAMd. Como visto no capítulo 1, a parte central da distribuição de votos brasileira segue uma lei de potência, com expoente próximo de -1; leis de potências ajustam a região central das distribuições de outros países, fornecendo expoentes diferentes uns dos outros.

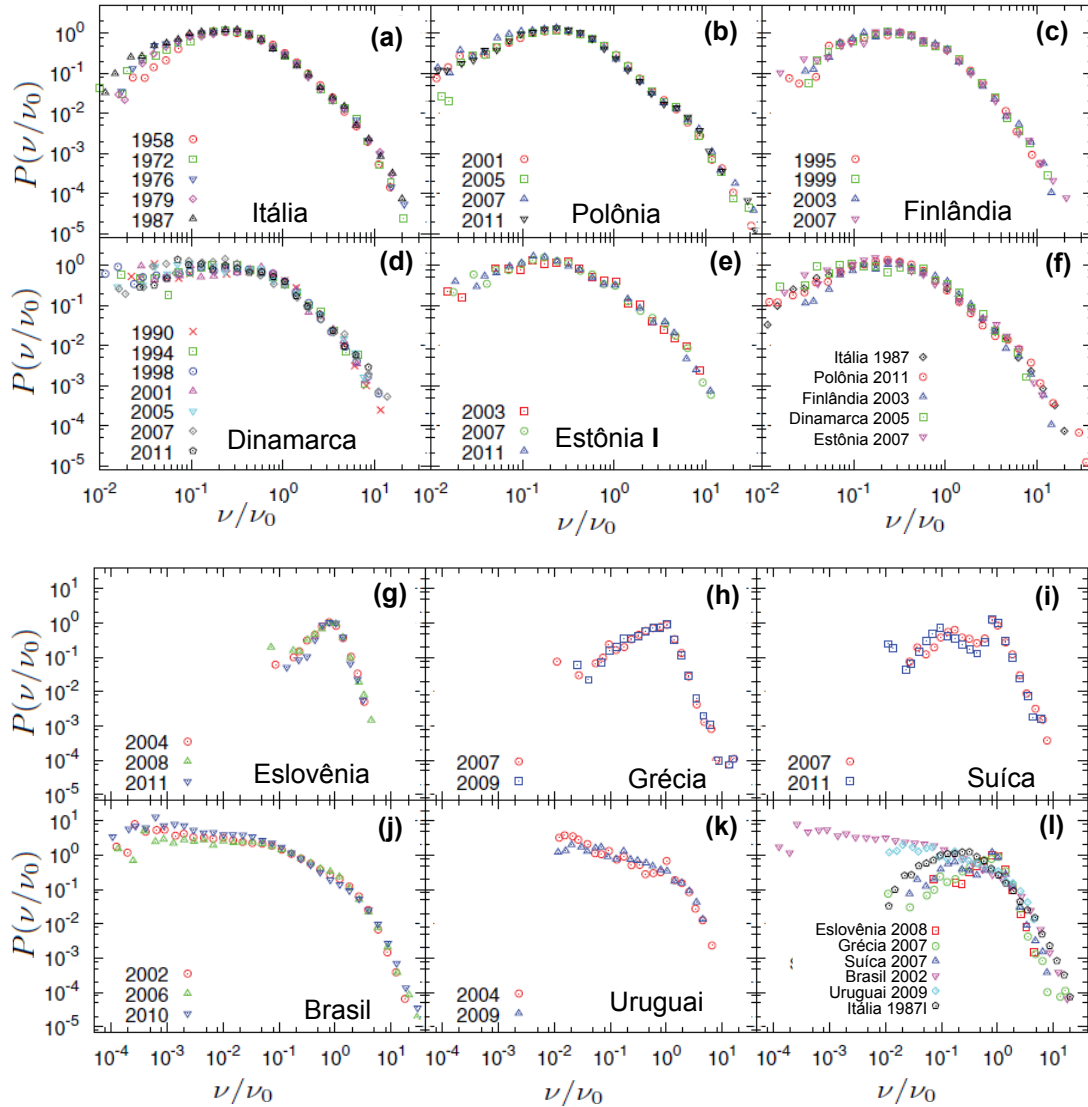


Figura 3.1: **Escala FC e listas abertas** - Distribuição do desempenho eleitoral de candidatos em eleições proporcionais com listas abertas usando a escala FC. (a) Itália (até 1992), (b) Polônia, (c) Finlândia, (d) Dinamarca, (e) Estônia (depois de 2002) e o colapso dos dados para esses países no painel (f). (g) Eslovênia, (h) Grécia, (i) Suíça, (j) Brasil e (k) Uruguai, com tentativa de colapso dos dados para esses países no painel (l). Adaptada de Chatterjee *et al.* [40].

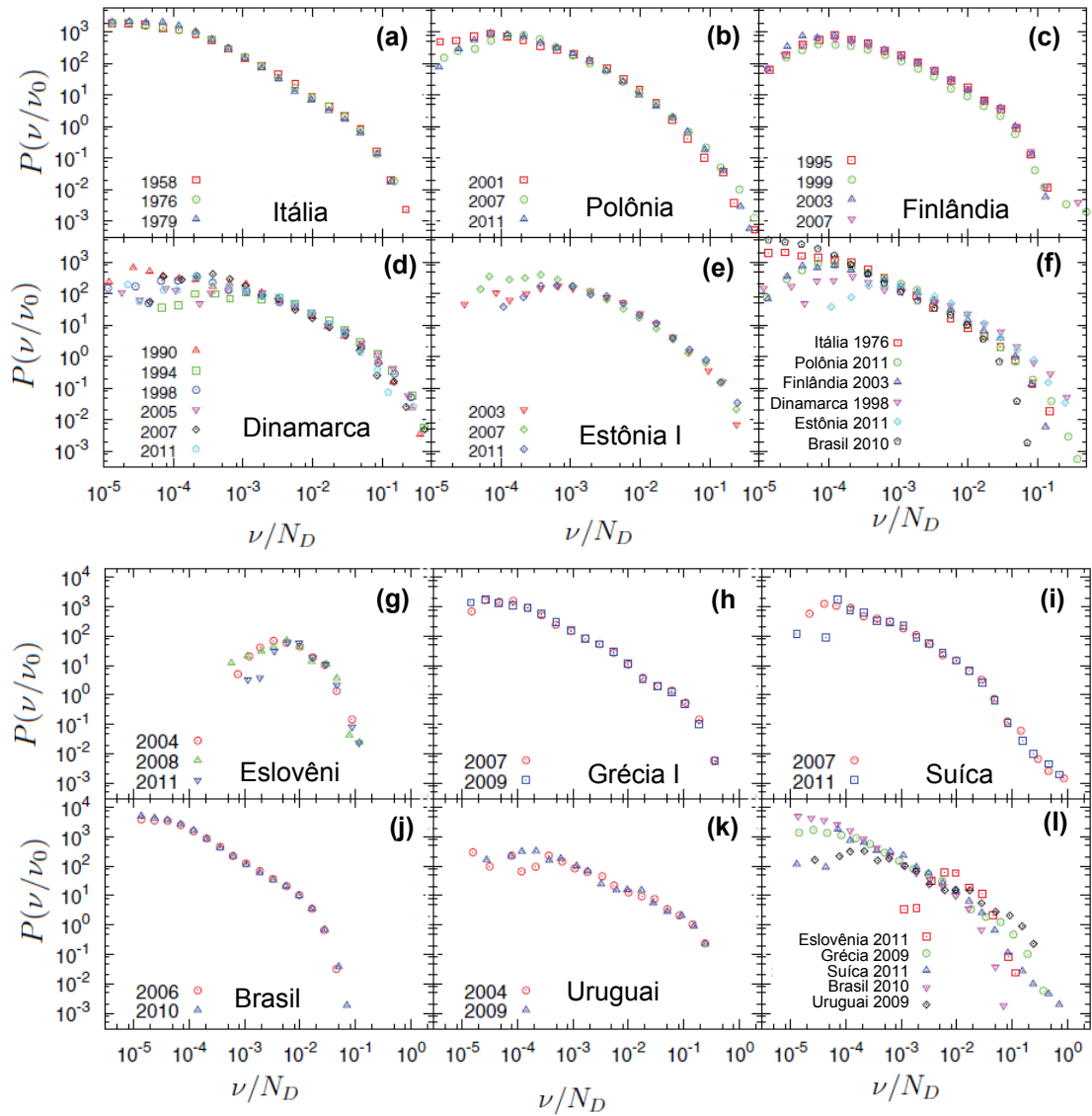


Figura 3.2: **Escala CAAMd e listas abertas** - Análise para os mesmos países da Figura 3.1, com escala a CAAMd. As curvas são estáveis a nível nacional, mas elas não se assemelham entre todos os países. Adaptada de Chatterjee *et al.* [40].

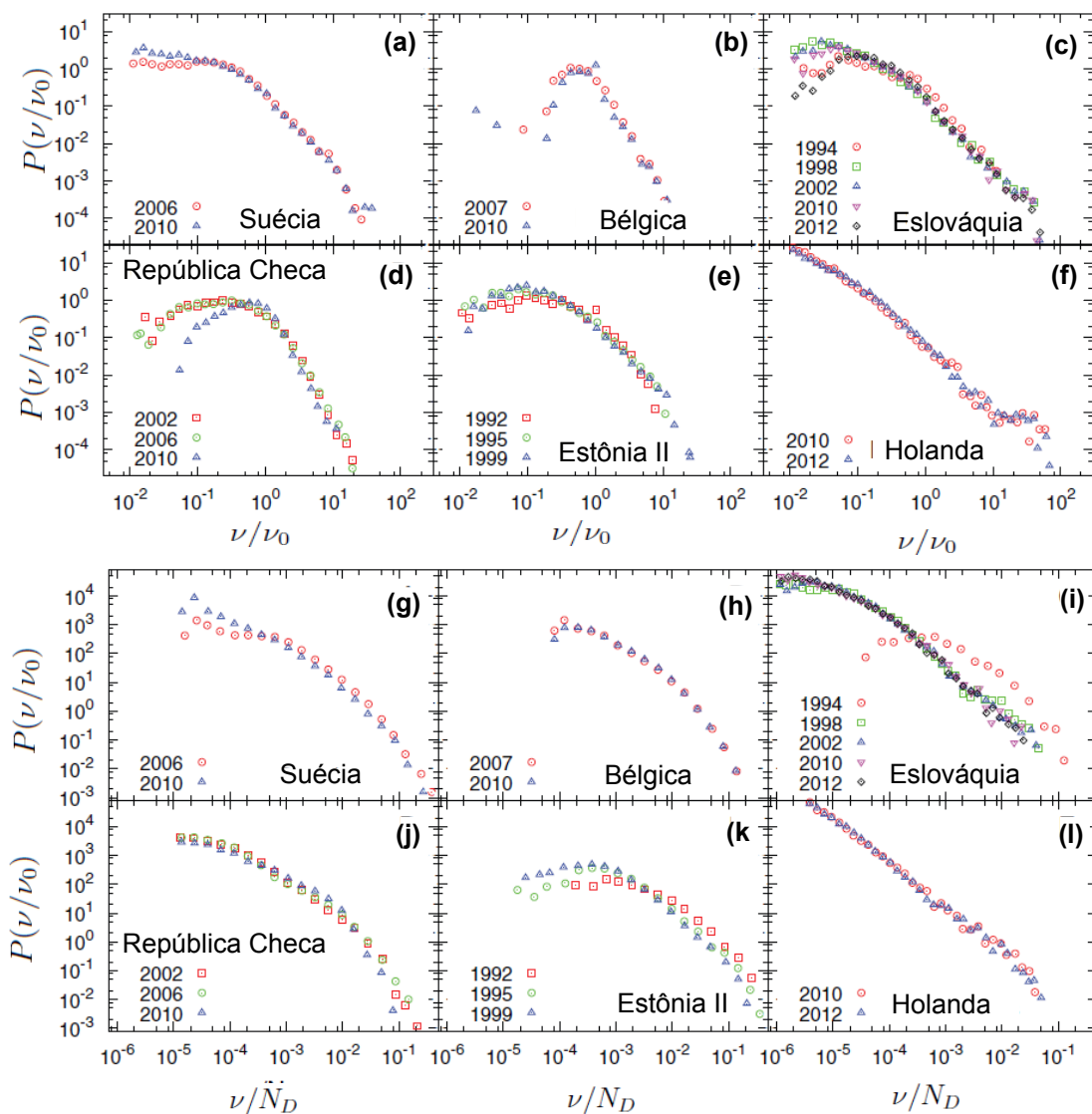


Figura 3.3: **Escala FC e listas semiabertas** - Distribuição do desempenho eleitoral de candidatos em eleições proporcionais com listas semiabertas, usando a escala FC. (a) Suécia, (b) Bélgica, (c) Eslováquia, (d) República Checa, (e) Estônia e (f) Holanda. Em (g)-(l), é mostrado, a distribuição do desempenho eleitoral com escala CAAMd para os mesmos países em (a)-(f). Adaptada de Chatterjee *et al.* [40].

Conclui-se que a fração de votos ν/N_D recebidos por um candidato em seu distrito

eleitoral não segue a mesma distribuição de probabilidade em diferentes países, ou melhor, não há classes de universalidade, como no caso da escala FC. A Figura 3.3 mostra os resultados para listas semiabertas.

3.3 Similaridade entre distribuições

Visando a estimar, quantitativamente, similaridades entre as distribuições, duas matrizes foram construídas, cujas entradas são os valores médios das distâncias D_{avg} e os máximos das D_{max} entre as distribuições para todos os pares de países para os quais têm-se os dados eleitorais. A distância usada foi a de Kolmogorov-Smirnov (K-S) [42], definida como $D = \max_x |S_{N_1}(x) - S_{N_2}(x)|$, em que S_{N_1} e S_{N_2} são as distribuições acumuladas para os dois conjuntos de dados N_1 e N_2 . Os valores das similaridades para eleições no mesmo país são encontrados na diagonal das matrizes. Por serem adotados três diferentes tipos de escalas para o desempenho eleitoral dos candidatos, FC, CAAMd e CAAMn, são obtidas seis matrizes, ilustradas na Figura 3.4. Foram construídas matrizes 16×16 , apesar de ter 15 países no banco de dados, as eleições da Estônia são consideradas separadamente em dois conjuntos, listas semiabertas (Ee II) e listas abertas (Ee I).

O colapso dos dados para os países do grupo da Figura 3.1 é ilustrado em (a) e (b) da Figura 3.4. É interessante verificar que as eleições da Estônia (Ee I), somente depois de 2002, são similares às outras curvas do grupo da Figura 3.1; antes de 2002, a Estônia usava listas semiabertas. As curvas correspondentes não são similares à curva universal. As distribuições brasileiras e uruguaias são bastante similares,

principalmente na cauda, embora difiram consideravelmente na parte inicial. A forte similaridade entre os resultados de eleições das nações do grupo da Figura 3.1 persiste mesmo quando se considera o máximo das distâncias. Observando as matrizes para a escala CAAMd (painéis (c) e (d)) o cenário é pior, as curvas muito menos similares do que as obtidas com a escala FC. Para a escala CAAMn (painéis (e) e (f)), as curvas são ainda mais diferentes. Essas comparações indicam que a escala FC é mais conveniente quando se tem como objetivo obter um comportamento mais próximo do universal para as distribuições de votos dos países.

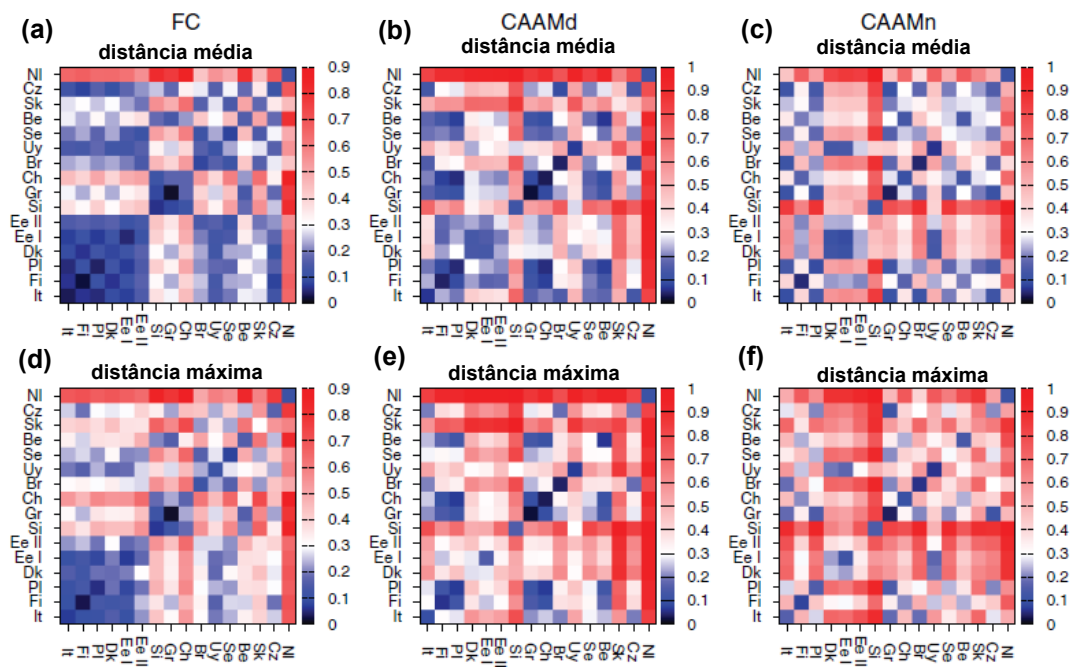


Figura 3.4: **Avaliação quantitativa da semelhança entre as distribuições em nível nacional e entre países.** As matrizes na linha superior indicam os valores da distância média K-S entre os conjuntos de dados de diferentes países. Na linha de abaixo, estão relacionadas as distâncias máximas. Cada coluna corresponde a um dos três tipos de distribuições examinadas, usando FC, CAAMd e CAAMn. Um código de cor é adoptado para distinguir melhor os baixos valores das distâncias (indicada por cores escuras). Adaptada de Chatterjee *et al.* [40].

Capítulo 4

Comparecimento às urnas

Além dos aspectos já discutidos nos capítulos anteriores, outro que chama a atenção na democracia é o comparecimento às urnas. Neste capítulo, discute-se o decaimento logarítmico das correlações espaciais da taxa de comparecimento e da fração de votos vencedores. Analisam-se ainda as frações de abstenções, brancos e nulos e votos válidos via entropia relacionada a essas três razões. Foi também considerado um modelo capaz de detectar e quantificar mecanismos de fraudes em eleições.

4.1 Decaimento logarítmico da taxa de correlação

Um estudo sobre correlações espaciais [41] investigou dados de eleições francesas de 1992 a 2010, com 36.000 resultados locais correspondendo às comunidades (cidades) com populações de tamanhos variados. Para cada eleitor i , definiu-se $S_i = 1$ se o eleitor compareceu para votar (ou votou no partido, ou no candidato vencedor) e

$S_i = 0$ corresponde ao não comparecimento às urnas (ou não votou no partido, ou não votou no candidato vencedor). Para uma cidade α , o número total de eleitores é N_α , o número de votos é $V_\alpha = \sum_{i=1}^{N_\alpha} S_i$, a taxa de comparecimento é $p_\alpha = V_\alpha/N_\alpha$ e o número total de votos do vencedor é W_α . Por conveniência, serão usadas as formas logarítmicas $\tau_\alpha = \ln[V_\alpha/(N_\alpha - V_\alpha)] = \ln[p_\alpha/(1 - p_\alpha)]$ e $\rho_\alpha = \ln[W_\alpha/(V_\alpha - W_\alpha)]$ para taxa de comparecimento e de votos vencedores, respectivamente. Embora a taxa de comparecimento média $m = \langle \tau \rangle$ varia substancialmente entre as eleições, os gráficos da Figura 4.1 indicam que a forma da distribuição de $\hat{\tau} = \tau - m$ é aproximadamente estável. Mais do que isso, o teste Kolmogorov-Smirnov permite concluir que a distribuição $P((\tau - \langle \tau \rangle)/\sigma)$ é praticamente a mesma (σ é o desvio padrão), exceto para 2009. Note que a distribuição de τ é claramente não Gaussiana, com significante assimetria positiva e curtose.

Para a análise da correlação espacial das taxas de comparecimento, emprega-se a função

$$C_\tau(r) = \frac{\langle (\tau_\alpha - m_\alpha)(\tau_\beta - m_\beta) \rangle_{|r_{\alpha\beta}=r}}{\langle (\tau_\alpha - m_\alpha)^2 \rangle}, \quad (4.1)$$

A cada comunidade corresponde uma coordenada espacial R_α , sendo $r_{\alpha\beta} = |R_\alpha - R_\beta|$ a distância entre duas comunidades e m_α a média de τ sobre todas as comunidades de tamanho α . Verificou-se, para todas as eleições, que $C_\tau(r)$ é de longo alcance, decaindo logaritmicamente com a distância: $0 < r < L$, $C_\tau(r) = -\lambda^2 \ln(r/L)$ e $C_\tau(r > L) \approx 0$, como pode ser visto na Figura 4.2. A inclinação logarítmica λ^2 depende da eleição e varia por um fator de no máximo 2. O mesmo foi feito para as estatísticas de votos vencedores, porém a distribuição do logaritmo das taxas de votos vencedores $P(\rho)$ é muito menos universal que $P(\tau)$.

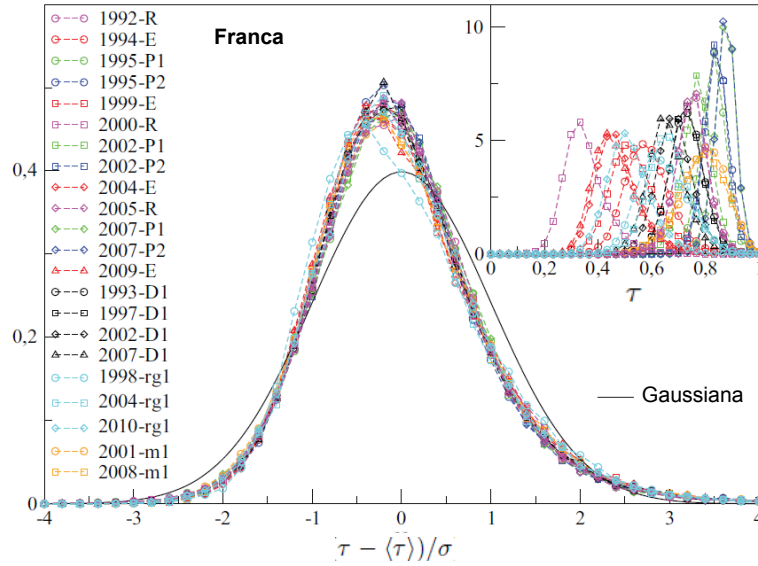


Figura 4.1: **Distribuição da taxa de comparecimento às urnas para eleições francesas** - Distribuição $P((\tau - \langle \tau \rangle)/\sigma)$ sobre comunidades para várias eleições, em que τ é a taxa logarítmica de comparecimento e $\langle \tau \rangle$ é a média sobre todas as comunidades. A inserção mostra as distribuições da taxa de comparecimento p . Adaptada de Borghesi e Christian [43].

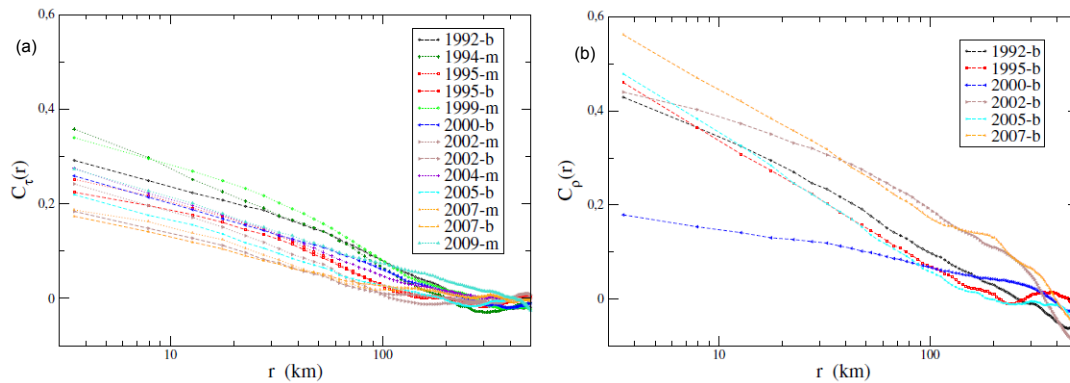


Figura 4.2: **Correlações espaciais das taxas logarítmicas** - Em (a) a correlação da taxa de comparecimento $C_\tau(r)$ e, em (b), a correlação dos votos vencedores $C_\rho(r)$, mostrando a dependência logarítmica em r com um corte $L = 300\text{km}$. A inclinação média logarítmica da taxa de comparecimento é $\lambda^2 = 0,065 \pm 0,01$ e para a taxa de votos vencedores $\lambda^2 = 0,11 \pm 0,02$. Adaptada de Borghesi e Bouchaud [41].

A dependência logarítmica das funções de correlação $C_\tau(r)$ e $C_\rho(r)$ lembra muito o comportamento da função de correlação de um campo de difusão livre em duas dimensões. Mais precisamente, seja $\phi(\mathbf{R}, t)$ obedecendo à seguinte equação estocástica:

$$\frac{\partial \phi(\mathbf{R}, t)}{\partial t} = D\Delta\phi(\mathbf{R}, t) + \eta(\mathbf{R}, t), \quad (4.2)$$

em que Δ é o Laplaciano bidimensional, D é uma constante de difusão e η o ruído de Langevin com média zero, variância σ_η^2 e correlações de curto alcance no tempo e no espaço. Nesse caso, as correlações de ϕ em tempos iguais (em equilíbrio) é dada por [41]:

$$C_\phi(r) = \frac{\langle \phi(r)\phi(0) \rangle}{\langle \phi(0)^2 \rangle} \approx -\Lambda^2 \ln \frac{r}{L}, \quad l_c \ll r \ll L. \quad (4.3)$$

Aqui, l_c é uma escala curta de corte (por exemplo, o comprimento de correlação do ruído η), L é o tamanho linear do sistema e Λ^2 uma constante. O tempo para alcançar o equilíbrio, para o qual o resultado acima é válido, é $T_{eq} = L^2/D$.

Essa semelhança relatada foi usada para modelar o comportamento logarítmico encontrado, relacionando um campo “cultural” que se difunde seguindo uma equação do tipo (4.2).

Esses resultados foram ampliados [43] com uma base de dados bem maior, 77 eleições em 11 países. Uma análise detalhada revela várias características interessantes; por exemplo, diferentes países têm diferentes graus de heterogeneidades locais que parecem ser caracterizadas por diferentes propensões individuais em conformidade com suas normas culturais. Ademais, as decisões são fortemente correlacionadas a níveis individuais em alguns países, mas não em outros.

A base de dados foi composta por várias eleições em diversos países: 22 da França (36.000 municípios da França continental), 13 da Áustria (2.400 municípios), 11 da Polónia (2.500 municípios), 7 da Alemanha (12.000 municípios), enquanto que para outros, o conjunto de dados é menor: 5 do Canadá (7.700 municípios), 4 da Espanha (8.000 municípios da Espanha continental), 4 da Itália (7.200 municípios da Itália continental), 4 da Romênia (3.200 municípios), 3 do México (2.400 municípios), 3 da Suíça (2.700 municípios) e 1 de República Checa (6.200 municípios).

A função de correlação na Figura 4.2 com a equação (4.3) pode ser comparada com as funções de correlação de outros países, representadas na Figura 4.3. Essa comparação indica que o comportamento logarítmico da função de correlação aproxima o de várias eleições de diversos países.

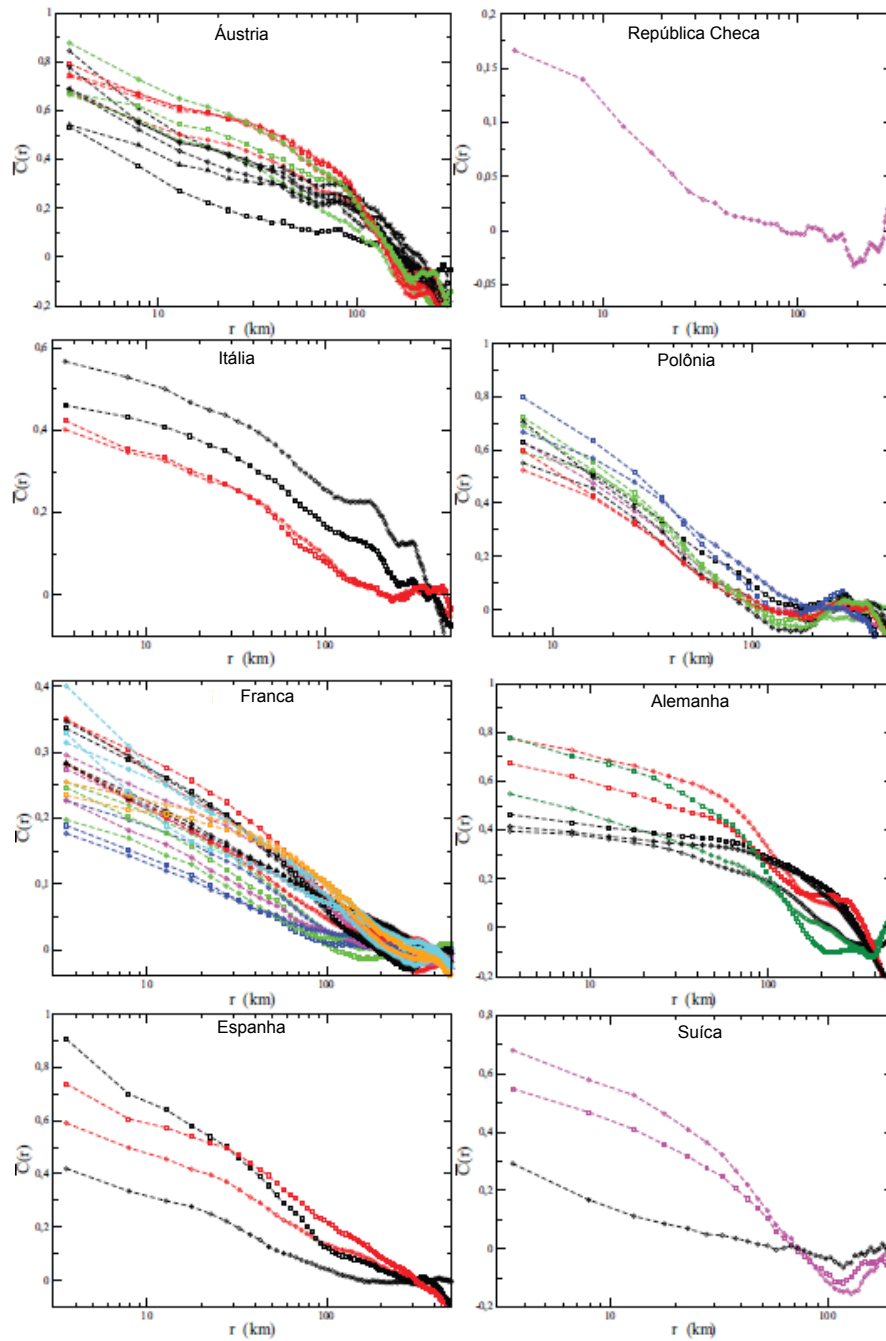


Figura 4.3: **Correlações espaciais normalizadas** - $\tilde{C}(r)$ para todos os países para os quais as posições geográficas das cidades estão disponíveis. A correlação é normalizada pela variância de τ' tal que $\tilde{C}(r=0) \equiv 1$. Adaptada de Christian *et al.* [43].

4.2 Entropia do envolvimento cívico do eleitorado

As discussões anteriores reportam-se a uma variedade de estudos que usam conceitos de Mecânica Estatística, e nada mais emblemático do que usar o conceito de entropia. Direcionado a isso, um estudo das proporções de abstenções, votos brancos e nulos e aqueles de acordo com a lista de opções, em um grande número de eleições nacionais em diferentes países, foi realizado em [44]. Para tal, foram considerados 11 países sem voto compulsório.

Os dados investigados foram agregados em três escalas diferentes:

(1) Em escala municipal. Nessa escala, os fenômenos estudados referem-se ao tamanho da população. As 76 eleições estudadas em nível municipal são recentes, após 1990, e em 11 países diferentes (Alemanha, Áustria, Canadá, República Checa, França, Itália, México, Polônia, Romênia, Espanha e Suíça).

(2) Em grande escala, por exemplo, nacional, provincial, etc. Aqui, a análise é focada na evolução temporal. Os países são estudados por seus aspectos históricos e são os mesmos da escala municipal. O estudo começa no início dos chamados regimes democráticos, após a Segunda Guerra Mundial, e mesmo antes, para alguns casos.

(3) Em nível de estações de votos, sobre as 100 cidades mais populosas, sempre que possível (ou seja, para o Canadá, França, México, Polônia e Romênia).

Para cada eleição e cada escala específica (município, província, país etc.) distinguem-se: o número total N de votos registrados; N_a o número de abstenções, os eleitores que não compareceram para votar; N_{bn} votos brancos ou nulos; N_c votos em favor da lista de escolhas. Com isso, têm-se: $N = N_a + N_{bn} + N_c$. O envolvimento

cívico dos eleitores é então medido por meio das três razões

$$p_a = \frac{N_a}{N}, \quad p_c = \frac{N_c}{N}, \quad p_{bn} = \frac{N_{bn}}{N}, \quad (4.4)$$

representando abstenções, votos válidos e votos brancos e nulos, respectivamente e, portanto, $p_a + p_c + p_{bn} = 1$.

As razões (4.4), relacionaram-se em uma quantidade onipresente em física estatística: a entropia. Nesse contexto, ela é definida por

$$S(p_a, p_c, p_{bn}) = -p_a \log p_a - p_c \log p_c - p_{bn} \log p_{bn}. \quad (4.5)$$

S será chamada entropia do envolvimento cívico ou envolvimento entrópico, e considerada como uma medida da desordem do envolvimento cívico em um nível coletivo. A ordem máxima, $S = 0$, é obtida quando uma das razões é igual a 1 (e então as outras duas são zero). Em contraste, a desordem máxima corresponde a $p_a = p_c = p_{bn} = 1/3$ e a entropia máxima, $S = \log 3 = 1,58$.

A razão p_a de abstenções é tipicamente muito menor para pequenos municípios e cresce com o tamanho do município, N . Espera-se então um crescimento do envolvimento entrópico com o tamanho do município. Isso foi observado para eleições de primeiro turno em 2001 e 2008 na França, como pode ser visto na Figura 4.4. Essa figura identifica também um valor limiar do número de eleitores de uma cidade, 3.500, a partir do qual o envolvimento entrópico torna-se constante. A entropia do envolvimento cívico também foi calculada para outros países, como mostra a Figura 4.5.

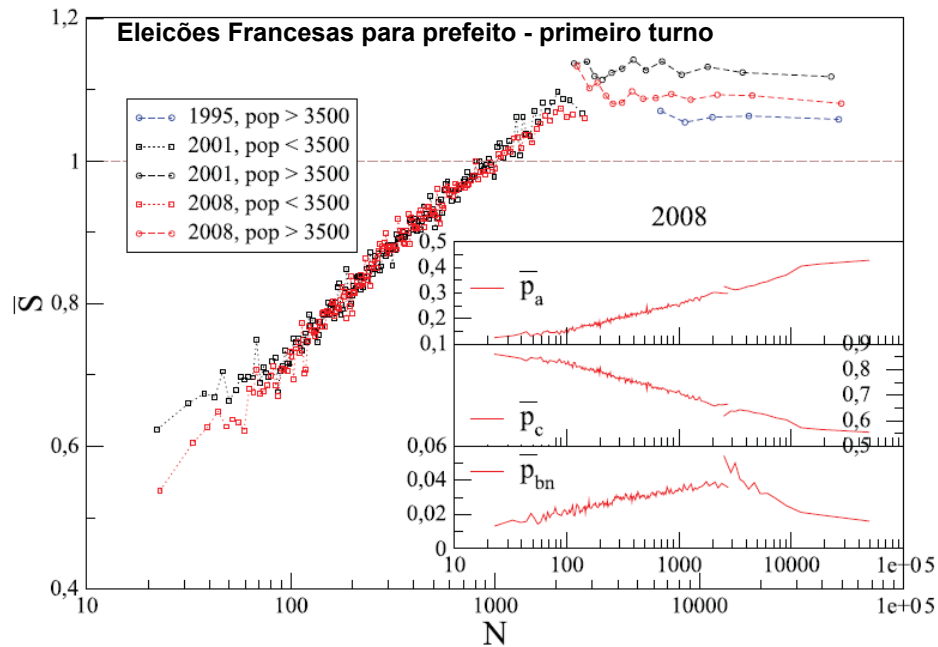


Figura 4.4: **Valores médios do envolvimento entrópico para a França** - S é tomado como uma função do número de eleitores N , para o primeiro turno de eleições para prefeito na França. Existem dois tipos de regras de votação, as quais dependem do tamanho da população, maior ou menor do que 3.500 habitantes. A inserção mostra os valores médios de p_a , p_c e p_{bn} como funções de N para eleições municipais em 2008 (que leva, para municípios altamente populosos, a um platô de S , apesar das variações em p_a , p_c e p_{bn}). Para cada N , os valores médios \bar{S} , \bar{p}_a , \bar{p}_c e \bar{p}_{bn} são calculados sobre ≈ 200 municípios de tamanho $\approx N$. Adaptada de Christian *et al.* [44].

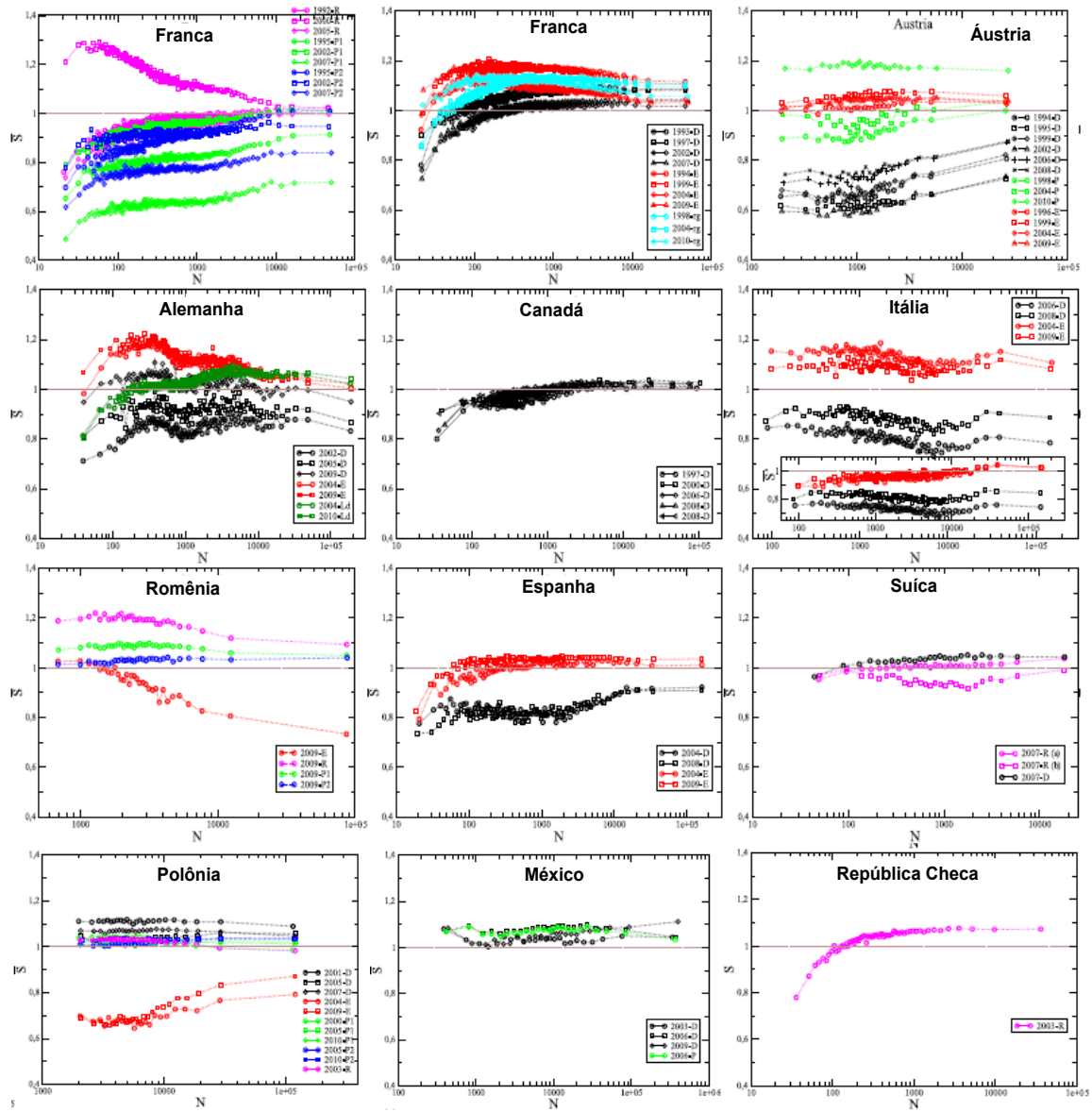


Figura 4.5: Valores médios do envolvimento entrópico de vários países - Valores médios \bar{S} do envolvimento entrópico para municípios como uma função do número de eleitores registrados N . Cada ponto resulta da média sobre uma amostra de ≈ 100 (200 para França) municípios de tamanho $\approx N$. No gráfico da Itália, a inserção mostra um variante de S em que votos brancos são agrupados a votos válidos. Adaptada de Christian *et al.* [44].

Outra análise é realizada agregando-se países da base de dados em dois grupos: um com eleições mais antigas (antes do ano 2000) e outro com eleições mais recentes (depois do ano 2000), com um total de 321 eleições divididas entre os dois grupos com aproximadamente o mesmo número de eleições em cada um. O histograma do envolvimento entrópico S é comparado para esses dois grupos na Figura 4.6. Os histogramas para o grupo das eleições mais recentes mostram um acentuado pico em $S \approx 1$, enquanto o grupo de eleições mais antigas tem ampla distribuição. Há poucas eleições atuais com envolvimento entrópico pequeno (digamos $S \leq 0,8$), porém há muitas eleições com $S \approx 1$.

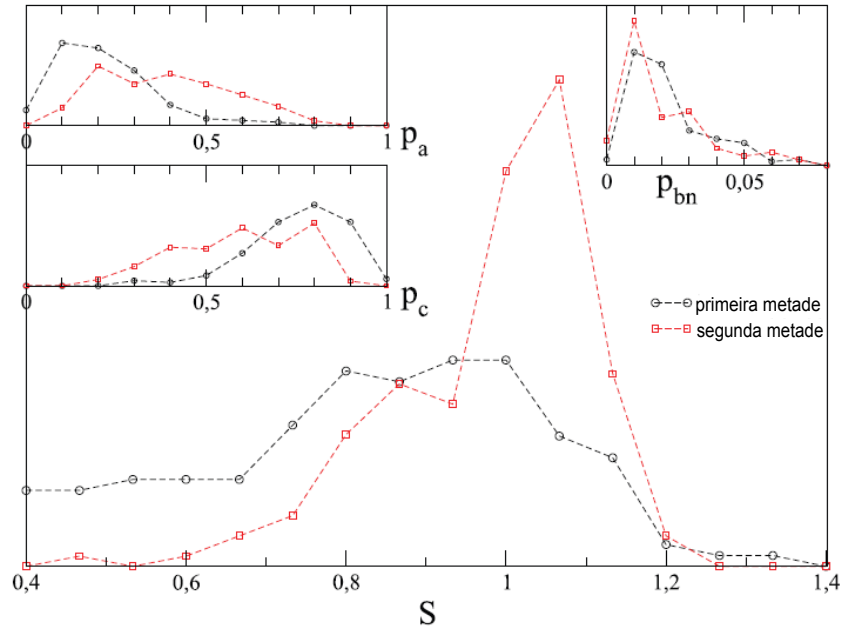


Figura 4.6: **Envolvimento entrópico para dois grupos de eleições** - Histogramas do envolvimento entrópico S em grande escala (nacional, provincial etc.) de 321 eleições, para dois grupos de eleições: primeira metade (antes do ano 2000) e segunda metade (depois do ano 2000). As inserções mostram os histogramas de p_a , p_c e p_{bn} para os dois grupos de eleições. Adaptada de Christian *et al.* [44].

4.3 Detecção de fraude em eleições

Olhar para a taxa de comparecimento às urnas pode ser muito útil na busca de irregularidades em eleições. O que se espera de uma eleição justa é que a distribuição da quantidade relativa de votos tenha uma distribuição “homogênea” nos locais de votação. Ou seja, não é esperado que a quantidade de votos de dado candidato seja distribuída de forma anômala, grande demais em certas unidades eleitorais. Em um estudo investigando esse tipo de anomalia, observou-se que distribuições de votos em eleições com suposta fraude mostram uma curtose muito superior a de eleições normais, dependendo do nível de agregação dos dados [45]. Um modelo paramétrico identifica a presença e quantifica a intensidade de tais mecanismos fraudulentos. Para isso, foram usados dados de eleições parlamentares na Áustria, Canadá, República Checa, Finlândia, Rússia (em 2011), Suíça e Espanha, eleições parlamentares europeias na Polônia e eleições presidenciais na França, Romênia, Rússia (em 2012) e Uganda. Chamou-se unidade a toda divisão de fronteira administrativa (distritos, municípios, províncias, etc.).

Para analisar a distribuição nas unidades em função das variáveis taxa de comparecimento às urnas e número de votos do partido vencedor, a primeira variável foi reescalada como na seção 4.1. De forma análoga, para reescalar a segunda variável, foram considerados: V_i o número de votos válidos, N_i o número de eleitores e W_i o número de votos do partido vencedor (ou candidato) na unidade i . Assim, a segunda variável reescalada foi dada pela taxa logarítmica $\nu_i = \log[(N_i - W_i)/W_i]$. Um histograma bidimensional, para a distribuição do número de unidades com uma dada taxa comparecimento às urnas e com um dado número de votos para o partido

vencedor, chamado de impressões digitais eleitorais é mostrado na Figura 4.7.

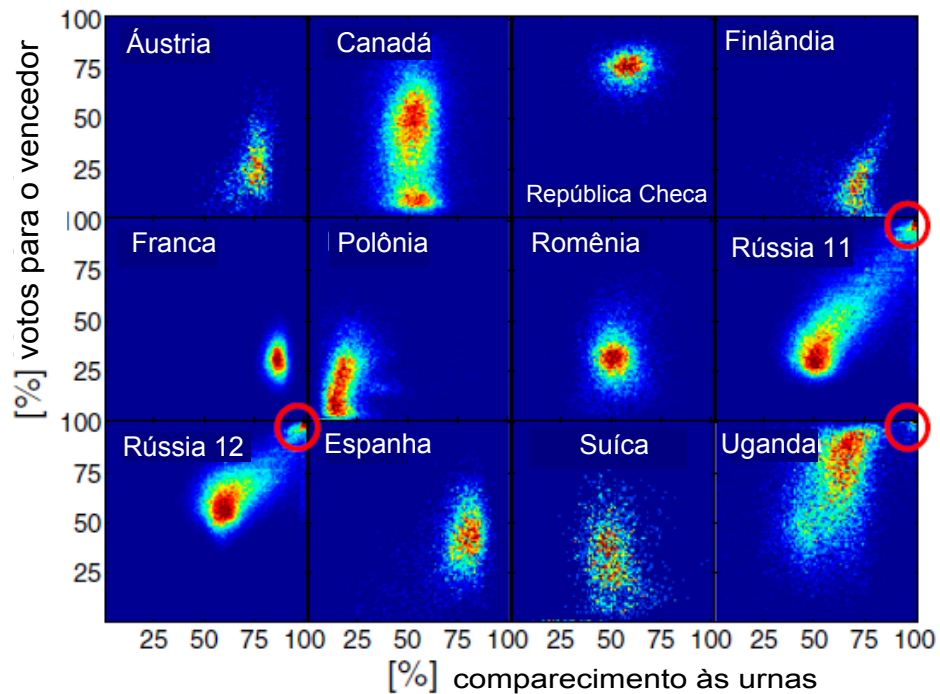


Figura 4.7: **Impressões digitais eleitorais** - Histogramas bidimensionais do número de unidades para uma dada taxa de comparecimento às urnas (eixo x) e a porcentagem de votos para o partido vencedor (ou candidato) (eixo y) em eleições recentes para diferentes países (Áustria, Canadá, República Checa, Finlândia, França, Polônia, Romênia, Rússia (em 2011), Rússia (em 2012), Espanha, Suíça e Uganda). As cores representam o número de unidades com correspondente taxa de votos para o partido vencedor e taxa de comparecimento às urnas. Adaptada de Klimek *et al.* [45].

Observe que, para Rússia e Uganda, a distribuição de votos e da taxa de comparecimento são bimodais, em ambos, votos e comparecimento. Um aglomerado maior ocorre para valores intermediários e outro ocorre na vizinha de 100% dos votos e 100% da taxa de retorno. Esse pico é consistente, com dois mecanismos de fraude. Primeiro, fraude incremental, quando em uma dada proporção cédulas para um partido são adicionadas à urna e/ou votos para outro partido são retirados. Essa fraude

ocorre em uma fração f_i das unidades. Segundo, fraude extrema corresponde ao comparecimento completo de eleitores às urnas e quase todos os votos para um único partido. Essa fraude acontece em uma fração f_e das unidades, formando um segundo aglomerado próximo a 100% do comparecimento e dos votos.

Um modelo para reproduzir os padrões observados na Figura 4.7 foi proposto em [45], considerando que em cada unidade a taxa de comparecimento e votos para o partido vencedor (ou candidato) são distribuídos gaussianamente, com média e desvio padrão tomados como os valores que maximizam a semelhança com os dados empíricos. Na simulação do modelo, o primeiro passo foi encontrar as distribuições empíricas V_i/N_i e W_i/N_i para estimar os parâmetros necessários. Os mecanismos de fraude são inseridos no modelo como segue: fraude incremental - com probabilidade f_i , cédulas são retiradas de ambos, não votantes e oposição, e adicionadas ao partido vencedor. O número de cédulas transferidas para o partido vencedor também foi estimado dos dados empíricos. Fraude extrema - com probabilidade f_e , quase todos os votos dos não votantes e da oposição são adicionados ao partido vencedor. A Figura 4.8 mostra e compara os resultados reais com simulações do modelo empregando ou não os mecanismos de fraude.

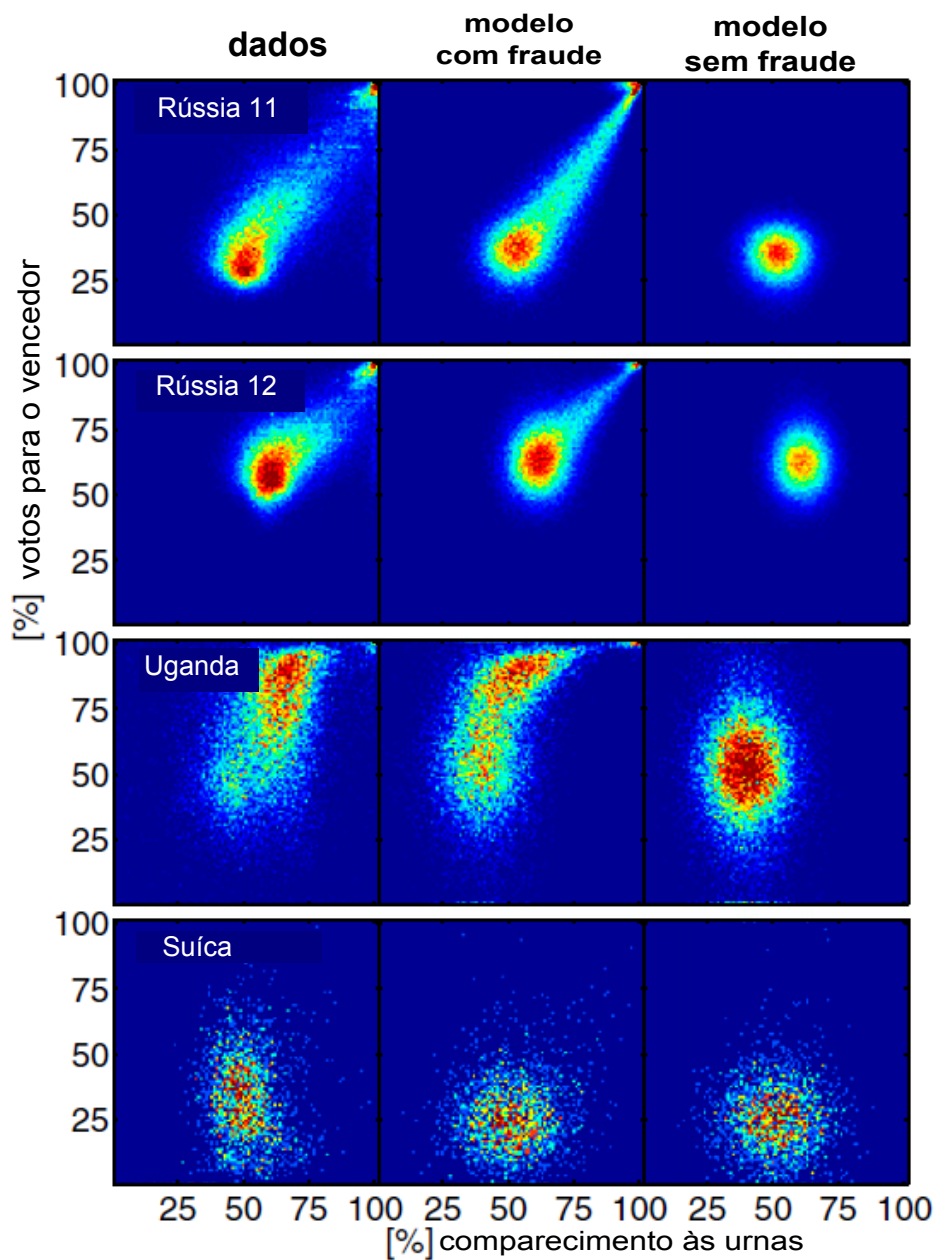


Figura 4.8: **Comparação dos dados empíricos com o modelo** - Comparação do modelo com os resultados empíricos observados para Rússia em 2011 e em 2012, Uganda e Suíça. Do lado esquerdo, estão os dados reais; no centro, a simulação do modelo considerando os mecanismos de fraude, e à direita, o modelo considerando eleições justas (isto é, a ausência do mecanismo de fraude). Adaptada de Klimek *et al.* [45].

Para descrever o aglomerado principal observado para Rússia e Uganda, foi necessário usar $f_i = 0,64$ para Rússia em 2011 e $f_i = 0,39$ para Rússia em 2012 no mecanismo de fraude incremental. Isso significa uma fraude em torno de 64% das unidades em 2011 e de 39% em 2012. O segundo pico no canto superior direito, próximo a 100% de comparecimento, em que existe aproximadamente 3.000 unidades com 100% de votos para Rússia nos dados de 2011, representa um eleitorado de mais de 2 milhões de pessoas. Para o melhor ajuste do modelo, considerou-se $f_e = 0,033$ para 2011 e $f_e = 0,021$ para 2012, isto é, 2 e 3% das unidades eleitorais tiveram fraude extrema.

Capítulo 5

Escolha de candidatos para eleições

Os capítulos anteriores contêm uma revisão de resultados relacionados a distribuições de votos em eleições. Entretanto, um aspecto essencial vinculado a eleições não foi considerado: a escolha de candidatos para eleições. Neste capítulo, apresenta-se uma investigação relativa à escolha de candidatos para eleições [72], encontrando-se um comportamento alométrico (lei de potência) médio entre o número de candidatos e o tamanho do eleitorado (número de eleitores) por cidade. Em torno desse comportamento médio, verificou-se que as flutuações dos dados apresentam uma distribuição lognormal e desvio padrão aproximadamente constante. Portanto, a escolha de candidatos para eleições pode ser interpretada como um processo multiplicativo. Além disso, uma abordagem de rede complexa é empregada para a simulação dos padrões encontrados.

5.1 Apresentação dos dados

Como base de dados para as investigações apresentadas neste capítulo, foram utilizadas 16 eleições [74], Brasil (1996, 2000, 2004, 2008 - eleições para prefeitos e vereadores), Itália (2010 - eleição para prefeito), Inglaterra (2006, 2010 - eleições para vereadores), Canadá (2006, 2008 - eleições parlamentares) e Austrália (2004, 2007, 2010 - eleições parlamentares). Assim, foram 5 eleições para prefeito, 6 eleições para vereadores e 5 eleições nacionais. Os dados referem-se ao número de candidatos e ao número de eleitores em cada divisão política local. No Brasil, eleições gerais são realizadas de quatro em quatro anos e em cada cidade elege-se um representante executivo (prefeito) e um número de representantes do legislativo (vereadores) de acordo com o tamanho da população da cidade. No entanto, para vereadores foram tomadas apenas cidades com até 47.600 habitantes, para as quais o número de vereadores é 9, correspondendo a 90% das cidades brasileiras (~ 5000 cidades). O não uso do restante das cidades não afeta o resultado. Para eleições italianas, foram usadas somente eleições para prefeito (provincial e municipal). No caso de eleições inglesas, os dados são para vereadores em todos os 624 bairros de Londres, onde a maioria dos bairros tem 3 vagas (somente 9 bairros possuem um número menor). Nos casos do Canadá e da Austrália, os dados são das eleições para *House of Commons* e *House of Representatives*, cada uma com 308 e 160 membros, respectivamente. Em ambos os casos, somente um único candidato é eleito em eleições disputadas em cada distrito.

5.2 Análise dos dados

Iniciou-se investigando como o número de candidatos n_c cresce com o número de eleitores N , por município. A Figura 5.1 mostra o gráfico de $\log_{10} n_c$ versus $\log_{10} N$ para eleições de vereadores no Brasil em 2008. Nota-se que uma clara tendência emerge, sugerindo que n_c cresce com N seguindo uma lei de potência, isto é, $n_c \sim N^\alpha$. Relacionado ao possível comportamento linear do gráfico em log-log, tem-se o coeficiente de correlação de Pearson $r = 0,63$. Esse valor de r e o aspecto visual à dispersão dos dados indicam que a correspondente lei de potência é subjetivada pelas flutuações.

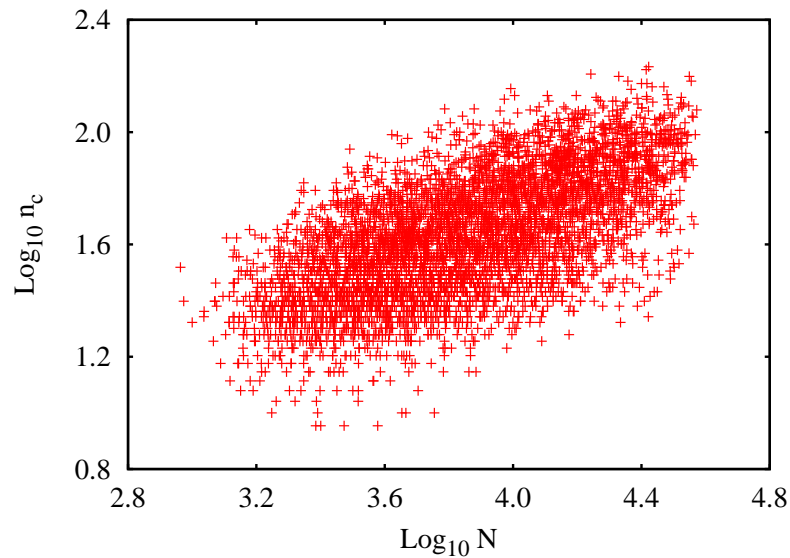


Figura 5.1: **Espalhamento dos dados** - Gráfico do número de candidatos n_c versus o número de eleitores N (cruzes vermelhas), tomados o logaritmo na base-10 para os dados de cada cidade com até 47.600 eleitores para eleições brasileiras de vereadores em 2008.

Uma maneira de superar as flutuações é tomar janelas igualmente espaçadas no eixo da variável $\log_{10} N$. A seguir, correspondente a cada janela, calcula-se os valores médios $\langle \log_{10} N \rangle$ e $\langle \log_{10} n_c \rangle$. Com isso, para cada uma delas, o ponto $(\langle \log_{10} N \rangle, \langle \log_{10} n_c \rangle)$ representa a tendência média dos dados. Esse procedimento é mostrado na Figura 5.2, representado por círculos abertos. Para os valores médios, a relação de lei de potência torna-se evidente, em que

$$\langle \log_{10} n_c \rangle = A + \alpha \langle \log_{10} N \rangle, \quad (5.1)$$

com $\alpha = 0,36$ para essa eleição e $\langle \dots \rangle$ representando o valor médio em cada janela w .

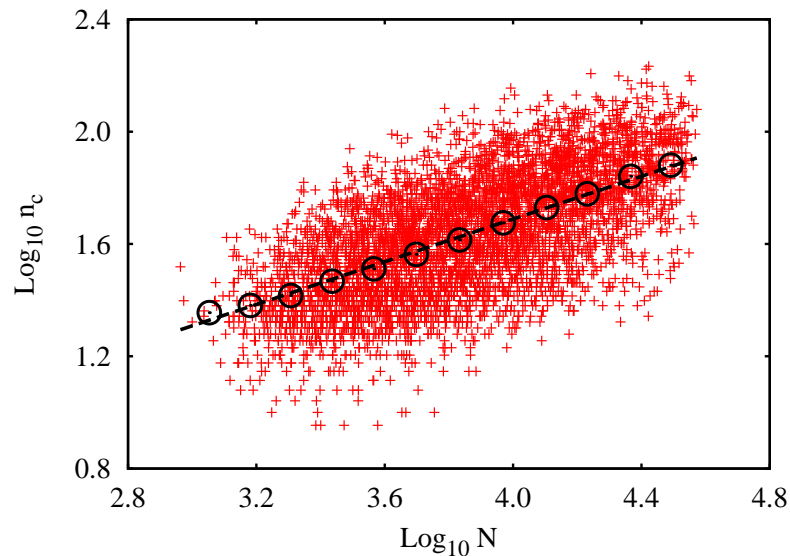


Figura 5.2: **Comportamento médio** - Gráfico mostrando os dados empíricos em log-log (cruzes vermelhas) para eleições brasileiras de vereadores em 2008. Os círculos abertos são os valores médios calculados em janelas logaritmicamente espaçadas e a linha tracejada é o ajuste linear para os valores médios, encontrando-se $n_c \sim N^{0,36}$.

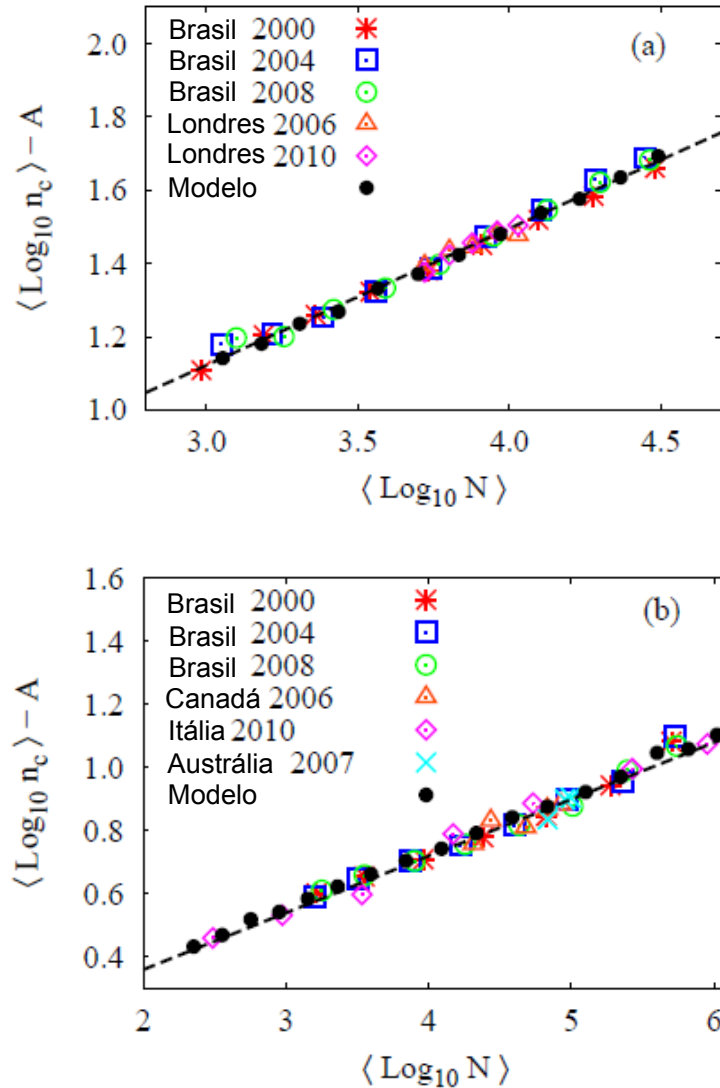


Figura 5.3: **Comportamento médio para vereadores e prefeitos** - Valores médios do logaritmo do número de candidatas n_c versus o logaritmo do número de eleitores N , descontado a constante A , para (a) 5 eleições de multi-membros e (b) 6 eleições de um único-membro. As linhas tracejadas representam leis de potências para os valores médios, cujos expoentes médios α são $\alpha = 0,36$ para eleições de multi-membros e $\alpha = 0,18$ para eleições de um único-membro. Os círculos pretos são os dados simulados com o modelo de rede apresentado ao final do deste capítulo, tomando a rede de Barabási-Albert com $\beta = 0,32$ ($\beta = 0,41$) para multi-membros (único-membro); os outros símbolos representam os dados empíricos.

A Figura 5.3(a) mostra os valores médios para 5 eleições de vereador e a Figura 5.3(b) faz o mesmo para 4 eleições de prefeito e 2 eleições parlamentares. Cada um desses gráficos foi construído descontando o valor da constante A , proveniente do ajuste dado pela Equação (5.1), objetivando o colapso dos dados. A boa qualidade do colapso e os valores de α sugerem duas classes de universalidades: uma para eleições de um único-membro, caracterizada por $\alpha = 0,18$, e outra para eleições de vereadores (multi-membros) com $\alpha = 0,36$.

Direciona-se a discussão para a questão das flutuações em torno do comportamento médio, considerando primeiramente a variância do logaritmo do número de candidatos. Para tal, emprega-se o mesmo procedimento, janelas logaritmicamente espaçadas, usados para os valores médios, denotando-se σ_w^2 a variância em cada janela w . Os resultados encontrados indicam basicamente que a variância não depende do número de eleitores N , fato que pode ser observado na Figura 5.4. Para as eleições da Figura 5.3, as variâncias também são aproximadamente constantes, como a mostrada na Figura 5.4 para o caso de candidatos a vereador em 2008.

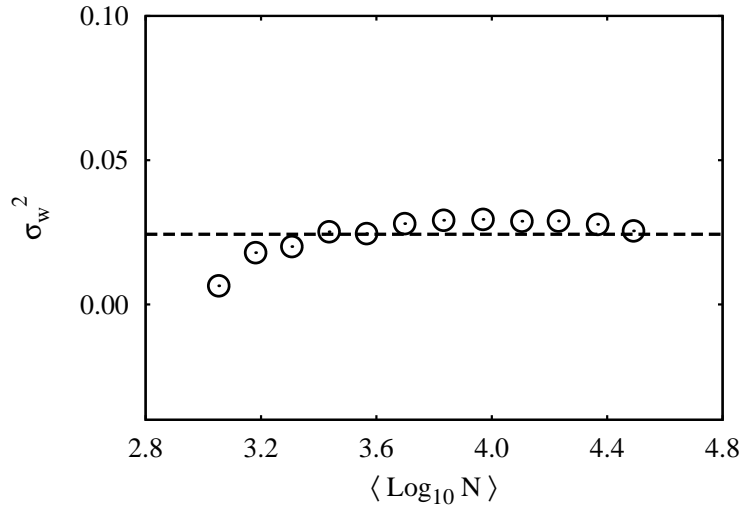


Figura 5.4: **Variância para os dados empíricos** - Gráfico da variância do logaritmo do número de candidatos n_c em cada uma das janelas logaritmicamente espaçadas para eleições brasileiras de vereadores em 2008. A linha tracejada representa o valor médio de σ_w^2 .

De maneira geral, investigações das flutuações em torno da relação média levam em conta a variável

$$\xi = \frac{\log_{10} n_c - f_w(N)}{\sigma_w}, \quad (5.2)$$

em que $f_w(N) = A + \alpha \langle \log_{10} N \rangle_w$ representa a função ajuste para os valores médios considerando cada janela w . A Figura 5.5 mostra as funções distribuições de probabilidades de ξ para as eleições mostradas na Figura 5.3. Observando as Figuras 5.5 (a) e (b), nota-se que a distribuição de probabilidades de ξ é bem ajustada pela Gaussiana padrão. Além disso, igualmente aos comportamentos médios, as distribuições não dependem do número de janelas w .

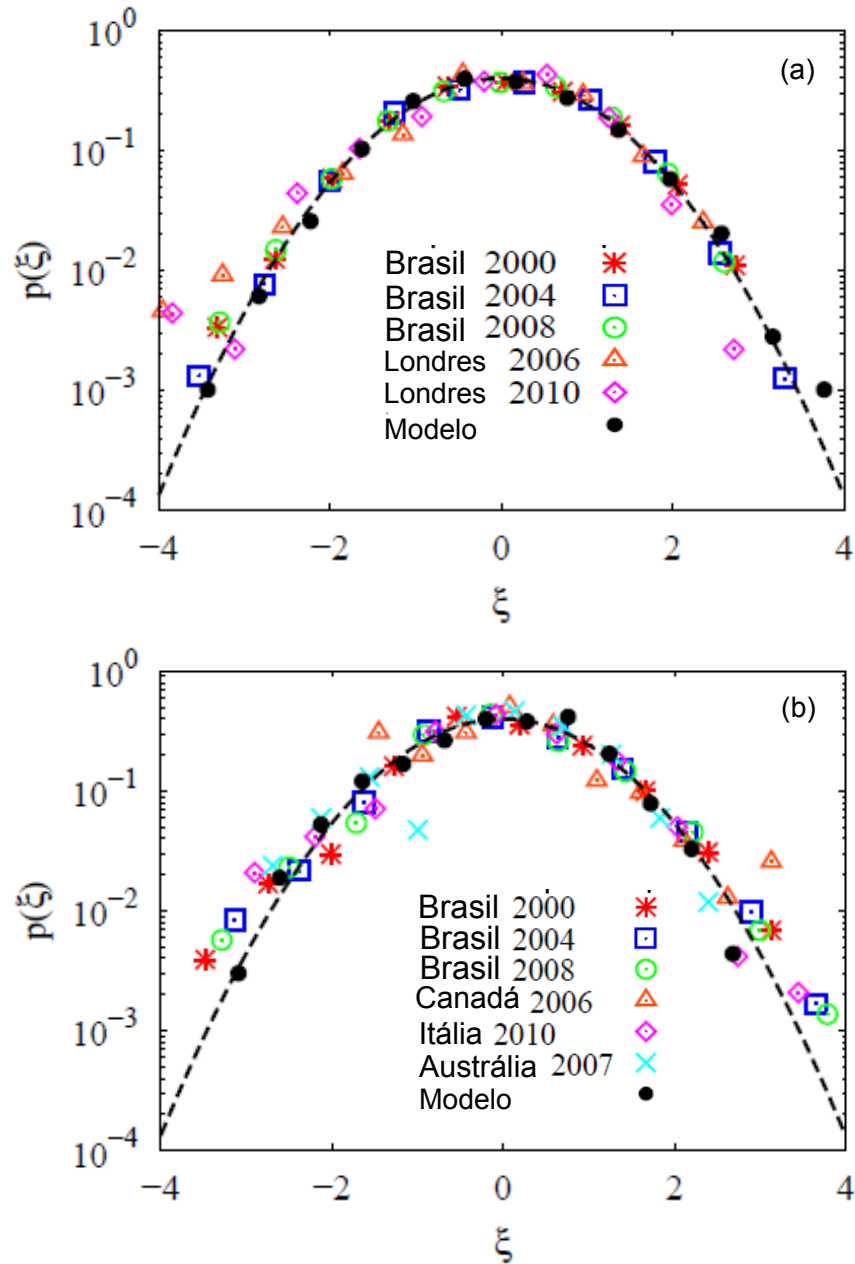


Figura 5.5: **Distribuições das flutuações** - Distribuições das flutuações para as eleições mostradas na Figura 5.3. As linhas tracejadas representam a distribuição Gaussiana padrão. Os círculos pretos são os dados simulados no modelo de rede apresentado no final deste capítulo tomando a rede de Barabási-Albert com $\beta = 0,32$ ($\beta = 0,41$) para multi-membros (único-membro) e os outros símbolos representam os dados empíricos.

5.3 Processo multiplicativo

Supor que as flutuações ξ estão distribuídas gaussianamente equivale a ter uma distribuição lognormal para a variável n_c ; portanto, um processo multiplicativo (Apêndice B.2). Levando em conta a lei de potência média e o processo multiplicativo relacionado às flutuações dos dados, empregou-se a seguinte expressão:

$$n_c = \mathcal{A} \zeta(N) N^\alpha \quad (5.3)$$

ou equivalentemente

$$\log_{10} n_c = \log_{10} \mathcal{A} + \alpha \log_{10} N + \log_{10} \zeta(N). \quad (5.4)$$

Para que ocorra o ajuste dos dados pela Equação (5.3), à Equação (5.4) deve corresponder a Equação (5.2) e conduzir a Equação (5.1), ou seja, deve-se ter $\log_{10} \mathcal{A} = A$ e $\log_{10} \zeta(N) = \sigma_w \xi(N)$ com $\langle \xi(N) \rangle = 0$. Usou-se a variável N para indicar que $\zeta(N)$ e $\xi(N)$ são variáveis tipo estocásticas. Pode-se observar que supor $\xi(N)$ normalmente distribuída e σ_w constante equivale à $\zeta(N)$ ser distribuída de acordo com uma lognormal. Portanto, a Equação (5.3) representa um processo estocástico em que o número de candidatos n_c tem uma relação média lei de potência com N conduzida por um ruído estocástico multiplicativo seguindo uma distribuição lognormal. Naturalmente, simulações usando Equação (5.3) são muito semelhantes aos dados empíricos.

5.4 Modelo de rede

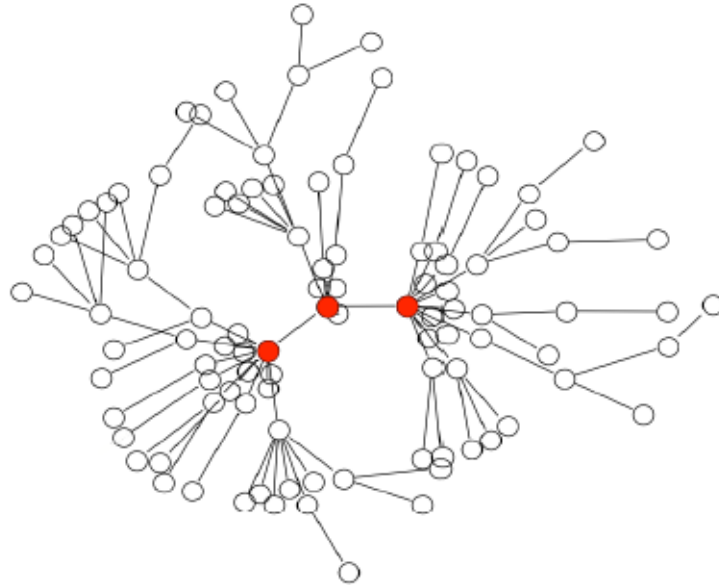


Figura 5.6: **Uma ilustração esquemática do modelo de rede** - Cada nó do grafo representa um eleitor e cada ligação entre dois nós significa que um eleitor conhece o outro de uma maneira não negativa. Um eleitor pondera sua popularidade (aqui, o número de ligações) na comunidade levando-o ou não a candidatura. Nessa rede, os nós destacados (círculos preenchidos de vermelho) são os mais populares e, conseqüentemente, potenciais candidatos.

Uma possibilidade para tentar simular o comportamento encontrado, além de usar a Equação (5.3), é empregar o formalismo de redes complexas. Nesse contexto, supõe-se que os agentes sociais são representados por nós em um grafo e as interações (relações) são expressas por ligações entre eles. Mais especificamente, cada agente social será um eleitor e a ligação entre dois eleitores significará que eles se conhecem. Em uma primeira aproximação, pode-se relacionar o grau (número de ligações partindo de um nó) com um tipo de medida de popularidade. Assim, quanto mais

ligações (maior o grau) um eleitor possui, maior é sua popularidade e, conseqüentemente, maior sua chance de ser candidato. Isso porque, em geral, espera-se que uma pessoa pondere ser candidato se avaliar que é bem conhecido (popular). Por esse motivo, empregou-se o modelo de rede mostrado na Figuras 5.6.

Naturalmente, vários modelos de redes podem ser empregados para simular a dinâmica descrita. No entanto, foi usado apenas um modelo minimal focando somente a distribuição de grau $p(k)$. Diante das considerações feitas, é razoável considerar que um eleitor se torna um candidato se sua popularidade (seu grau) excede um valor característico k_c . Conseqüentemente, o número de candidatos n_c em um dado eleitorado N deve ser tal que

$$n_c \simeq N \int_{k_c}^{\infty} p(k) dk. \quad (5.5)$$

Nesse ponto, é interessante observar que não existe razão para supor que $p(k)$ dependa sensivelmente de N e, assim, a única maneira de n_c se comportar de forma não linear com relação a N é considerar que o valor limiar de k_c dependa de N . Para motivar uma possível dependência de k_c em termos de N , observou-se uma característica notável de muitas redes sociais, que é a do grau distribuição livre de escala. Em outras palavras, a distribuição do número de ligações (a distribuição de grau) é, muitas vezes, bem descrita por uma lei de potência:

$$p(k) \sim k^{-\gamma}. \quad (5.6)$$

Essa característica é geralmente relacionada com a existência de ligações preferenciais

em redes sociais [7] e tem sido encontrada em vários sistemas sociais (tabela II em [65]). Portanto, uma rede livre de escala em que $p(k) \sim k^{-\gamma}$ soa como uma escolha natural para modelar o sistema social. Assim, usa-se as Equações (5.5) e (5.6) para obter

$$n_c \sim N k_c^{-\gamma+1}. \quad (5.7)$$

Comparando a Equação (5.7) com o resultado para os dados empíricos, $n_c \sim N^\alpha$, o valor característico k_c deve depender do número de eleitores N seguindo outra lei de potência:

$$k_c \sim N^\beta, \quad (5.8)$$

com $\beta = \frac{\alpha-1}{1-\gamma}$.

É importante ressaltar que outras opções para $p(k)$ levam em geral a outras dependências funcionais para k_c , geralmente mais complicadas que uma lei de potência. Por exemplo, quando se considera o modelo de Erdős-Rényi (Apêndice B.4.2), tem-se

$$p(k) \simeq \frac{e^{-\langle k \rangle} \langle k \rangle^k}{\Gamma(k+1)} \quad (5.9)$$

e k_c deve ser obtido a partir da equação transcendental

$$\mathcal{A}N^{\alpha-1} = \int_{k_c}^{\infty} \frac{e^{-\langle k \rangle} \langle k \rangle^k}{\Gamma(k+1)} dk \quad (5.10)$$

para a qual empregou-se $n_c = \mathcal{A}N^\alpha$ na Equação (5.5).

Retornando a discussão da Equação (5.8), cabe observar que k_c não deve ser puramente determinístico. Assim, ao invés da Equação (5.8) pode ser empregada a

fórmula

$$k_c = c\varphi_\sigma(N)N^\beta, \quad (5.11)$$

em que c é uma constante positiva e $\varphi_\sigma(N)$ é um número aleatório com média unitária e desvio padrão σ . Uma justificativa quantitativa desse aspecto não determinístico é que candidatos não são sempre pessoas populares, pois não existem muitas restrições para ser um candidato. Para ser mais específico, investigou-se o cenário com a Equação (5.11) empregando a rede livre de escala de Barabási-Albert (Apêndice B.4.2) para a qual $\gamma = 3$ [7, 66].

Para as simulações, construíram-se redes em que o número de nós são iguais aos valores empíricos de N e os valores de β são obtidos diretamente dos valores de α . Os parâmetros c e σ foram ajustados para cada eleição, minimizando as diferenças entre os valores simulados e empíricos das quantidades $\langle n_c \rangle$ e $\langle \sigma_w^2 \rangle$, como ilustra a Figura 5.7. Mostrou-se que σ_w^2 é sempre um número pequeno (menor do que 1) e para eleições de um único membro, esses parâmetros não precisam ser considerados, uma vez que as flutuações intrínsecas ao modelo Barabási-Albert são capazes de reproduzir $\langle \sigma_w^2 \rangle$, Figuras 5.7 e 5.8, e as distribuições de probabilidades das flutuações, Figura 5.5. Esses fatos indicam que o modelo Barabási-Albert pode imitar o ruído multiplicativo existente nos dados.

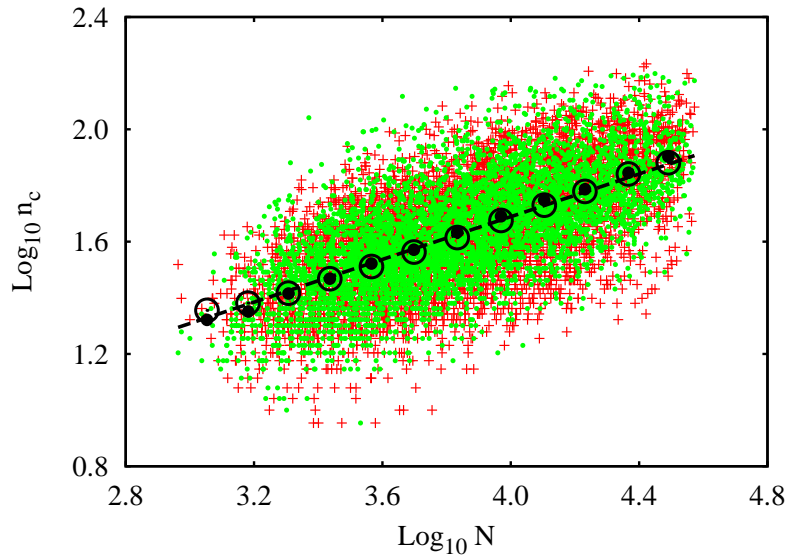


Figura 5.7: **Espalhamento dos dados empíricos e simulados** - Gráfico do número N de eleitores versus o número de candidatas n_c tomados o logaritmo na base-10 para eleições brasileiras de vereador em 2008. As cruzes vermelhas representam os dados empíricos e os pontos verdes são os resultados da simulação obtidos do modelo na rede de Barabási-Albert com $c = 1,12$, $\alpha = 0,196$ e $\beta = 0,32$ (o eleitorado de cada cidade é simulado em uma rede e os dados obtidos são resultantes de uma única simulação). Os círculos abertos (preenchidos) são os valores médios, calculados em janelas logarítmicas, para os dados (resultados simulados) e a linha pontilhada é o ajuste linear para os valores médios dos dados empíricos, encontrando $n_c \sim N^{0,36}$.

É também interessante notar que ambos os parâmetros c e σ não são importantes quando se empregam as variáveis escaladas das Figuras 5.3 e 5.5. Diante dessas figuras, pode-se observar que o modelo reproduz o gráfico do espalhamento muito bem, assim como os valores médios e as distribuições de probabilidades.

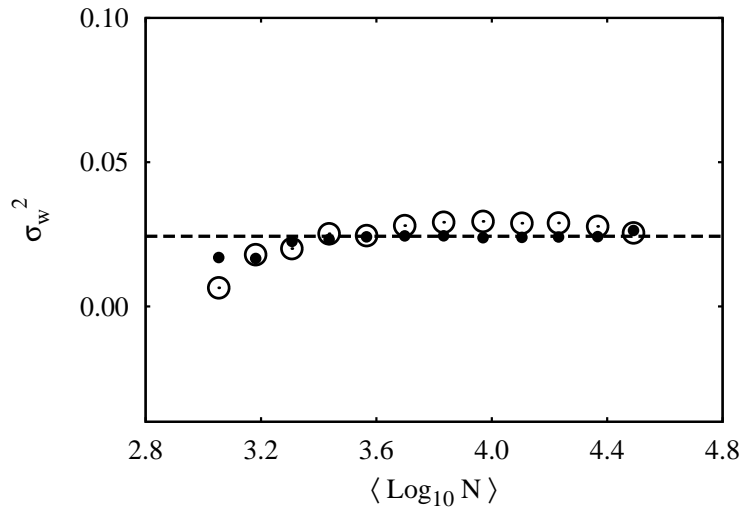


Figura 5.8: **Variância para os dados simulados** - Gráfico da variância para os dados simulados (círculos preenchidos) em rede de Barabási-Albert, comparados aos valores calculados para os mesmos dados empíricos (círculos abertos) da Figura 5.4.

5.5 Discussões

Neste capítulo, foi apresentada uma análise de como o número de candidatos n_c está relacionado com o número de eleitores N de cada distrito eleitoral. Com dados de 16 eleições, mostrou-se que n_c cresce com N seguindo uma relação média lei de potência, $n_c \sim N^\alpha$. O expoente α encontrado foi $\alpha = 0,18$ para eleições de um único membro e $\alpha = 0,36$ para eleições de multi-membros, indicando que no processo de candidatura há efetivamente uma ponderação do número de vagas e da quantidade de eleitores. No entanto, é interessante notar que há três vagas para cada bairro de Londres, enquanto nas eleições brasileiras foram consideradas cidades com 9 vagas para vereadores em cada uma e os valores dos expoentes são aproximadamente o

mesmo. Isso indica que as responsabilidades do cargo público têm também um papel central no processo de candidatura, uma vez que o prefeito ou funções parlamentares estão em um patamar diferente do representante local. Foram também examinadas as flutuações existentes em torno dos valores médios e observado que as variâncias são aproximadamente constantes. Além disso, as flutuações revelam que as candidaturas são dirigidas por um ruído multiplicativo, com uma distribuição de probabilidade lognormal.

Um modelo em uma rede livre de escala foi empregado para representar as relações sociais, reproduzindo muito bem o comportamento encontrado. Os resultados mostraram ser robustos no sentido de que são independentes da situação complexa relacionada com diferenças culturais e sociais entre os países, bem como de atributos psicológicos individuais dos candidatos. Essa robustez é frequentemente encontrada em sistemas fora do equilíbrio ou sistemas prestes à criticalidade, em que poucos mecanismos podem levar à estatística livre de escala e à criticalidade auto-organizada [68]. Em particular, observou-se que medir a “popularidade” (o parâmetro de corte k_c do modelo de rede) pode depender do tamanho do sistema.

Capítulo 6

Ocupação de posições políticas

Neste capítulo, apresenta-se uma análise de dados de filiados a partidos políticos, a qual, juntamente com as investigações abordadas no capítulo anterior, fornece um cenário quantitativo do envolvimento dos eleitores na política partidária, tornando-se um candidato ou um membro de um partido político [73]. O objetivo dessa análise foi tentar compreender como o eleitor pondera sua candidatura a uma posição política. Analisando-se dados das cidades brasileiras, observou-se, assim como nos casos investigados no capítulo 5, uma forma funcional bem descrita por uma lei de potência (alometria) cercada por um ruído multiplicativo com distribuição lognormal. Os expoentes das leis de potências exibem uma clara hierarquia em que a posição mais influente apresenta o menor expoente. Foram também encontradas as distribuições de probabilidades do número de candidatos (prefeito e vereador) de membros de partidos políticos e de eleitores.

6.1 Apresentação dos dados

Os dados para pesquisa foram acessados gratuitamente dos sítios eletrônicos do Tribunal Superior Eleitoral (TSE) brasileiro [74] e do Instituto Brasileiro de Geografia e Estatística (IBGE) [75]. Os dados eleitorais consistem do número de candidatos em eleições para prefeito e vereador em todas as cidades brasileiras nos anos 2000, 2004 e 2008. Para o número de candidatos a vereador, foram consideradas somente cidades com menos de 47.600 habitantes ($\sim 90\%$ das cidades brasileiras), nas quais o número de vagas para vereador é 9. No caso do número de membros de partidos políticos, os dados coletados são referentes aos anos de 2006, 2008 e 2011 para todos os 27 partidos políticos brasileiros em cada cidade brasileira. As populações de eleitores foram coletadas para os anos correspondentes aos dados dos números de candidatos e dos número de membros de partidos políticos. Durante o período dos dados coletados, o número de cidades quase não mudou (5.564 em 2006 e 5.565 em 2011) e o número de partidos políticos permaneceu constante. Seguindo a notação do capítulo anterior, para cada cidade considera-se N o número de eleitores e n_c o número de candidatos (prefeito ou vereador) ou o número de membros de partidos políticos.

6.2 Análise dos dados

Da mesma forma que na investigação apresentada no capítulo 5, primeiro investigou-se a relação entre $\log_{10} n_c$ e $\log_{10} N$. A Figura 6.1 mostra o espalhamento dessas relações quando se considera o número de membros de partidos políticos em (a), o

número de candidatas a vereador em (b) e o número de candidatas a prefeito em (c).

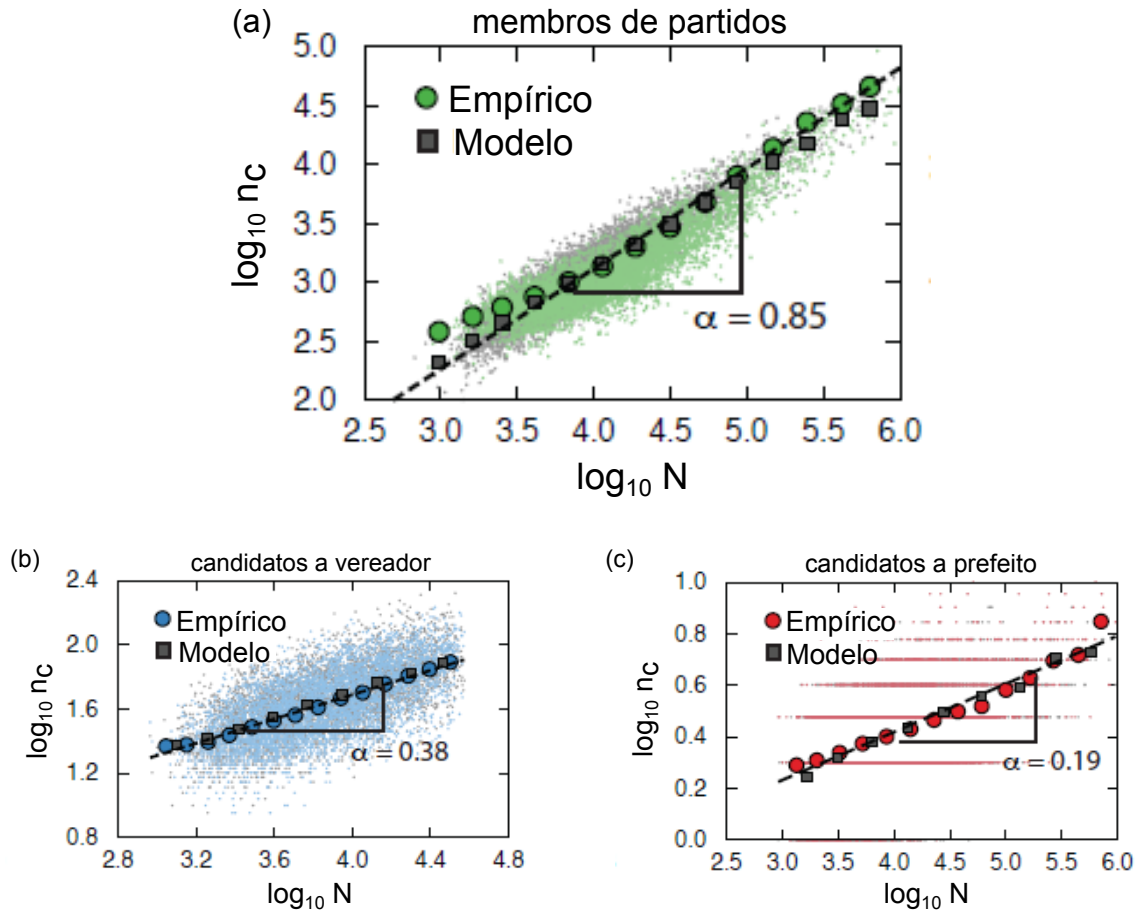


Figura 6.1: **Gráficos dos espalhamentos dos dados** - Relação entre $\log_{10} n_c$ e $\log_{10} N$, N número de eleitores e n_c , o número de membros de partidos políticos em (a), o número de candidatas a vereador em (b) e a prefeito em (c). Os dados correspondem ao ano de 2008. Nota-se em todos os casos que as relações aproximam-se de leis de potências cercadas por flutuações. Os pequenos pontos coloridos são os valores de $\log_{10} n_c$ e $\log_{10} N$ para cada cidade e os círculos são os valores médios desses pontos depois de separar os dados em janelas w igualmente espaçadas em $\log_{10} N$. As linhas tracejadas são os ajustes lineares para os valores médios e os expoentes das leis de potências são mostrados em cada gráfico. Os pontos cinza são os resultados da simulação realizada com a equação (6.3).

Apesar de haver flutuações, é possível observar uma tendência a um regime de lei de potência. Visando a superar as flutuações, isto é, identificar um comportamento médio, os dados $\log_{10} N$ foram separados em w janelas igualmente espaçadas e calculados os valores médios de $\log_{10} n_c$ e $\log_{10} N$ em cada janela. Os círculos na Figura 6.1 representam esses valores médios para os quais nota-se claramente uma relação alométrica (lei de potência) que pode ser escrita como a relação linear

$$\langle \log_{10} n_c \rangle = A + \alpha \langle \log_{10} N \rangle, \quad (6.1)$$

a qual descreve muito bem esses valores médios em escala log-log. Vale a pena observar que os valores A e α praticamente não mudam quando se varia o número de janelas w . Entretanto, na Figura 6.1, verificam-se os diferentes valores de α (expoente da lei de potência) em cada um dos casos: membro de partidos políticos, candidato a vereador e candidato a prefeito. Com o objetivo de verificar se os valores dos α dependem somente da posição política e se eles não mudam com o tempo, encontraram-se as leis de potências médias para todos os anos da base de dados. A Figura 6.2 mostra essas análises dos comportamentos médios das médias de $\log_{10} n_c$ versus $\log_{10} N$, identificando que os valores de α praticamente não mudam com o tempo para membros de partidos políticos, candidatos a vereador e a prefeito. Para membros de partidos políticos, encontraram-se $\alpha = 0,87 \pm 0,02$, $\alpha = 0,86 \pm 0,02$ e $\alpha = 0,85 \pm 0,02$, respectivamente aos anos 2006, 2008 e 2011; para candidatos a vereador, $\alpha = 0,37 \pm 0,02$, $\alpha = 0,37 \pm 0,02$ e $\alpha = 0,38 \pm 0,02$, respectivamente aos anos 2000, 2004 e 2008 e para candidatos a prefeito, $\alpha = 0,19 \pm 0,01$, $\alpha = 0,20 \pm 0,01$ e $\alpha = 0,19 \pm 0,01$, respectivamente aos anos 2000, 2004 e 2008. Logo, não se pode

rejeitar a hipótese de que os valores de α são constantes no tempo para cada posição política depois de levar em conta os intervalos de confiança.

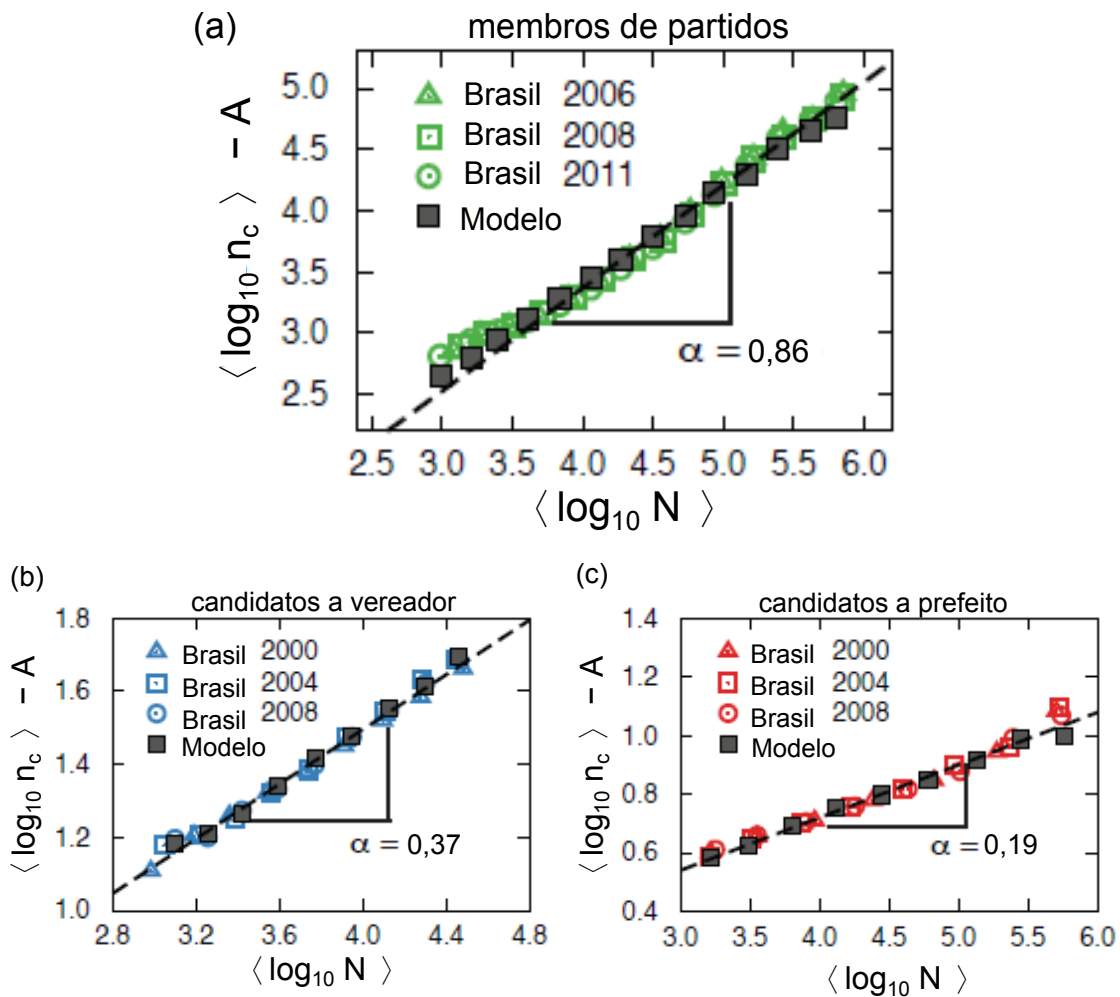


Figura 6.2: **Leis de potências dos dados** - Relações alométricas médias entre $\log_{10} n_c$ e $\log_{10} N$, N número de eleitores e n_c , o número de membros de partidos políticos em (a), o número de candidatos a vereador em (b) e a prefeito em (c). Para cada posição política, os valores de α (a inclinação das retas ajustadas) praticamente não mudam com o tempo. O valor médio de α é mostrado em cada gráfico. As linhas pontilhadas representam ajustes dos valores médios dos dados empregando os expoentes médios.

Em adição aos valores médios analisados, estudaram-se também as flutuações em torno das leis de potências, calculando a variância σ^2 de $\log_{10} n_c$ dentro de cada janela w em função do valor médio de $\log_{10} N$. A Figura 6.3 mostra que as variâncias são, em uma primeira aproximação, funções constantes de $\log_{10} N$. Os valores das variâncias médias sobre os anos são $\sigma^2 = 0,027 \pm 0,002$ para membros de partidos políticos, $\sigma^2 = 0,025 \pm 0,001$ para candidatos a vereador e $\sigma^2 = 0,017 \pm 0,001$ para candidatos a prefeito. Foram também construídas as distribuições das flutuações em torno das leis de potências considerando os resíduos normalizados

$$\xi = \frac{\log_{10} n_c(N) - \langle \log_{10} n_c \rangle_w}{\sigma_w}, \quad (6.2)$$

em que $\langle \log_{10} n_c \rangle_w$ representa o valor médio do número de candidatos n_c e σ_w seu desvio padrão, calculados em cada janela w . A Figura 6.4 mostra a distribuição de ξ , indicando claramente que essas distribuições são bem aproximadas pela distribuição Gaussiana padrão para toda a base de dados. E ainda, depois de calcular o teste de hipótese de Pearson, conclui-se que a hipótese de tomar a distribuição de ξ como a distribuição Gaussiana padrão deve ser aceita em todos os casos (para todos os dados).

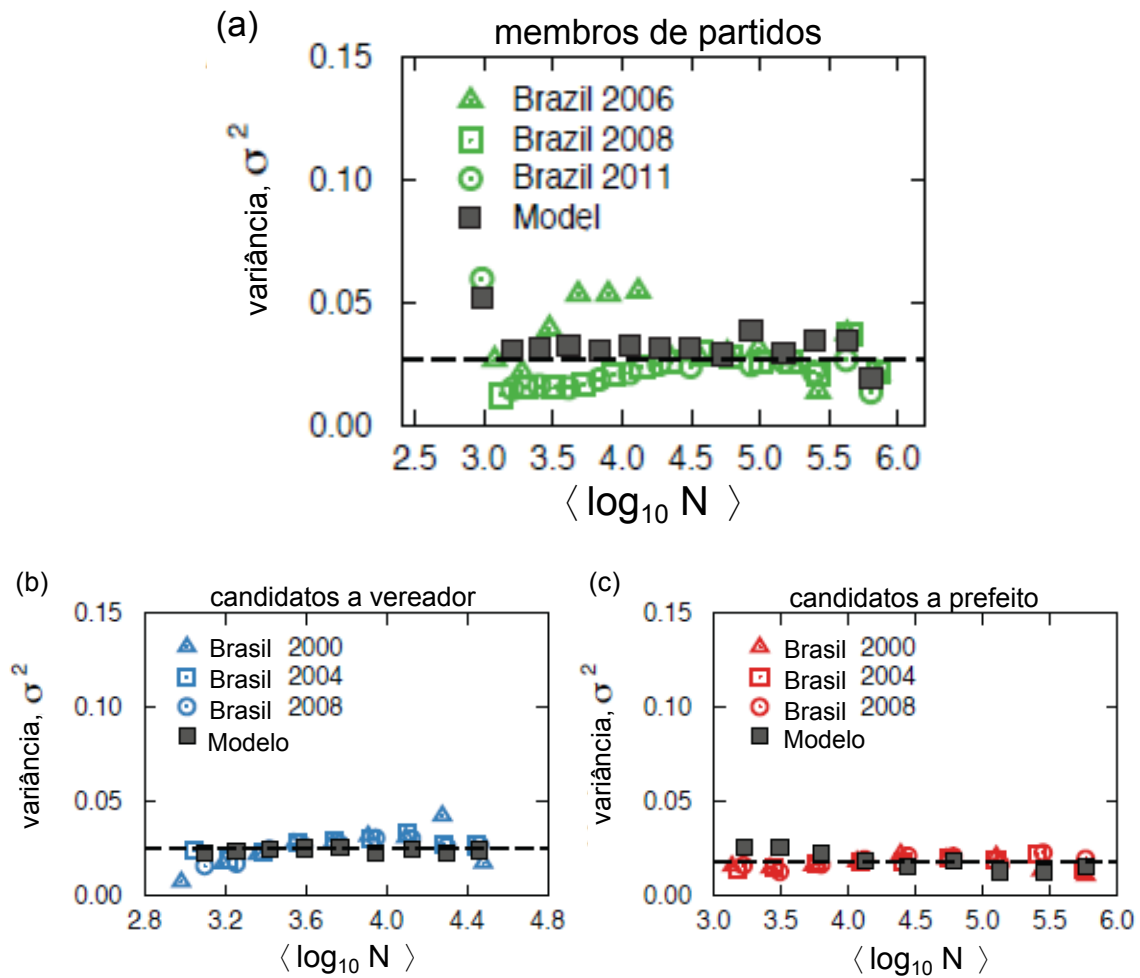


Figura 6.3: **Variâncias dos dados** - Variância σ^2 de $\log_{10} n_c$ versus $\langle \log_{10} N \rangle$, calculada em janelas igualmente espaçadas em $\log_{10} N$, para o número de membros de partidos políticos em (a), o número de candidatos a vereador em (b) e a prefeito em (c). Notam-se, em uma primeira aproximação, valores constantes de σ^2 para todas as três posições políticas (as linhas horizontais tracejadas são os valores médios de σ^2 sobre $\log_{10} N$ e sobre todos os anos). O modelo empregado corresponde à equação (6.3).

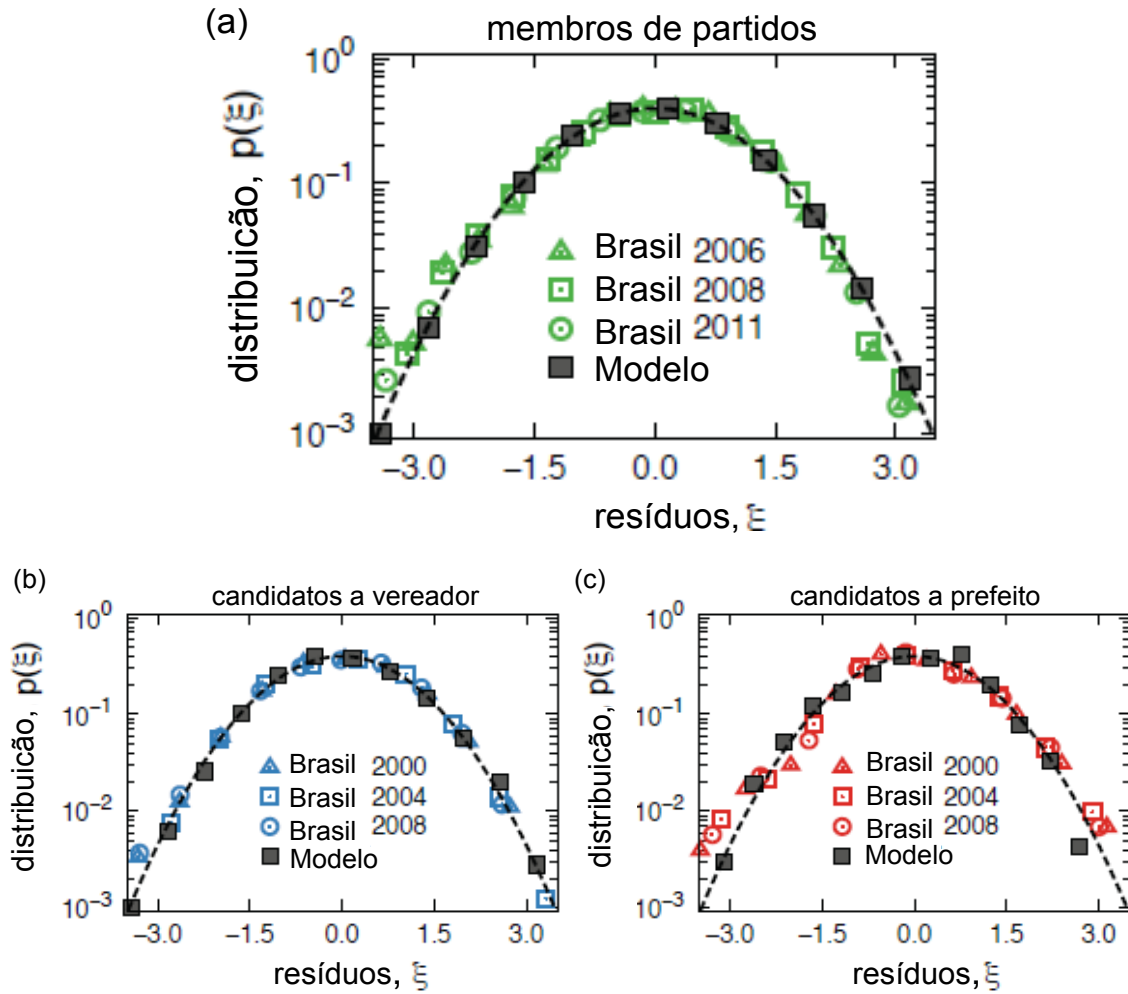


Figura 6.4: **Flutuações em torno das leis de potências** - Distribuições dos resíduos normalizados ξ em torno das leis de potências, para o número de membros de partidos políticos em (a), o número de candidatos a vereador em (b) e a prefeito em (c). As linhas tracejadas são a distribuição Gaussiana padrão. Em todos os gráficos, os quadrados cinzas representam os resultados das simulações com a equação (6.3).

6.3 Processo Multiplicativo

De forma análoga ao processo multiplicativo empregado no capítulo 5, aqui também os resultados encontrados para os dados empíricos são modelados empregando-se a equação estocástica

$$n_c = \mathcal{A}\zeta(N)N^\alpha, \quad (6.3)$$

em que \mathcal{A} é uma constante, $\zeta(N)$ é o ruído estocástico e α é o expoente da lei de potência entre n_c e N . Note que se for aplicado \log_{10} em ambos os lados da equação (6.3) e comparando com as equações (6.1) e (6.2), obtêm-se $\log_{10} \mathcal{A} = A$, e $\log_{10} \zeta(N) = \xi(N)$ e $\langle \xi(N) \rangle = 0$. Observando que $\xi(N)$ é normalmente distribuído, então $\zeta(N)$ deve ser distribuído de acordo com uma lognormal. Portanto, a equação (6.3) representa um processo múltiplicativo dirigido por um ruído lognormal.

Nas simulações da equação (6.3), foram usados os valores médios de α e σ para membros de partidos políticos, candidatos a vereador e a prefeito. Os resultados das simulações são mostrados nas Figuras 6.1, 6.2, 6.3 e 6.4 por quadrados cinzas. Observa-se que a equação (6.3) descreve muito bem todos os dados empíricos apresentados na análise.

6.4 Regra de ocupação das posições políticas

Os resultados apresentados neste capítulo mostraram padrões na decisão dos eleitores em participar de processos eleitorais em diferentes posições políticas. Para o número de membros de partidos políticos e de candidatos a vereador e a prefeito, os resultados empíricos mostraram que emergem robustas leis de potências médias com

a população de eleitores e que essas tendências médias são subjetivadas por flutuações com variâncias aproximadamente constantes e distribuições lognormais. Além disso, como pode ser visto na Figuras 6.5, existe uma clara hierarquia para os valores do expoente da lei de potência média α , em que a posição mais influente apresenta o menor expoente. Esse resultado pode ser relacionado também com o número de vagas para dada posição política, pois existe uma única vaga para prefeito de cada cidade, 9 para vereadores em cada cidade com menos de 47.600 habitantes e, em princípio, pode-se considerar que não há restrições para ser membro de um partido político (se filiar a um partido).

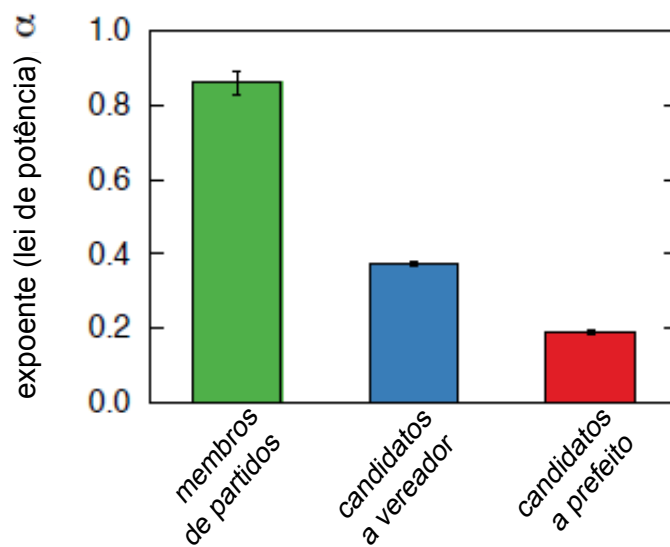


Figura 6.5: **Regra de ocupação das posições políticas** - Valores médios dos expoentes α das leis de potências médias para cada posição política investigada. Nota-se que o valor médio de α cresce com o número de vagas para a posição a ser ocupada. As barras de erro são os desvios padrão do valor médio de α .

Outro comportamento de contraste entre a ocupação de posições políticas aparece ao se analisar as distribuições de probabilidades dos números de candidatos a prefeito, candidatos a vereador, membros de partidos políticos e eleitores. A Figura 6.6 mostra a distribuição acumulada desses números.

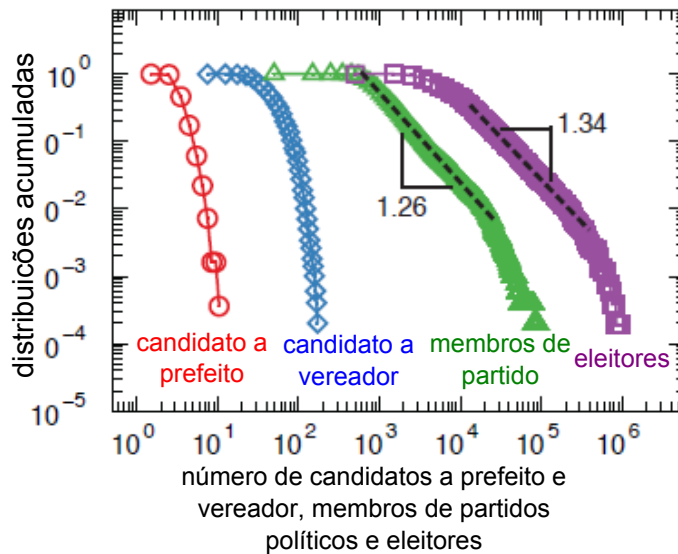


Figura 6.6: **Distribuições acumuladas** - para o número de candidatos a prefeito (círculos), candidatos a vereador (diamantes), membros de partidos políticos (triângulos) e população de eleitores (quadrados), considerando dados de 2008 (distribuições muito similares são obtidas para os outros anos). Note um deslocamento das distribuições com o crescimento do número de cadeiras disponíveis. Observa-se ainda que as distribuições para prefeito e vereador são caracterizadas por um comportamento assintótico exponencial, enquanto que as distribuições para membros de partidos políticos e eleitores apresentam um comportamento aproximadamente lei de potência seguido por um decaimento tipo exponencial. As linhas tracejadas mostram as regiões onde as distribuições podem ser aproximadas por leis de potência (os expoentes são mostrados no gráfico).

Note que, para candidatos a vereador e a prefeito, as distribuições são caracterizadas por caudas curtas, as quais têm aproximadamente um decaimento tipo exponencial. Por outro lado, para membros de partidos políticos e para eleitores, as

distribuições exibem um comportamento intermediário que pode ser aproximado por uma lei de potência (linha tracejada na Figura 6.6) seguida de um comportamento tipo exponencial. Verificou-se também que o número de eleitores cresce linearmente com o tamanho da população, o que seria de se esperar por causa da votação compulsória para os maiores de 18 anos.

6.5 Discussões

Neste capítulo, foi apresentada uma investigação sobre o envolvimento dos eleitores nos processos eleitorais em todas as cidades do Brasil, levando em consideração o número de candidatos a prefeito e vereador, bem como o número de membros de partidos políticos. Os resultados encontrados apontam um notável comportamento sub-linear entre as variáveis consideradas (números de candidatos a vereador e a prefeito e membros de partidos políticos) e o número de eleitores para cada cidade brasileira. Um modelo que contém uma lei de potência multiplicada por um ruído com distribuição lognormal apresentou resultado muito próximo ao espalhamento dos dados empíricos. Em seguida, observou-se que os expoentes dessas leis de potências apresentam uma regra clara: uma hierarquia em que as posições mais influentes exibem expoentes menores. O comportamento sub-linear intrínseco e o fato que α torna-se bem menor do que um para a maioria das posições políticas (em sequência) pode de alguma forma estar ligado a efeitos de sub-representação, ou seja, nas grandes cidades, os obstáculos para os indivíduos se envolverem diretamente nos processos eleitorais aumentam.

Outro aspecto investigado foi as distribuições de probabilidades acumuladas dos números de candidatos (a prefeito e a vereador), de membros de partidos políticos e de eleitores. Observou-se que as posições mais influentes (prefeito e vereadores) apresentam distribuições de caudas muito curtas, enquanto as posições menos influentes (membros de partidos políticos e eleitores) exibem um comportamento aproximadamente lei de potência para valores intermediários e seguido de um decaimento rápido. Essa mudança de comportamento pode ser relacionada à limitação no número de cadeiras, em que somente o decaimento tipo exponencial aparece para posições com o número de cadeiras limitado (prefeito e vereador) e o intermediário aproximadamente lei de potência aparece para as posições que não têm limitações estabelecidas para o número de cadeiras (membros de partidos políticos e eleitores).

Conclusões

Do capítulo 1 ao 4, foi apresentada uma revisão de resultados obtidos usando conceitos e métodos de Física Estatística na análise de dados de eleições. Inicialmente, considerou-se a distribuição dos votos em eleições proporcionais brasileiras, encontrando um comportamento hiperbólico em parte considerável da distribuição dos votos. Em seguida, outro ajuste foi empregado usando uma generalização da lei de Zipf, modelando uma quantidade maior dos dados. Dando continuidade ao estudo de distribuições de votos, porém em eleições pluralistas brasileiras, o ajuste foi baseado na distribuição de Laplace, notando-se uma polarização dos votos entre os dois primeiros colocados, independente do número de candidatos. Para o comportamento hiperbólico em eleições proporcionais, foram revistas duas dinâmicas de opinião usando diferentes modelos de redes, sugerindo que propriedades de “mundo pequeno” são necessárias para simular esse comportamento. Dando continuidade ao estudo de distribuições de votos em eleições proporcionais, porém considerando a ‘força’ dos partidos e dados de vários países, encontrou-se uma distribuição lognormal para cinco dos quinze analisados.

O último capítulo de revisão apresenta resultados relacionados à taxa de com-

parecimento às urnas. O primeiro mostra que a correlação da taxa logarítmica de comparecimento às urnas decai logaritmicamente com a distância entre as cidades, seguido por uma análise entrópica levando em conta a proporção de abstenções, brancos e nulos e votos válidos. Ainda usando a taxa de comparecimento, foi proposto um modelo capaz de identificar e quantificar fraudes em eleições. Esses resultados revelam o sucesso da aplicação da Física Estatística no estudo de processos eleitorais, um caso particular do sucesso também obtido em outros sistemas sociais e, mais geral, nos sistemas complexos.

Como deve estar claro, os estudos que acabaram de ser citados foram desenvolvidos a partir de análises de resultados de eleições. Entretanto, no momento de uma votação, os eleitores têm apenas um conjunto de candidatos a partir do qual fazem sua escolha. Nesta tese, foram investigados aspectos relacionados à composição do conjunto de eleitores que se tornam candidatos e, de forma mais ampla, o comportamento do eleitorado na ocupação de posições políticas.

Em uma primeira investigação, apresentada no capítulo 5, foi abordada uma análise de como ocorre o processo de candidatura, ou seja, a formação do conjunto de candidatos a serem votados no dia da eleição. Foi averiguado, para dados de 16 eleições em 5 países, que número de candidatos n_c cresce com o número de eleitores N em cada cidade, seguindo uma relação média lei de potência, $n_c \sim N^\alpha$, em que os expoentes encontrados foram: $\alpha = 0,18$ para eleições de um único membro (prefeito) e $\alpha = 0,36$ para eleições de multi-membro (vereador). Esse fato aponta para uma ponderação do número de vagas e da quantidade de eleitores no processo de candidatura. No entanto, pode-se observar que há três vagas para cada bairro de

Londres, enquanto que nas eleições brasileiras foram consideradas cidades com 9 vagas para vereadores em cada uma e o os valores dos expoentes são aproximadamente o mesmo. Isso indica que as responsabilidades do cargo público também são levadas em conta no processo de candidatura, uma vez que prefeito ou funções parlamentares estão em um patamar diferente do representante local.

Foram também examinadas as flutuações existentes em torno dos valores médios e observado que as variâncias são aproximadamente constantes. Ademais, as flutuações revelam que as candidaturas são dirigidas por um ruído multiplicativo com uma distribuição de probabilidade lognormal. Foi também empregado um modelo de rede livre de escala para representar as relações sociais, usando um tipo de “medida de popularidade” (o parâmetro de corte k_c do modelo de rede), mostrando-se robusto no sentido de que reproduziu muito bem tanto o comportamento médio quanto as flutuações. A análise mostrou que, independente da situação complexa relacionada com diferenças culturais e sociais entre os países, bem como de atributos psicológicos individuais dos candidatos, é possível observar regularidades sugerindo universalidades nos processos de candidatura.

Dando continuidade a este estudo, uma segunda investigação, apresentada no capítulo 6, foi direcionada a investigar o envolvimento dos eleitores nos processos eleitorais em todas as cidades do Brasil. Os resultados encontrados apontaram comportamentos sub-lineares (alometrias) entre as variáveis números de candidatos a vereador e a prefeito e membros de partidos políticos versus o número de eleitores para cada cidade brasileira. Um modelo lei de potência acoplado a um ruído multiplicativo com distribuição lognormal apresentou resultado muito próximo ao

espalhamento dos dados empíricos. Em seguida, observou-se que os expoente dessas leis de potências apresentam uma regra clara: uma hierarquia em que as posições mais influentes exibem expoentes menores. O fato desses expoentes serem menores do que um indica que cidades maiores podem apresentar maiores dificuldades para o envolvimento político.

Outro aspecto investigado mostrou que posições mais influentes (prefeito e vereadores) apresentam distribuições de caudas muito curtas, enquanto que as posições menos influentes (membros de partidos políticos e eleitores) exibem um comportamento aproximadamente lei de potência para valores intermediários e seguido de um decaimento rápido. Essa mudança de comportamento pode ser relacionada à limitação no número de cadeiras, em que somente o decaimento tipo exponencial aparece para posições com o número de cadeiras limitado (prefeito e vereador) e o intermediário aproximadamente lei de potência aparece para as posições que não têm limitações estabelecidas para o número de cadeiras (membros de partidos políticos e eleitores).

Apêndice A

Sistemas eleitorais

Sistema eleitoral é o regulamento que define as regras de apuração, contagem, agregação de votos e sua conversão em mandatos, ou seja, são as ações que convertem votos em mandatos [69, 70, 71]. Para os brasileiros, os sistemas utilizados são majoritário e proporcional.

A.1 Sistema proporcional

O princípio básico no sistema proporcional é que todos os eleitores merecem representação e todos os grupos políticos merecem ser representados nas legislaturas na proporção de sua força no eleitorado. Para alcançar essa representação “justa”, cada partido define internamente sua lista de candidatos (no Brasil, candidatos a deputados federal, estadual e vereadores). Após a contagem dos votos, as cadeiras a serem ocupadas em dado posto serão atribuídas aos partidos proporcionalmente ao número de votos que o partido recebeu.

Dentro de cada partido, a decisão dos candidatos que assumirão as cadeiras alocadas no partido ainda depende do sistema de listas, que são três:

Listas abertas - permitem que os eleitores expressem sua preferência não só entre os partidos, mas também entre os candidatos. O partido apresenta uma lista dos candidatos e os eleitores escolhem seu candidato preferido ou apenas o partido (na legenda do partido). A posição de cada candidato em relação à votação do partido depende unicamente do número de votos que ele recebe. Assim, se um partido alocou um número n de cadeiras, os n candidatos mais votados dentro do partido assumirão essas cadeiras; com isso, a eleição de um candidato depende do seu número de votos e do número total de votos do seu partido. Países como Itália (antes de 1994, quando um novo sistema foi introduzido), Polônia, Finlândia, Dinamarca, Estônia (desde 2002), Grécia, Suíça, Eslovênia, Brasil e Uruguai utilizam o sistema de listas abertas.

Listas semiabertas - impõe ao eleitor algumas restrições direta ou indiretamente. O eleitor vota para um partido ou para um candidato dentro do partido. Os partidos geralmente formulam uma lista de candidatos, de acordo com uma ordem determinada por critérios dentro do partido, e os eleitores podem contribuir para a alteração dessa ordem. Os candidatos conquistam cadeiras parlamentares na ordem em que são classificados na lista. No entanto, se um candidato recebe um número de votos que excede um limiar, sobe sua classificação na lista. Assim, a eleição de um candidato depende dessa ordem de classificação na lista e do número de votos recebidos pelo candidato. Suécia, Eslováquia, República Checa, Bélgica, Estônia (até 2002) e Holanda são alguns países que aderiram a esse sistema.

Listas fechadas - nesse sistema, o partido fixa a ordem na qual o candidato é listado e eleito. O eleitor lança um voto para um partido como um todo e não pode expressar sua preferência por qualquer candidato. Os representantes são então selecionados como aparecem na lista. Rússia, Itália (desde 2006), Espanha, Angola, África do Sul, Israel, Sri Lanka, Hong Kong e Argentina são alguns países que definem seus representantes parlamentares por esse sistema.

A.2 Sistema majoritário

No sistema majoritário, também conhecido como eleições pluralistas, o candidato que obtém o maior número de votos entre os competidores é eleito. Sistema pelo qual, no Brasil, se elege o presidente da República, senador, governador e prefeito. Para eleições de presidente, governador e prefeito em municípios com mais de 200 mil eleitores, é realizado segundo turno entre os dois candidatos mais votados e um deles é eleito se receber a maioria absoluta dos votos válidos (50 % mais um). Porém haverá segundo turno caso nenhum candidato tenha conquistado a maioria absoluta dos votos válidos no primeiro turno.

Apêndice B

Física Estatística e sistemas complexos

A seguir, serão abordados, de forma breve, alguns conceitos e ferramentas básicas de Mecânica Estatística utilizados nas investigações discutidas neste trabalho.

B.1 Leis de potências

Apesar de não haver uma definição precisa de sistemas complexos, em geral eles apresentam propriedades emergentes que decorrem em grande parte da interação não-linear entre suas partes. Assim, suas propriedades apenas se tornam notáveis quando vistos como um todo. Uma das propriedades marcantes de muitos desses sistemas é a presença de leis de escala ou leis de potência [50]. Entre os sistemas complexos estão os sistemas sociais, os quais emergem das dinâmicas sociais e dizem respeito aos comportamentos de indivíduos isolados, ou seja, como cada indivíduo

escolhe uma convenção, toma uma decisão, agenda tarefas, e mais geralmente decide realizar uma dada ação.

Uma lei de potência (ou lei de escala) pode ser escrita como

$$P(x) = ax^k + o(x^k), \quad (\text{B.1})$$

em que a e k (expoente de escala) são constantes e $o(x^k)$ é uma função assintoticamente pequena em relação x^k . A equação acima é uma relação polinomial que exhibe a propriedade de invariância de escala (ou livre de escala), que significa que $P(cx) \propto P(x)$. Note que, em um gráfico log-log da função de escala $P(x)$, tem-se uma reta com inclinação k . De fato, se $P(x) = ax^k$ então $\log P(x) = \log a + k \log x$.

B.2 Distribuição gaussiana e o teorema do limite central

É usual, na investigação de distribuições empíricas simétricas em relação à média, compará-las com uma distribuição de probabilidades chamada gaussiana (ou normal) [50]. Dois parâmetros identificam a distribuição de probabilidades gaussiana de uma variável aleatória x : a média μ e o desvio-padrão σ . Essa distribuição é da forma

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right]. \quad (\text{B.2})$$

Geralmente, utiliza-se a distribuição gaussiana padronizada, em que a variável aleatória x é normalizada, ou seja, considera-se uma nova variável $X = (x - \mu)/\sigma$, com média

zero e desvio-padrão unitário e, conseqüentemente, sua distribuição pode ser escrita como segue:

$$\tilde{P}(X) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}X^2\right). \quad (\text{B.3})$$

Basicamente, quando um grande número de eventos aleatórios e independentes conduzem a determinado evento (o que aproxima muitos fenômenos da natureza), este poderá ter uma distribuição de probabilidades gaussiana (B.2). Formalmente, se x_i são variáveis aleatórias independentes de média zero e desvio padrão finito σ , a função densidade de probabilidade $X(n) = \sum_{i=1}^n x_i/\sigma\sqrt{N}$ tende à distribuição gaussiana padronizada à medida que n cresce. Esse resultado é chamado Teorema do Limite Central.

Supondo que um processo x possa ser decomposto no produto de n subprocessos x_i , ou seja, $x = x_1 \cdot x_2 \cdot \dots \cdot x_n$ em que cada x_i se comporta como uma variável aleatória não negativa e independente, pelo Teorema do Limite Central, para n suficientemente grande, a distribuição de probabilidades da variável aleatória $\log x = \log x_1 + \log x_2 + \dots + \log x_n$ é normal. Assim, para variável x tem-se a distribuição

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right], \quad (\text{B.4})$$

chamada lognormal e diz-se que x segue um processo multiplicativo.

B.3 Modelo de Sznajd

O modelo de Sznajd [51, 52] parte do princípio de que convencer alguém é mais fácil para duas ou mais pessoas do que para um único indivíduo. Na versão original do modelo de Sznajd [11], o modelo é proposto para descrever um mecanismo de tomada de decisão em uma comunidade fechada a partir do modelo de Ising. O modelo de spin de Ising [53] considera uma coleção de N spins (agentes) s_i que podem assumir dois valores, ± 1 . Cada spin é energeticamente forçado a alinhar-se com seus vizinhos próximos. A energia total é dada por $H = -\frac{J}{2} \sum_{\langle i,j \rangle} s_i s_j$, em que a soma é sobre os pares de vizinhos próximos e J a constante de acoplamento.

A versão original é chamada Sznajd B. Nele, agentes ocupam sítios de uma cadeia linear e têm opiniões binárias, denotadas pelas variáveis de spins de Ising. Um par de agentes vizinhos i e $i + 1$ determinam a opinião dos seus vizinhos próximos $i - 1$ e $i + 2$, de acordo com a seguinte regra:

$$s_i = s_{i+1} \implies s_{i-1} = s_i = s_{i+1} = s_{i+2} \quad (\text{B.5})$$

$$s_i \neq s_{i+1} \implies s_{i-1} = s_i \text{ e } s_{i+1} = s_{i+2} \quad (\text{B.6})$$

As opiniões são atualizadas em uma ordem sequencial aleatória a partir de uma configuração inicial totalmente aleatória, em que ambas as opiniões são igualmente distribuídas. Dois tipos de estados estacionários, correspondendo ao consenso, são encontrados, com todos os spins para cima $m = 1$ ou todos os spins para baixo $m = -1$, e um impasse com o mesmo número de spins para cima e para baixo $m = 0$. Cada caso de consenso ocorre com probabilidade $1/4$. O tempo de relaxação

do sistema para um dos possíveis atratores segue uma distribuição lognormal [55]. O número de agentes que nunca mudam de opinião decai primeiro como uma lei de potência do tempo, em seguida atinge um valor constante [56].

A regra de Sznajd B, dada por (B.5) e (B.6), foi considerada pouco realística. Uma alternativa, em estudos posteriores, foi tomar como regra apenas a (B.5), chamada de Sznajd A. Extensões do modelo Sznajd, diferentes substratos geralmente adotam essa prescrição.

B.4 Redes complexas

Atualmente, vários estudos de sistemas sociais têm utilizado redes para analisar e simular dinâmicas de sistemas complexos [65]. Em geral, são redes cuja estrutura é irregular e com dinâmica evolutiva no tempo.

A seguir, serão descritas algumas propriedades comuns observadas nas topologias de redes reais e é apresentada também uma breve revisão dos principais modelos que têm sido propostos para abordar sistemas sociais (grafos aleatórios, redes de mundos pequenos e redes livres de escala).

Um grafo $G(\mathcal{N}, \mathcal{L})$ consiste de dois conjuntos: um conjunto \mathcal{N} de N nós (ou sítios, ou vértices) e um conjunto \mathcal{L} de L ligações (ou arestas) entre os nós. Uma ligação entre dois nós i e j é denotada por (i, j) ou l_{ij} , dois nós unidos por uma ligação são ditos adjacentes ou vizinhos próximos. Um grafo é dito direcionado quando a direção da ligação é relevante, ou seja, l_{ij} não é igual a l_{ji} para qualquer (i, j) . O número máximo de ligações em um grafo $G(\mathcal{N}, \mathcal{L})$ é $N(N - 1)/2$.

B.4.1 Distribuição de grau

O grau de um nó i denotado por k_i , é o número de ligações partindo desse nó que pode ser calculado usando a matriz adjacência $\mathfrak{A} = [a_{ij}]$, em que $a_{ij} = 1$ se há uma ligação entre i e j e $a_{ij} = 0$ se não há ligação entre i e j . Nesse caso, $k_i = \sum_{j=1}^N a_{ij}$. Se o grafo for direcionado, o grau terá duas componentes, $k_i^{out} = \sum_{j=1}^N a_{ij}$ para as ligações partindo de i e $k_i^{in} = \sum_{j=1}^N a_{ji}$ para as ligações chegando em i . Informações da distribuição de grau podem ser obtidas fazendo o gráfico de $P(k)$ ou calculando os seus momentos. O n -ésimo momento de $P(k)$ é definido como $\langle k^n \rangle = \sum_k k^n P(k)$. O primeiro momento representa o valor médio do grau. A distribuição de grau determina completamente as propriedades estatísticas em redes não correlacionadas. No entanto, um grande número de redes reais são correlacionadas, ou seja, a probabilidade que um nó de grau k seja conectado a outro nó de grau k' depende de k . Nesse caso, é necessário introduzir a probabilidade condicional $P(k'|k)$ [54].

B.4.2 Grafos aleatórios e mundo pequeno

O estudo de grafos aleatórios foi iniciado por Erdős e Rényi em 1959 com propósito de estudar, por meio de métodos probabilísticos, as propriedades de grafos como uma função do número crescente de ligações aleatórias. Em seu primeiro artigo, Erdős e Rényi propuseram um modelo para gerar grafos aleatórios com N nós e K ligações, que passou a ser chamado de grafo aleatório de Erdős and Rényi (ER). Começando com N nós desconectados, o grafo é gerado conectando nós selecionados aleatoriamente, até que o número de ligações seja igual a K . Um modelo alternativo

para o grafo aleatório ER consiste em ligar cada par de nós com uma probabilidade $0 < p < 1$. Esse procedimento descreve grafos com diferentes números de ligações, grafos com K ligações devem aparecer com uma probabilidade $p^K(1-p)^{N(N-1)/(2-K)}$ [59, 60, 61]. Note que $N \rightarrow \infty$ é tomado para $\langle k \rangle$ fixado, o que corresponde a fixar $2K/N$ no primeiro modelo e $p(N-1)$ no segundo modelo. Para N grande e $\langle k \rangle$ fixado, a distribuição de grau é aproximada por uma distribuição de Poisson

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}. \quad (\text{B.7})$$

O estudo de vários processos dinâmicos em redes reais apontou a existência de atalhos, ou seja, ligações que conectam diferentes áreas das redes. Na maior parte das redes reais, apesar de seu tamanho grande, há um caminho relativamente curto entre quaisquer dois nós. Essa propriedade é conhecida como “mundo pequeno” e é matematicamente caracterizada pelo comprimento médio de trajetória curta. Essa propriedade foi primeiramente investigada, no contexto social, por Milgram na década de 1960 [62, 63].

O modelo Watts e Strogatz (WS) é um método para construir grafos com propriedades de “mundos pequenos” e alto coeficiente de aglomeração [57]. O ponto de partida é um anel com N nós, no qual cada nó é simetricamente conectado a seus $2m$ vizinhos próximos, produzindo um total de $K = mN$ ligações. Para cada nó, cada ligação a um vizinho no sentido horário é religada a um nó escolhido aleatoriamente com uma probabilidade p e preservado com uma probabilidade $1 - p$. Note que para $p = 0$, tem-se a malha regular, enquanto para $p = 1$ o modelo reproduz um grafo aleatório. Para valores intermediários de p , o procedimento gera grafos com

propriedades de “mundos pequenos” e coeficiente de aglomeração não trivial. Para a distribuição de grau, quando $p = 0$, ela é uma função delta, enquanto para $p = 1$ ela é a obtida para ER. No caso de p intermediário, a distribuição de grau é dada por [64]:

$$P(k) = \sum_{i=0}^{\min(k-m, m)} \binom{m}{i} (1-p)^i p^{m-i} \frac{(pm)^{k-m-i}}{(k-m-i)!} e^{-pm}, \quad (\text{B.8})$$

para $k \geq m$, e é igual a zero se $k < m$.

B.4.3 Redes livres de escala

O modelo de Barabási-Albert (BA) é um modelo de rede evoluindo no tempo, inspirado na formação da World Wide Web (WWW) e baseado em dois ingredientes: crescimento e ligação preferencial [7, 65]. A ideia básica é que na WWW, nós (páginas) com alto conectividade adquirem novos nós em uma taxa mais alta do que nós com baixo conectividade. O modelo começa com m_0 nós isolados em cada intervalo de tempo $t = 1, 2, 3, \dots, N - m_0$ um novo nó j é adicionado a rede com $m \leq m_0$ ligações. A probabilidade que o nó j seja ligado a um nó já existente i é proporcional a conectividade atual de i :

$$\prod_{j \rightarrow i} = \frac{k_i}{\sum_l k_l}. \quad (\text{B.9})$$

Cada novo nó tem m ligações; no tempo t , a rede deve ter $N = m_0 + t$ nós e $K = mt$ ligações, correspondendo a um grau médio $\langle k \rangle = 2m$ para tempos grandes. No limite de $t \rightarrow \infty$, o modelo produz uma distribuição de grau $P(k) \propto k^{-\gamma}$, com um expoente

$\gamma = 3$. Usualmente, diz-se que uma rede é livre de escala se sua distribuição de grau é uma lei de potência, $P(k) \propto k^{-\gamma}$.

Um modelo generalizado de Barabási-Albert é parametrizado por duas probabilidades: de ligar λ e religar ρ , tal que $\lambda + \rho < 1$. Em cada passo, uma das seguintes ações são tomadas: com probabilidade $1 - \lambda - \rho$, um novo nó é inserido; com probabilidade ρ , m arestas são religadas, desconectando-o de um de seus vizinhos, e reconectando a outro nó com ligação preferencial; com probabilidade λ , m arestas são adicionadas escolhendo-se um nó aleatoriamente e um novo vizinho para ele com probabilidade preferencial [67]. A distribuição de grau segue uma lei de potência $k^{-\gamma}$, com expoente

$$\gamma = 1 + \frac{2m(1 - \rho) + 1 - \lambda - \rho}{m}. \quad (\text{B.10})$$

Referências Bibliográficas

- [1] S. Picoli Jr., R. S. Mendes, Universal features in the growth dynamics of religious activities, *Phys. Rev. E* 77 (2008) 036105.
- [2] M. Ausloos, F. Petroni, Statistical dynamics of religion evolutions, *Physica A* 388 (2009) 4438-4444.
- [3] R. N. Mantegna, Hierarchical structure in financial markets, *Eur. Phys. J. B* 11 (1999) 193-197.
- [4] Y. Lee, L. A. N. Amaral, D. Canning, M. Meyer, H. E. Stanley, Universal features in the growth dynamics of complex organizations, *Phys. Rev. Lett.* 81 (1998) 15 (3275-3278).
- [5] P. G. Lind, L. R. da Silva, J. S. Andrade Jr., H. J. Herrmann., The Spread of Gossip in American Schools, *Europhys. Lett.* 78 (2007) 68005.
- [6] P. G. Lind, L. R. da Silva, J. S. Andrade Jr., H. J. Herrmann., Spreading Gossip in Social Networks, *Phys. Rev. E* 76 (2007) 036117.

- [7] A. L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (1999) 509-512.
- [8] P. L. Krapivsky, S. Redner, Statistics of changes in lead node in connectivity-driven networks, *Phys. Rev. Lett.* 89 (2002) 25803.
- [9] M. E. Ravasz, Z. Toroczkai, Centrality scaling in large networks, *Phys. Rev. Lett.* 105 (2010) 038701.
- [10] E. Ben-Naim, P. L. Krapivsky, Popularity-driven networking, *Europhys. Lett.* 97 (2012) 48003.
- [11] K. Sznajd-Weron, J. Sznajd, Opinion evolution in closed community, *Int. J. Mod. Phys. C* 11 (2000) 1157-1165.
- [12] F. Slanina, H. Lavicka, Analytical results for the Sznajd model of opinion formation, *Eur. Phys. J. B* 35 (2003) 279-288.
- [13] V. Sood, S. Redner, Voter model on heterogeneous graphs, *Phys. Rev. Lett.* 94 178701 (2005) 178701.
- [14] A. Pluchino, V. Latora, A. Rapisarda, Compromise and synchronization in opinion dynamics, *Eur. Phys. J. B* 50 (2006) 169-176.
- [15] S. Galam, Application of statistical physics to politics, *Physica A* 274 (1999) 132-139.
- [16] T. M. Gradowski, R. A. Kosinski, The effect of an electoral campaign for election results in an open community, *Int. J. Mod. Phys. C* 9 (2006) 1327-1338.

- [17] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics, *Phys. Rep.* 424 (2006) 175-308.
- [18] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, *Rev. Mod. Phys.* 81 (2009) 591-646.
- [19] S. Galam, *Sociophysics: a physicist's modeling of psycho-political phenomena*, Springer, Heidelberg, 2012.
- [20] W. Weidlich, The statistical description of polarization phenomena in society, *Br. J. Math. Stat. Psychol.*, 24 (1971) 251-266.
- [21] S. Galam, Y. Gefen and Y. Shapir, Sociophysics: a mean behavior model for the process of strike, *J. Math. Sociol.* 9 (1982) 1-13.
- [22] S. Galam, S. Moscovici, Towards a theory of collective phenomena: consensus and attitude changes in groups, *Eur. J. Soc. Psychol.* 21 (1991) 49-74.
- [23] P. Clifford, A. Sudbury, A model for spatial conflict, *Biometrika* 60 (1973) 581-588.
- [24] R. Holley, T. M. Liggett., Ergodic theorems for weakly interacting systems and the voter model, *Ann. Probab.* 3 (1975) 643-663.
- [25] S. Galam, Minority opinion spreading in random geometry, *Eur. Phys. J. B* 25 (2002) 403-406.
- [26] B. Latané, The psychology of social impact, *Am. Psychol.* 36 (1981) 343-356.

- [27] A. Nowak, J. Szamrej, B. Latané, From private attitude to public opinion: a dynamic theory of social impact, *Psychol. Rev.* 97 (1990) 362-376.
- [28] M. Lewenstein, A. Nowak, B. Latané, Statistical mechanics of social impact, *Phys. Rev. A* 45 (1992) 763-776.
- [29] R. N. Costa Filho, M. P. Almeida, J. S. Andrade, J. E. Moreira, Scaling behavior in a proportional voting process, *Phys. Rev. E* 60 (1999) 1067-1068.
- [30] R. N. Costa Filho, M. P. Almeida, Brazilian elections: voting for a scaling democracy, *Physica A* 322 (2003) 698-700.
- [31] M. L. Lyra, U. M. S. Costa, R. N. Costa Filho, J. S. Andrade, Generalized Zipf's law in proportional voting processes, *Europhys. Lett.* 62 (2003) 131-137.
- [32] L. E. Araripe, R. N. Costa Filho, H. J. Herrmann, J. S. Andrade, Plurality voting: the statistical laws of democracy in Brazil, *Int. J. Mod. Phys. C* 17 (2006) 1809-1813.
- [33] N. A. M. Araújo, J. S. Andrade, H. J. Herrmann, Tactical voting in plurality elections, *PLoS ONE* 5 (2010) e12446.
- [34] A. T. Bernardes, D. Stauffer, J. Kertész, Election results and the Sznajd model on Barabási network, *Eur. Phys. J. B* 25 (2002) 123-127.
- [35] M. C. González, A. O. Sousa, H. J. Herrmann, Opinion formation on a deterministic pseudo-fractal network, *Int. J. Mod. Phys. C* 16 (2004) 45-57.

- [36] J. S. Andrade Jr., H. J. Herrmann., R. F. S. Andrade, L. R. da Silva, Apollonian Networks: Simultaneously Scale-Free, Small World, Euclidean, Space Filling, and with Matching Graphs, *Phys. Rev. Lett.* 94 (2005) 018702.
- [37] G. Travieso, L. F. Costa, Spread of opinions and proportional voting, *Phys. Rev. E* 74 (2006) 036112.
- [38] S. Fortunado, C. Castellano, Scaling and universality in proportional elections, *Phys. Rev. Lett.* 99 (2007) 138701.
- [39] L. E. Araripe, R. N. Costa Filho, Role of parties in the vote distribution of proportional elections, *Physica A* 388 (2009) 4167-4170.
- [40] A. Chatterjee, M. Mitrovic, S. Fortunato, Universality in voting behavior: an empirical analysis, *Sci. Rep.* 3 (2013) 1049.
- [41] C. Borghesi, J. P. Bouchaud, Spatial correlations in vote statistics: a diffusive field model for decision-making, *Eur. Phys. J. B* 75 (2010) 395-404.
- [42] G. Marsaglia, W. W. Tsang, J. Wang, Evaluating Kolmogorov's distribution, *Journal de Estatistical Software* 1 (2003) 8.
- [43] C. Borghesi, J.-C. Raynal, J.-P. Bouchaud, Election turnout statistics in many countries: similarities, differences, and a diffusive field model for decision-making, *PLoS ONE* 7 (2012) e36289.
- [44] C. Borghesi, J. Chiche, J.-P. Nadal, Between order and disorder: a 'weak law' on recent electoral behavior among urban voters, *PLoS ONE* 7 (2012) e39916.

- [45] P. Klimek, Y. Yegorov, R. Hanel, S. Thurner, Statistical detection of systematic election irregularities, PNAS 109 (2012) 16469-16473.
- [46] B. Derrida, H. Flyvbjerg, Statistical properties of randomly broken objects and of multivalley structures in disordered systems, J. Phys. A 20 (1987) 5273-5288.
- [47] A. T. Bernardes, U. M. S. Costa, A. D. Araujo, D. Stauffer, Damage spreading coarsening dynamics and distribution of political notes in Sznajd model on square lattice, Int. J. Mod. Phys. C 12 (2001) 159-167.
- [48] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, J. Stat. Phys. 52 (1988) 479-487.
- [49] F. Y. Wu, The Potts model, Rev. Mod. Phys., 54 (1982) 235-268.
- [50] D. Sornette, Critical phenomena in natural sciences, Springer, Heidelberg, 2006.
- [51] D. Stauffer, A. O. Sousa, S. Moss de Oliveira, Generalization to square lattice de Sznajd sociophysics model, Int. J. Mod.Phys. C 11 (2000) 1239-1245.
- [52] K. Sznajd-Weron, Sznajd model and its applications, Acta Phys. Pol. B 36 (2005) 2537-2547.
- [53] J. Binney, N. Dowrick, A. Fisher, M. Newman, The theory of critical phenomena: an introduction to the renormalization group, Oxford University Press, Oxford, UK, 1992.
- [54] M. E. J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.

- [55] L. Behera, F. Schweitzer, On spatial consensus formation: Is the Sznajd model different from a voter model?, *Int. J. Mod. Phys. C* 14 (2003) 1331-1354.
- [56] D. Stauffer, D., P. M. C. de Oliveira, Persistence of opinion in the Sznajd consensus model: computer simulation, *Eur. Phys. J. B* 30 (2002) 587-592.
- [57] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393 (1998) 440-442.
- [58] D. J. Watts, *Small worlds: the dynamics of networks between order and randomness*, Princeton University Press, Princeton, NJ, 1999.
- [59] P. Erdős, A. Rényi, On random graphs, I. *Publ. Math.* 6 (1959) 290-297.
- [60] P. Erdős, A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* 5 (1960) 17-61.
- [61] Z. Burda, J. Jurkiewicz, A. Krzywicki, *Statistical mechanics and its applications*, *Physica A* 344 (2004) 56-61.
- [62] S. Milgram, The small-world problem, *Psychol. Today* 1 (1967) 60-67.
- [63] J. Travers, S. Milgram, An experimental study of the small world problem, *Sociometry* 32 (1969) 425-443.
- [64] A. Barrat, M. Weigt, On the properties of small-world network models, *Eur. Phys. J. B* 13 (2000) 547-560.
- [65] R. Albert, A.-L. Barabási, *Statistical mechanics of complex networks*, *Rev. Mod. Phys.* 74 (2002) 47-97.

- [66] A. L. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks, *Physica A* 272 (1999) 173-187.
- [67] R. Albert, A.-L. Barabási, Topology of Evolving Networks: Local Events and Universality, *Phys. Rev. Lett.* 85 (2000) 5234-5237.
- [68] P. Bak, *How Nature Works: the Science of Self Organized Criticality* (Springer-Verlag, New York) 1996.
- [69] <http://www.senado.gov.br/atividade/materia/getPDF.asp?t=86989tp=1>
- [70] <http://www.jornaljovem.com.br/edicao4/politica03.php>
- [71] <http://www.guiadedireitos.org>
- [72] M. C. Mantovani, H. V. Ribeiro, M. V. Moro, S. Picoli Jr., R. S. Mendes, Scaling laws and universality in the choice of election candidates, *Europhys. Lett.* 96 (2011) 48001.
- [73] M. C. Mantovani, H. V. Ribeiro, E. K. Lenzi, S. Picoli Jr., R. S. Mendes, Engagement in the electoral processes: scaling laws and the role of the political position, *Phys. Rev. E* 88 (2013) 024802.
- [74] <http://www.tse.gov.br>, <http://elezionistorico.interno.it>, <http://data.london.gov.uk/taxonomy/keywords/elections>, <http://www.aec.gov.au>, <http://www.elections.ca/>, (accessed in January 2011).
- [75] <http://www.ibge.gov.br/home/>

Agradecimentos

A Deus, nada seria possível sem a minha fé;

Ao meu orientador Renio dos Santos Mendes, por ter me acolhido e me direcionado em tudo de forma precisa e com muita dedicação e paciência;

A todos os colegas do grupo de pesquisa Complex UEM, pelos momentos agradáveis em nossas reuniões e por tudo o que venho aprendendo com eles. E em especial ao colega Haroldo, que me ajudou em todos os momentos que necessitei;

A um grupo muito especial de colegas de trabalho da Universidade Tecnológica Federal do Paraná (Adilandri, Wellington, Priscila, Viviani e Sara), que tornaram possível a minha liberação para o doutorado;

Aos professores do Departamento de Física da Universidade Estadual de Maringá, que me receberam muito bem e me ajudaram cumprir os requisitos exigidos para a conclusão do doutorado;

Aos meus filhos Lucas, Guilherme e Giovanna, meus parceiros nos momentos alegres e principalmente nos momentos difíceis.

“Dedico esse trabalho em memória dos meus pais: Maria Eunice e David Genésio, por tudo o que fizeram por mim.”