Pontifícia Universidade Católica
do Rio de Janeiro



Marlon Ferreira Ramos

Dinâmicas de opinião em redes complexas

PUC-Rio - Certificação Digital Nº 1122074/CA

Tese de Doutorado

Tese apresentada ao Programa de Pós–graduação em Física do Departamento de Física da PUC–Rio como requisito parcial para obtenção do título de Doutor em Física

Orientadora: Prof.ª Celia Beatriz Anteneodo de Porto

Rio de Janeiro
Agosto de 2015

Pontifícia Universidade Católica
do Rio de Janeiro



Marlon Ferreira Ramos

Dinâmicas de opinião em redes complexas

Tese apresentada ao Programa de Pós–graduação em Física do Departamento de Física do Centro Técnico Científico da PUC–Rio como requisito parcial para obtenção do título de Doutor em Física. Aprovada pela comissão examinadora abaixo assinada.

Prof.ª Celia Beatriz Anteneodo de Porto
Orientadora
Departamento de Física — PUC–Rio

Prof. Roberto Fernandes Silva Andrade
UFBA

Prof. Welles Antonio Martinez Morgado
Departamento de Física — PUC-Rio

Prof. Thadeu Josino Pereira Penna
UFF

Prof. Juan Carlos Gonzalez Avella
Departamento de Física — PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro Técnico Científico PUC-Rio

Rio de Janeiro, 7 de Agosto de 2015

PUC-Rio - Certificação Digital Nº 1122074/CA

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Marlon Ferreira Ramos

Possui graduação - bacharelado e licenciatura - em Física pela Universidade Federal Fluminense (2009), mestrado em Física pela Universidade Federal Fluminense (2011). Tem experiência na área de Física, com ênfase em Física Estatística e Sistemas Complexos, atuando principalmente nos seguintes temas: modelagem estatística e computacional de sistemas complexos, redes complexas e simulação de Monte Carlo.

PUC-Rio - Certificação Digital Nº 1122074/CA

Ficha Catalográfica

Ramos, Marlon Ferreira

Dinâmicas de opinião em redes complexas / Marlon Ferreira Ramos; orientador: Celia Beatriz Anteneodo de Porto. — Rio de Janeiro : PUC–Rio, Departamento de Física, 2015.

v., 142 f: il. ; 29,7 cm

1. Tese (Doutorado em Física) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Física.

Inclui referências bibliográficas.

1. Física – Tese. 2. Redes Complexas. 3. Modelos de Opinião. 4. Simulações computacionais. I. Anteneodo de Porto, Celia Beatriz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Física. III. Título.

CDD: 510

PUC-Rio - Certificação Digital Nº 1122074/CA

Agradecimentos

À minha orientadora Celia Anteneodo por todo seu apoio, dedicação e trabalho, pelas discussões, por estar sempre presente, por tudo que me ensinou e pelo carinho, atenção e paciência durante meu doutorado.

Ao meu co-orientador no período de doutorado sanduíche, Hernán Makse, pela receptividade e suporte durante esse período.

Às Agências de fomento: CNPq, CAPES e FAPERJ, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos coautores dos trabalhos que fiz durante esse período: Angelo Calvão, Saulo Reis, Jia Shao, José Andrade Jr e Shlomo Havlin.

Aos colegas da PUC–Rio e do Levich Institute of New York, e aos ex-colegas da UFF. Em especial à Fernanda Deus, Fillipe Siqueira e Angelo Calvão pelas sugestões e revisão do manuscrito; à Vitor Lara e Alexandre Lima pelas discussões frutíferas, pelas valiosas sugestões e ajuda nos trabalhos.

Aos meus pais pelo seu amor incondicional e aos meus sogros por todo cuidado e apoio.

À minha esposa Gabriela Coutinho, futura doutora em Cosmologia, pelas suas contribuições para os trabalhos desenvolvidos e para a tese. Pelo seu amor, carinho, atenção, por todo o suporte emocional e por estar do meu lado em todos os momentos.

Resumo

Ramos, Marlon Ferreira; Anteneodo de Porto, Celia Beatriz. **Dinâmicas de opinião em redes complexas**. Rio de Janeiro, 2015. 142p. Tese de Doutorado — Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese aborda diversos problemas que podem ser tratados mediante modelos de dinâmica de opiniões, segundo os quais os indivíduos, conectados de acordo com redes complexas, interagem mediante regras que moldam as preferências e o posicionamento desses indivíduos com relação a uma determinada questão. A metodologia utilizada para investigar os padrões emergentes dessas interações consiste na utilização de diversas técnicas da física estatística. A tese está organizada em torno de quatro problemas distintos, com uma questão particular a ser respondida em cada caso, buscando sempre a validação empírica dos resultados teóricos e computacionais. No primeiro trabalho, é respondida a seguinte questão básica sobre propriedades da rede que podem ter impacto sobre os processos de propagação: quais são os valores típicos das distâncias, coeficiente de aglomeração e outras grandezas estruturais da rede, quando considerado o ensemble de redes aleatórias com uma assortatividade fixa? No segundo trabalho, investigamos os padrões que surgem na avaliação de filmes, considerando como fonte o IMDb (Internet Movie Database). Encontramos que a distribuição de votos apresenta um comportamento livre de escala com um expoente muito próximo de 3/2. Curiosamente, esse padrão é robusto, independente de atributos dos filmes como nota média, idade ou gênero. A análise empírica aponta para um mecanismo de propagação de adoções simples, que gera uma dinâmica de avalanches de campo médio. No terceiro trabalho, abordamos o problema de múltiplas escolhas por meio de um modelo que inclui a possibilidade de indecisão e onde as escolhas dos indivíduos evoluem segundo uma regra de pluralidade. Mostramos que essa dinâmica em redes com a propriedade de mundo pequeno produz diferentes estados estacionários realísticos, que dependem do número de alternativas e da distribuição de graus: consenso, distribuição de adoções larga similar à reais e situações onde a indecisão predomina, quando o número de alternativas é suficientemente grande. Por último, investigamos o surgimento de posições extremas na sociedade, mediante pesquisas em uma ampla gama de questões. O aumento de atitudes extremas tem como precursor uma relação não linear entre a fração de extremistas e a de moderados. Propomos um modelo, com regras de ativação baseadas na teimosia dos indivíduos, que permite interpretar o início da não linearidade em termos de uma transição abrupta do tipo percolação de inicialização onde

PUC-Rio - Certificação Digital Nº 1122074/CA

acontecem cascatas de extremismo. Como conclusão geral, destacamos que esta tese ilustra como os modelos de opinião, aliados às enormes bases de dados, fornecem resultados com poder de interpretação e predição dos padrões empíricos.

PUC-Rio - Certificação Digital Nº 1122074/CA

Palavras–chave

Redes Complexas ; Modelos de Opinião ; Simulações computacionais .

Abstract

Ramos, Marlon Ferreira; Anteneodo de Porto, Celia Beatriz (advisor). **Opinion dynamics in complex networks**. Rio de Janeiro, 2015. 142p. PhD Thesis — Departamento de Física, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis addresses several problems that can be treated through models of opinion dynamics, according to which individuals, connected according to complex networks, interact through rules that shape their preferences and opinions in relation to a particular issue. The methodology used to investigate the patterns that emerge from those interactions relies on the use of various techniques of statistical physics. The thesis is organized around four distinct problems, with a particular question to be answered in each case, always looking for empirical validation of the theoretical and computational results. In the first work, it is answered the following basic question about network properties that can have impact on the spreading processes: what are the typical values of the distances, clustering coefficient and other structural quantities, when considering the ensemble of random networks with fix assortativity? In the second study, we investigated the patterns that emerge in the ratings of films, considering as source IMDb (Internet Movie Database). We found that the distribution of votes has a scale-free behavior with a exponent close to 3/2. Interestingly, this pattern is robust, independently of movie attributes such as average note, age or gender. The empirical analysis points to a simple mechanism of adoption propagation, that generates mean-field avalanches. In the third study, we discuss the problem of multiple choices by means of a model which includes the possibility of indecision and where the choices of individuals evolve according to a plurality rule. We show that this dynamics on top of networks with the small-world property produces different stationary states that depend on the number of alternatives and on the degree distribution: consensus, wide adoption distributions similar to actual ones and situations where indecision prevails when the number of alternatives is large enough. Finally, we investigate the appearance of extreme positions in society, through the polls on a wide variety of questions. The increase of extreme opinions has as precursor a non-linear relationship between the fraction of extremists and that of moderates. We propose a model with activation rules, based on the stubbornness of the individuals, which enables interpreting the beginning of the non-linearity in terms of an abrupt transition of the class of bootstrap percolation, where activation cascades occur. As a general conclusion, we emphasize that this

PUC-Rio - Certificação Digital Nº 1122074/CA

thesis illustrates how opinion models, combined with huge databases, provide results with power of interpretation and prediction of empirical patterns.

PUC-Rio - Certificação Digital Nº 1122074/CA

Keywords

Complex Networks ; Opinion Models ; Computer Simulations .

Sumário

1 Prefácio **11**

2 Introdução **16**
2.1 Redes complexas 16
2.2 Sociofísica 30
2.3 Correlações de longo alcance e universalidade 33
2.4 Metodologia 37

3 Correlações em redes complexas **40**
3.1 Introdução 40
3.2 Redes e ensembles 44
3.3 Resultados 45
3.4 Discussão e considerações finais 55

4 Padrões estatísticos na avaliação de produtos **59**
4.1 Introdução 59
4.2 Resultados 61
4.3 Discussão 73

5 Implicações da regra da pluralidade em situações com múltiplas opções **76**
5.1 Introdução 76
5.2 A regra da pluralidade aplicada a um modelo de agentes 79
5.3 Resultados 81
5.4 Considerações Finais 90

6 Modelo não consensual com opiniões contínuas **92**
6.1 Introdução 92
6.2 Resultados 95
6.3 Considerações finais 112

7 Conclusão e considerações finais **113**

A Correlação em redes complexas - Cálculos **115**
A.1 Coeficiente de correlação de Pearson 115
A.2 Decomposição de r 116

B Base de dados do IMDb **119**

PUC-Rio - Certificação Digital Nº 1122074/CA

C Regressão não paramétrica **120**

D Descrição dos levantamentos usados na Fig.6.2 **121**

Referências Bibliográficas **128**

PUC-Rio - Certificação Digital Nº 1122074/CA

1
Prefácio

Aos olhos do público, físicos se ocupam em explicar fenômenos "naturais" como o comportamento de átomos, ou o movimento dos planetas. Poucas pessoas arriscariam dizer que a física poderia também acrescentar algo relacionado ao comportamento dos seres humanos. No fundo, a ideia de se aplicar uma modelagem matemática ou estatística que explicaria como humanos se comportam e se organizam parece ficção científica, e de fato, esse é um tema que já foi explorado em algumas obras, como em *Fundação* de Isaac Asimov. Nessa obra clássica, o cientista Hari Seldon apresenta uma nova ciência chamada de psico-história que é capaz de predizer o comportamento de uma sociedade [1].

> "O ser humano individual é imprevisível, porém as reações das multidões humanas, descobriu Seldon, podem ser tratadas estatisticamente. Quanto maior a multidão, tanto maior a precisão que pode conseguir-se. E a grandeza das massas humanas com que Seldon trabalhava era nada menos do que a população da Galáxia que, no seu tempo, se contava por quintilhões. Foi Seldon, pois, quem previu, contra todo o senso comum e a crença popular, que o brilhante Império que parecia tão forte achava-se num estado de decadência e declínio irremediáveis. Previu (ou resolveu as suas equações e interpretou os seus símbolos, o que vem a dar na mesma) que, entregue a si mesma, a Galáxia viria a atravessar um período de trinta mil anos de miséria e anarquia antes de se estabelecer mais uma vez um governo unificado".

A série *Fundação* inspirou muitos escritores e cientistas. Paul Krugman, ganhador do prêmio Nobel de economia, uma vez declarou que seu sonho na adolescência era ser um psico-historiador, mas como essa profissão não existia, decidiu se tornar economista, que na época ele acreditava ser a ciência real que mais se aproximava da psico-história [2]. É fato que a economia tem muito a dizer sobre o comportamento dos seres humanos, mas acredito que a área da física que trata de fenômenos sociais (ou sociofísica) se aproxima mais da ciência idealizada por Asimov. Por mais improvável que pareça ser, nos últimos anos tem surgido leis "físicas" sobre o comportamento humano [3]. Os esforços para se chegar a essas leis fazem parte de uma tradição antiga. Na verdade, a ideia de uma modelagem "física" para problemas sociais é até mais antiga do

PUC-Rio - Certificação Digital Nº 1122074/CA

que a própria física estatística. Nomes como Hobbes, Laplace, Comte, Stuart Mill fizeram incursões nessa área [3, 4].

Em 1831, o astrônomo belga Adolphe Quetelet, considerado o fundador do que chamamos hoje de sociofísica [5], publicou o trabalho *Physique sociale* [6], uma tentativa de encontrar "leis da sociedade" análogas às leis de Newton. Nesse trabalho, Quetelet traça analogias diretas entre a ordem apresentada pelo sistema solar com a ordem observada na sociedade. Quetelet intenta mostrar que o estudo da sociedade pode ser feito com regras baseadas nas leis da astronomia. O trabalho de Quetelet foi bastante importante na época, tendo influenciado nomes como Karl Marx [4]. Foi também uma influência para que Henry Thomas Buckle, historiador inglês, publicasse sua obra *História da Civilização na Inglaterra* [7] em 1857.

O livro *História da Civilização na Inglaterra* ajudou a moldar o pensamento dos intelectuais ingleses no século XIX. Nesta obra, Buckle se aprofunda em vários exemplos que também foram estudados por Quetelet, como taxas de natalidade e mortalidade, estatísticas de crime, suicídio, casamento, etc. Surpreendentemente um dos primeiros leitores do trabalho de Buckle foi James Clerk Maxwell. Nos seus estudos sobre gases, cujos constituintes são moléculas em constante colisão umas com as outras, Maxwell reconheceu certas similaridades com os problemas analisados por Buckle [8, 4]:

> "those uniformities which we observe in our experiments with quantities of matter containing millions of millions of molecules are uniformities of the same kind as those explained by Laplace and wondered at by Buckle arising from the slumping together of multitudes of causes each of which is by no means uniform with the others".

Os trabalhos de Maxwell sobre a distribuição de velocidades de um gás foram aprofundados e colocados dentro de um arcabouço maior, que hoje chamamos de mecânica estatística, por Ludwig Boltzmann em 1872. Boltzmann também fez analogias entre as partículas que eram o objeto de seu estudo e os indivíduos nas pesquisas sociais que fundamentavam o trabalho de Buckle [9, 4]:

> "The molecules are like to many individuals, having the most various states of motion, and the properties of gases only remain unaltered because the number of these molecules which on the average have a given state of motion is constant".

Mesmo que o tópico não seja nenhuma novidade para a comunidade física, somente nos últimos anos o ferramental da física estatística passou de uma

PUC-Rio - Certificação Digital Nº 1122074/CA

possibilidade quase filosófica para uma alternativa real de se estudar problemas de sociofísica, com um número crescente de pesquisadores na área [3]. Esta tese, se encaixa dentro deste cenário: através de uma abordagem de física estatística, exploramos alguns mecanismos de interação em sistemas compostos de "agentes sociais" que se relacionam através de redes com topologia aleatória e que dão origem a comportamentos emergentes e auto-organização.

A estrutura da tese A tese está organizada em torno de quatro problemas distintos que desenvolvemos ao longo do doutorado. Esses trabalhos são auto-contidos e de certa forma independentes. Organizamos a tese de tal forma que pudesse haver uma conexão entre eles, e para que o texto como um todo tivesse coerência. Ainda assim, pode-se lê-los separadamente. Em outras palavras, cada capítulo tem uma pergunta diferente a ser respondida, buscando sempre a validação empírica dos resultados teóricos e computacionais.

Apesar disso, há uma unidade na tese, há conceitos e metodologias que permeiam todo o texto. Primeiro, o contexto da pesquisa tem um enfoque em problemas sociais, onde os constituintes do sistema são pessoas. Segundo, temos a aplicação de conceitos de física estatística e por fim, usamos o computador como principal ferramenta, através de simulações computacionais. O tema principal da tese está vinculado a "dinâmicas de opinião": como as interações entre as pessoas moldam as preferências e o posicionamento dos indivíduos em relação a alguma questão.

Posto isso, vamos desenvolver os conceitos e ideias mais gerais, bem como parte da metodologia empregada, no Cap. 2, *Introdução*. A utilização de redes complexas, por exemplo, está presente ao longo de todo texto e por isso esse tópico merece ser detalhado a parte, como faremos no Cap. 2. Outros tópicos como modelos de agentes, correlações de longo alcance e transições de fase também serão apresentados no Cap. 2. Por outro lado, cada capítulo faz uso de certos conceitos e ferramentas específicos, que são aplicados principalmente no problema que está em foco, e são menos usados em outras partes da tese, por esse motivo serão desenvolvidos a medida em que eles sejam necessários.

No primeiro trabalho, apresentado no Cap. 3, *Correlações em redes complexas* é respondida a seguinte questão básica sobre propriedades da rede que podem ter impacto sobre os processos de propagação: quais são os valores típicos das distâncias, coeficiente de aglomeração e outras grandezas estruturais da rede, quando considerado o ensemble de redes aleatórias com uma assortatividade fixa? Esse capítulo é baseado no artigo da Ref. [10].

No segundo trabalho, investigamos os padrões que surgem na avaliação de filmes, considerando como fonte o IMDb (Internet Movie Database). Este

PUC-Rio - Certificação Digital Nº 1122074/CA

trabalho é exposto no Cap. 4, *Padrões estatísticos na avaliação de produtos*. Encontramos que a distribuição de votos apresenta um comportamento livre de escala com um expoente muito próximo de 3/2. Curiosamente, esse padrão é robusto, independente de atributos dos filmes como nota média, idade ou gênero. A análise empírica aponta para um mecanismo de propagação de adoções simples, que gera uma dinâmica de avalanches de campo médio.

No terceiro trabalho, apresentado no Cap. 5, *Implicações da regra da pluralidade em situações com múltiplas opções*, abordamos o problema de múltiplas escolhas por meio de um modelo que inclui a possibilidade de indecisão e onde as escolhas dos indivíduos evoluem segundo uma regra de pluralidade. Mostramos que essa dinâmica em redes com a propriedade de mundo pequeno produz diferentes estados estacionários realísticos, que dependem do número de alternativas e da distribuição de graus: consenso, distribuição de adoções larga similar à reais e situações onde a indecisão predomina, quando o número de alternativas é suficientemente grande.

Por último, no Cap. 6, *Modelo não consensual com opiniões contínuas*, investigamos o surgimento de posições extremas na sociedade, mediante levantamentos em uma ampla gama de questões. O aumento de atitudes extremas tem como precursor uma relação não linear entre a fração de extremistas e a de moderados. Propomos um modelo, com regras de ativação baseadas na teimosia dos indivíduos, que permite interpretar o início da não linearidade em termos de uma transição abrupta do tipo percolação de ativação onde acontecem cascatas de extremismo. Este trabalho foi desenvolvido parcialmente durante um período de doutorado sanduíche no Levich Institute of New York, onde fui co-orientado pelo Prof. Hernán Makse e pode ser encontrado na Ref. [11].

Essa não é uma tese em física teórica, também não é uma tese de física experimental. Trabalhamos numa espécie de interface entre esses dois ramos, onde o computador é a principal ferramenta. De fato, essa é uma nova vertente da física, que damos o nome de física computacional. Junto com os dois ramos tradicionais da física, integra-se a física computacional como uma maneira adicional de se fazer física e compreender a natureza. Como uma tese de física computacional esse texto tem algumas particularidades, o leitor vai perceber que há menos equações e mais gráficos que em uma tese de física teórica, por exemplo. O campo de especialização é sistemas complexos, subárea da física estatística. Uma caraterística marcante de sistemas complexos é a interdisciplinaridade; além dos temas estudados nesta tese, redes e dinâmicas de opinião, também estão dentro do escopo dos sistemas complexos temas como

PUC-Rio - Certificação Digital Nº 1122074/CA

linguística, ecologia, geofísica, ecologia e trânsito [3, 12].

PUC-Rio - Certificação Digital Nº 1122074/CA

2
Introdução

2.1
Redes complexas

Existem muitos sistemas de interesse físico, tecnológico e social que podem ser modelados com redes ou grafos. Exemplos típicos são a internet, as cadeias alimentares e as redes sociais online. A definição mais simples para uma rede é um conjunto de pontos (vértices ou nós) conectados por linhas. As redes podem ser uma representação útil de sistemas compostos por muitos elementos que interagem entre si, os componentes do sistema são representados pelos pontos e as conexões entre os elementos são representadas pelas linhas.

A estrutura de tais redes pode ter um grande efeito sobre o comportamento do sistema como um todo. Por exemplo, as conexões entre computadores na internet afetam como os dados são transmitidos. Em uma rede social cada sítio representa uma pessoa ou grupo de pessoas, e as ligações representam alguma forma de interação entre elas. As conexões em redes sociais influenciam como as pessoas aprendem, formam suas opiniões e até contaminam umas às outras numa epidemia. A menos que saibamos alguma informação a respeito da estrutura dessas redes, não podemos ter a esperança de entender como os sistemas correspondentes funcionam.

Tradicionalmente, o estudo de redes sociais (e tecnológicas) é feito usando a modelagem do grafo aleatório, conhecido como grafo de Erdős e Rényi: considera-se uma rede com N nós, onde cada par de nós está conectado independentemente com uma certa probabilidade p. Esse paradigma começou a mudar com o trabalho publicado na revista Nature em 1998 por Watts e Strogatz [13] sobre redes de mundo pequeno, e com o trabalho de Barabási e Albert publicado na revista Science [14] um ano depois sobre redes livres de escala. Uma rede complexa é um grafo com propriedades topológicas não triviais, e as redes estudadas nos trabalhos supracitados se encaixam nesta classificação. A partir daí o número de publicações em revistas de física sobre redes complexas disparou, passando a ser um dos tópicos mais populares na mecânica estatística. A Fig. 2.1 mostra o número de artigos publicados no ArXiv na categoria “Physics” com as palavras “complex networks”, o gráfico mostra o surgimento desse novo tópico e sua crescente popularidade nos anos posteriores às publicações mencionadas acima.

PUC-Rio - Certificação Digital Nº 1122074/CA

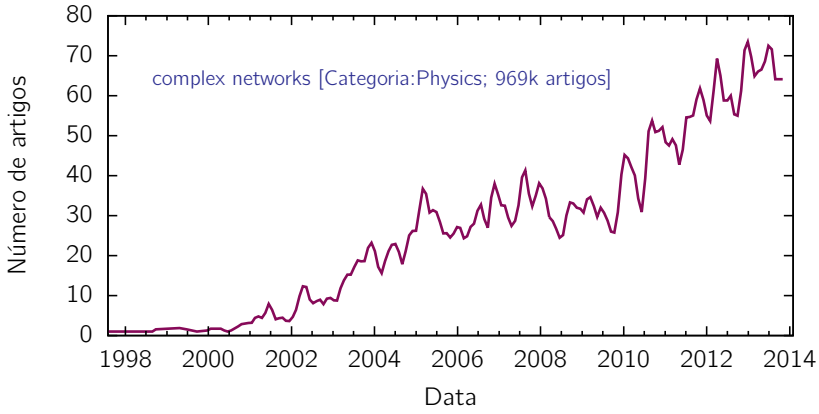


Figura 2.1: Número de artigos publicados no ArXiv dentro da categoria "Physics", contendo as palavras "complex networks". Os dados desse gráfico foram coletados usando a ferramenta *Bookworm* do site `http://arxiv.culturomics.org`

PUC-Rio - Certificação Digital Nº 1122074/CA

Percebeu-se, no final dos anos 90, que o grafo de Erdős e Rényi não descrevia de maneira satisfatória muitas das propriedades das redes reais (nesse momento, as informações a respeito de tais redes eram muito maiores), por isso novos modelos foram, e continuam sendo, propostos como alternativa.

O número de ligações de cada nó, *i.e.*, o número de vizinhos, é chamado de conectividade, ou grau. Existe uma variabilidade no número de ligações de cada vértice, que é quantificada por uma função chamada de distribuição de conectividade (ou de graus) $P(k)$, essa função nos dá a probabilidade de que um vértice da rede, escolhido ao acaso, tenha exatamente k ligações. Uma das características mais marcantes das redes reais é a distribuição de conectividade. No grafo de Erdős e Rényi, como veremos na sequência, a distribuição de conectividade segue uma distribuição de Poisson e o que se observa em muitas redes reais são distribuições largas, na forma de lei de potência. O modelo de Barabási ficou muito famoso justamente por propor um mecanismo de construção da rede que produz uma distribuição de conectividade mais realista.

O conjunto $\{k_1, k_2, k_3, \ldots\}$ dos graus de cada um dos vértices é chamado de *sequência de graus* e contem basicamente a mesma informação que a distribuição de conectividade. Vale dizer, que o conhecimento da distribuição de conectividade ou da sequência de graus não garante o conhecimento da estrutura total da rede, *i.e.*, redes diferentes podem ter a mesma sequência de graus [15].

De maneira geral, ligações entre os vértices podem ter uma direção, *i.e.*, uma ligação *aponta* de um vértice para o outro. Neste caso dizemos que a

rede é direcionada (ou orientada). Também podemos ter casos onde as ligações têm um peso. Por exemplo, ligações na internet poderiam ter um peso que representaria a quantidade de dados que passam por uma ligação. Além disso, algumas redes podem apresentar ligações múltiplas, onde há mais de uma ligação entre dois vértices, e até autoligações, (*self-loops* em inglês) onde um vértice está ligado com ele mesmo [15]. Aqui, vamos considerar *grafos simples*, *i.e.*, sem ligações multiplas ou autoligações. Também consideraremos redes não direcionadas e ligações sem peso.

Uma das formas de se representar matematicamente uma rede é através da matriz de adjacência $\mathbf{A}$, que para uma rede simples, é definida como

$$A_{ij} = \begin{cases} 1 & \text{se existe uma ligação entre os sítios } i \text{ e } j \\ 0 & \text{caso contrário.} \end{cases} \tag{2-1}$$

A matriz de adjacência permite calcular o grau k_i do sítio i por meio das equações

$$k_i = \sum_j A_{ij} = \sum_j A_{ji}, \quad \text{daí} \quad \langle k \rangle = \frac{1}{N}\sum_i k_i = \frac{1}{N}\sum_{ij} A_{ij}. \tag{2-2}$$

Como o somatório de k_i em todos os sítios da rede é igual a $2M$, também podemos usar a matriz $\mathbf{A}$ para calcular o número de ligações da rede,

$$M = \frac{1}{2}\sum_i k_i = \frac{1}{2}\sum_{ij} A_{ij}. \tag{2-3}$$

Grafo de Erdős e Rényi Há duas maneiras equivalentes de se definir o grafo de Erdős e Rényi. Uma delas, que chamaremos de modelo $\mathrm{G}(M, N)$, é fixando o número de vértices N e o número de ligações M: tomamos N vértices e colocamos M ligações, ao acaso, entre eles, ou seja, escolhemos M pares de vértices aleatoriamente de todos os possíveis pares e os conectamos com uma ligação. A conectividade média do grafo resultante é dada por $\langle k \rangle = 2M/N$.

A outra maneira é fixar a *probabilidade* de conexão entre os vértices. Definimos uma rede com N vértices e conectamos cada par de sítios, de maneira independente, com probabilidade p. Essa formulação é conhecida como modelo $\mathrm{G}(N, p)$. Nessa formulação, a conectividade média é dada pela equação

$$\langle k \rangle = (N-1)p, \tag{2-4}$$

ou seja, o número esperado de ligações de um vértice é igual à probabilidade de conexão entre esse vértice e outro vértice da rede vezes o número $N-1$ de outros vértices. No modelo $\mathrm{G}(N, p)$ alguns cálculos são mais tratáveis, como o cálculo de $P(k)$. Um vértice i da rede é ligado de maneira independente com

PUC-Rio - Certificação Digital Nº 1122074/CA

probabilidade p com os outros $N-1$ vértices da rede. Portanto, a probabilidade do sítio i estar conectado com outros k vértices *em particular* e não estar conectado com nenhum outro é $p^k(1-p)^{N-1-k}$. Temos $\binom{N-1}{k}$ maneiras de se escolher k outros vértices dos $N-1$ disponíveis. Consequentemente, a probabilidade do sítio i estar conectado com exatamente k outros sítios é dada por

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \tag{2-5}$$

ou seja, o modelo $\mathrm{G}(N,p)$ tem distribuição de conectividade binomial. Em problemas de dinâmica social podemos considerar que o número de sítios N, que irá representar o número total da população, é grande. Além disso, em muitas redes sociais reais a conectividade média é independente do tamanho da rede [15], ou posto de outra forma: o número típico de amigos de uma pessoa não depende de maneira significativa do tamanho total da população. Dadas essas observações, podemos simplificar a Eq. 2-5. No limite de $N \to \infty$, $p = \langle k \rangle / (N-1)$ tende a um valor pequeno, o que justifica escrever,

$$\begin{aligned} \ln[(1-p)^{N-1-k}] &= (N-1-k)\ln\left(1 - \frac{\langle k \rangle}{N-1}\right) \\ &\simeq -(N-1-k)\frac{\langle k \rangle}{N-1} \simeq -\langle k \rangle . \end{aligned} \tag{2-6}$$

Na equação acima expandimos o logaritmo em série de Taylor e tomamos a exponencial dos dois lados. Como resultado temos que $(1-p)^{N-1-k} = \mathrm{e}^{-\langle k \rangle}$. Para N grande, obtemos

$$\binom{N-1}{k} = \frac{(N-1)!}{(N-1-k)!k!} \simeq \frac{(N-1)^k}{k!} \tag{2-7}$$

e, com isso, a Eq. 2-5, fica

$$\begin{aligned} P(k) = \frac{(N-1)^k}{k!} p^k \mathrm{e}^{-\langle k \rangle} &= \frac{(N-1)^k}{k!}\left(\frac{\langle k \rangle}{N-1}\right)^k \mathrm{e}^{-\langle k \rangle} \\ &= e^{-\langle k \rangle}\frac{\langle k \rangle^k}{k!}. \end{aligned} \tag{2-8}$$

Vemos então que o modelo $\mathrm{G}(N,p)$ tem distribuição de conectividade de Poisson no limite de $N \to \infty$. A maior parte dos nós tem a mesma conectividade, igual a $\langle k \rangle$, e conectividades mais altas, ou mais baixas, têm probabilidade consideravelmente menor. A Fig. 2.2, mostra a distribuição resultante do modelo $\mathrm{G}(N,M)$ para uma simulação com $N = 10^5$ sítios e diferentes conectividades médias.

PUC-Rio - Certificação Digital Nº 1122074/CA

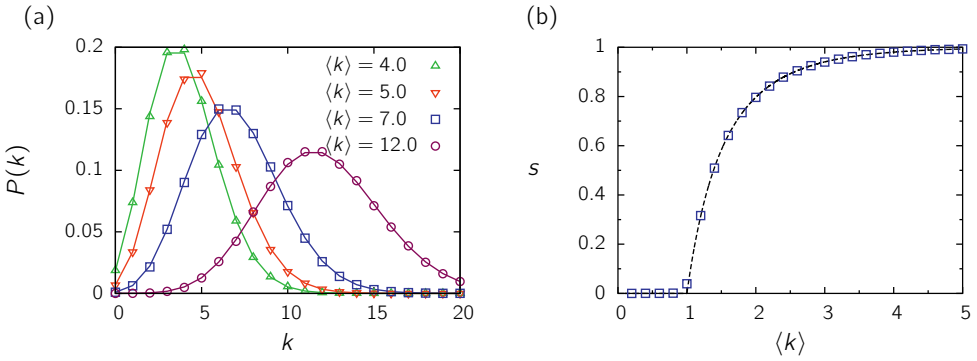


Figura 2.2: (a) Distribuição de conectividade para o grafo de Erdős e Rényi. Na figura é mostrado pontos resultantes de uma simulação com uma rede de tamanho $N = 10^5$ e diferentes valores de $\langle k\rangle$. As linhas foram calculadas a partir da distribuição de Poisson, Eq. 2-8 e conectando os pontos. (b) Fração de sítios s pertencentes à maior componente para redes de ER, a linha pontilhada é a solução da Eq. 2-9, os pontos são resultados de simulação para uma rede de ER de tamanho $N = 10^5$.

Componentes Um *caminho* em uma rede é definido como uma sequência de vértices tal que cada par de vértices nessa sequência está conectado por uma ligação da rede. No grafo de ER pode não existir um caminho entre cada par de vértices da rede, já que as ligações são feitas de maneira aleatória. Nessa situação dizemos que a rede é desconectada. No caso oposto, onde existe um caminho entre cada um dos pares de sítios, a rede é dita conectada.

Cada subgrupo de sítios conectados é chamado de *componente*. De maneira formal: uma componente é um subconjunto de vértices da rede tal que existe pelo menos um caminho de cada membro do subconjunto para todos os outros membros, e tal que nenhum outro vértice da rede pode ser adicionado ao subconjunto sem que seja violada essa propriedade [15].

Vamos considerar dois casos limites da formulação $\mathrm{G}(N, p)$. Se $p = 0$ não existe nenhuma ligação e a rede é totalmente desconectada, ou seja, cada vértice é uma componente e a rede tem, portanto, N componentes, todas com um único vértice. No outro extremo, se $p = 1$, cada ligação possível está presente, e cada um dos vértices está conectado com todos os outros. Portanto, todos os vértices da rede fazem parte de uma só componente. Note o seguinte: enquanto no primeiro caso, para $p = 0$, a maior componente S tem tamanho 1, no segundo caso, para $p = 1$, o tamanho da maior componente tem o mesmo tamanho da rede N. A diferença qualitativa crucial é que no primeiro caso o tamanho da maior componente é independente do tamanho da rede. No segundo caso, S é proporcional a N e, se aumentarmos o valor de N, a maior componente vai acompanhar o crescimento da rede. Neste último

PUC-Rio - Certificação Digital Nº 1122074/CA

caso, podemos dizer que S é uma grandeza extensiva.

Em modelos de opinião é pertinente termos uma componente que cobre a maior parte da rede. Consequentemente, é importante entender como acontece a transição entre os dois regimes acima. Começando de $p = 0$ e gradualmente aumentando o valor de p, e consequentemente de $\langle k \rangle$, o tamanho da maior componente passa por uma transição de fase; de um tamanho constante (independente de N) para um tamanho extensivo. A transição ocorre para o valor $\langle k \rangle = 1$.

Uma componente da rede que cresce em proporção ao número de sítios é chamada de *componente gigante*. No limite $N \to \infty$ é possível calcular S exatamente para a rede de ER. Pode-se mostrar que a *fração* s de vértices que pertencem à maior componente do grafo de ER pode ser computada pela equação

$$s = 1 - \mathrm{e}^{-\langle k \rangle s}. \tag{2-9}$$

A Eq. 2-9 não apresenta uma solução fechada, mas podemos entender o comportamento de s através de uma solução gráfica, ou iterando a equação no computador. Na figura 2.2b, mostramos o resultado dessa iteração, e também o resultado de simulações para uma rede de ER de tamanho $N = 10^5$.

Para valores pequenos de $\langle k \rangle$, menores que 1, não existe componente gigante. Aumentando-se gradativamente o valor $\langle k \rangle$, há o surgimento de uma componente gigante. Para $\langle k \rangle = 5$, s já ocupa mais de 99% da rede.

Rede de Barabási e Albert Como foi mencionado acima, muitas redes reais têm uma distribuição de graus que se desvia da distribuição homogênea do grafo de ER, boa parte delas tem distribuição $P(k)$ com lei de potência, comumente referidas como *redes livres de escala*.

Em 1999, Barabási e Albert estudaram os mecanismos que poderiam ser a origem desse tipo de distribuição. Ao contrário do grafo de ER, onde o número N de nós é fixo, a rede de Barabási é um modelo de rede onde nós são continuamente adicionados. Esse primeiro mecanismo é baseado no argumento de que redes reais não são estáticas, mas estão em constante crescimento [16]. O segundo mecanismo é conhecido como *anexação preferencial* (*preferential attachment* em inglês); a probabilidade de que dois sítios estejam conectados não é uniforme e independente da conectividade dos sítios. No modelo de Barabási, leva-se o grau do sítio em consideração, de tal forma que vértices que tenham maior grau, têm maior chance de receber uma nova ligação. Implementa-se o modelo de Barabási da seguinte maneira: começamos com um número pequeno (m_0) de nós e a cada passo de tempo é adicionado um

PUC-Rio - Certificação Digital Nº 1122074/CA

novo sítio que se conecta com outros sítios já presentes na rede, através de $m(\leqslant m_0)$ ligações. A probabilidade Π de que um novo nó se conecte a um vértice i já existente é proporcional ao grau k_i do sítio i, ou seja

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{2-10}$$

Após t passos de tempo da iteração da receita acima temos uma rede com $N = t + m_0$ sítios e $M = mt$ ligações. Na Fig. 2.3, mostramos a distribuição resultante de uma simulação da rede de Barábasi com $N = 10^6$ sítios, para diferentes valores do parâmetro m. A distribuição de graus $P(k)$ segue uma lei de potência com expoente $\gamma = 3$. Note que o expoente não depende do parâmetro m.

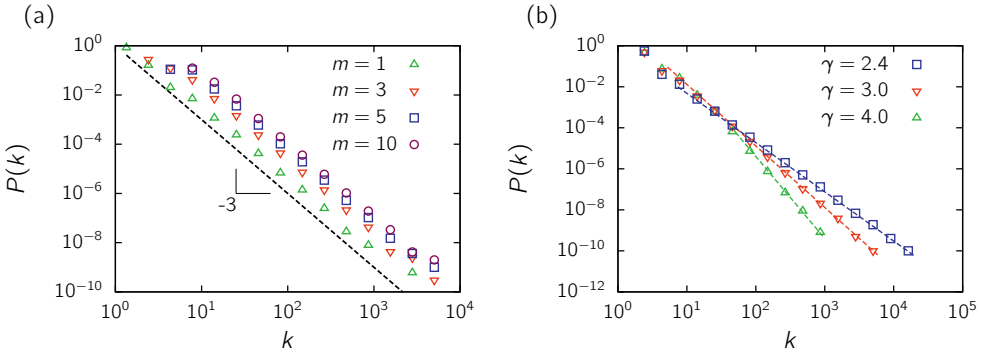


Figura 2.3: Distribuição de conectividade para simulações do modelo de Barabási (a) e do modelo de Price (b). Na figura (a) usamos uma rede de tamanho $N = 10^6$ e diferentes valores de m. A linha pontilhada com inclinação -3 é apresentada para comparação. Na figura (b), escolhemos o valor do parâmetro A de tal forma a ajustar o valor dos expoentes mostrados na figura, usamos uma rede de $N = 10^7$ e $m = 2$.

Existem algumas abordagens analíticas para estudar o modelo de Barábasi (e outros modelos semelhantes), sendo as principais: a teoria contínua proposta no trabalho original de Barabási e Albert [14], a abordagem através de equações mestras [17] e a abordagem mediante uma equação para a taxa [18]. Os resultados desses diferentes métodos são equivalentes na maior parte dos casos. Aqui vamos seguir a solução introduzida por Barabási.

Sendo um modelo dinâmico, vamos começar calculando a dependência de k_i (grau do sítio i) com o tempo. Sempre que um novo sítio se conecta ao sítio i, o valor de k_i aumenta. A probabilidade de ocorrência desse evento é $\Pi(k_i)$. Vamos fazer a aproximação de que k_i é uma variável contínua (daí o nome desse método). Sendo assim, a taxa com a qual k_i aumenta é proporcional a $\Pi(k_i)$ e, portanto, k_i satisfaz a seguinte equação

PUC-Rio - Certificação Digital Nº 1122074/CA

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}. \tag{2-11}$$

A soma do denominador inclui todos os nós com exceção dos sítios que acabaram de ser adicionamos. Supondo que no instante de tempo considerado já foram adicionados muitos sítios no sistema, podemos escrever $\sum_j k_j = 2M - m = 2mt - m \approx 2mt$. Com isso, a Eq. 2-11 toma a forma

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \tag{2-12}$$

A solução da equação acima é obtida utilizando a condição inicial $k_i(t_i) = m$, ou seja, no instante de tempo t_i quando o nó i é introduzido na rede, sua conectividade é m e, consequentemente

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}, \qquad \text{com} \quad \beta = \frac{1}{2}. \tag{2-13}$$

A expressão acima diz que o grau de todos os nós evolue seguindo uma lei de potência.

A probabilidade de que um sítio com grau $k_i(t)$ seja menor que k, $P(k_i(t) < k)$, pode ser escrita, usando a Eq. 2-13, como

$$P(k_i(t) < k) = P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right). \tag{2-14}$$

Como a cada passo de tempo introduzimos um sítio na rede, a densidade de probabilidade de t_i é constante e dada por

$$P(t_i) = \frac{1}{m_0 + t} = \text{constante}. \tag{2-15}$$

Substituindo a equação acima na Eq. 2-14, obtemos o resultado

$$P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta}t}{k^{1/\beta}(t + m_0)}, \tag{2-16}$$

onde também usamos que $P(t_i > t') + P(t_i < t') = 1$ e $P(t_i < t') = \int_0^{t'} P(t_i)dt_i$. Com o resultado da Eq. 2-16, podemos obter a distribuição de graus

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^{1/\beta}t}{m_0 + t}\frac{1}{k^{1/\beta+1}}. \tag{2-17}$$

No limite de $t \to \infty$, obtemos

$$P(k) \sim 2m^{1/\beta}k^{-\gamma}, \quad \text{com} \quad \gamma = \frac{1}{\beta} + 1 = 3. \tag{2-18}$$

A Eq. 2-18 mostra que a rede, apesar de crescer continuamente, atinge uma distribuição de graus estacionária, *i.e.*, independente do tamanho N. Note também que o expoente $\gamma = 3$ é independente de m, como nos resultados numéricos (Fig. 2.3).

PUC-Rio - Certificação Digital Nº 1122074/CA

A anexação preferencial do modelo de Barabási é também conhecida como *vantagem cumulativa* e aparece em outras áreas. Na economia, por exemplo, a taxa com que as pessoas ganham dinheiro é proporcional à quantidade de dinheiro que elas já possuem. Vale dizer que nesse caso a vantagem cumulativa foi usada para explicar alguns dados sobre a distribuição de riquezas que também seguem uma lei de potência [19]. Traduzindo a vantagem cumulativa em outros contextos: uma página da internet tem maior probabilidade de ter um novo link que aponta para ela se a página *já* tem muitos links; outro exemplo: um artigo recém publicado tem maior probabilidade de citar artigos que já são conhecidos e têm muitas citações do que citar artigos poucos conhecidos e com poucas citações.

Poderíamos ter também um modelo de anexação preferencial onde o número de novas ligações que um sítio recebe é proporcional ao grau do sítio *mais* uma constante A. Pensando ainda em redes de citações: seria como se uma fração das novas citações fossem uniformemente distribuídas entre os artigos sem que seja levado em consideração o número de citações que o artigo já tem; a outra parte seria escolhida de acordo com a contagem de citações. Curiosamente, o modelo acima foi proposto por Derek Price nos anos de 1970 [20]. Nessa situação a probabilidade de anexação $\Pi(k_i)$ é dada por

$$\Pi = \frac{k_i + A}{N(\langle k \rangle + A)}. \tag{2-19}$$

É possível mostrar, fazendo uso de equações mestras [15], que com a probabilidade de anexação acima, a distribuição resultante de graus apresenta uma cauda em lei de potência

$$P(k) \sim k^{-\gamma}, \quad \text{com} \quad \gamma = 3 + \frac{A}{m}, \tag{2-20}$$

ou seja, ajustando o valor do parâmetro A, podemos construir redes com distribuição de graus que têm decaimento em lei de potência e expoentes diferentes de 3. Na Fig. 2.3b, apresentamos o resultado de simulações para o modelo de Price. Em relação às componentes, vale dizer, que pela natureza da regra de criação dessas redes dinâmicas, só teremos uma única componente que engloba toda a rede.

Redes livres de escala têm características bem diferentes de redes de ER, enquanto a grande maioria dos vértices tem conectividade baixa, existe uma cauda na distribuição com vértices que apresentam conectividade excepcionalmente grande. Esses vértices são chamados de *hubs*. Por exemplo, na Fig. 2.3b, para $\gamma = 2,4$, onde a rede tem 10^7 sítios, está presente um vértice com 203725 conexões. Uma observação importante é que muitas redes reais contêm um número pequeno, mas significante, de hubs. Redes sociais, por

PUC-Rio - Certificação Digital Nº 1122074/CA

exemplo, frequentemente contêm alguns indivíduos com um número enorme de amigos; na internet, existem alguns sites com uma quantidade extremamente grande de links. De fato, a maior parte das redes do mundo real apresentam distribuição de graus heterogênea, enviesada à direita, com uma cauda de vértices com grande conectividade. A Tabela 2.1 mostra o valor de γ para algumas redes reais. Os hubs têm um papel importante em muitos modelos dinâmicos que usam a rede como substrato, inclusive modelos de dinâmica social.

A distribuição de graus não é a única propriedade interessante de redes complexas. Na sequência, vamos discutir brevemente e de maneira qualitativa algumas dessas outras propriedades.

Fenômeno mundo-pequeno Redes sociais, como as consideradas aqui, têm um grande número de pessoas. No entanto, a distância típica $\langle L \rangle$ entre os indivíduos é relativamente pequena. Esse fenômeno é conhecido como mundo pequeno (*small-world* em inglês). Em 2011, foi feita uma estimativa de que a distância média entre os membros do Facebook é de 4,74 ligações. Essa pesquisa foi feita pelo time de pesquisadores do Facebook e pode ser encontrada na Ref. [56]. No ano de 2011, a rede do Facebook contava com $7,21 \times 10^6$ usuários ativos e $6,9 \times 10^9$ ligações (ou "amizades") entre os usuários. A Tabela 2.1 mostra outros exemplos desse fenômeno. O experimento feito em 2011 com usuários do Facebook pode ser vista como uma versão moderna do famoso experimento de Milgram que popularizou o conceito de "6 graus de separação". Matematicamente, o fenômeno de mundo pequeno significa um crescimento lento entre a menor distância média $\langle L \rangle$ entre sítios e o tamanho da rede. A maior parte das redes reais, e também as artificiais, exibem essa característica de mundo pequeno. No grafo de ER, por exemplo, temos que $\langle L \rangle \sim \log(N)$.

Transitividade A transitividade é outra propriedade marcante das redes sociais. Podemos definir relações entre os vértices de uma rede como, por exemplo, a relação "conectado por uma ligação". A definição matemática de relação transitiva é a seguinte: uma relação "$\circ$" é chamada transitiva quando $a \circ b$ e $b \circ c$ juntos implicam que $a \circ c$. O primeiro exemplo que vem a mente é a igualdade, se $a = b$ e $b = c$ segue que $a = c$. Usando esse conceito de transitividade, teríamos que a relação "conectado por uma ligação" é transitiva quando um vértice a conectado a um vértice b, e b conectado a um vértice c implicam que a está conectado ao vértice c. Se todos os sítios da rede estivessem conectados por uma ligação, teríamos uma rede perfeitamente transitiva. Em

PUC-Rio - Certificação Digital Nº 1122074/CA

	Rede	Tipo	N	$\langle k \rangle$	s	$\langle L \rangle$	γ	C	r	Ref.
Social	Film Actors	ND	449.913	113,43	0,980	3,48	2,3	0,20	0,208	[21, 13]
	Companing directors	ND	7.673	14,44	0,876	4,60	-	0,59	0,276	[22, 23]
	Math coauthorship	ND	253.339	3,92	0,822	7,57	-	0,15	0,120	[24, 25]
	Physics coauthorship	ND	52.909	9,27	0,838	6,19	-	0,45	0,363	[26, 27]
	Biology coauthorship	ND	1.520.251	15,53	0,918	4,92	-	0,088	0,127	[26, 27]
	APC	D	16706	14,5	0,89	5,63	1,0	0,42	0.23	[27, 28]
	Telephone call graph	ND	47.000.000	3,16			2,1			[29, 30]
	Email messages	D	59.812	1,44	0,952	4,95	1,5/2,0			[31]
	Email address books	D	16.881	3,38	0,590	5,22	-	0,17	0,092	[32]
	Student dating	ND	573	1,66	0,503	16,01	-	0,005	-0,029	[33]
	Sexual contacts	ND	2.810				3,2			[34, 35]
Informação	www nd.edu	D	269.504	5,55	1,000	11,27	2,1/2,4	0,11	-0,067	[36, 37]
	www AltaVista	D	203.549.046	7,20	0,914	16,18	2,1/2,7			[38]
	Citation Network	D	783.339	8,57			3,0/-			[39]
	PGP	ND	10680	4,55	1,0	6.77	-	0.38	0.24	[40]
	Roget's Theasauus	D	1.022	4,99	0,977	4,87	-	0,13	0,157	[41]
	Word co-occurrence	ND	460.902	66,96	1,00		2,7			[42, 43]
Tecnológica	Internet	ND	10.697	5,98	1,000	3,31	2,5	0,035	-0,189	[44, 45]
	Power grid (PGR)	ND	4.941	2,67	1,000	18,99	-	0,10	-0,003	[13, 28]
	Train routes	ND	587	66,79	1,000	2,16	-		-0,033	[46]
	Software packages	D	1.439	1,20	0,998	2,42	1,6/1,4	0,070	-0,016	[47]
	Software classes	D	1.376	1,61	1,000	5,40	-	0,033	-0,119	[48]
	Electronic circuits	ND	24.097	4,34	1,000	11,05	3,0	0,010	-0,154	[49]
	Peer-to-peer (P2P)	ND	10876	7,35	1,0	4,43	4.0	0,005	-0.013	[50]
Biológica	Metabolic network	ND	765	9,64	0,996	2,56	2,2	0,090	-0,240	[51]
	Potrein interactions	ND	2.115	2,12	0,689	6,80	2,4	0,072	-0,156	[52]
	Marine food web	ND	134	4,46	1,000	2,05	-	0,160	-0,263	[53]
	Freshwater food web	D	92	10,84	1,000	1,90	-	0,200	-0,326	[54]
	Neural network	ND	307	7,68	0,967	3,97	-	0,180	-0,226	[13, 55]

Tabela 2.1: Propriedades básicas de algumas redes reais. São mostradas, o tipo de rede (direcionada, D, ou não direcionada, ND), o número de sítios N, a conectividade média $\langle k \rangle$, a fração de sítios na maior componente s, a menor distância média $\langle L \rangle$, o expoente γ (nos casos em que $P(k)$ segue uma lei de potência), o coeficiente de aglomeração C e a assortatividade. Onde não há dados disponíveis, temos entradas em branco. As redes destacadas em cores serão utilizadas na análise do Cap. 3. A Tabela acima foi adaptada da Ref. [15].

geral isto não ocorre, e as redes são só parcialmente transitivas já que, se a conhece b e b conhece c, nada garante que a conheça c. De qualquer forma, é mais provável que um amigo de um amigo meu seja meu amigo do que alguma pessoa escolhida ao acaso da população. Nesse contexto poderíamos dizer: "amigo de amigo meu é meu amigo também".

Quantifica-se nível de transitividade de uma rede através do *coeficiente de agrupamento* C (*clustering coefficient* em inglês). A maneira mais usual de se definir o coeficiente de agrupamento é baseada na observação de que se temos um caminho de tamanho 2, *e.g.*, abc, então é verdade que os vértices a e c têm um vizinho em comum, que nesse caso seria o vértice b. Se abc for um loop fechado, então a e c são vizinhos também. O coeficiente de agrupamento pode ser pensado então a fração de pares de pessoas que tem um vizinho em

PUC-Rio - Certificação Digital Nº 1122074/CA

comum que são também vizinhos. Traduzindo matematicamente:

$$C = \frac{3 \times \#\text{ de triângulos}}{\#\text{ de trios conectados}} \tag{2-21}$$

Na definição acima um "trio conectado" significa um grupo de 3 vértices *abc* com as ligações (a, b) e (b, c), podendo a ligação (a, c) estar ou não presente. O fator 3 no numerador vem do fato que cada triângulo é considerado 3 vezes quando se conta o número de trios conectados. Exemplo: o triângulo *abc* contém os trios *abc*, *bca* e *cab*. Se $C = 1$, temos uma rede perfeitamente transitiva e, se $C = 0$, não temos nenhum triângulo, o que, de fato, ocorre em alguns tipos de redes. Por exemplo, numa rede do tipo árvore, como nos modelos de crescimento que vimos anteriormente, ou numa rede quadrada. Vale dizer que o valor de C de uma rede social típica (alguns exemplos também podem ser encontrados na Tabela 2.1) é bem maior do que o valor de C de um grafo aleatório com o mesmo número de nós e ligações.

Além de definir o coeficiente de agrupamento para a rede, também podemos definir uma quantidade similar, o coeficiente de agrupamento local C_i para um único vértice através da equação

$$C_i = \frac{\#\text{ de pares de vizinhos de } i \text{ que estão conectados entre si}}{\#\text{ de pares de vizinhos de } i}. \tag{2-22}$$

O coeficiente de agrupamento local mede a probabilidade de que um par de amigos do sítio i sejam também amigos entre si. A média de C_i sobre todos os sítios da rede,

$$\bar{C} = \frac{1}{N}\sum_{i}^{N} C_i, \tag{2-23}$$

é muitas vezes usada como definição do coeficiente de agrupamento da rede como um todo. A definição acima foi proposta originalmente no trabalho de Watts e Strogatz [13].

Assortatividade por caraterísticas escalares É comum redes sociais exibirem o que chamamos de assortatividade: uma tendência de vértices que apresentam características semelhantes estarem conectados. Por exemplo, uma pessoa costuma ter amigos com idades próximas à dela; se isso for uma tendência geral da rede, *i.e.*, se existe uma *correlação* entre os vértices tal que vértices vizinhos exibem uma idade próxima, dizemos que a rede exibe assortatividade por idade.

Uma maneira de se quantificar o nível de assortatividade de uma rede é através da covariância. Suponha que a quantidade x_i seja o valor de alguma caraterística do vértice i (a idade, por exemplo). Vamos considerar que as quantidades (x_i, x_j) estão associadas a vértices i e j conectados por uma

PUC-Rio - Certificação Digital Nº 1122074/CA

ligação. Antes de calcular a covariância, vamos definir o valor médio $\langle x \rangle_e$ de x_i na extremidade de cada ligação,

$$\langle x \rangle_e = \frac{\sum_{ij} A_{ij} x_i}{\sum_{ij} A_{ij}} = \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2M} \sum_i k_i x_i. \tag{2-24}$$

Perceba que $\langle x \rangle_e$ é diferente do valor médio de x_i sobre vértices. O símbolo $\langle ... \rangle_e$ denota média sobre ligações. Se o vértice i tem k_i vizinhos, esse vértice estará presente na extremidade de k_i ligações e, por essa razão, temos um fator k_i na soma. A covariância de x_i e x_j sobre *ligações* é definida como:

$$\begin{aligned} \mathrm{cov}(x_i, x_j) &= \frac{\sum_{ij} A_{ij} (x_i - \langle x \rangle_e)(x_j - \langle x \rangle_e)}{\sum_{ij} A_{ij}} \\ &= \frac{1}{2M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) x_i x_j. \end{aligned} \tag{2-25}$$

Nos passos acima usamos a Eq. 2-3 e Eq. 2-24 (veja também o Apêndice A.1). Teremos uma covariância positiva (assortatividade) se, na média, x_i e x_j exibem uma tendência a ter valores próximos (ambos exibem valores altos, ou ambos exibem valores baixos). Na situação oposta temos uma covariância negativa. É conveniente normalizar a covariância pelo seu maior valor possível, assim teremos uma grandeza que varia entre -1 e $+1$. Se tivéssemos uma assortatividade perfeita em todas as ligações da rede, o valor de x_i e x_j seriam iguais. Daí, fazendo $x_j = x_i$ na Eq. 2-25, temos o valor máximo para a covariância

$$\frac{1}{2M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) x_i^2 = \frac{1}{2M} \sum_{ij} \left(k_i \delta_{ij} - \frac{k_i k_j}{2M} \right) x_i x_j, \tag{2-26}$$

normalizando a Eq. 2-25 pela equação acima temos o *coeficiente de assortatividade* r

$$r = \frac{\sum_{ij} \left(A_{ij} - k_i k_j / 2M \right) x_i x_j}{\sum_{ij} \left(k_i \delta_{ij} - k_i k_j / 2M \right) x_i x_j}. \tag{2-27}$$

Coeficientes deste tipo, onde temos o covariância no numerador e a variância no denominador, são conhecidos como coeficiente de correlação de Pearson, que é uma medida da correlação linear [57, 15].

Estamos interessados na assortatividade por *grau*, mas vale citar que poderíamos ter também correlações envolvendo outras característica dos sítios, como idade, renda, etc [15].

Assortatividade por grau A assortatividade por grau, ou correlação grau-grau é de particular interesse na tese, e será estudada com detalhe no Cap. 3. Numa rede com correlações grau-grau, os vértices com conectividade

PUC-Rio - Certificação Digital Nº 1122074/CA

alta possuem uma tendência a se conectarem com outros vértices que também exibem uma conectividade alta, e vértices com conectividade baixa conectam-se preferencialmente com outros vértices com conectividade baixa. Daqui por diante vamos usar simplesmente *redes assortativas* para nos referir as redes que têm assortatividade por grau. No caso oposto, onde vértices com conectividade alta têm uma inclinação a conectar-se com vértices com conectividade baixa, e vice-versa, diremos que temos uma *rede disassortativa*.

Ao contrário de outros atributos de um vértice como idade ou renda, o grau de um sítio já é uma propriedade da rede. Esse é um dos motivos pelos quais a correlação grau-grau é uma propriedade de interesse: temos uma propriedade da rede, nesse caso o grau, influenciando outra propriedade que é a posição das ligações. Para quantificar a correlação grau-grau, usaremos também o coeficiente de Pearson. Neste caso, o atributo x_i é a conectividade k_i, assim

$$\mathrm{cov}(k_i, k_j) = \frac{1}{2M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) k_i k_j. \tag{2-28}$$

Normalizando a Eq. 2-28 pela variância temos o coeficiente de assortatividade de Pearson

$$r = \frac{\sum_{ij} \left(A_{ij} - k_i k_j / 2M \right) k_i k_j}{\sum_{ij} \left(k_i \delta_{ij} - k_i k_j / 2M. \right) k_i k_j}. \tag{2-29}$$

A partir da matriz de adjacência $\mathbf{A}$ podemos deduzir uma outra relação importante:

$$\langle k^n \rangle_e = \frac{\sum_{ij} A_{ij} k_i^n}{\sum_{ij} A_{ij}} = \frac{\sum_i k_i k_i^n}{2M} = \frac{1}{\langle k \rangle} \frac{\sum_i k^{n+1}}{N} = \frac{\langle k^{n+1} \rangle}{\langle k \rangle} \tag{2-30}$$

onde $\langle ... \rangle$ denota, como antes, média feita sobre sítios. Na equação acima usamos novamente a Eq. 2-3 e $\langle k \rangle = 2M/N$. A Eq. 2-30 faz uma conexão entre as médias de potências da conectividade calculadas em relação a ligações e em relação aos sítios. Usando a relação 2-30 podemos reescrever a Eq. 2-29 como (mais detalhes no Apêndice A.1).

$$r = \frac{\langle k k' \rangle_e - \langle k \rangle_e^2}{\langle k^2 \rangle_e - \langle k \rangle_e^2}, \tag{2-31}$$

que é a forma mais encontrada na literatura para o coeficiente de assortatividade. A Tabela 2.1 também inclui o valor da assortatividade para algumas redes reais. Podemos notar que existe uma propensão de redes sociais terem um valor positivo para r, enquanto as outras redes apresentam, tipicamente, valores negativos de r. No Cap. 3 vamos abordar novamente essa questão.

O tópico de redes complexas é interessante por si só. No entanto, boa parte da motivação do estudo deste tópico é analisar o papel que as redes complexas

PUC-Rio - Certificação Digital Nº 1122074/CA

podem ter em processos dinâmicos, como em modelos de dinâmica social. A seguir discutiremos algumas das características mais relevantes desses modelos.

2.2 Sociofísica

Ordem e desordem em sistemas sociais Em ciências sociais, o termo *influência social* se refere ao fato de que as emoções, opiniões ou comportamentos de um indivíduo são afetados pelas outras pessoas [58, 59]. É através da imitação e influência de nossos pares que aprendemos uma linguagem, as normas pelas quais devemos nos comportar, aprendemos uma cultura, e desenvolvemos valores. Sem esse mecanismo fundamental, ou seja, se cada pessoa agisse sozinha, sem levar em conta a opinião ou o comportamento das outras, nada disso seria possível. Não existiria, por exemplo, nenhum consenso em relação a uma questão política, já que cada um teria seu próprio posicionamento individual. No entanto, o compartilhamento de opiniões, cultura e linguagem existem em todos os grupos humanos. Vale dizer que o que é chamado de consenso, concordância ou uniformidade por cientistas sociais é chamado de *ordem* no jargão dos físicos. O termo oposto usado pelos cientistas sociais, fragmentação ou discordância, é traduzido como *desordem* pelos físicos.

A interação entre os agentes tende a torná-los mais semelhantes, sendo esse o principal ingrediente para que surja a ordem no sistema. A homogeneidade (total ou parcial) observada em sistemas reais é fruto das repetidas interações entre os membros da população ao longo do tempo. Há um comportamento global emergente fruto dessa interação. Na física, as interações entre agentes seriam o equivalente às interações ferromagnéticas entre magnetos. Também poderíamos pensar numa interação anti-ferromagnética, ou seja, que faz com que as pessoas adotem um estado diferente do estado de seus vizinhos.

Um dos principais objetivos da aplicação da física estatística em problemas de dinâmica social é justamente entender o surgimento das propriedades globais originadas da interação entre pares de indivíduo, ou seja, como essas interações podem criar ordem a partir de uma situação inicialmente desornada. Busca-se , por exemplo, entender os mecanismos fundamentais que favorecem o consenso, ou o surgimento de uma linguagem, e também quais os mecanismos que dificultariam esse processo.

PUC-Rio - Certificação Digital Nº 1122074/CA

Modelagem de sistemas sociais O problema clássico da mecânica estatística é inferir o comportamento macroscópico de um sistema a partir de sua dinâmica microscópica. E esse problema continua presente quando fazemos a transposição das técnicas da física estatística para a sociofísica. No entanto, na sociofísica temos um problema adicional. Na mecânica estatística os componentes individuais do sistema são entidades "simples", como átomos ou moléculas, e sabemos as leis básicas do funcionamento dessas entidades, também conhecemos exatamente como essas entidades interagem entre si. Por outro lado, na sociofísica, as entidades básicas são seres humanos, e não temos leis para o funcionamento de seres humanos, e nem existem leis bem definidas de como devem ser as interações entre eles. Se em algum momento no futuro viermos a desvendar tais leis, elas seriam, muito provavelmente, mais complicadas e com mais variáveis do que o comportamento de átomos e moléculas. Essa é uma crítica comum à modelagem física de sistemas sociais: os modelos são simples demais. Felizmente, dentro do escopo de conceitos da física estatística encontra-se a ideia de universalidade: os *detalhes* microscópicos da interação entre os elementos do sistema não são fundamentais para se entender as propriedades emergentes. Algumas poucas características como simetrias, dimensionalidade e leis de conservação são relevantes. Sendo assim, sistemas com interações microscópicas distintas podem exibir o mesmo comportamento global. Vamos discutir um pouco mais sobre universalidade na Sec. 2.3. A justificativa para se usar modelos simples segue essa ideia: tentamos incluir no modelo somente as propriedades mais relevantes dos indivíduos e das interações entre eles e, a partir daí, estudamos com as ferramentas da física estatística, o comportamento qualitativo global do sistema.

Ainda assim, há sempre um grande número de detalhes negligenciados na modelagem de sistemas sociais e muitos fatores são desconhecidos. Pode-se levar em conta esses detalhes com a inclusão de algum tipo de ruído, por exemplo: a heterogeneidade dos indivíduos pode ser incluída através de um ruído independente do tempo nos parâmetros do modelo. Pode-se considerar também um ruído dependente do tempo, assim introduzimos transições espontâneas entre os estados de um indivíduo [3].

Modelos de opinião Faz parte das nossas vidas tomar algum posicionamento em relação a uma questão, política, por exemplo; ou se posicionar a favor ou contra um tema polêmico. Também temos, constantemente, que tomar decisões de compra ou expressar nossa opinião sobre algum bem de consumo ou filme, ou outro produto cultural. Em geral, não tomamos essas decisões sozinhos, mas levamos em conta a opinião de pessoas conhecidas, concordando ou discordando

PUC-Rio - Certificação Digital Nº 1122074/CA

delas.

Essa dinâmica de concordância/discordância entre pessoas pode ser bastante complicada, o que reflete o fato, mencionado acima, de que as pessoas não são entidades simples. A abordagem da física estatística nesses casos tem como objetivo definir quais são os estados (opiniões) dos agentes (pessoas) e definir também as regras (relações entre as pessoas) que podem levar a transição entre esses estados (mudanças de opinião).

Quando dizemos "definir os possíveis estados", estamos dizendo que a opinião será representada por uma variável (ou conjunto de variáveis), que no fundo, são apenas números. Essa abordagem parece redutiva, mas em muitas situações, temos que escolher entre (ou se posicionar em relação a) um certo número de alternativas. Podemos ter só duas alternativas, exemplos: contra ou a favor da redução da maioridade penal; iOS ou Android; assistir ou não ao novo filme da franquia *Exterminador do futuro*, etc. Em outras situações o número de alternativas é maior, o que não necessariamente representa uma vantagem, já que podemos cair no problema conhecido como sobrecarga de escolhas, esse tema será discutido no Cap. 5 (o problema de assistir ou não ao novo filme da franquia *Exterminador do futuro* será discutido no Cap. 4). Podemos ter também variáveis contínuas que, nesse caso, representam o nível de convicção em relação a um tópico (mais sobre isso no Cap. 6). Definidos os estados do sistema, o desafio é descrever as regras dinâmicas, ou seja, os mecanismos de interação responsáveis pela mudança entre os estados. Via de regra, a dinâmica é tal que tende a reduzir a variabilidade de estados iniciais conduzindo a um consenso, onde todos (ou quase todos) os indivíduos compartilham da mesma opinião, ou a um estado fragmentado onde existe coexistência de opiniões.

Considera-se que o modelo proposto Weidlich em 1971 [60] foi o primeiro modelo de opinião proposto por um físico [3]. Apesar do modelo de Ising (ver Sec. 2.3) ser anterior ao modelo de Weidlich, sua interpretação como modelo de opinião é posterior [61, 62].

Nas últimas décadas, os físicos têm trabalhado intensamente nessa área, e muitos modelos foram propostos. Podemos destacar o modelo do votante, o modelo da maioria, o modelo de Sznajd e os modelos de confiança delimitada. Esses modelos serão melhor explorados nos capítulos subsequentes.

Fenômenos de propagação em redes complexas Em modelos de dinâmica de opinião temos a competição de duas ou mais opções que competem entre si. As alternativas são equivalentes, no sentido de que têm o mesmo grau de plausibilidade, de tal forma que um agente pode influenciar o outro, nesse

PUC-Rio - Certificação Digital Nº 1122074/CA

sentido as interações são simétricas. Por outro lado, em certos problemas, como o espalhamento de notícias ou a propagação de adoções, a interação é assimétrica. Existe um "fluxo de informação" de um agente que tem a informação para aquele que não tem. Funciona como numa epidemia, onde uma pessoa transmite a doença para outra. Estamos usando aqui a palavra "informação" num sentido amplo, podemos também ter a propagação de atitudes, ou propagação de adoções, etc. Diz-se que os sítios que já detêm a informação estão num estado "ativo".

Há também nessa área, a participação de físicos estatísticos. Muitos modelos propostos por físicos pertencem à classe de *modelos de ativação por limiar* (*threshold models* em inglês): é preciso que uma certa fração mínima dos vizinhos de um sítio estejam ativos para que ocorra a propagação da informação. Existem também os *modelos sem limiar*: o acesso a informação por um único vizinho é suficiente para que haja a transmissão da informação de um agente para outro. Em geral, a pergunta clássica que se quer responder nesse tipo de problema é: no final da dinâmica, quantas pessoas serão atingidas pela informação? Ao contrário de situações envolvendo epidemias, aqui considera-se que, quanto maior esse número, melhor.

O entendimento desses mecanismos de espalhamento pode ter aplicações práticas, por exemplo, em campanhas de marketing. No Cap. 6, consideraremos regras de ativação semelhantes a do modelo de ativação de Watts [63] e da percolação de inicialização [64]. Regras desse tipo também são influentes na modelagem que usamos no Cap. 4. Embora, de maneira geral, não fazemos uma distinção rígida dos problemas abordados aqui como pertencendo a uma classe de modelos ou outra, ou seja, não classificamos os problemas como dinâmica de opinião ou propagação de informação.

2.3 Correlações de longo alcance e universalidade

Mencionamos acima que em problemas de dinâmica social, uma das questões mais fundamentais, senão a questão mais fundamental, é compreender a transição de um estado inicialmente desordenado para uma configuração ordenada (ou parcialmente ordenada). Esse tipo de transição é super recorrente na física estatística tradicional. Para ilustrar alguns conceitos e ferramentas, vamos utilizar o modelo de Ising, que é o exemplo paradigmático de sistema que exibe uma transição ordem-desordem.

Considera-se uma rede regular $L \times L \times L$ onde a cada vértice está associada uma variável $s_i = \pm 1$ que representa o spin de cada átomo. A hamiltoniana do modelo de Ising (sem campo externo) é [65]

PUC-Rio - Certificação Digital Nº 1122074/CA

$$H = -\frac{1}{2}\sum_{\langle i,j\rangle} s_i s_j, \tag{2-32}$$

onde $\langle i, j\rangle$ significa que a soma é feita sobre pares de primeiros vizinhos, ou seja, no modelo de Ising cada elemento do sistema interage diretamente apenas com seus vizinhos mais próximos. Dito de outra forma: um componente do sistema não interage diretamente com todos os outros. Por outro lado, uma mudança no estado de um spin i pode modificar o estado de seu vizinho j, o spin j pode modificar o estado de um dos seus vizinhos, e assim por diante.

Com isso, mesmo que não haja interação direta entre spins que não são vizinhos, eles podem estar correlacionados. Medimos a correlação entre spins separados por uma distância r através da função de correlação definida como $C(r,t) = \langle s_i(t)s_{i+r}(t)\rangle - \langle s_i(t)\rangle^2$. Pode-se calcular o comprimento de correlação ξ, que é uma medida do alcance espacial das correlações citadas acima a partir de $C(r)$. O comprimento de correlação é uma medida fundamental e nos diz qual a distância típica que uma perturbação feita em um componente do sistema é sentida pelos demais [12].

Já adiantando, o modelo de Ising próximo de uma temperatura crítica T_c exibe correlações de longo alcance, em que todos os componentes do sistema estão correlacionados (mesmo que a interação *direta* entre eles seja local e de curto alcance). Podemos nos referir ao modelo de Ising próximo de T_c como um sistema crítico.

Por um lado, correlações de longo alcance vão contra a abordagem reducionista da física, em que o *modus operandi* é isolar e entender o funcionamento de uma parte do sistema para depois incluir a influência das outras partes. A abordagem reducionista falha quando considerada no estudo de sistemas críticos, já que não podemos estudar o comportamento global partindo da análise separada de suas componentes. Nesse caso é comum dizer que a soma das partes é diferente do todo.

Felizmente, temos também um lado positivo, as correlações de longo alcance estão associadas a um princípio fundamental chamado *universalidade*. Esse princípio diz que as características globais, ou emergentes, de um sistema são originárias das correlações indiretas entre os elementos do sistema; correlações tais que estão associadas a todas as escalas de comprimento internas e não só as interações diretas entre um componente e outro. Com isso, dois sistemas que aparentemente não têm nenhuma característica em comum, apresentam o mesmo comportamento global, porque, perto de seus respectivos pontos críticos, as correlações de longo alcance têm a mesma natureza.

Voltando ao modelo de Ising: como cada spin está disposto em um vértice

PUC-Rio - Certificação Digital Nº 1122074/CA

de uma rede regular, a distância entre eles é sempre a mesma. Cada spin s_i tem uma direção preferencial, apontando para cima ($s_i = +1$) ou para baixo ($s_i = -1$). As interações ferromagnéticas da Eq. 2-32 induzem os spins a ficarem emparelhados, apontando na mesma direção, e favorecendo a ordem no sistema. Simultaneamente, o ruído térmico introduz flutuações que tendem a destruir essa ordem. Abaixo da temperatura crítica $T < T_c$, o sistema apresenta ilhas de spins apontando na mesma direção. Para fins de discussão, diremos que o número de spins apontando para cima é igual a n e, portanto, teremos $N-n$ spins apontando na direção oposta. A magnetização $m = 1/N \sum_i s_i$ é uma medida da diferença $N-2n$ entre spins que apontam para cima ou para baixo. Se $T < T_c$, uma das direções prevalece, digamos $+1$, e então $m(T) > 0$. Se aumentarmos a temperatura, o número n diminui ao passo que $N-n$ aumenta; esses valores se aproximam a medida que aumentamos T, e se tornam iguais em T_c. Nesse ponto o número de spins que apontam para cima ou para baixo é, em média, igual a $N/2$. Para qualquer temperatura acima disso a magnetização é zero.

Seguindo a ideia da Ref. [12], ilustraremos a universalidade usando outro sistema físico bem diferente do modelo magnético que acabamos de ver: a transição de fase que ocorre em um líquido, digamos a água, para ficar num caso familiar. A interação entre as moléculas da água é descrita pela Mecânica Quântica, sendo fruto da interação entre cargas elétricas de prótons e elétrons. Essa interação faz com que as moléculas se atraiam ou se repilam. Não temos uma estrutura de rede como no modelo de Ising, e as partículas executam um movimento caótico introduzido pelo ruído térmico. No entanto, existe uma distância típica entre moléculas vizinhas. Dependo da temperatura, podemos ter uma fase líquida onde a distância d entre moléculas é pequena (temperaturas baixas), ou podemos ter uma distância D maior se o sistema se encontra na fase gasosa (temperaturas altas). É possível ter também a coexistência de fases em uma certa temperatura. Vamos considerar que o volume V do sistema é fixo e estamos na temperatura onde temos coexistência. Nesse caso, se aumentarmos a temperatura, o líquido se expande e d aumenta. Como estamos supondo que o volume é fixo, a outra distância, D, diminui. A diferença entre essas distâncias típicas é quantificada através da diferença de densidade entre líquido e vapor $\Delta\rho = \rho_{\text{agua}} - \rho_{\text{vapor}}$, que faz o mesmo papel que a magnetização no modelo de Ising. Se continuarmos o processo de aumentar a temperatura até chegarmos numa temperatura crítica T_c, a diferença entre as densidades se anulam. Acima dessa temperatura temos um único fluido, e não há distinção do que é líquido ou vapor.

As duas descrições acima são praticamente à mesma mudando-se as

PUC-Rio - Certificação Digital Nº 1122074/CA

variáveis. Esses sistemas diferentes, a água e o magneto, pertencem a mesma classe de universalidade e apresentam o mesmo comportamento global. De maneira explícita, as caraterísticas importantes para o comportamento global desses sistemas são: *(i)* a dimensionalidade do sistema: nos dois casos temos sistemas em 3 dimensões. *(ii)* a magnetização m e a diferença entre as densidades $\Delta\rho$ são grandezas escalares. *(iii)* o alcance das interações: em ambos os casos as interações são de curto alcance. Essas similaridades entre os sistemas são suficientes para que eles sejam da mesma classe de universalidade. As semelhanças entre os dois sistemas não é apenas qualitativa: com T tendendo a T_c, m e $\Delta\rho$ se anulam de acordo com uma lei de potência $(T_c - T)^\beta$ e, surpreendentemente, o valor do expoente β é o mesmo. Resumindo: os sistemas (reais ou artificiais) que apresentam as propriedades destacadas acima vão exibir esse mesmo comportamento. A conclusão é que os detalhes finos do tipo de interação não são fundamentais e pode-se estudar o comportamento emergente de sistemas da mesma classe de universalidade de uma só vez.

O lado ruim é que a solução analítica de sistemas interagentes nem sempre é possível, salvo algumas excessões. Em uma dimensão podemos resolver exatamente o modelo de Ising, no entanto, nesse caso, não temos uma transição de fase propriamente dita. A solução para o modelo em 2D é bastante sofisticada e foi obtida por Onsager em 1944. Em 3 dimensões não há uma solução analítica, mas sabemos, através de simulações computacionais, que nesse caso, $m \sim (T_c - T)^\beta$ com $\beta = 0,3269 \pm 0,0005$. Há vários métodos computacionais para o estudo de sistemas como o modelo de Ising, sendo o mais famoso deles o algoritmo de Metropolis que discutiremos na sequência.

Algoritmo de Metropolis Para um sistema em equilíbrio térmico com um reservatório a temperatura T, a distribuição de probabilidade para um estado μ com energia E_μ é dada por

$$P_\mu = \frac{1}{Z} \mathrm{e}^{-E_\mu / k_B T}, \tag{2-33}$$

onde k_B é a constante de Boltzmann e a contante de normalização $Z = \sum_\mu \mathrm{e}^{-E_\mu/k_B T}$ é a função de partição [66].

Na dinâmica de Metropolis estados do modelo de Ising são automaticamente gerados de acordo com a distribuição 2-33. A ideia é gerar um grande número de estados de acordo com a distribuição de Boltzmann e a partir desses estados calcular médias das grandezas de interesse como, por exemplo, a magnetização [67]. Considere que o sistema se encontra num estado com energia E_μ. Modificamos esse estado fazendo uma perturbação aleatória, trocando o sinal de um spin. Dessa forma geramos um novo estado com energia

PUC-Rio - Certificação Digital Nº 1122074/CA

E_ν. Esse novo estado será aceito no cálculo das grandezas de interesse com probabilidade $p = \min\{1, \exp(-\Delta E/k_B T)\}$, onde $\Delta E = E_\nu - E_\mu$. Isto é, se a energia do estado tentativa diminui $E_\nu < E_\mu$, o estado é aceito, estados com energia $E_\nu > E_\mu$ são aceitos com maior probabilidade quanto maior for a temperatura. O algoritmo de Metropolis assegura que estados com energia $E_\nu > E_\mu$ serão gerados com probabilidade menor que aqueles com energia E_μ por um fator $\exp(-\Delta E/k_B T)$, como exigido pelos fatores de Boltzmann correspondentes $\exp(-E_\mu/k_B T)$ e $\exp(-E_\nu/k_B T)$ [68].

Na simulação do modelo de Ising, temos um ensemble de sistemas (uma coleção de estados) com a mesma temperatura, mas que é aleatório em outros aspectos. Grandezas como a magnetização, o calor específico e a susceptibilidade são amostradas segundo a distribuição de Boltzmann. Vamos usar essa técnica no Cap. 3 para gerar um ensemble de redes que mantém uma propriedade estrutural fixa, no caso, a assortatividade da rede. A partir daí calcularemos outras grandezas como o coeficiente de aglomeração fazendo médias sobre o ensemble resultante.

Por fim, vale uma nota sobre efeitos de tamanho finito. Uma transição de fase só acontece, rigorosamente, no limite termodinâmico, quando o número N de partículas vai a infinito. Somente no limite $N \to \infty$ temos uma verdadeira divergência nas grandezas termodinâmicas, como a susceptibilidade ou calor específico [65, 66]. Obviamente sistemas sociais não têm um número infinito de pessoas, apesar desse número, em geral, ser bastante grande. Ainda assim o número de pessoas numa população não é nem ao menos próximo do número de partículas em um sistema costumeiramente estudado pela física estatística. A finitude de N deve ser, portanto, levada em consideração na análise de sistemas sociais [3]. Via de regra, estudamos os modelos para diversos valores de N e as vezes inferimos o comportamento do sistema no limite $N \to \infty$. Isto nos ajuda a caracterizar melhor os comportamentos qualitativos do sistema, e a separar os resultados mais robustos daqueles que são mais dependentes dos detalhes microscópicos, por exemplo.

2.4 Metodologia

Modelos baseados em agentes Em muitos problemas de dinâmica social é inviável (ou, pelo menos, muito difícil) usar a abordagem mais tradicional da física, onde descrevemos o sistema através de um conjunto de equações que são resolvidas numericamente ou analiticamente. Como alternativa para

PUC-Rio - Certificação Digital Nº 1122074/CA

a resolução de problemas nessa área, encontram-se os modelos baseados em agentes (ABM) que, na tese, foi uma das principais ferramentas utilizadas.

Nesse tipo de modelagem o sistema é composto por uma coleção de *agentes*: entidades independentes que tomam "decisões" baseadas em determinadas regras [69]. Em alguns contextos, o termo "atores" também é usado. Outros ingredientes comuns na ABM são: a inclusão de processos adaptativos ou "aprendizado" por parte dos agentes e a interação dos agentes com um ambiente.

A origem da ABM pode ser atribuída a introdução dos autômatos celulares feita por Von Neumann e Ulam na década de 1940. Vale destacar, como um dos exemplos mais clássico de ABM, o jogo da vida de Conway [70]. Algoritmos análogos são populares no estudo de populações em biologia.

Aplicações da ABM podem ser encontradas em economia, ciências sociais, biologia e inteligência artificial. Nos dias de hoje a modelagem baseada em agentes adquiriram um papel fundamental na modelagem de sistemas complexos. Nos nossos problemas, o estado interno dos agentes representa opiniões e as regras de tomada de decisão correspondem às interações sociais. O foco não é nas características de cada indivíduo em específico, mas em propriedades globais que são calculadas realizando uma média sobre o estado de toda a população de agentes.

Análise de dados Sendo a física uma ciência experimental, um passo crucial dentro de qualquer estudo, inclusive o estudo de sistemas sociais, é a comparação dos resultados teóricos com dados empíricos. Talvez ainda mais nesse caso. Como foi discutido, não existem "primeiros princípios" quando o objeto de estudo são pessoas. Através do confronto com dados reais, validamos se o comportamento obtido através de um modelo é compatível com as tendências observadas em dados reais. Sabemos assim se a modelagem é plausível e consistente, ou se é preciso adicionar novas variáveis ou modificar o modelo. Apesar disso, a relação entre modelos teóricos e validação empírica é desbalanceada. Há muitos mais modelos teóricos do que evidências empíricas [3].

Nos dias de hoje, a profusão de redes sociais, e a facilidade que indivíduos têm de expressar seu posicionamento em relação a diversos temas em diferentes plataformas podem significar uma mudança nesse paradigma.

Os dados empíricos, que coletamos e analisamos na tese, são uma parte importante do trabalho. Usamos o resultado dessas análises para validar alguns resultados, como no Cap. 5. Nos capítulos 4 e 6 os dados empíricos têm um papel ainda mais relevante pois o estudo foi diretamente guiado pelos padrões

PUC-Rio - Certificação Digital Nº 1122074/CA

que observamos, que serviram como ponto de partida.

Além de fazer uso de dados de pesquisa de opinião, utilizamos também um *web crawler* para coleta de dados. De forma técnica: um crawler é um programa que visita sites da internet de maneira automática e sistematizada [71, 72]. Cada crawler tem um conjunto de regras específico, de acordo com seu objetivo. Normalmente, iniciamos com uma lista de URLs para serem visitadas (*seeds* em inglês). Conforme visita-se essas URLs, identificamos todos os links (ou os links de interesse) da página e os adicionamos à lista de URLs a serem visitadas (chamada de *crawl frontier* em inglês), eventualmente coletando dados nesse processo. Assim, pode-se visitar recursivamente as URLs, simultaneamente copiando e salvando as informações das páginas. Essa técnica foi usada para coletar os dados dos sites Netshoes e GooglePlay que utilizamos no Cap. 5, e para coletar dados do IMDb utilizados nos Cap. 4 e 6.

PUC-Rio - Certificação Digital Nº 1122074/CA

3
Correlações em redes complexas

Correlações podem afetar processos de propagação em redes complexas. Para analisar seu efeito, é útil construir *ensembles* de redes que tenham a restrição de ter um certo valor de uma medida estrutural, tal como a correlação grau-grau, ou assortativade r, mas que sejam aleatórios em outros aspectos e que preservem a sequência de graus. Isso pode ser conseguido através de procedimentos de optimização de Monte Carlo. No entanto, na medida em que ajustamos o valor de r, outras propriedades da rede podem mudar concomitantemente.

Então, neste capítulo da tese, analisamos, para o ensemble de redes com assortatividade r, o impacto dessa grandeza em propriedades tais como a transitividade, a ramificação e as distâncias características, que são relevantes quando investiga-se fenômenos de propagação nessas redes.

A presente análise é realizada para redes com dois tipos de distribuição de conectividade: distribuições localizadas em torno de um conectividade típica (com decaimento exponencial assintótico) e distribuições largas (com decaimento em lei de potência). Também investigamos efeitos de tamanho.

3.1
Introdução

Redes complexas são substratos realísticos para simular muitos fenômenos naturais e sociais. Para entendermos a influência da topologia da rede, primeiramente, podemos considerar diferentes classes de distribuições de conectividade $P(k)$. No entanto, para uma dada distribuição de conectividade, correlações podem dar origem a importantes efeitos, relacionados com a estrutura da rede, em determinado problema [73–78].

Esses efeitos estruturais podem ter consequências importantes, por exemplo, correlações podem deslocar o limiar epidêmico [78]. Embora esses efeitos possam estar ausentes em alguns casos, em outros as correlações não podem ser desprezadas.

Diversas quantidades podem ser consideradas para caracterizar as correlações de uma rede. O mais natural seria considerar a distribuição condicional $P(k'|k)$, de que um vértice com k ligações esteja conectado a outro com conectividade k'. No entanto, histogramas para $P(k'|k)$ são muito afetados

PUC-Rio - Certificação Digital Nº 1122074/CA

por flutuações estatísticas, sendo muito ruidosos em redes de tamanho pequeno e moderado, e, portanto, na prática são raramente usados.

No caso onde o interesse é estudar a correlação entre dois vértices, pode-se considerar a conectividade média dos primeiros vizinhos de um nó com conectividade k, $\bar{k}_{nn}(k)$ (em inglês *average nearest neighbors degree* ou ANND). Se ANND é uma função crescente de k tem-se uma rede assortativa: nós têm a propensão de se associar a outros nós com conectividade semelhante; no caso contrário, onde $\bar{k}_{nn}(k)$ decresce com k, temos uma rede dissortativa: conexões entre nós com conectividade alta e baixa são favorecidas.

É possível, inclusive, obter o nível de correlações grau-grau de uma rede usando uma única quantidade escalar através do coeficiente de assortatividade de Pearson [79]. Portanto, mesmo que algumas variantes tenham sido definidas na literatura, nós iremos considerar a assortatividade como medida da tendência de sítios vizinhos terem conectividade similar (ou diferente). O coeficiente de assortatividade de Pearson pode ser definido como (veja Eq. 2-31):

$$r = \frac{\langle kk' \rangle_e - \langle k \rangle_e^2}{\langle k^2 \rangle_e - \langle k \rangle_e^2},$$

onde $\langle \cdots \rangle_e$ denota média sobre ligações e k e k' são as conectividades dos vértices em cada extremidade de uma ligação. Como foi dito no Cap. 2, valores positivos de r correspondem a redes assortativas e são encontrados principalmente em redes sociais. Já as redes tecnológicas, de informação e biológicas são em geral disassortativas, apresentando valores negativos de r. A Fig. 3.1a mostra uma ilustração de rede assortativa; temos um núcleo denso, onde os nós com maior conectividade ficam concentrados e em volta desse núcleo encontram-se os nós com conectividade menor. Uma rede disassortativa é visualmente bem diferente: a estrutura é parecida a uma "estrela". A Fig. 3.1b ilustra este tipo de estrutura, veja que não há uma separação tão grande entre o núcleo e a periferia da rede como numa rede assortativa.

Sabe-se que o coeficiente de Pearson apresenta alguns inconvenientes, mas por outro lado é uma quantidade padrão e de uso comum e portanto vale a pena ser analisada. Além do mais, tem a vantagem de ser uma quantidade simples, podemos através dela quantificar as correlações com um único número, e é mais fácil de ser controlada do que grandezas que são funções com mais variáveis.

Para analisar a influência de correlações, bem como qualquer outra característica estrutural é útil construir ensembles de redes que mantêm essa propriedade, e que mantenham fixas a sequência de graus. Como iremos

PUC-Rio - Certificação Digital Nº 1122074/CA

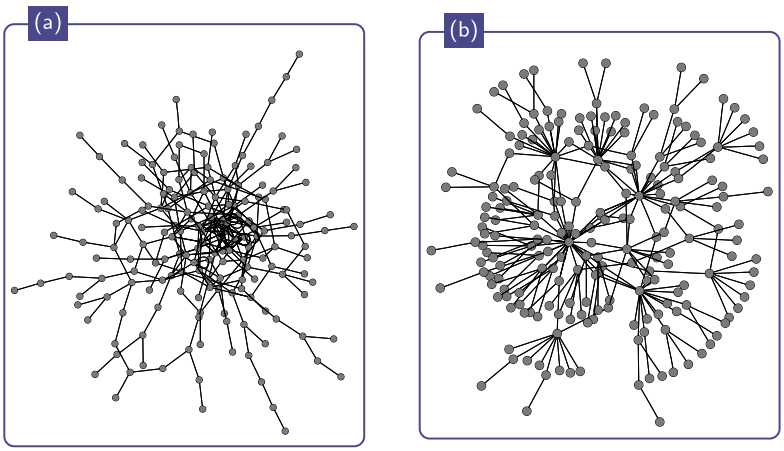


Figura 3.1: Exemplos de redes assortativas (a) e disassortativas (b). A rede assortativa apresenta um núcleo denso, onde os nós com maior conectividade ficam concentrados e em volta desse núcleo ficam os nós com conectividade menor. Já a rede disassortativa tem uma estrutura parecida com uma estrela. As redes foram geradas artificialmente. Essa figura foi adaptada da Ref.[80].

PUC-Rio - Certificação Digital Nº 1122074/CA

discutir na seção 3.2, este tipo de ensemble pode ser obtido por intermédio de religações apropriadas entre os nós. As religações são feitas através de um procedimento padrão de Monte Carlo (MC) para minimizar uma quantidade semelhante à energia (abordagem do ensemble de máxima entropia), que seja função da propriedade do grafo que busca-se controlar (r, no nosso caso) [81–83].

Uma vez que tenhamos ajustado o valor de r, é importante caracterizar como outras propriedades da rede são alteradas como consequência. Algumas interdependências entre certas propriedade já foram numericamente mostradas na literatura, tanto para redes reais quanto para grafos aleatórios simulados [82]. Algumas relações analíticas também já foram deduzidas [84–86]. Por causa do seu papel crucial em problemas de espalhamento [27], vamos focar aqui no efeito de r sobre (i) as distâncias típicas entre nós, (ii) sobre a transitividade, e (iii) sobre a ramificação.

O número de ligações em um caminho que conecta os vértices i e j é chamado de *comprimento* desse caminho. Já o comprimento do *menor caminho* que conecta os vértices i e j é chamado de distância geodésica d_{ij}.

Como uma medida da separação entre os nós da rede, vamos considerar a distância geodésica média ou menor caminho médio, que é definido pela expressão [13],

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \tag{3-1}$$

onde N é o número de sítios da rede. Um problema com a expressão acima é

definir qual seria a distância d_{ij} se os vértices i e j pertencem a componentes diferentes, já que nesse caso d_{ij} é indefinido. Se considerarmos que $d_{ij} = \infty$, teríamos uma divergência em L. Para evitar esse problema pode-se ignorar os pares de nós que pertencessem a componentes diferentes [87], ou seja, somente pares de nós i e j para os quais existe um caminho seriam considerados no cálculo da Eq. 3-1.

Nos cálculos subsequentes, usaremos a expressão,

$$L = \frac{\sum_{i=1}^{n} \langle L_i \rangle w_i}{\sum_{i=1}^{n} w_i} \quad \text{com} \quad w_i = \frac{N_i(N_i-1)}{2} \tag{3-2}$$

e

$$\langle L_i \rangle = \frac{\sum_{j,k=1}^{N_i} d_{kj}}{N_i(N_i-1)}\,, \tag{3-3}$$

onde n é o número de componentes do grafo e N_i é o número de sítios na componente i. A Eq. 3-2 é a média de $\langle L_i \rangle$ (distância geodésica média da componente i) ponderada pelo número de pares de sítios da componente i. No caso em que só há uma componente na rede, a Eq. 3-2 se reduz à definição usual, Eq. 3-1. Alternativamente, para evitar o problema da divergência na distância entre nós desconectados, nós consideramos o inverso, $1/E$, da grandeza conhecida como eficiência global [88, 89]

$$E = \frac{1}{N(N-1)} \sum_{\substack{1 \leqslant i,j \leqslant N \\ i \neq j}} \frac{1}{d_{ij}}\,. \tag{3-4}$$

A grandeza $h = 1/E$ representa a média harmônica ao invés da média aritmética das distâncias geodésicas. Também calculamos o diâmetro $D = \max\{d_{ij}\}$.

A transitividade ou coeficiente de agrupamento C será mensurada [90, 23] através da equação Eq. 2-21,

$$C = \frac{6n_{\triangle}}{\sum_{i=1}^{N} k_i(k_i-1)}\,, \tag{3-5}$$

onde $n_{\triangle}$ é o número de triângulos e usamos que o número de trios conectados da rede é dado por $1/2\sum_{i=1}^{N} k_i(k_i-1)$. Também medimos o valor médio, $\bar{C}$, do coeficiente de agrupamento local C_i (veja Eq. 2-22), definido como $C_i = 2e_i/(k_i(k_i-1))$, onde e_i é o número de conexões entre os vizinhos do sítio i [13]. Consideramos $C_i = 0$ quando $k_i = 0$ ou 1.

Outras medidas que surgem da decomposição de r [84] também serão consideradas. Além de detectar interdependências entre propriedades estruturais, é também importante saber como essas propriedades dependem do tamanho do sistema N. Analisaremos essas questões para duas classes principais de distribuições de grau (Poisson e com decaimento em lei de potência). Por fim, investigamos sequências de grau de redes reais.

PUC-Rio - Certificação Digital Nº 1122074/CA

3.2
Redes e ensembles

Para cada classe de rede, vamos considerar diferentes valores do tamanho, N, e da conectividade média, $\langle k \rangle$, dentro de intervalos realísticos. Como paradigma da classe de redes com uma distribuição de graus centrada, com todos os momentos finitos, consideraremos o grafo aleatório de Erdős e Rényi [91, 16]. Como vimos no Cap. 2, neste modelo, uma rede de N nós é construída selecionando M pares diferentes de nós ao acaso e conectando cada par. A distribuição resultante dos graus é uma distribuição de Poisson $P(k) = e^{-\langle k \rangle}\langle k \rangle^k / k!$, onde a conectividade média é dada por $\langle k \rangle = 2M/N$. Também analisamos redes do tipo lei de potência, i.e., com $P(k) \sim k^{-\gamma}$, $\gamma > 2$, correspondendo a uma distribuição normalizável, mas onde os momentos de ordem $n \geqslant \gamma - 1$ são divergentes. Construímos essas redes usando o modelo configuracional [92, 15]. Seguindo esse modelo, começamos escolhendo N números aleatórios k, sorteados a partir de uma distribuição $P(k)$. Esses números representam o número de ligações saindo de cada um dos nós. Tais ligações têm uma extremidade conectada ao nó enquanto a outra ainda é livre. No próximo passo, duas extremidades abertas são escolhidas e conectadas de tal maneira que ligações múltiplas são permitidas, enquanto autoligações são proibidas.

Esse segundo estágio é repetido até que cada nó tenha atingido a conectividade atribuída no primeiro passo. Se eventualmente uma ligação tiver uma extremidade aberta, ela é descartada. Para redes grandes, no entanto, a fração dessas ligações descartadas é desprezível. Para sortear o conjunto de números k com probabilidade $P(k) = \mathcal{N}k^{-\gamma}$, com $k_{\min} \leqslant k \leqslant k_{\max}$ (e, portanto, com fator de normalização $\mathcal{N} = 1/\sum_{k_{\min}}^{k_{\max}} k^{-\gamma}$), usamos o método da transformação inversa [93, 67]. Note que $k_{\max} \leqslant N - 1$ e $k_{\max} >> k_{\min}$, então determinamos $k_{\min}$ para ajustar o valor selecionado de $\langle k \rangle$ (dentro de uma tolerância de no máximo 1%), tal que

$$\langle k \rangle = \frac{\sum_{k_{\min}}^{k_{\max}} k^{-\gamma+1}}{\sum_{k_{\min}}^{k_{\max}} k^{-\gamma}} \simeq \frac{\gamma - 1}{\gamma - 2} \frac{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \simeq \frac{\gamma - 1}{\gamma - 2} k_{\min}. \tag{3-6}$$

Vale a pena mencionar que o valor $k = N - 1$ não é usualmente obtido, sendo o cut-off natural igual à $k_c \sim N^{\frac{1}{\gamma-1}}$ [94]. Para obter o valor desejado de r, nós seguimos uma abordagem que usa religamentos entre os nós.

Queremos construir o ensemble de redes $\{\mathrm{G}\}$ como um dado valor de r (ensemble r), mas que seja aleatório em qualquer outro aspecto, i.e, fazendo o menor número de suposições possível sobre a distribuição $P(G)$. Para conseguir isso, utilizamos o modelo do grafo aleatório exponencial, tal que o conjunto

PUC-Rio - Certificação Digital Nº 1122074/CA

de redes {G} tenha distribuição $P(G) \propto e^{-H(G)}$, onde $H(G)$ é uma quantidade do tipo hamiltioniana (ou energia) [81]. Para conseguir um ensemble r, com $r = r_\star$, consideramos [82]

$$H(G) = \beta |r - r_\star| \,, \tag{3-7}$$

onde β é um parâmetro real. O ensemble pode ser simulado usando um procedimento de MC: a cada passo de tempo, uma tentativa de religamento é aceita com probabilidade $\min\{1, e^{-[H(G')-H(G)]}\}$.

Cada passo de religamento é realizado selecionando duas ligações que conectam os vértices a, b e c, d, respectivamente, e substituindo essas ligações por duas novas que que conectam a, c e b, d, como ilustrado na Fig. 3.2. Esse movimento tentativa é equivalente a trocar o sinal do spin em uma simulação de Monte Carlo do modelo de Ising. Movimentos que levam a ligações duplas são proibidas. Note que esse processo preserva o grau de cada nó [95].

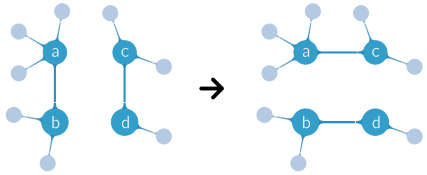


Figura 3.2: Ilustração do processo de religamento. Estão representados os sítios (a, b, c e d) e seus primeiros vizinhos.

Começamos a simulação tomando $\beta = 0$ (durante no máximo 100 passos de Monte Carlo (MCS), onde cada MCS corresponde a N tentativas de religamento). O efeito desse estado é basicamente destruir múltiplas conexões. Não observamos efeitos de histerese como os encontrados na Ref. [96], onde foi usado uma hamiltoniana diferente para se controlar o número de triângulos do grafo. Subsequentemente, o valor de β é aumentado (em incrementos de $\Delta\beta = 1000$), a cada 50 MCS, até que r se estabilize, tipicamente atingindo o valor prescrito. Então, as quantidades de interesse são calculadas e todo o processo é repetido, começando com uma nova sequência de graus. Para redes com distribuição em lei de potência, observamos que o processo é não ergódico, e calculamos os valores da média e do desvio padrão sobre 100 realizações do protocolo descrito acima. Verificamos que a escolha de outras expressões para $H(G)$, que também vão a zero em $r_\star$, não alteram significativamente os resultados, mas apenas o tempo de convergência.

3.3 Resultados

PUC-Rio - Certificação Digital Nº 1122074/CA

3.3.1
Efeitos de r sobre o coeficiente C

Vamos começar reportando os efeitos de r sobre o coeficiente de agrupamento C. Para o caso de Poisson, nós mostramos na Fig. 3.3(a) o comportamento de C como função de r mantendo o número de nós fixos ($N = 8000$) mas para diferentes valores da conectividade média $\langle k \rangle$.

Os valores de C que surgem são pequenos. A transitividade C cresce monotonicamente com r, o que é consistente com os resultados da Ref. [82] (a qual se restringe a $r \geqslant 0$) para esse tipo de rede. Observamos dois regimes com um *crossover* em $r \simeq 0,5$: um crescimento lento com r abaixo do crossover e um aumento mais pronunciado na região acima dele.

A existência de dois regimes pode estar relacionada ao caráter assimétrico de r, o qual não mede assortatividade e disassortatividade da mesma maneira. Abaixo do crossover, C aumenta linearmente com $\langle k \rangle$ aproximadamente uma ordem de magnitude dentro do intervalo analisado. Por sua vez, acima do crossover, C permanece com a mesma ordem de magnitude quando a conectividade média aumenta, até mesmo para valores pequenos de $\langle k \rangle$ (veja também o inset da Fig. 3.3(a) onde C é plotado vs $\langle k \rangle$ para valores selecionados de r). Na Sec. 3.4, vamos discutir essas questões em mais detalhes.

Para o coeficiente de agrupamento local $\bar{C}$ (não mostrado), nós obtivemos uma dependência qualitativa com r similar à que observamos para o coeficiente de agrupamento C. No entanto, o aumento de $\bar{C}$ com $\langle k \rangle$ é linear para qualquer valor fixo de r. Na figura Fig. 3.3(b), efeitos de tamanho são exibidos para $\langle k \rangle = 4$, conectividade representativa dos outros valores considerados. A medida em que o número de sítios aumenta, C decai como $C \sim 1/N$ para todos os valores de r (como é mostrado no gráfico inserido). Portanto, para o ensemble r de redes de Poisson, a transitividade é somente um efeito de tamanho finito e vai a zero no limite de uma rede infinita, com a mesma lei assintótica $C \sim 1/N$ que um grafo aleatório não correlacionado [27, 28].

Para a classe de redes com distribuição em lei de potência, o intervalo de valores permitidos de r é limitado. Isto é, valores arbitrariamente diferentes de zero não são atingidos numa realização típica do protocolo de MC descrito na Sec. 3.2. Para determinar os valores máximos (mínimos) típicos, $r_{\max}$ ($r_{\min}$), nós impomos $r_\star = 1$ (-1) e detectamos qual o valor estacionário de r. A evolução temporal de r para $r_\star = 1$ (-1) é ilustrada na Fig. 3.4(a) para $\gamma = 3,5$; $N = 5000$ e $\langle k \rangle \simeq 4$. Note os grandes desvios entre os valores estacionários de diferentes realizações, principalmente para o limite superior. Nós verificamos que esse quadro não muda através da implementação de outras definições de $H(G)$ na Eq. (3-7), e.g., $\beta|r - r_\star|^\alpha$, com $\alpha \neq 1$. Os valores extremos

PUC-Rio - Certificação Digital Nº 1122074/CA

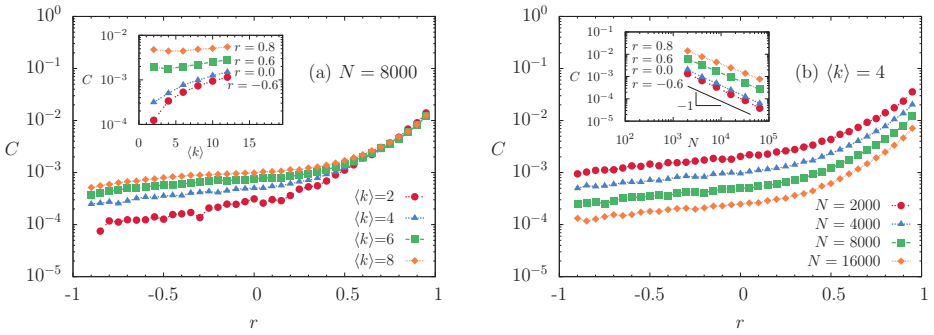


Figura 3.3: Coeficiente de agrupamento C como função de r para as redes de Poisson: (a) $N = 8000$ e diferentes valores de $\langle k \rangle$ (indicados na figura). O gráfico mostra um aumento monotônico de C com r. Existem dois regimes: um ligeiro aumento abaixo de $r \simeq 0,5$ e um aumento mais pronunciado acima desse valor. (b) $\langle k \rangle = 4$ e diferentes valores de N, também indicados na figura. À medida que o número de nós aumenta, C decresce segundo a lei assintótica $C \sim 1/N$, característica de gráficos aleatórios não correlacionadas. Os erros padrão são aproximadamente 10%. As linhas pontilhadas são um guia para os olhos. Os gráficos inseridos mostram C vs $\langle k \rangle$ (a) e N (b) para valores selecionados de r (-0,6; 0,0; 0,6 e 0,8).

médios (sobre 100 realizações, após 3×10^4 MCS) são exibidos como função do tamanho, para diferentes valores de γ. Para um tamanho fixo, quanto menor γ, mais estreito é o intervalo de valores permitidos para r. De fato, em redes

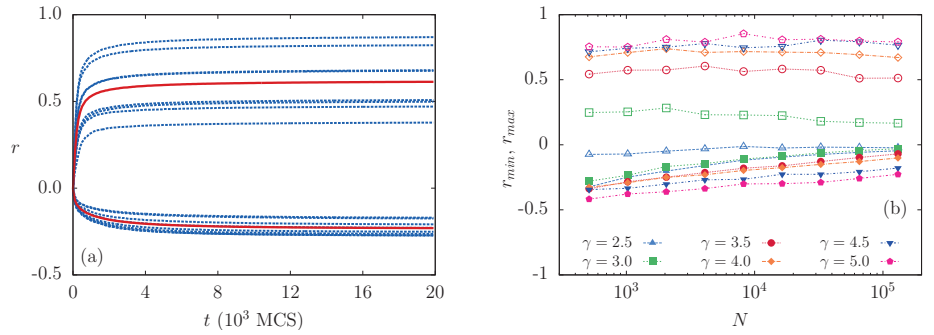


Figura 3.4: O intervalo de valores permitidos de r é restrito para a classe de redes com lei de potência: (a) evolução temporal de r, após definirmos $r_\star = 1$ (-1) a fim de obter $r_{\max}$ ($r_{\min}$), para redes com distribuição de grau em lei de potência (com $\gamma = 3,5$ e $N = 5000$). São mostradas oito amostras individuais (linhas finas) e suas respectivas médias (linhas grossas). (b) Valores extremos médios [$r_{\max}$ (símbolos abertos) e $r_{\min}$ (símbolos cheios), os erros padrão são, no máximo, 50%] vs o tamanho do sistema N para os diferentes valores de γ indicados na figura. Linhas pontilhadas e tracejadas são guias para os olhos para $r_{\max}$ e $r_{\min}$, respectivamente. Em todos os casos $\langle k \rangle = 4,00 \pm 0,04$. Para um dado tamanho, o intervalo permitido de r é tão mais estreito quanto menor for o γ.

PUC-Rio - Certificação Digital Nº 1122074/CA

restritas a uma dada sequência de graus, podem surgir limitações estruturais ou correlações negativas [97]. Por exemplo, a exclusão de múltiplas ligações dificulta a tendência natural de *hubs* (vértices com grau substancialmente alto, bem maior que a conectividade média) se conectarem uns aos outros, daí diminuindo a assortatividade. A medida em que diminuímos γ esse efeito é cada vez mais pronunciado.

Aqui vale relembrar um comentário que fizemos no Cap. 2.1. Com exceção das redes sociais, a maior parte das redes apresenta correlações negativas. Embora as razões para isso não sejam totalmente entendidas, um dos argumentos é que essa tendência está relacionada com o fato de que boa parte das redes reais são grafos simples [98]. Redes que têm somente ligações simples, na ausência de qualquer outro viés, apresentam correlação negativa porque o número de ligações entre *hubs* é limitada.

Restaria entender o porquê de redes sociais exibirem valores de r positivos. Uma das possíveis explicações é que as redes sociais exibem uma tendência a se dividir em grupos [99]. Esses grupos são altamente conectados entre si e praticamente desconectados do restante da rede. Os vértices dos grupos menores, em geral, possuem conectividade menor do que os vértices dos grupos maiores. Sendo assim, as pessoas dentro dos grupos menores vão estar conectadas com outras pessoas que *também* apresentam conectividade baixa; e o mesmo acontece no caso das pessoas com conectividade alta. Essa organização das redes sociais seria a origem dos valores positivos para a assortavidade. Podemos resumir a discussão dizendo que há uma predisposição natural das redes reais exibirem correlações negativas pelo fato de serem redes simples. Por outro lado, as redes sociais desfazem essa predisposição porque exibem uma tendência de se organizar em grupos [15, 47].

PUC-Rio - Certificação Digital Nº 1122074/CA

3.3.2 Efeitos de γ

Para um γ fixo, o intervalo de valores possíveis de r diminui com o tamanho do sistema, para $\gamma < 3$, devido à divergência das flutuações no limite de N grande [85]. Na ref. [100], uma restrição semelhante também foi observada para $2 < \gamma < 3$, embora ao invés de manter a conectividade média, $k_{\min}$ foi mantido constante ($k_{\min} = 6$). Nesse caso foi relatado que ambos os limites, superior e inferior, tendem para zero, portanto, $r \to 0$ no limite de uma rede infinita. De fato, observamos que nesse intervalo de γ (e.g., $\gamma = 2,5$) ambos os limites são negativos e à medida que N aumenta, o intervalo permitido colapsa para um valor negativo que tende a zero. Também notamos essas restrições nos limites para a correlação quando são considerados valores de γ maiores que 3.

A medida em que o valor de N cresce, o limite inferior também aumenta tendendo para zero ou, pelo menos, para um valor finito pequeno. Simultaneamente, o limite superior parece mais estável, porém seu comportamento assimptótico não é claro mesmo considerando redes com $N > 10^5$. Lembrando que conforme N aumenta leva cada vez mais tempo para se atingir o estado estacionário. O intervalo permitido de valores de r é bastante restrito para redes livres de escala. Contudo, analisamos sistematicamente casos com $\gamma > 3$ ($\gamma = 3,5$; $4,0$ e $4,5$), onde temos segundo momento finito. Mesmo nesse intervalo de γ, o intervalo acessível de r é limitado, procedemos, portanto, da seguinte forma: se o valor desejado $r_\star$ não for atingido dentro de uma tolerância de 10^{-3}, em 2×10^4 MCS, essa instância da rede é descartada e fazemos um nova tentativa. Se não tivermos pelo menos 100 sucessos em 200 tentativas, o procedimento é interrompido. Alternativamente ao modelo configuracional, também geramos as redes usando o modelo de anexação preferencial [16, 17], e obtivemos resultados semelhantes.

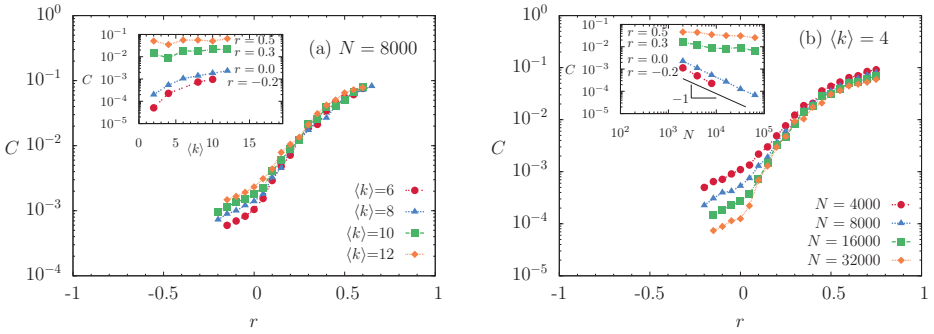


Figura 3.5: Coeficiente de agrupamento C como função de r, como na Fig. 3.3, mas para as redes com lei de potência com $\gamma = 4,0$: (a) $N = 8000$ e diferentes valores de $\langle k\rangle$. Na região assortativa, C atinge valores maiores do que para a classe de Poisson. (b) $\langle k\rangle = 4$ e diferentes números de nós N. Note que para as redes assortativas um grau finito de agrupamento parece persistir para grandes redes. Os erros padrão chegam a 50% para os valores menores de C. No gráfico inserido, os valores que estão ausentes são resultado da limitação em atingir os valores prescritos de r.

Os resultados para a classe de redes com lei de potência com $\gamma = 4,0$ são exibidos na Fig. 3.5. Os erros padrão são maiores para essa classe de redes, provavelmente devido à variabilidade na cauda das distribuições de nós de amostra para amostra. Resultados para os outros dois valores de γ estudados (3,5 e 4,5) exibem características similares ao caso $\gamma = 4,0$ usado como exemplo representativo, mesmo que o terceiro momento se torne divergente no caso $\gamma = 4,0$.

PUC-Rio - Certificação Digital Nº 1122074/CA

Dois regimes também são observados, com um crossover agora próximo de $r = 0$, mas, por outro lado diferenças qualitativas aparecem em comparação ao caso de Poisson. C aumenta rapidamente com r, atingindo, para redes assortativas, valores maiores que para a classe de redes de Poisson. Esses valores maiores são causados pela inclusão de links entre os hubs, ausentes no caso de Poisson. Um valor de $r_\star$ grande faz com que haja uma tendência desses nós se conectarem, o que contribui não só para o aumento de r como também de C. Com relação aos efeitos de tamanho finito, abaixo do crossover, os valores de C são causados ao tamanho finito da rede. No entanto, na região assortativa (acima do crossover), parece que um grau finito de agrupamento persiste para redes grandes (veja o gráfico inserido da Fig. 3.5(b)), em contraste com o caso de Poisson e com a região disassortativa. De fato, note que, quando N aumenta uma ordem de magnitude, C decresce também uma ordem de magnitude na região disassortativa, enquanto C permanece da mesma ordem de magnitude na região assortativa. Mesmo que C vá a zero no limite de uma rede infinita, já que o decaimento é bastante lento, um coeficiente de agrupamento efetivo pode permanecer para redes de tamanho moderado/grande. Discutiremos a relação entre C e r com mais detalhes na Sec. 3.4. Para o coeficiente de agrupamento local $\bar{C}$, uma dependência qualitativa similar é observada, mas com valores menores. Ademais, $\bar{C}$ aumenta linearmente com $\langle k \rangle$, no intervalo analisado, para qualquer valor de r, não somente em redes disassortativas, e $\bar{C}$ decai com N para qualquer r. Para $r = 0$, $C = \bar{C} = (\langle k^2 \rangle - \langle k \rangle)^2/(N\langle k \rangle^3)$ [86], como esperado.

3.3.3
Efeitos de r sobre as distancias características

Vamos analisar agora a influência de r em distâncias características da rede. A dependência das medidas $1/E$, L e D em r é exibida na Fig. 3.6, para redes de Poisson e lei de potência, com $N = 8000$ e $\langle k \rangle \simeq 4$. $1/E$ e L têm valores próximos, sistematicamente deslocados. Numa primeira aproximação, ambos tipos de rede fornecem resultados similares para $1/E$ (daí também para L), dado um N e $\langle k \rangle$. No entanto, o diâmetro D é mais dependente do tipo de rede. Ele é maior e mais fortemente afetado por r no caso homogêneo de Poisson.

Gráficos para $1/E$ vs r para diferentes valores de N e $\langle k \rangle$ são mostrados nas Figs. 3.7 e 3.8 para redes de Poisson e lei de potência, respectivamente. Nos dois casos, as redes exibem a propriedade de mundo pequeno [13] (até mesmo menor para redes com lei de potência) com um lento crescimento (logarítmico) com N e um decrescimento suave com $\langle k \rangle$ (veja os gráficos inseridos das Figs.

PUC-Rio - Certificação Digital Nº 1122074/CA

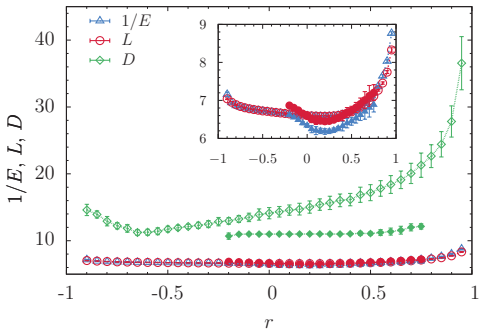


Figura 3.6: Medidas de distância $1/E$, L e diâmetro D, como função de r para redes Poisson (símbolos abertos) e lei de potência com expoente $\gamma = 4,0$ (símbolos cheios). Em todos os casos $N = 8000$ e $\langle k \rangle = 4,00 \pm 0,04$. O gráfico inserido é um amplificação do gráfico principal. À primeira vista, para um dado N e $\langle k \rangle$, os dois tipos de redes exibem valores similares de $1/E$ e L. O diâmetro é mais sensível ao tipo de rede e é mais afetado por r no caso de Poisson.

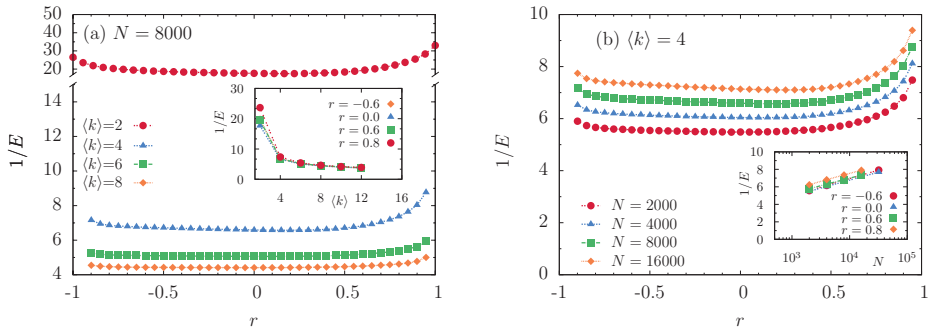


Figura 3.7: Distância média $1/E$ como função de r para redes de Poisson: (a) $N = 8000$ e diferentes valores de $\langle k \rangle$ (indicados na figura). As redes apresentam a propriedade de mundo pequeno. (b) $\langle k \rangle = 4,00 \pm 0,04$ e diferentes valores de N (indicados na figura). Os efeitos de r sobre o caminho médio são significativos apenas para $\langle k \rangle$ pequeno devido à fragmentação da rede.

3.7 e 3.8).

No caso de Poisson, a distancia média não depende significantemente de r quando $\langle k \rangle$ não é muito pequeno ($\geqslant 6$), como indicado pelos gráficos relativamente planos na Fig. 3.7(a). Somente para $\langle k \rangle$ pequeno existem efeitos importantes para correlações assortativas. Por exemplo, para $\langle k \rangle \simeq 4$ (Fig. 3.7(b)), $1/E$ aumenta aproximadamente duas unidades de $r \simeq 0$ para $r \simeq 1$, em todos os tamanhos analisados. Este efeito é ainda maior para $\langle k \rangle = 2$ onde L aumenta quase por um fator dois no mesmo intervalo de r, como é mostrado na Fig. 3.7(a) para $N = 8000$.

PUC-Rio - Certificação Digital Nº 1122074/CA

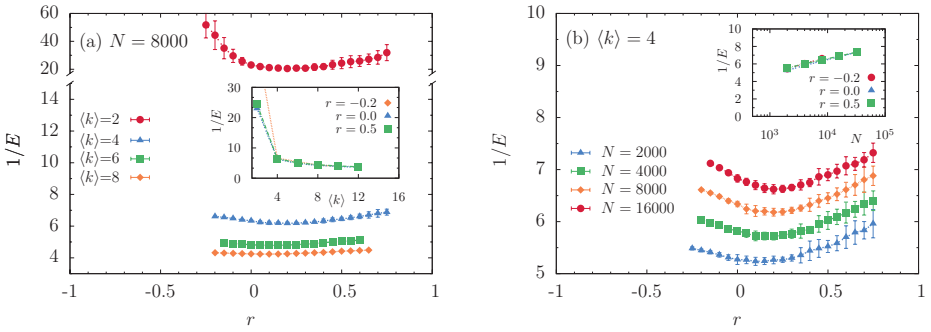


Figura 3.8: Distância média $1/E$ como função de r, como na Fig. 3.7, mas para rede com lei de potência com $\gamma = 4,0$. Neste caso, as distâncias são mais curtas do que nas redes de Poisson.

3.3.4
Estrutura das componentes

A fim de interpretar melhor esses resultados, nós investigamos a estrutura das componentes resultantes do processo de religamento. Medimos o tamanho da maior componente (que iremos chamar de N_1), o número de componentes e o tamanho médio de componentes com tamanho maior que 1. Os gráficos são apresentados na Fig. 3.9 para $\langle k\rangle = 2$ e 4. Para $\langle k\rangle = 4$, o tamanho relativo da maior componente (círculos) é aproximadamente 0.98 para a maior parte do intervalo de r, note no entanto que ele decai levemente próximo de $|r| = 1$ (o que é mais evidente para $\langle k\rangle = 2$). Conforme $\langle k\rangle$ aumenta, o número de fragmentos decai rapidamente e o tamanho médio S (triângulos) tende a unidade, o que significa que somente nós isolados estão desconectados (lembrando que $P(0) = \mathrm{e}^{-\langle k\rangle}$). Vemos portanto que o aumento da distância média relativamente a $|r| = 1$, observado para $\langle k\rangle$ pequeno na Fig. 3.7, pode ser simplesmente um reflexo da fragmentação da rede. Para valores altos da assortatividade, a rede tende a se fragmentar em grupos de vértices que têm a mesma conectividade. De mais a mais, para valores de $\langle k\rangle$ aproximadamente maiores que 1, a análise da percolação para redes de Poisson feitas por Noh [83] mostra que o tamanho da maior componente é menor para redes assortativas do que para redes disassortativas e neutras. Entretanto, à medida que $\langle k\rangle$ aumenta, a fração de vértices que não pertencem ao maior cluster se torna desprezível, mais lentamente quanto maior for a assortividade da rede. Por isso, no limite de $\langle k\rangle$ grande, a distância média permanece invariante sob mudanças em r. Assim, os processos de transporte modelados nessas redes podem sofrer um impacto importante somente quando r é grande e $\langle k\rangle$ pequeno. Quanto maior for a separação típica entre os nós, mais lenta é a propagação.

Note que para redes com lei de potência, para N fixo e $\langle k\rangle > 2$, existe

PUC-Rio - Certificação Digital Nº 1122074/CA

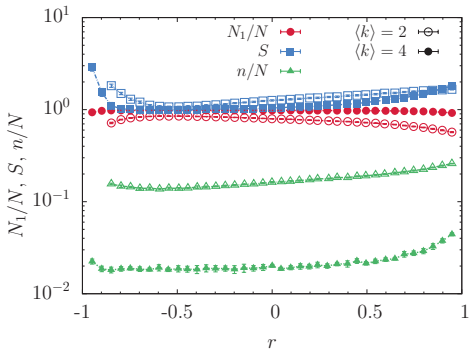


Figura 3.9: Análise de clusters. Gráficos de N_1/N (onde N_1/N é o tamanho do maior cluster) [círculos], tamanho médio dos clusters de tamanho finito, S [quadrados], e número de clusters, n [triângulos], como função de r para as redes de Poisson com $\langle k\rangle = 2$ (símbolos abertos) e 4 (símbolos cheios). A fragmentação da rede observada para altos valores da assortatividade reflete o agrupamento de vértices com o mesmo grau. A figura mostra os resultados para $N = 8000$. Resultados para tamanhos os $N = 4000$, 8000 e 16000 colapsam em uma curva única (não exibida).

um intervalo de r onde as distâncias são menores no que para redes de Poisson (Fig. 3.6), e são ainda menores para valores de γ menores (não mostrado). Além do mais, $1/E$ se torna mais sensível ao coeficiente r (gráficos com formato de sorriso), na região onde esses gráficos são planos em redes de Poisson. Note também que o valor mínimo do caminho médio ocorre para regiões ligeiramente assortativas ($r \gtrsim 0$), aumentando lentamente com N (Fig. 3.8(b)). A análise da distribuição de componentes para $\gamma = 4,0$, mostra que para $\langle k\rangle \geqslant 4$, há somente uma componente, de tamanho N, para todo r. Somente para $\langle k\rangle = 2,0$, observamos fragmentação com $N_1/N \simeq 0,7 - 0,8$, $n/N \simeq 0,06$, $S \simeq 4$ para todo $N > 2000$ (gráficos não mostrados).

Para o caminho médio L, observamos similaridades qualitativas nos resultados, embora deslocados para valores levemente maiores, como ilustrado na Fig. 3.6.

3.3.5 Redes reais

Também aplicamos o procedimento de religamento descrito na Sec. 3.2 à sequências de grau reais. As redes usadas foram simetrizadas e os pesos entre as ligações foram ignorados. Na Fig. 3.10 mostramos o comportamento de C e $1/E$ vs r, para as seguintes redes: PGP (encrypted communication network using Pretty Good Privacy encryption algorithm) [40], P2P (Gnutella peer-to-peer network) [50], PGR (electrical power grid of the western United States) e APC (astrophysics collaboration network) [27, 28].

PUC-Rio - Certificação Digital Nº 1122074/CA

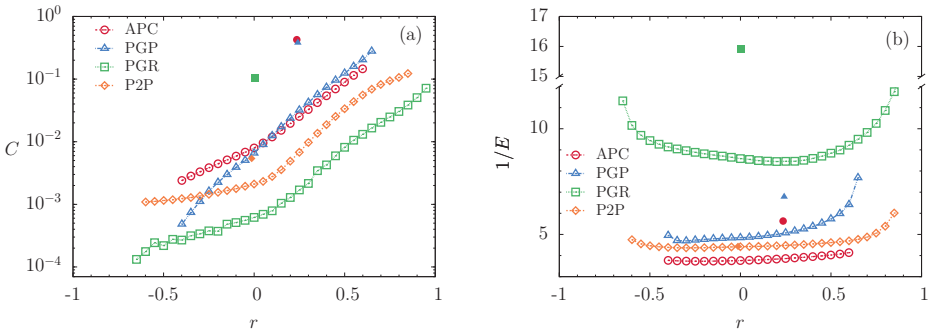


Figura 3.10: C e $1/E$ vs r, para redes reais. Valores originais antes do processo de religação, estão também indicados na figura (símbolos cheios). PGP (Pretty Good Privacy encrypted communication network) [40]: $N = 10680, \langle k \rangle \simeq 4,55$; P2P (Gnutella peer-to-peer network) [50]: $N = 10876$, $\langle k \rangle \simeq 7,35$; PGR (power grid) [13, 28], $N = 4941, \langle k \rangle \simeq 2,67$; APC (astrophysics collaboration) [27, 28]: $N = 16706$, $\langle k \rangle \simeq 14,5$. Além dos detalhes diferentes das sequências de conectividade reais, podemos interpretar as principais características destas redes em termos dos valores observados nas classes de Poisson e lei de potência.

Primeiro, note que, em todos os casos, o coeficiente de agrupamento C é maior nas redes reais do que nas redes randomizadas (ensemble r), como já foi observado em outros exemplos na Ref. [82]. A distância média é também tipicamente maior em redes reais. Uma exceção é a rede P2P, caracterizada por um valor de $1/E$ típico do ensemble r. As redes reais "religadas" apresentam alguns comportamentos típicos observados nos casos artificiais. Segue alguns comentários que surgem da comparação:

(i) PGR (power grid) apresenta plots de C vs r e $1/E$ vs r que estão em bom acordo como aqueles observados para parâmetros similares no caso de Poisson. De fato, a distribuição de conectividade para PGR decai exponencialmente.

(ii) P2P apresenta um decaimento em lei de potência na sua distribuição de graus para $k > 10$ com expoente próximo de $\gamma = 4$. Ambos os gráficos para P2P estão em concordância com aqueles obtidos para a classe $\gamma = 4$ com valores similares de N e $\langle k \rangle$, apesar da única coisa em comum entre as distribuições ser a cauda em lei de potência.

(iii) PGP (que tem tamanho similar à P2P) apresenta uma distribuição que decai com expoente $\gamma < 3$ para $k < 50$ e $\gamma \simeq 4$ para valores maiores de k [40]. O gráfico para $1/E$ vs r apresenta valores maiores de $1/E$ que a rede P2P, o que é consistente com seu $\langle k \rangle$. No entanto, o gráfico de C vs r da rede PGP desvia daquilo que seria o comportamento padrão, apresentando maiores valores de C que aumentam com r em um regime

PUC-Rio - Certificação Digital Nº 1122074/CA

único. O intervalo de valores permitidos de r é sensível também a outras características além da cauda da distribuição de graus. Esses desvios podem ser atribuídos a diferentes inícios do regime de lei de potência

(iv) Finalmente, APC tem um decaimento em lei de potência com expoente $\gamma \simeq 1$ e cut-off exponencial para $k > 50$ [27, 28]. O valor baixo e aproximadamente constante do gráfico $1/E$ vs r é esperado para uma rede com $\langle k \rangle$ grande, quase que independente da classe da distribuição de graus. Os valores altos de C também são consistentes com distribuições heterogêneas com $\langle k \rangle$ grande.

Portanto, apesar dos detalhes diferentes das distribuições de grau reais, as características principais observadas podem ser interpretadas em termos das classes analisadas com valores correspondentes dos parâmetros $\langle k \rangle$ e N.

PUC-Rio - Certificação Digital Nº 1122074/CA

3.4 Discussão e considerações finais

Para todas as classes de rede consideradas, C aumenta com r em todo o intervalo permitido de valores de r. Esse comportamento foi observado também na Ref. [82], onde somente valores positivos de r tinham sido analisados e também na Ref. [85] embora definições diferentes para a transitividade e para a correlação tenham sido usadas.

No entanto, observamos que, no ensemble r para redes de ER, o valor finito do coeficiente de agrupamento é devido tipicamente a efeitos de tamanho finito, tal que, no limite de uma rede infinita, a transitividade vai a zero com $1/N$ como em um grafo sem correlação.

Em contraste, para redes com lei de potência, acima de um certo valor de r, surge uma transitividade não nula significativa, aparentemente persistindo para redes com N grande.

Em todo caso, desde que o processo de religamento no ensemble r retorna valores tipicamente pequenos de C, a transitividade não parece contribuir para alcançarmos o valor pre-estabelecido de r.

Para identificar quais são os fatores que contribuem para r, é conveniente considerar a expressão dada pela Eq. A-11.

Seguindo a decomposição feita por Estrada [84] (veja Apêndice A.2), note que a quantidade $\tilde{P}_3 \equiv \sum_{(k,k')}(k-1)(k'-1)$ (com a soma feita sobre todos os diferentes pares de vértices vizinhos) é o número de caminhos de tamanho 3 (abertos e fechados). Portanto $\tilde{P}_3 = 3n_\triangle + P_3$, onde P_3 é o número de caminhos não triangulares de tamanho 3 (envolvendo 4 vértices). Assim como foi feito na Eq. (3-5) para $3n_\triangle$, re-escalamos também P_3 pelo número de caminhos

de tamanho 2 (cunhas), definindo a grandeza *ramificação escalada* B (*scaled branching* em inglês)

$$B = \frac{P_3}{P_2}\,. \tag{3-8}$$

Assim, segundo desenvolvido detalhadamente no Apêndice A.2, a Eq. (2-31) pode ser escrita como

$$r = \frac{\langle k\rangle(\langle k^2\rangle - \langle k\rangle)\left(B + 1 - \frac{\langle k^2\rangle}{\langle k\rangle} + C\right)}{\langle k\rangle\langle k^3\rangle - \langle k^2\rangle^2}\,. \tag{3-9}$$

A expressão (3-9) é determinada pelos três primeiros momentos de $P(k)$, e pelas quantidades B e C que são as variáveis que incorporam a informação sobre as correlações entre as conectividades de dois vértices.

Na rede de Erdös e Rényi, os 3 primeiros momentos $\mu_n = \sum x^n P(x)$ da distribuição de graus (distribuição de Poisson) são dados por

$$\mu_1 = \langle k\rangle\,, \quad \mu_2 = \langle k\rangle\,(1 + \langle k\rangle) \quad \text{e} \quad \mu_3 = \langle k\rangle\,(1 + 3\langle k\rangle + \langle k\rangle^2). \tag{3-10}$$

Substituindo as Eqs. 3-10 em A-21, obtemos a expressão

$$r = B - \langle k\rangle + C\,. \tag{3-11}$$

Claramente, correlações disassortativas são favorecidas por um coeficiente de agrupamento que tende a zero. Somente para r positivo, o crescimento de C é conveniente, mas $B - \langle k\rangle$ pode variar num intervalo maior que C (duas vezes maior nesse caso). A existência de dois regimes no crescimento de $C(r)$, observado na Fig. 3.3, é consistente com esse quadro. Em outras palavras, a Eq. (3-11) indica que, no processo de religamento em redes de Poisson, r é governado predominantemente por B, já que $P(k)$ é conservado e C apresenta valores bem pequenos.

Para redes com distribuição em lei de potência, alguns efeitos qualitativos similares ocorrem já que, como veremos a seguir, a relação de r com B é sempre linear, e C é restrito a um intervalo menor que o intervalo de valores permitidos para B. A formação de triângulos nesse caso também contribui somente para valores assortativos de r (acima do crossover), com valores de C maiores que no caso de Poisson, mas ainda pequenos. Então também nesse caso o aumento da ramificação deve conduzir o religamento.

Nestas redes livres de escala, o m-ésimo momento μ_m de uma distribuição com decaimento em lei de potência $P(k) \sim k^{-\gamma}$, limitada entre $k_{\min}$ e $k_{\max}$, é dado pela expressão

$$\mu_m = \frac{1-\gamma}{1+m-\gamma}\,\frac{k_{\max}^{1+m-\gamma} - k_{\min}^{1+m-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}\,, \tag{3-12}$$

onde vamos considerar que $k_{\max} \sim N$. Assim obtemos

PUC-Rio - Certificação Digital Nº 1122074/CA

$$\mu_m = \frac{1-\gamma}{1+m-\gamma}\,\frac{N^{1+m-\gamma} - k_{\min}^{1+m-\gamma}}{N^{1-\gamma} - k_{\min}^{1-\gamma}}\,. \tag{3-13}$$

Em todos os casos $\gamma > 1$, então o denominador é da ordem de $k_{\min}^{1-\gamma} \sim \mathcal{O}(1)$.

Em contraste com o caso de Poisson, os outros termos na Eq. (3-9), relacionados com os momentos de $P(k)$, podem ter uma influência crucial em r por causa da divergência, no limite de uma rede infinita, dos n-ésimos momentos para $\gamma \geqslant n+1$. Veja que, se $1+m-\gamma > 0$ então $\gamma < 1+m$ e $k^\star \equiv N^{1+m-\gamma}$ cresce com N, caso contrário $k^\star$ vai a zero.

Vamos analisar o limite de $k_{\max}$ (daí N) grande, deixando de lado casos marginais (logarítmicos). Cabe notar que os mesmos resultados acima podem ser obtidos usando o corte natural $k_{\max} \sim N^{1/(\gamma-1)}$.

Considerando as expressões para os momentos: (e.g., Eq. (3-6)), teríamos, para $3 < \gamma < 4$, que μ_3 diverge. Logo

$$r \sim \frac{B - \mathcal{O}(1)}{\mathcal{O}(N^{4-\gamma})}\,.$$

onde $\mathcal{O}(x^\alpha) \sim ax^\alpha$, com $a > 0$.

Entretanto, para $2 < \gamma < 3$, μ_2 e μ_3 divergem. Neste caso

$$r \sim \frac{\mathcal{O}(N^{3-\gamma})(B - \mathcal{O}(N^{3-\gamma}))}{\mathcal{O}(N^{4-\gamma}) - \mathcal{O}(N^{6-2\gamma})} \sim \frac{B - \mathcal{O}(N^{3-\gamma})}{\mathcal{O}(N)}\,.$$

Para se chegar no limite inferior $r = -1$, precisamos ter o mínimo B possível (sendo B não negativo). Se B for de ordem maior que os outros termos no numerador, então não seria possível obter valores negativos de r já que B é uma grandeza não negativa e irá, nesse caso, dominar o numerador. Daí, valores negativos de r podem surgir somente se B é da mesma ordem ou menor que outros termos, mas nesse caso $r \to 0$ no limite de N grande. Isso explica porque o limite inferior tende a 0 quando $\gamma \leqslant 4$ (see Fig. 3.4(b)). Seguindo essa linha, no entanto, não é esperado que $r_{\min}$ vá a zero quando $\gamma > 4$, mas tenda a um pequeno valor finito. Veja que, se $\gamma > 4$, então nenhum termo diverge. Logo

$$r \sim B - \mathcal{O}(1)\,.$$

Similarmente, para atingir um limite superior não nulo para r, B precisa crescer como o denominador, caso contrário, o limite superior será negativo e também irá a zero quando $N \to \infty$, levando também ao colapso do limite superior. No entanto, isso não necessariamente acontece se o religamento faz com que B cresça o suficiente, o que parece acontecer de acordo com a Fig. 3.4(b).

PUC-Rio - Certificação Digital Nº 1122074/CA

A conexão entre r e as distâncias medidas não é tão direta analiticamente. Resultados numéricos mostram que, para redes com uma distribuição localizada de graus, mudar r modifica significativamente o caminho médio somente quando correlações são assortativas ($r > 0.5$) e $\langle k \rangle$ é pequeno. Essas mudanças podem estar relacionadas com a fragmentação induzida pelo religamento, que diminui com o aumento de $\langle k \rangle$. Portanto, o impacto de r se torna menos importante conforme $\langle k \rangle$ aumenta. Por outro lado, a influência no diâmetro é mais dramática. Em redes com lei de potência, a modificação no menor caminho médio por r é um pouco mais marcada, mesmo que a fragmentação esteja ausente para $\langle k \rangle \geqslant 4$, enquanto o diâmetro não é fortemente afetado. Em ambos os casos, a modificação nas distâncias caraterísticas que ocorrem quando se varia r podem afetar processos dinâmicos e devem ser levados em consideração na interpretação e modelagem de experimentos numéricos feitos nessas redes. Nos capítulos subsequentes, usamos redes sem correlação. Como temos modelos novos, primeiro consideramos o impacto de $P(k)$ nos resultados. Trabalhos futuros (em andamento) podem fazer uso dos resultados apresentados aqui.

PUC-Rio - Certificação Digital Nº 1122074/CA

4
Padrões estatísticos na avaliação de produtos

Nos dias de hoje, os usuários e consumidores podem fazer resenhas e avaliar produtos através de serviços online, o que fornece imensas bases de dados que podem ser usadas para explorar as preferências de uma dada população e revelar padrões comportamentais. Nesse trabalho, nós investigamos os padrões que surgem da avaliação de filmes, considerando como fonte o IMDb (Internet Movie Database), um dos sites mais visitados do mundo.

Nós encontramos que a distribuição de votos apresenta um comportamento livre de escala sobre diversas décadas, com um expoente muito próximo de 3/2 com corte exponencial. É surpreendente que esse padrão manifeste-se independente de atributos do filme, como a nota média, idade ou gênero, com exceção de alguns gêneros e de filmes com grande orçamento.

Esses resultados apontam para um mecanismo subjacente bastante geral para a propagação de adoções em audiências potenciais que é independente das caraterísticas intrínsecas do filme e que pode ser entendida através um modelo simples de espalhamento que está relacionado a uma dinâmica de avalanche de campo médio.

4.1
Introdução

Vimos que a física estatística tem contribuído para o estudo de sistemas sociais através de modelos teóricos, fornecendo *insights* e descobrindo leis que governam muitos fenômenos, como o espalhamento de informação, rumores e opiniões [3, 101]. Enquanto existem progressos no desenvolvimento de modelos teóricos, sua validação através da confrontação direta com dados reais ainda não foi obtida.

Graças a popularização de redes sociais online e, mais recentemente, de sites que permitem ao usuário avaliar e escrever resenhas de produtos, novas possibilidades têm surgido para explorar esse campo. Por um lado, tais avaliações fornecem informações relevantes sobre como as preferências de uma sociedade estão distribuídas. Por outro, os dados empíricos permitem aferir ou validar modelos teóricos que podem então ser usados para interpretar

PUC-Rio - Certificação Digital Nº 1122074/CA

e/ou predizer os resultados observados. Além de ter aplicações práticas que intentam, por exemplo, melhorar sistemas de recomendação.

Aqui, pretendemos explorar padrões que surgem da avaliações de filmes, como uma fonte de informação para a distribuição de preferências das pessoas. Analisar o número de votos (onde um voto consiste em atribuir uma nota a um filme do catálogo através de um número de estrelas) ao invés, por exemplo, do número total de espectadores no cinema, é uma maneira conveniente de mensurar a popularidade de um filme. As pessoas podem assistir a um filme de diversas maneiras; portanto, o número de espectadores no cinema fornece somente uma informação parcial. Analisar o ganho/renda total de um filme sofre do mesmo problema: é difícil somar todas as fontes de ganho de um filme. Como alternativa, o número de votos é independente dos meios que foram usados para assistir ao filme. Além do mais, é uma fonte de informação direta e acessível, bem como muito mais rica, dado que podemos extrair a opinião do público em relação ao filme (veja a Fig. 4.1).

Ademais, enquanto alguns modelos teóricos indicam que a distribuição de cascatas de adoção segue uma lei de potência [63, 64, 102, 103], existe uma falta de evidências empíricas que validem esses resultados. Nesse sentido, o trabalho presente pode contribuir de maneira significativa.

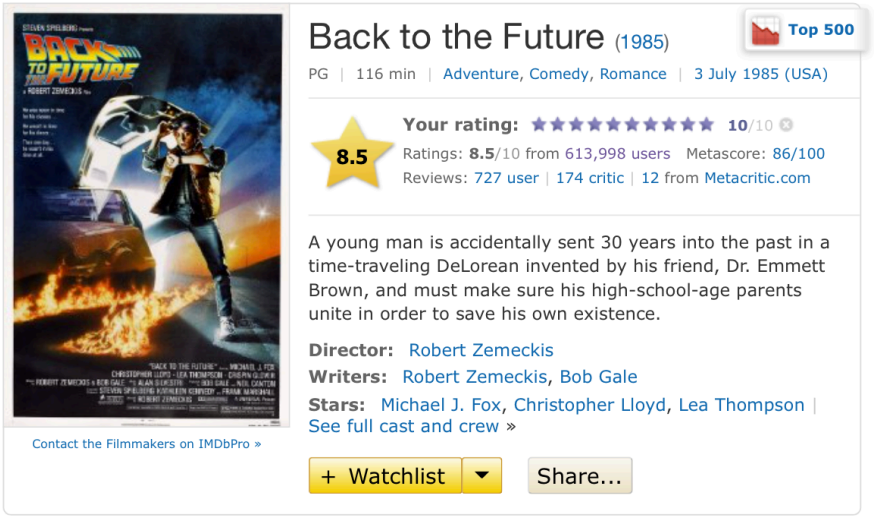


Figura 4.1: Ilustração de um item do catálogo do IMDb, usamos como exemplo o filme *De volta para o futuro*. Podemos extrair de maneira direta informações sobre o filme, como número de votos dos usuários n_v, data de lançamento, gênero, etc. Para coletar os dados dos itens de maneira sistemática usamos um *Web crawler* como descrito no Cap. 2. Ref: `http://www.imdb.com/title/tt0088763`.

PUC-Rio - Certificação Digital Nº 1122074/CA

4.2
Resultados

Resultados empíricos

Analisamos os dados do IMDb (Internet Movie Database), uma fonte de informação sobre filmes e material relacionado que permite que os visitantes façam resenhas e avaliem online filmes e outros itens de entretenimento (Apêndice B). É um dos sites mais visitados do mundo e o primeiro nesse campo [104].

Coletamos o número de votos de cada item, onde um voto consiste em atribuir uma classificação numa escala de 1 a 10 estrelas; 1 estrela significa "terrível" e 10 estrelas significa "excelente". A partir disso, estimamos a distribuição de probabilidade $P(n_v)$ do número de votos n_v construindo um histograma normalizado, como mostrado na Fig. 4.2. Observamos um comportamento livre de escala notável que se estende por aproximadamente 4 ordens de magnitude em n_v, com uma lei de potência com expoente $1,51 \pm 0,02$ e um corte exponencial. Como um teste para um possível viés, consideramos como alternativa a fonte de dados fornecida pelo Netflix [105], e observamos o mesmo comportamento de escala.

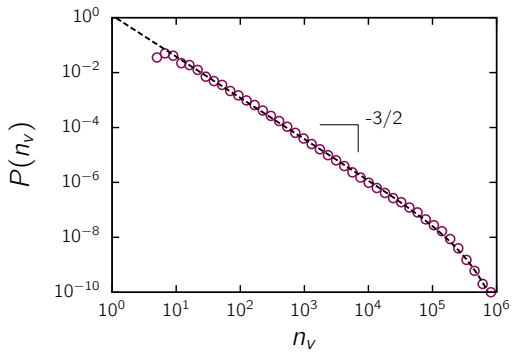


Figura 4.2: Distribuição do número de votos para os filmes do IMDb. A linha pontilhada ajusta os dados com a função $P(n_v) = An_v^{-\alpha} \exp(-\lambda n_v)$ onde A é a constante de normalização, $\alpha = 1,51$ e $\lambda = 4.0\ 10^{-6}$.

Existem vários processos que podem gerar leis de potência [106, 107], em particular, uma mistura estatística de diferentes escalas. A utilização de uma mistura de processos diferentes para explicar o comportamento de escala é, de fato, uma possibilidade nesse caso, já que os dados envolvem diferentes categorias de itens, e.g., filmes para TV, longas-metragens, e itens de diferentes

PUC-Rio - Certificação Digital Nº 1122074/CA

idades e com diferentes notas. Portanto, devemos conduzir uma análise mais refinada separando os dados em subconjuntos por diferentes critérios.

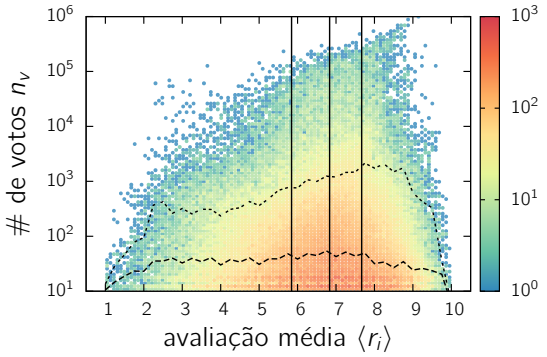


Figura 4.3: Mapa de cores do número de votos n_i vs nota média $\langle r_i \rangle$ para cada filme i do IMDb. Nessa figura e nos mapas de cores subsequentes, cada ponto contém o número de filmes indicado pela escala de cores. As regiões brancas indicam nenhum filme. As linhas pontilhadas e tracejadas representam a média aritmética e geométrica, respectivamente. As linhas verticais indicam os quartis, os quais dividem o conjunto de dados em 4 grupos $(G^r_{4,1}, \ldots, G^r_{4,4})$ de mesmo tamanho.

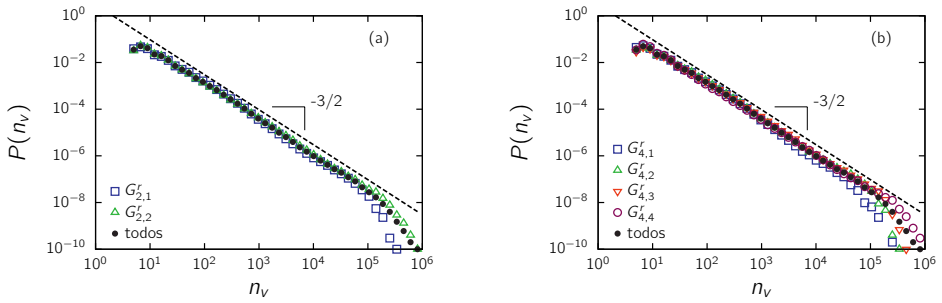


Figura 4.4: Impacto das avaliações na distribuição de votos $P(n_v)$ nos filmes do IMDb, para (a) os dois grupos $G^r_{2,1}$ e $G^r_{2,2}$ separados pela mediana e (b) os quatro grupos $G^r_{4,1}, \ldots, G^r_{4,4}$ determinados pelos quartis, como indicado na Fig. 4.3. Nesta e noutras figuras equivalentes, a linha tracejada com inclinação -3/2 é exibida para comparação com a distribuição $P(n_v)$ do conjunto de dados total.

Classificação Começamos por inspecionar a relação entre o número de votos e a nota média $\langle r_i \rangle$ de cada um dos filmes i. Na Fig. 4.3, nós representamos um gráfico de dispersão de n_i contra $\langle r_i \rangle$. Embora os pontos estejam muito espalhados, algumas tendências podem ser identificadas. A escala logarítmica no eixo das ordenadas implica uma distribuição larga do número de votos para

PUC-Rio - Certificação Digital Nº 1122074/CA

cada intervalo de nota média, como observado para o conjunto total (Fig. 4.2). A distribuição de votos dentro de um dado intervalo de $\langle r_i \rangle$ é assimétrica, com um viés na direção dos valores positivos, o que difere do que seria esperado de um perfil aleatório. Vemos que o número máximo de votos aumenta com a nota média, indicando que números de votos muito grandes estão associados com filmes bem avaliados. Um padrão muito similar foi reportado para avaliação do Yahoo music [108, 109].

O número de votos deve aumentar com o número de pessoas que assistiram o filme, e espera-se que esse número seja maior quando muitas pessoas gostaram do filme, o que também conduz a valores de $\langle r_i \rangle$ maiores. Portanto, esperamos que haja uma correlação positiva entre a avaliação média e o número de votos. Embora o corte aumente com a nota média, um grande número de filmes bem avaliados recebem poucos votos. A média geométrica do número de votos como função da avaliação média apresenta um perfil plano (Fig. 4.3, linha tracejada), embora a média aritmética aumente ligeiramente porque é mais influenciada por valores extremos.

A pergunta natural que surge é: como a distribuição normalizada de votos é afetada pela classificação dos filmes? Para examinar esta questão, computamos $P(n_v)$ para grupos distintos de dados, dividindo o conjunto total pela mediana em relação à avaliação média. Isto é, consideramos separadamente a metade superior e inferior dos dados $G^r_{2,1}$ e $G^r_{2,2}$, respectivamente. Também consideramos os quartis, e subdividimos os dados nos grupos $G^r_{4,1}, \ldots, G^r_{4,4}$. Os resultados são apresentados na Fig. 4.4. Note que $P(n_v)$ é quase insensível ao fato das avaliações serem favoráveis ou não, ou seja, os dados acima e abaixo da mediana apresentam o mesmo padrão, coincidindo por quase 4 décadas no regime de lei de potência (Fig. 4.4a). A principal discrepância aparece no corte exponencial acima de 10^5 votos, onde o decaimento ocorre mais rápido para a metade dos dados com avaliação média menor. Isto é consistente com o fato de que os filmes que foram pior avaliados não recebem um número de votos tão grande quanto os filmes com avaliação melhor. No entanto, o mesmo comportamento livre de escala permanece para as duas metades dos dados sobre diversas ordens de magnitude, apontando para um mecanismo independente dos atributos que possam ser medidos através da nota. As mesmas tendências são observadas no nível dos quartis, como mostrado na Fig. 4.4b, e dividindo os dados por estrelas (não mostrado).

Ano de lançamento Em seguida, investigamos o impacto da idade dos itens em $P(n_v)$. A Fig. 4.5 mostra o gráfico de dispersão para o número de votos contra ano de lançamento para cada filme. Nota-se que o número de votos

PUC-Rio - Certificação Digital Nº 1122074/CA

depende da idade do filme. O corte no número de votos aumenta conforme a idade diminui. Concomitantemente, filmes mais recentes tendem a receber mais votos em média. Todas essas tendências são consistentes com a expectativa de que filmes novos recebem substancialmente mais votos que os antigos. Além do mais, filmes antigos podem receber votos somente retroativamente (o sistema de registro de usuários do IMDb foi lançado em 1997 [110]), enquanto um filme novo recebe votos contemporaneamente a sua fase mais efervescente.

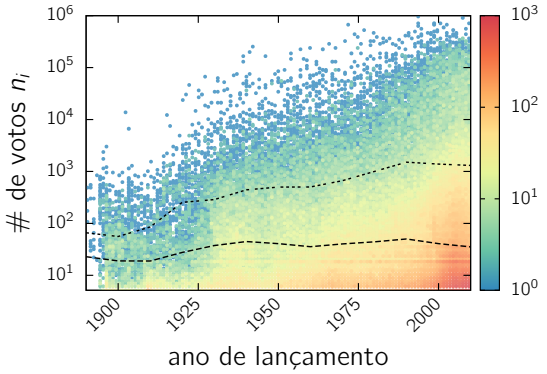


Figura 4.5: Mapa de cores do número de votos n_i vs ano de lançamento para os filmes do IMDb. Como na Fig. 4.3, linhas pontilhadas e tracejadas representam a média aritmética e geométrica, respectivamente.

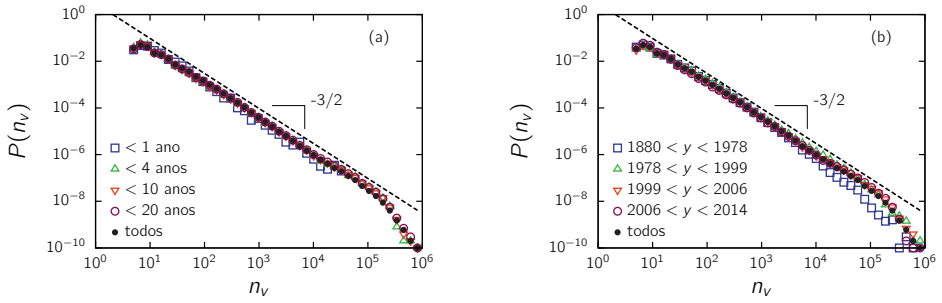


Figura 4.6: Impacto da idade do filme na a distribuição de votos $P(n_v)$ para os filmes do IMDb (a) filmes com menos do que um dado número de anos, e (b) filmes lançados dentro do intervalo indicado na figura. Os intervalos foram escolhidos para conter o mesmo número de filmes.

Também neste caso, analisamos $P(n_v)$ para subconjuntos separados pelo ano de lançamento. Conjuntos de filmes mais novos que y anos, mesmo aqueles lançados dentro de um ano e portanto com pior estatística, apresentam a mesmo padrão sobre todo o intervalo. Inspecionamos os dados ainda mais de perto, agrupando os itens pelo intervalo de tempo decorrido desde o lançamento (Fig. 4.6b). Uma discrepância significativa existe somente para filmes mais

PUC-Rio - Certificação Digital Nº 1122074/CA

antigos, na cauda acima de 10^4 votos, em concordância com a Fig. 4.5, a qual mostra que o corte ocorre em valores menores de n_v à medida que a idade aumenta. No entanto, a região de escala ainda cobre mais do que 3 ordens de magnitude de n_v, mesmo neste caso.

Gênero Representamos também $P(n_v)$ para séries de TV e longas-metragens separadamente (Fig. 4.7). Ambas categorias comportam-se de maneira similar ao conjunto total, com alguma discrepância ocorrendo apenas no corte.

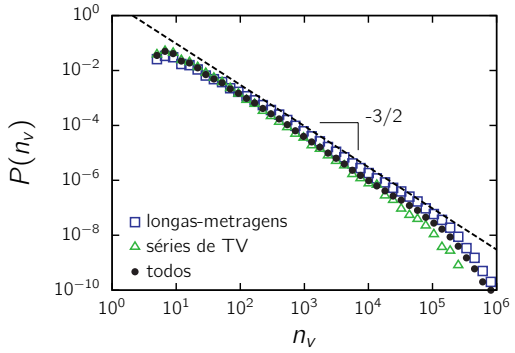


Figura 4.7: $P(n_v)$ para séries de TV e longas-metragens.

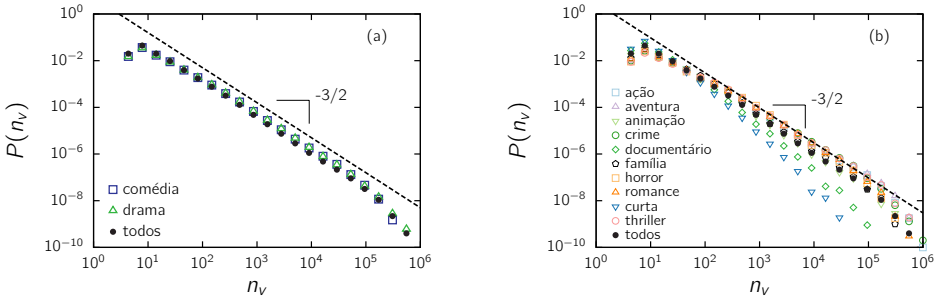


Figura 4.8: Impacto do gênero dos filme sobre a distribuição de votos $P(n_v)$ para (a) dramas e comédias e (b) outros gêneros. Alguns filmes podem pertencer a mais de um gênero.

Dividimos ainda mais os itens, agora por gênero. Consideramos somente os gêneros com mais de 15 mil filmes (Apêndice B). Os dois gêneros mais numerosos, comédia ($\alpha = 1,50 \pm 0,03$) e drama ($\alpha = 1,51 \pm 0,03$), têm o mesmo comportamento de escala, dentro das barras de erro, que o conjunto de dados total e praticamente coincide sobre todo o intervalo (Fig. 4.8a). Em adição, o mesmo comportamento de escala vale, dentro do erro ($\pm 0,04$ para os conjuntos menores), para os gêneros: animação ($\alpha \simeq 1,56$), família ($\alpha \simeq 1,55$),

PUC-Rio - Certificação Digital Nº 1122074/CA

horror ($\alpha \simeq 1,45$) e romance ($\alpha \simeq 1,46$), como mostrado na Fig. 4.8b. Um expoente ligeiramente menor aparece para filmes de ação ($\alpha \simeq 1,42$), aventura ($\alpha \simeq 1,38$), crime ($\alpha \simeq 1,40$) e thrillers ($\alpha \simeq 1,40$), tendo cada categoria menos de 20 mil filmes (Apêndice B). Curtas ($\alpha \simeq 2,1$) e documentários ($\alpha \simeq 1,9$), os quais constituem subconjuntos relativamente grandes, apresentam uma discrepância apreciável, com um expoente próximo de 2, o que indica que as audiências desses filmes têm uma resposta diferenciada.

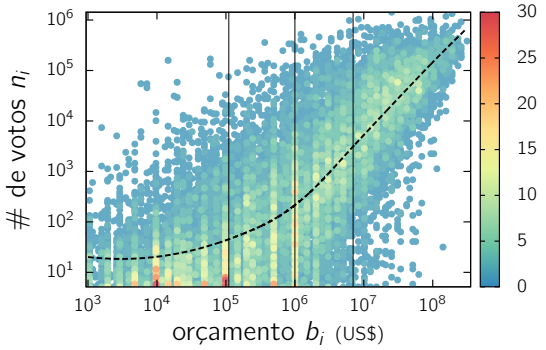


Figura 4.9: Mapa de cores do número de votos n_i vs orçamento b_i para cada filme i do IMDb. As linhas verticais indicam os quartis. A linha tracejada foi obtida por meio da regressão não paramétrica rLOESS [111].

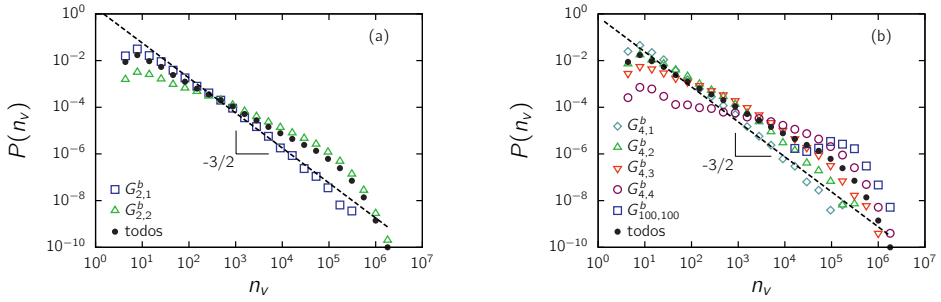


Figura 4.10: Impacto do orçamento dos longas-metragens sobre a distribuição de votos $P(n_v)$ para (a) os dois grupos $G^b_{2,1}$ e $G^b_{2,2}$ separados pela mediana (em relação ao orçamento) e (b) os quatro grupos $G^b_{4,1}, \ldots, G^b_{4,4}$ separados usando os quartis, como indicado na Fig. 4.9 e o último percentil. A linha pontilhada com inclinação -3/2 é exibida por comparação, bem como a distribuição de todos os filmes que têm informações orçamentárias (apenas longas-metragens com $b_i \geqslant 10^3$ US$ foram considerados).

Orçamento Por último, mas não menos importante, foi investigada a dependência entre o número de votos e o orçamento de produção b_i de cada

PUC-Rio - Certificação Digital Nº 1122074/CA

filme. As informações sobre orçamentos estão disponíveis para um conjunto reduzido de longas-metragens, e nós só consideramos os filmes que tiveram orçamento acima de 10^3 US$ (18 mil filmes de um total de aproximadamente 228 mil). Para esse conjunto, nós plotamos o número de votos contra orçamento (Fig. 4.9). Esse gráfico mostra que os pontos estão espalhados, mas apresentam uma correlação positiva após o primeiro quartil, indicando que acima de um ponto de inflexão, em média, o número de votos aumenta com o orçamento. Embora existam filmes com o alto custo, mas com pouco apelo e com número de votos pequeno e também filmes de baixo orçamento com uma resposta moderada.

A distribuição de votos desse conjunto não tem o decaimento em lei de potência com expoente 3/2, que é característico do conjunto total e foi mostrado nas figuras anteriores. Para analisar essa questão mais profundamente, nós dividimos o conjunto usando a mediana e observamos que a metade dos filmes com maior orçamento são os responsáveis por esse desvio, enquanto a metade com orçamento menor preserva a lei de potência 3/2. Em seguida, nós analisamos os outros quartis. Na Fig. 4.10b (e na Fig. 4.9), vemos que após o nível da mediana (onde a curva da regressão não paramétrica apresenta um crescimento mais pronunciado), a distribuição perde completamente a região de escala. E mais: a probabilidade de se ter n_v alto aumenta com o orçamento e até desenvolve um pico, como observado para o último percentil. Isto é, o custo de produção praticamente se torna o fator determinante para o número médio de votos que um filme recebe. Esse resultado mostra que mecanismos geradores diferentes governam a distribuição de votos dado para super produções, o que não é surpreendente porque orçamentos altos incluem propaganda e ações de publicidade para alcançar grandes audiências. O efeito do orçamento do filme pode explicar porque alguns gêneros que estão tipicamente associados com grandes produções, como ação e thrillers, apresentam expoentes ligeiramente menores do que a maioria dos filmes.

Cascatas de adoção

As pessoas podem escolher os filmes que vão assistir baseado no gênero, no tema, no elenco de atores, diretores, produtores, etc. O público também é certamente influenciado pela publicidade e propaganda, que são maiores em super-produções; no entanto filmes com grande orçamento serão deixados de lado na discussão a seguir.

Novos filmes são lançados constantemente, com atores e diretores que

PUC-Rio - Certificação Digital Nº 1122074/CA

nem sempre são conhecidos ou populares, e sinopses que nem sempre são suficientes para ajudar as pessoas a decidir se vão assistir ou não a um filme. Portanto, em muitos casos, se torna mais prático adotar a opinião de outras pessoas [112], através de sistemas de recomendação, ou sugestões de amigos e colegas antes de ir ao cinema ou alugar um filme. É também comum as pessoas escolherem um filme que alguém próximo tenha assistido e esteja comentando, possivelmente para evitar se sentir excluído de conversas do dia-a-dia (esse fenômeno é conhecido como *fear of missing out* [113]). Em um contexto mais geral, sabe-se que quanto mais pessoas adotam um item, se torna mais provável que outras pessoas queiram também adotá-lo [112]. Imitação é um processo comum em muitos cenários sociais e é uma estratégia de decisão muito útil. De fato, na Fig. 4.11a, nós observamos que o incremento $\Delta n_v = n_v(t_2) - n_v(t_1)$ para um dado intervalo de tempo $\Delta t = t_2 - t_1$ aumenta com n_v, indicando algum tipo de vantagem cumulativa (o mesmo tipo de mecanismo usado no modelo de redes de Barabási, que chamamos no Cap. 2 de anexação preferencial).

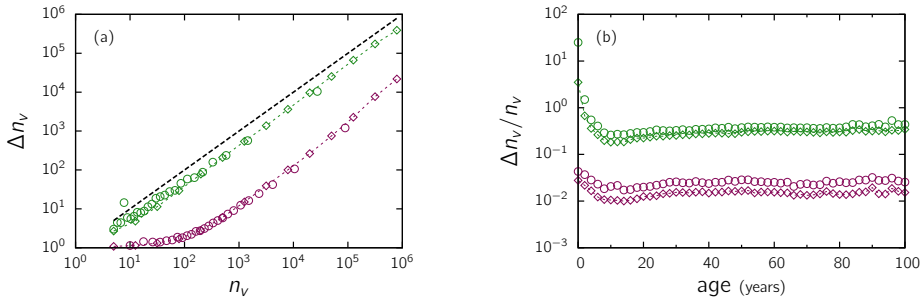


Figura 4.11: (a) Incremento do número de votos $\Delta n_v = n_v(t_2) - n_v(t_1)$ como função de n_v e (b) incremento relativo $\Delta n_v/n_v = [n_v(t_2) - n_v(t_1)]/n_v(t_1)$ como função da idade do filme para $\Delta t = t_2 - t_1 \simeq 1$ mês (vermelho, obtido com os conjuntos de dados 2 e 3) e 21 meses (verde, obtido com conjuntos 1 e 2). A mesma lista de filmes, com pelo menos 5 votos em t_1, foi considerada (veja o Apêndice B). Os símbolos representam a média aritmética (círculos) e t_1 geométrica (diamante). As linhas pontilhadas são um guia; a linha tracejada no painel (a) com um inclinação 1 foi desenhada por comparação.

Na situação que estamos considerando, os indivíduos precisam decidir entre duas alternativas, assistir ou não a um filme, portanto seu estado pode ser caracterizado por um variável binária, e. g., 1 (ativo) ou 0 (inativo), como é considerado em muitos modelos de contágio de opinião [114–117, 63]. Para definir as regras de contágio, lembre que o valor empírico de α é muito próximo do valor de campo médio padrão, 3/2, para a distribuição dos tamanhos das avalanches, como em criticalidade auto-organizada com vizinhos aleatórios [118–120], inicialização por limiar [63] e percolação de inicialização (*bootstrap*

PUC-Rio - Certificação Digital Nº 1122074/CA

percolation em inglês) em redes aleatórias (com segundo momento finito na distribuição de graus) [64, 102, 103, 11]. O mesmo comportamento de escala também ocorre para a distribuição de tamanhos de componentes em redes aleatórias no ponto crítico percolação [27, 121].

O carácter de campo médio sugere alguma aleatoriedade no processo de disseminação, independente do padrão preciso dos contatos dos indivíduos. Os resultados empíricos indicam que, tipicamente, as regras de contágio não precisam incorporar caraterísticas intrínsecas dos filmes ou de uma audiência em particular, embora, como contra-exemplo, as audiências de curtas e documentários parecem se comportar de maneira diferente das demais. Algumas regras de disseminação propostas na literatura requerem que um nó esteja em contato com outros nós ativos para que haja ativação, tal como o modelo de ativação por limiar desenvolvido por Watts [63], onde uma fração mínima de consenso entre vizinhos é necessária para que haja ativação do nó. No entanto, filmes podem ser consumidos através de diversas mídias, e a escolha de ver um filme não exclui outras; portanto, acreditamos que, no caso presente, um único contato entusiasta pode ser suficiente para induzir a decisão de assistir a um filme. Portanto, podemos considerar o caso mais simples no qual um único nó ativo é suficiente para ativar (infectar) alguns dos seus vizinhos.

O processo de avaliação de um filme como um todo ou seja, atribuir um certo número de estrelas ao filme, é um processo complexo. A opinião do indivíduo sobre o filme é certamente influenciada pelas interações sociais e preferências pessoais. Entretanto, com base nas evidências empíricas, podemos modelar a estatística do número de pessoas que se interessaram em assistir a um determinado filme independentemente de sua pontuação, já que é razoável supor que o número de votos é proporcional ao número de pessoas que assistiu ao filme. Esta simplificação do processo é motivada pela observação de que as principais características da estatística do número de votos são independentes da avaliação média, bem como de outros atributos dos filmes.

Na nossa modelagem, alguns iniciadores ativos propagam a ideia de assistir a um determinado filme, convencendo ou "infectando" alguns de seus contatos que, por sua vez, podem infectar outros, e assim por diante. Ou seja, entre os k contatos de um sítio ativo, um número aleatório j deles, escolhido a partir da distribuição de probabilidade $\{p_j$, com $0 \leqslant j \leqslant k < \infty\}$, torna-se ativa (veja a Fig. 4.12a; mostramos nessa figura uma rede com conectividade pequena por uma questão de clareza).

O processo de espalhamento é interrompido, quando, em um determinado passo de tempo, nenhum novo sítio é ativado. O número total de sítios ativados

PUC-Rio - Certificação Digital Nº 1122074/CA

no final do processo, o qual vamos nos referir como uma *avalanche* ou *cascata*, corresponde ao número de pessoas que decidiram assistir ao filme.

Simulamos este processo de contágio em redes de Erdös-Rényi (ER) [91] e Barabási-Albert (BA) [14]. Assumimos que a conectividade média $\langle k \rangle$ é grande, como em redes sociais on-line, por exemplo, Twitter ou Facebook, ou mesmo em redes sociais do "mundo real". A Fig. 4.12d, mostra a distribuição de tamanhos de avalanche resultante das simulações, onde usamos uma distribuição discreta de probabilidade exponencial $\{p_j\}$, com valor médio $p \lesssim 1$[1]. O expoente 3/2 observado nos dados reais é bem reproduzido pelo modelo, independente da rede utilizada, desde que a conectividade da rede seja suficientemente grande ($\langle k \rangle >> p$).

A avalanche de contágio pode ser *representada* por um processo de ramificação: cada indivíduo ativo corresponde a um nó i, e se esse indivíduo convence um contato, então é criado um novo ramo da árvore a partir do sítio i; o número j de novos ramos de cada nó é sorteado de acordo com a distribuição de probabilidade $\{p_j\}$ independente da história do processo ou de outros ramos.

O protocolo acima descreve um processo de ramificação de Galton-Watson [122]. Uma avalanche de contágio se desenvolve e é extinta com uma certa probabilidade. A extinção é certa se o valor médio $p < 1,0$, caso contrário, existe uma probabilidade de que o processo continue indefinidamente.

O crescimento da árvore é análogo ao desenvolvimento de uma avalanche na rede. Este mapeamento é ilustrado na Fig. 4.12a-b. Cada filme desenvolve independentemente sua própria árvore, cujo número de nós reflete o número de pessoas que decidiram assistir ao filme. Na Figura Fig. 4.12c, mostramos uma realização do processo de ramificação, no caso particular em que a distribuição de probabilidade $\{p_j\}$ é exponencial.

É um resultado conhecido que o tamanho total da avalanche no processo de ramificaçao $n = \sum_{k \geqslant 1} m_k$, tem distribuição $P(n) \propto n^{-3/2}$ com corte exponencial [122]. O corte em $P(n)$ depende de detalhes da distribuição $\{p_j\}$, que nos exemplos foi escolhida como sendo exponencial, mas o expoente 3/2 é um resultado robusto desde que $\{p_j\}$ cumpra alguns critérios mínimos [122].

O mapeamento do espalhamento feito na rede com o processo de ramificação de GW envolve a simplificação que interferências ou sobreposições no espalhamento feito na rede podem ser ignoradas, uma vez que, no processo GW, os ramos são independentes. Quando o modelo é implementado em

[1]Usamos um distribuição de probabilidade exponencial discreta $P(n) = (1 - \mathrm{e}^{-\lambda})\mathrm{e}^{-\lambda n}$, com $\lambda = 0,7$, daí $p \equiv \langle n \rangle = \mathrm{e}^{-\lambda}/(1 - \mathrm{e}^{-\lambda}) \simeq 0,986$.

PUC-Rio - Certificação Digital Nº 1122074/CA

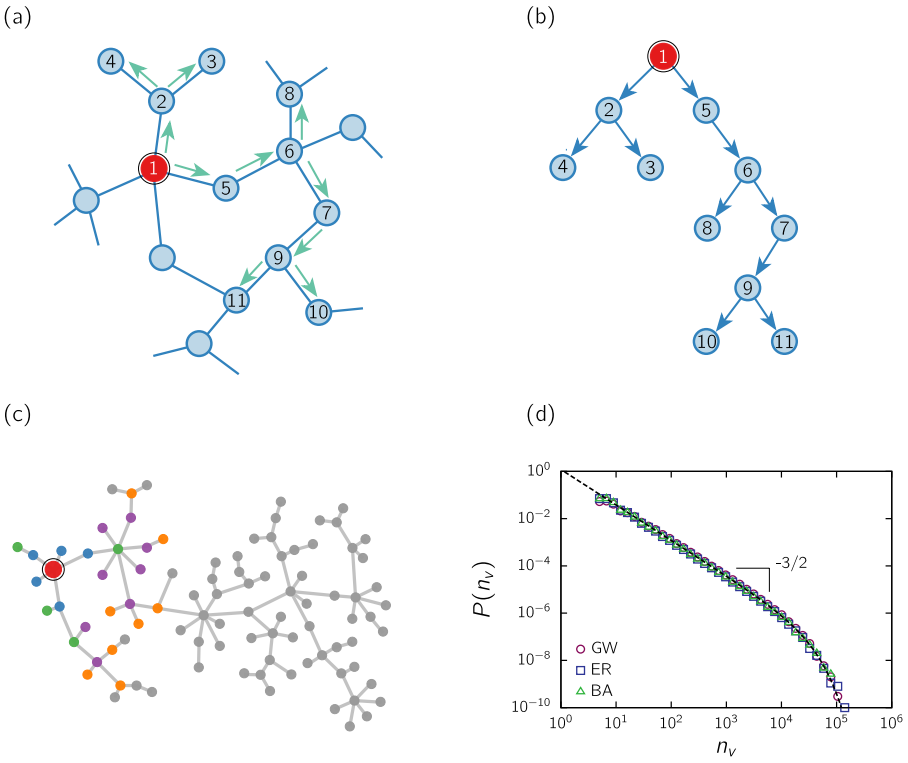


Figura 4.12: Representação pictórica do processo de contágio e o processo de ramificação equivalente. (a) Espalhamento na rede de contatos. O contágio começa em um nó iniciador (representado pelo maior nó na figura). O contágio ocorre (setas verdes) para alguns de seus sítios vizinhos (que sorteamos de uma distribuição de probabilidade) e assim por diante, desenvolvendo uma avalanche. (b) Uma árvore ramificada é construída a partir dos contatos que participam do processo de espalhamento. (c) Simulação de um processo de ramificação de Galton Watson. A maior nó representa o sítio iniciador, as primeiras gerações sucessivas da árvore são identificadas com cores; também é mostrada a árvore final, resultante de uma cascata que se extingue na 13^a geração. (d) Distribuição dos tamanhos de avalanche resultantes das simulações do processo de contágio: para um processo (sem rede) de Galton-Watson (GW), e para os processos de contágio equivalentes simulados em redes de Erdős-Rényi (ER) e Barabási-Albert (BA) com tamanho 10^6 e conectividade média $\langle k \rangle = 100$. Em todos os casos, a probabilidade p_j de influenciar j indivíduos foi arbitrariamente escolhida como exponencial com média $p \lesssim 1.0$. Foram feitas 10^6 realizações em cada caso.

uma rede, conforme o número de nós infectados aumenta, a probabilidade de contágio diminui, uma vez que, em princípio, o número de nós que podem ser ativados diminui. No entanto, a aproximação de que a probabilidade de contágio seja constante ao longo do tempo é razoável quando a conectividade da rede é grande (no exemplo usamos $\langle k \rangle = 100$). A distribuição dos tamanhos de

PUC-Rio - Certificação Digital Nº 1122074/CA

avalanche apresenta o mesmo comportamento de escala nas duas formulações, como podemos observar na Fig. 4.12d.

Além disso, os resultados parecem ser independentes do tipo de rede considerada, o que é consistente com o carácter de campo médio do processo. Apesar da simplicidade do modelo (e das aproximações envolvidas), ele nos permite capturar as características principais dos dados empíricos. Não é necessário fazer qualquer suposição específicas sobre a rede de contatos subjacente, exceto que a quantidade de contatos é grande o suficiente. Suposições adicionais não são relevantes para reproduzir o comportamento de escala com expoente 3/2. Por outro lado, as categorias "curtas" e "documentários" são contra-exemplos, exibindo um comportamento de escala diferente. Nesses casos, onde o público é, em geral, constituído por uma comunidade mais "cult", devemos incorporar mais ingredientes no modelo, como será discutido abaixo.

O tipo de processo de ramificação ilustrado na Fig. 4.12 mapeia-se diretamente em outros modelos, em particular, ao modelo de pilha de areia com vizinhos aleatórios e recozimento, um paradigma de criticalidade auto-organizada [118–120]. Em modelos aleatórios de pilha de areia, quando um sítio inicial desaba sobre seus vizinhos, novos ramos da árvore são desenhados para representar o desabamento. Esses vizinhos, por sua vez, também podem desabar continuando a cascata. Dessa maneira uma árvore é gerada. No modelo de pilha de areia, os vizinhos são escolhidos ao acaso a cada passo de tempo de modo que não temos uma rede de contatos subjacente. Mesmo quando existe explicitamente uma rede de contatos, temos situações, tais como quando a dimensionalidade crítica é excedida, em que o caráter de campo médio é mantido. Por exemplo, o expoente 3/2 de campo médio é preservado em modelos de pilha de areia considerados sobre redes exponenciais e em lei de potência com segundo momento finito. Por outro lado, o valor do expoente é maior se forem usadas redes muito heterogêneas [123–126].

Além disso quando há superposição de ramos no processo de disseminação, a estrutura da rede de contatos tornar-se importante e, portanto, o processo de ramificação associado se torna correlacionado e, concomitantemente, perde-se o carácter de campo médio. Algo similar acontece para certas audiências, onde uma distribuição de escala diferente é observada, tal como no caso de curtas e documentários. Nesses casos, provavelmente deveríamos levar em consideração os detalhes das redes de contatos e/ou não poderíamos desprezar correlações no processo de propagação. No entanto, tais informações não estão disponíveis nos dados empíricos.

Vale dizer aqui que curvas de adoção reais típicas (número de adoções

PUC-Rio - Certificação Digital Nº 1122074/CA

como função do tempo) experimentam primeiro um estágio de pré-crescimento após o lançamento, depois crescem rapidamente nos estágios iniciais, atingem um pico e finalmente decaem. A integral da curva de adoção fornece o tamanho da cascata. Usualmente, uma cascata real não se torna completamente extinta, em contraste com o modelo, mas permanece num nível pequeno após certo tempo porque o público continua assistindo filme antigos. Este efeito é ilustrado na Fig. 4.11b, onde plotamos o intervalo relativo $\Delta n_v/n_v = [n_v(t_2) - n_v(t_1)]/n_v(t_1)$ do número de votos como função da idade do filme (tempo transcorrido desde o lançamento) para dois intervalos de tempo $\Delta t = t_2 - t_1$. Em ambos os casos, enquanto o incremento é alto nos estágios iniciais, depois de alguns anos ele cai a um nível médio baixo.

Em relação às super-produções, não consideradas até o momento e que produzem uma estatística diferente, sabemos que altos orçamentos não somente implicam um apelo maior trazido por atores famosos ou alto custo de produção, mas também investimentos altos em publicidade e propaganda, os quais fazem o papel de um campo externo que ajuda a provocar cascatas muito grandes.

Portanto, não é surpreendente que nesses casos o mecanismo de propagação não dependa só das interações entre as pessoas. Por outro lado, sem grandes orçamentos, as interações podem se tornar o mecanismo dominante e surge auto-organização. Mesmo nesse último caso, o comportamento de escala em lei de potência indica que avalanches extremamente grandes são possíveis, com isso podemos ter *blockbusters* mesmo com custos de produção modestos.

4.3 Discussão

A distribuição de votos do IMDB apresenta uma cauda longa, com lei de potência com expoente próximo de 3/2. Esse é um resultado robusto que permanece válido para diversos subconjuntos definidos pelos atributos de um filme tais como avaliação média, idade e gênero, mas parece não valer para super-produções. Além disso, ele não depende do público alvo, sugerindo que o processo de disseminação ocorre independentemente da natureza da rede de contatos. As excessões são os gêneros curtas e documentários, nos quais o público parece apresentar um padrão menos heterogêneo. Surpreendentemente, independente do filme ser considerado "bom" o suficiente para ser bem avaliado, o caráter livre de escala da estatística parece não ser afetada. A universalidade desses resultados indicam que a modelagem não exige que se leve em conta as características intrínsecas dos filmes ou de uma particular audiência. O tipo de processo que pode-se associar mais naturalmente ao mecanismo subjacente de propagação de adoções de filmes é um processo

PUC-Rio - Certificação Digital Nº 1122074/CA

multiplicativo aleatório ou de ramificação, em que as ativações (ou adoções) se multiplicam de maneira aleatória, começando de uns poucos adeptos, como ilustrado na Fig. 4.12. No caso não correlacionado emergem cascatas de ativação com uma distribuição com decaimento em lei de potência com expoente 3/2. De fato, processos de ramificação são o arcabouço de muitos modelos de ativação, produzindo avalanches na mesma classe de universalidade [63, 64, 102, 103, 11]. Como consequência, os resultados empíricos se encaixam em um cenário de cascatas de imitação. Assumindo que os usuários são uma amostra representativa do público, nossos resultados representam uma evidência empírica desse tipo de cascata.

Para estimar a extensão no qual os usuários do IMDb constituem uma amostra do público com algum viés, nós consideramos outra fonte similar, Netflix [105]. Observamos o mesmo scaling 3/2 (não mostrado), mesmo que essa base de dados contenha principalmente filmes mais atrativos comercialmente, enquanto o IMDb é mais amplo nesse sentido. A possibilidade de falhas no sistema de avaliações tais como voto cego, votação massiva dos fãs para um item em particular ou imprecisões nas informações fornecidas pelo IMDb não pode ser descartada. No entanto, não se espera que casos isolados influenciem a estatística. Vale dizer também que o IMDb não é uma base de dados estática; suas informações são constantemente aperfeiçoadas com usuários acrescentando, refinando e corrigindo os dados através de uma seção chamada de "*Contributor Zone*".

Permanece como uma questão a ser investigada se outros itens de consumo ou culturais, que também podem ser avaliados online, seguem um processo de propagação semelhante. Diferentemente do IMDb, outros sites que permitem que os usuário façam avaliações, e.g., Apple App Store e Google Play, vendem os produtos que estão sendo avaliados, o que pode moldar as escolhas de maneira diferente. As pessoas também escolhem "itens" em processos políticos eleitorais. No entanto, como a distribuição de votos nesse caso segue uma log-normal ou uma lei de potência com expoentes diferentes [127–129], o processo de decisão nesse contexto parece ser de uma natureza diferente porque correlações na rede de interações, ou outros mecanismos de espalhamento, podem ser relevantes. Em contraste, no presente caso, a processo de disseminação parece ser não correlacionado. Essa aleatoriedade pode ser uma das raízes da dificuldade de se otimizar sistemas de recomendação [130].

Mais do que proporcionar uma explicação completa para a fenomenologia complexa relacionada com o espalhamento de adoções, nesse trabalho, objetivamos expandir as fontes de informação empírica relacionadas ao campo da dinâmica social. Vale lembrar que muito do progresso no campo das redes

PUC-Rio - Certificação Digital Nº 1122074/CA

complexas foi guiado por observações empíricas de redes reais. As ferramentas disponíveis para avaliar produtos podem relevar propriedades interessantes de sistemas sociais online, tais como os padrões observados nesse trabalho.

PUC-Rio - Certificação Digital Nº 1122074/CA

5
Implicações da regra da pluralidade em situações com múltiplas opções

Frequentemente enfrentamos circunstâncias onde temos que escolher uma entre diversas alternativas. Esse tipo de situação acontece nos mais diversos cenários, desde escolhas complexas, como as decisões políticas e religiosas, até situações mais simples, como nas compras corriqueiras no supermercado. Estudamos este cenário por meio das ferramentas da mecânica estatística. Seguindo esta abordagem, apresentamos e analisamos um modelo de dinâmica de opinião, usando uma variável de estado tipo Potts para representar as várias opções, incluindo um "estado indeciso" onde o indivíduo não tem uma opinião definida. Investigamos a dinâmica usando redes de Erdös-Rényi e Barabási-Albert, que como vimos, são duas classes paradigmáticas. Mostramos o impacto do tipo de rede sobre a dinâmica de opinião. Dependendo do número de opções disponíveis q e da distribuição de graus da rede de contatos, diferentes estados estacionários finais são acessíveis: podemos ter uma distribuição de opções larga ou o predomínio de uma determinada opção. A transição abrupta entre estes estados é consistente com a dominância súbita de uma determinada opção sobre outras opções similares. Além disso, validamos as distribuições de probabilidade obtidas do modelo através do confronto com dados empíricos. Finalmente, mostramos que o modelo contempla também a situação real de sobrecarga de escolhas (*overchoice* em inglês), onde um grande número de alternativas semelhantes torna o processo de escolha mais difícil e a indecisão prevalece.

5.1
Introdução

As pessoas frequentemente enfrentam situações que oferecem uma variedade ampla de opções, por exemplo, quando procuram por um restaurante, um hotel, um modelo de celular ou um artigo de consumo básico no supermercado. O número de produtos disponíveis aumenta a cada dia. É estimado que aproximadamente 50.000 novos produtos são introduzidos a cada ano nos EUA [131]. Mesmo dentro de cada categoria, podem existir muitas marcas e variações entre um item e outro, sem que esses apresentem um atrativo diferencial entre si. Isto leva ao problema onde nos deparamos com um excesso de opções, denominado sobrecarga de escolhas [132]. Tomar decisões

PUC-Rio - Certificação Digital Nº 1122074/CA

em tais situações pode ser custoso, e esse processo estressante pode levar a decisões ruins ou à não tomada de decisão. Com isso, as vantagens de se ter múltiplas escolhas podem ser canceladas pelas desvantagens de um processo de escolha mais complicado. Portanto, apesar de representar, aparentemente, uma evolução positiva, muitas opções realmente podem dificultar o processo de escolha. Por exemplo, o estudo apresentado na Ref. [133] mostra que pessoas com muitas opções de compra tendem a ter mais dificuldade em escolher e podem acabar não comprando nada.

Motivados por essas observações, nos questionamos em que medida a regra de decisão por si só, deixando de lado os fatores psicológicos, contribui para este cenário, introduzindo, por exemplo, conflito e frustração. Então, por meio de um modelo de dinâmica de opinião, investigamos a distribuição de adoções feitas por uma população quando confrontada com um número grande de escolhas.

Diversos modelos baseados em regras de imitação diferente tem sido propostos e estudados nos últimos anos [11, 62, 115–117, 134–146]. No entanto, poucos deles contemplam a possibilidade de várias opções (estados) [116, 142]. Na grande maioria dos modelos, a opinião de um agente é representada por uma variável binária, uma vez que muitas situações podem ser modeladas através de duas possíveis atitudes (opostas), por exemplo, sendo favoráveis ou desfavoráveis a uma determinada escolha/posição. Vimos um exemplo dessa situação no Cap. 4, onde os agentes decidem entre assistir ou não a um filme. Este tipo de variável binária muitas vezes é inspirada no modelo de Ising com spin 1/2. Há também modelos em que uma variável contínua é considerada [11, 137], de modo a representar o espectro contínuo de posições ou atitudes que um indivíduo pode tomar em relação a um determinado assunto. Uma variável contínua pode representar também o grau de convicção da escolha, apresentaremos um exemplo do uso de variável contínua no Cap. 6. No entanto, para o nosso presente propósito de estudar as situações de múltipla escolha, é natural considerar uma variável de estado tipo Potts que pode assumir vários valores (discretos).

Consideramos que as mudanças de um estado para outro são regidas por uma regra da pluralidade, baseada na idéia de que um indivíduo faz a escolha que é a mais popular entre os seus contatos. De fato, quando temos que escolher ou comprar alguma coisa, especialmente quando há tantas opções similares que não há uma favorita *a priori* e não é viável testar todas as opções, é razoável levar em conta as preferências das outras pessoas [112]. Naturalmente, quanto mais próximo é a pessoa em nossa rede de contatos, maior a importância que damos a sua opinião, uma vez que, normalmente, vizinhos mais próximos

PUC-Rio - Certificação Digital Nº 1122074/CA

na rede têm interesses e gostos semelhantes [147]. Usa-se essa estratégia não só em questões mais triviais do cotidiano, mas também em situações mais complexas, tais como nas eleições políticas onde muitos candidatos competem. Para mudar ou adotar uma nova opinião, no entanto, um mínimo de consenso entre os contatos é necessário. Isto é expresso em uma regra da pluralidade, segundo a qual um indivíduo é persuadido a adotar a opinião compartilhada pelo maior número de seus vizinhos próximos. Vale dizer que o conceito de pluralidade é diferente do conceito de maioria. Maioria é o subconjunto de um grupo cujo tamanho é superior a 50% do grupo total. Em contrapartida, a pluralidade é o maior subconjunto de um grupo, o que não necessariamente constitui a maioria, dado que o maior subconjunto pode ser menor que 50% do grupo como um todo.

Regras de evolução baseadas na opinião localmente dominante já foram considerados antes como, por exemplo, a regra da maioria para dois estados introduzida por Galam [143, 144]. Esta regra mais tarde foi estendida, incorporando múltiplas opiniões [116] e com uma interação do tipo todos-para-todos [148]: o indivíduo e seus vizinhos adotam simultaneamente a opinião majoritária do grupo. Também foi estudada uma variante baseada na regra da pluralidade, ao invés da maioria local [116]. Porém, em ambos os casos, não foram considerados indivíduos indecisos, e o foco desses trabalhos é no tempo para se atingir o consenso. Diferentemente, vamos considerar a presença de pessoas indecisas e que apenas um indivíduo, e não o grupo, muda de estado a cada passo de tempo da dinâmica. Em outro modelo que emprega muitas alternativas, introduzido na Ref. [142], a imitação direta de um vizinho escolhido ao acaso se alterna com religações na rede de contatos, produzindo uma transição de fase interessante ao se ajustar o balanço entre as regras. Ao invés de considerar o contágio aos pares, investigamos a regra da pluralidade aplicada sobre os contatos mais próximos do indivíduo. Embora a generalização do modelo usando uma rede dinâmica possa ser abordada em um trabalho futuro, assumimos aqui uma rede estática de contatos. Consideramos as duas classes de redes com a propriedade de mundo pequeno que viemos estudando: as paradigmáticas redes de Erdös-Rényi (ER) [91, 16] e Barabási-Albert (BA) [14]. Os detalhes do modelo serão definidos na Sec. 5.2.

O modelo engloba casos em que as diferentes alternativas têm atratividade inicial equivalente. Muitos produtos, vendidos na internet, com qualidades e preços semelhantes estão dentro do escopo do modelo, como, por exemplo, álbuns de música, sapatos, etc. Fizemos também a simplificação de que os indivíduos diferem um do outro apenas pelo número de contatos. Outras heterogeneidades dos agentes podem também ser introduzidas em

PUC-Rio - Certificação Digital Nº 1122074/CA

trabalhos futuros. Por ora, nos colocamos uma pergunta básica: como a regra da pluralidade molda o espectro de decisão neste caso mais simples?

Primeiro, abordamos a questão clássica neste tipo de problema, se um estado de consenso, em que todos (ou quase todos) os indivíduos compartilham a mesma preferência, pode ser alcançado ou não. Portanto, nas simulações do modelo, calculamos a fração de realizações que atingiram o consenso. Em situações não consensuais, analisamos a distribuição final das adoções. Também calculamos outras quantidades relevantes, tais como a fração final de pessoas indecisas e a fração final da população que opta pela alternativa mais popular. Os resultados serão apresentados na Sec. 5.3.

5.2
A regra da pluralidade aplicada a um modelo de agentes

A regra da pluralidade governa a dinâmica de opinião de N agentes que interagem por meio de sua rede de contatos. Cada agente i, representado por um nó na rede, tem uma variável do tipo Potts s_i, que pode assumir os valores $s_1, \ldots, s_q$ refletindo q alternativas (opções/escolhas, que enumeramos em uma ordem arbitrária) elegíveis, bem como um estado "indeciso" s_0, atribuído quando o indivíduo ainda não adotou uma das opções disponíveis. A adição do estado indeciso reflete o fato de que às vezes as pessoas não têm uma escolha favorita. Isto traz uma distinção importante em relação aos modelos já existentes, onde tal estado é ignorado.

Usamos redes de ER e BA, como representativas de redes mundo pequeno com distribuições de conectividade homogêneas e heterogêneas, respectivamente. Nós assumimos que a maior parte da população não tem uma opinião formada *a priori*, com exceção dos nós "iniciadores" que representam cada uma das escolhas oferecidas. Começamos a dinâmica com todos os nós no estado $S = s_0$, exceto para q nós escolhidos ao acaso (onde $q < N$), onde, para cada um deles, atribuímos uma opinião distinta $S = s_1, \ldots, s_q$. Estabelecemos o mesmo número de iniciadores para cada uma das alternativas (a saber, 1 iniciador), o que reflete a atratividade equivalente de todas as alternativas. Este tipo de estado inicial também é empregado em modelos de opinião para as eleições proporcionais [149].

Em cada passo de Monte Carlo (MC) da dinâmica visitamos todos os nós da rede em ordem aleatória. O estado da rede é atualizado de acordo com os seguintes passos, aplicados para cada sítio i:

(i) Definimos o conjunto de nós, $\mathcal{A}_i$, formado por i e seus primeiros vizinhos.

PUC-Rio - Certificação Digital Nº 1122074/CA

(ii) Determinamos o estado da pluralidade (EP) $S_i^p \neq s_0$, associado ao nó i, como o estado compartilhado pelo maior número de nós no conjunto $\mathcal{A}_i$.

(iii) O agente i adota então o estado da pluralidade correspondente.

As atualizações são repetidas até que o sistema atinja o estado estacionário. Observe que, quando medimos o estado S_i^p, ignoramos os nós em $\mathcal{A}_i$ que têm $S = s_0$ (ver Fig. 5.1), por outro lado, a opinião corrente do sítio i também é levada em consideração no cálculo de S_i^p.

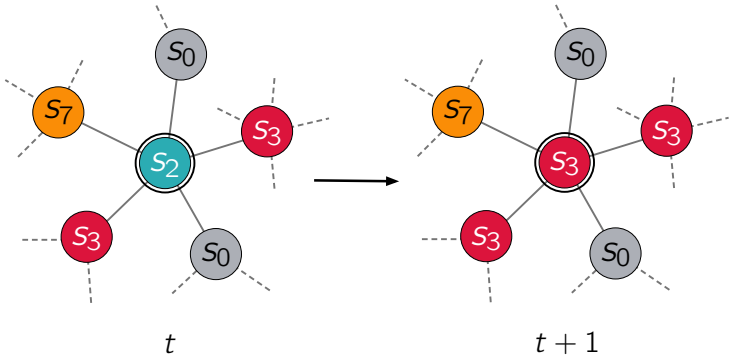


Figura 5.1: Ilustração da regra modelo. Em um dado instante de tempo t, um nó i (na figura, representado pelo nó central) e os seus vizinhos formam um grupo $\mathcal{A}_i$. No caso mostrado na figura, o estado da pluralidade é $S_i^p = s_3$, uma vez que este é o estado que é compartilhado por mais sítios. Assim, no próximo passo de tempo, o nó i muda sua opinião de s_2 para s_3. Observe que os nós com $S = s_0$ não são contabilizados no cálculo do estado da pluralidade S_i^p.

Vale ressaltar que esta dinâmica difere da dinâmica do tipo Sznajd [150, 151] em que dois ou mais indivíduos que partilham a mesma opinião, impõem essa opinião a todos os seus vizinhos. A dinâmica apresentada aqui também difere em vários aspectos da regra da pluralidade introduzida na Ref. [116]: *i)* enquanto no nosso caso, apenas o nó central é afetado, em [116] todo o grupo $\mathcal{A}_i$ muda sua opinião para o estado da pluralidade S_i^p a cada atualização *ii)* Aqui, o tamanho do grupo interagente $\mathcal{A}_i$ é dado por $G = k_i + 1$, onde k_i é a conectividade do sítio i, em contraste com [116] onde o tamanho do grupo é constante (em todo o caso, o parâmetro $\langle k \rangle + 1$ desempenha o papel de um G efetivo, e o tamanho dos grupos nos dois casos coincide no limite de uma rede altamente homogênea, ou regular); *iii)* enquanto o nosso modelo introduz o estado indeciso, a possibilidade de indecisão não é contemplada em [116]; *iv)* em um caso de empate, a opinião do nó i permanece inalterada, mas em [116] uma das opções dominantes é selecionada aleatoriamente. Com relação a este último item, a presente dinâmica é mais próxima da versão

PUC-Rio - Certificação Digital Nº 1122074/CA

da regra da maioria de [116], onde a dinâmica se torna estática por não ter um critério de desempate, ou seja, quando há empate, o estado do sítio não muda. *v)* Em termos da rede de contatos subjacente, aqui consideramos redes mundo pequeno. Por outro lado, a dinâmica da Ref. [116] foi investigada usando redes quadradas. *vi)* Finalmente, uma outra diferença importante é a condição inicial: consideramos que os indivíduos decididos estão inicialmente diluídos em um "mar" de nós indecisos, em contraste com a condição inicial onde todos os sítios estão, de maneira equiprovável, em um dos estados decididos.

5.3 Resultados

Acompanhamos a evolução de cada realização da dinâmica até o estado estacionário ser atingido. A Fig. 5.2 ilustra a evolução temporal de n_s, número de nós que compartilham a mesma opinião s, em algumas realizações representativas em redes ER de tamanho $N = 10^4$, com diferentes conectividades médias $\langle k \rangle$ e diferente número de alternativas q. Em alguns passos de Monte-Carlo chega-se ao estado estacionário. Padrões distintos são observados dependendo do valor dos parâmetros $\langle k \rangle$ e q:

(a) No primeiro exemplo, quando $\langle k \rangle = 4$ e $q = 20$, temos um estado final fragmentado, onde todos os valores iniciais de S têm adeptos. Nesta realização particular, o número de agentes não decididos é predominante, representando aproximadamente 20% da população.

(b) Para uma conectividade maior (e.g., $\langle k \rangle = 6$), várias opiniões ainda são possíveis no estado estacionário, mas a opinião vencedora s_w é amplamente dominante. O número de agentes não decididos no estado estacionário diminuí em relação ao caso *(a)*.

(c) Para uma conectividade ainda maior (por exemplo, $\langle k \rangle = 12$) é provável que haja consenso.

(d) No caso extremo de uma rede altamente conectada e um grande número de alternativas, pode-se observar a situação em que quase todos os nós permanecem no estado indeciso. Devido aos empates no cálculo de S_i^p a evolução fica estagnada.

Na presente versão do modelo, o número de agentes indecisos sempre diminui com o tempo já que esses indivíduos não surgem na dinâmica, eles são introduzidos apenas no estágio inicial. Padrões semelhantes também são observados na rede de BA (não mostrado), embora para valores diferentes dos parâmetros.

PUC-Rio - Certificação Digital Nº 1122074/CA

PUC-Rio - Certificação Digital Nº 1122074/CA

A dinâmica pode ser qualitativamente entendida da seguinte maneira: em uma primeira etapa, cada opinião propaga-se através dos nós vizinhos indecisos. Se os iniciadores estão muito diluídos ($q \ll N$), e a conectividade não é muito alta, então cada um dos grupos compostos por sítios com a mesma opinião pode se propagar de maneira quase independente durante vários passos de MC (regime não competitivo). Neste caso, o crescimento inicial é aproximadamente exponencial na rede de ER, sendo descrito por $dn_s/dt = \langle k \rangle n_s$. Em redes de BA, o crescimento é ainda mais rápido. Quando os grupos "colidem", um regime competitivo é iniciado. Dependendo dos parâmetros do modelo, a competição ocorre de maneira mais ou menos equilibrada e com isso os empates estancam a evolução evitando o domínio amplo de uma determinada opinião (como na Fig. 5.2.a). Este congelamento da dinâmica pode ocorrer também desde o início, como na Fig. 5.2.d. Caso contrário, predomina um mecanismo do tipo vantagem cumulativa ou "rich-get-richer". Nessa situação, o grupo com a opinião vencedora torna-se visivelmente maior, convencendo cada vez mais os indivíduos de outros grupos (Fig. 5.2.b), e

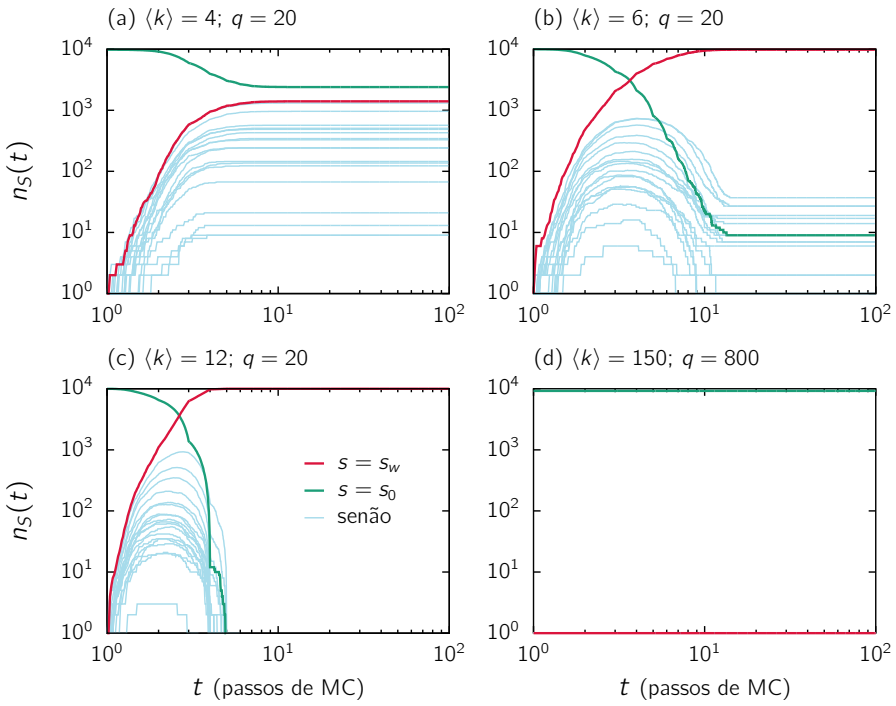


Figura 5.2: Evolução temporal de n_s, número de nós que compartilham a mesma opinião s, para redes de ER de tamanho $N = 10^4$. Todas as opiniões são mostradas. A opinião $S = s_0$ (estado indeciso) e a opinião vencedora $S = s_w$ estão destacadas em verde e vermelho, respectivamente. O estado estacionário é atingido em alguns poucos passos de MC e diferentes padrões surgem para diferentes valores de $\langle k \rangle$ e q.

podendo até dominar toda a rede (Fig. 5.2.c).

A fim de caracterizar os estados estacionários possíveis, nós monitoramos, para cada realização, a fração de pessoas decididas $f_d \equiv n_d/N \equiv (N - n_{s_0})/N$, e a fração de nós que partilha a opinião mais adotada, ou escolha vencedora, $f_w \equiv n_{s_w}/N$. A média dessas quantidades, calculadas sobre diferentes realizações da rede e condições iniciais, serão denotadas por $\overline{f_d}$ e $\overline{f_w}$, respectivamente. Também calculamos a fração f_c das simulações que atingem o consenso (operacionalmente definido como pelos menos 99% da população). Nos cálculos envolvendo redes de ER, somente a componente principal do grafo é considerada. Em situações onde temos um estado final fragmentado, ou seja, quando vários opiniões estão representadas no final da dinâmica, analisamos como essas opiniões se distribuem dentro da população. Por fim, validamos os resultados através da comparação com distribuições de adoção reais.

Efeito do número de alternativas q Primeiro, vamos analisar o impacto do número de alternativas q no estado estacionário, para uma distribuição de graus fixa. A Fig. 5.3 mostra as quantidades de interesse, $\overline{f_d}$, $\overline{f_w}$ e f_c, como função de q. Notamos que, quando $q \ll N$, os resultados são praticamente independentes de N em redes de ER (painéis da esquerda), enquanto as frações aumentam com o tamanho do sistema para redes de BA (painéis da direita).

A fração média de pessoas decididas $\overline{f_d}$ (triângulos), indicam que a abundância de alternativas dificulta o processo de decisão feito através da regra da pluralidade. O decaimento de $\overline{f_d}$ é mais pronunciado em redes de ER, enquanto a maioria das pessoas permanece decidida mesmo para q grande na rede de BA. A razão disso é que em redes heterogêneas a maior parte dos nós tem conectividade pequena e estes podem ser facilmente convencidos por um único vizinho no estado decidido. Em contrapartida, a ocorrência de uma pluralidade local é mais difícil em redes homogêneas (com a mesma conectividade média que uma rede heterogênea). Como consequência, empates são muito mais frequentes, a dinâmica "congela" e mais nós permanecem no estado indeciso. Portanto, se as decisões são tomadas de acordo com a regra da pluralidade, a heterogeneidade dos graus ajudaria a diminuir a indecisão, quando o número de alternativas permanece constante. O papel da conectividade será analisado com mais detalhe na próxima seção.

Embora tipicamente o valor de f_d diminua com q, contra intuitivamente, o comportamento oposto também pode acontecer como observado na Fig. 5.3.c. A fração $\overline{f_d}$ exibe um decaimento até um nível caraterístico não nulo que permanece aproximadamente constante (desde que q não se torne muito próximo de N). Este nível característico se mantém para q infinitamente grande

PUC-Rio - Certificação Digital Nº 1122074/CA

no limite de uma rede também infinita $N \to \infty$. Obviamente, quando $q \to N$, a fração $\overline{f_d}$ aumenta, se aproximando do seu valor máximo $\overline{f_d} = 1$, já que nesse caso a maior parte dos nós está associada a uma determinada escolha (diferente para cada sítio) desde o início da dinâmica. A existência de um platô significa que, dentro de um certo intervalo, a fração indecisa da população não é afetada de modo significativo pelo número de alternativas. No entanto, o valor desse nível característico muda com a conectividade da rede. Por exemplo, na rede de ER, o platô em $\overline{f_d}$ corresponde à maioria da população (aproximadamente 59%) para $\langle k \rangle = 4$ e a uma minoria (aproximadamente 42%) para $\langle k \rangle = 8$.

A fração média da população que adota a opinião vencedora, $\overline{f_w}$ (quadrados) também é uma quantidade significativa. Necessariamente $\overline{f_w} \leqslant \overline{f_d}$. A fração $\overline{f_w}$ é também maior na rede de BA. Isto é, a opinião vencedora conquista em média uma fração maior da população em redes de BA do que em redes de ER com o mesmo $\langle k \rangle$. De fato, a vantagem cumulativa que conduz

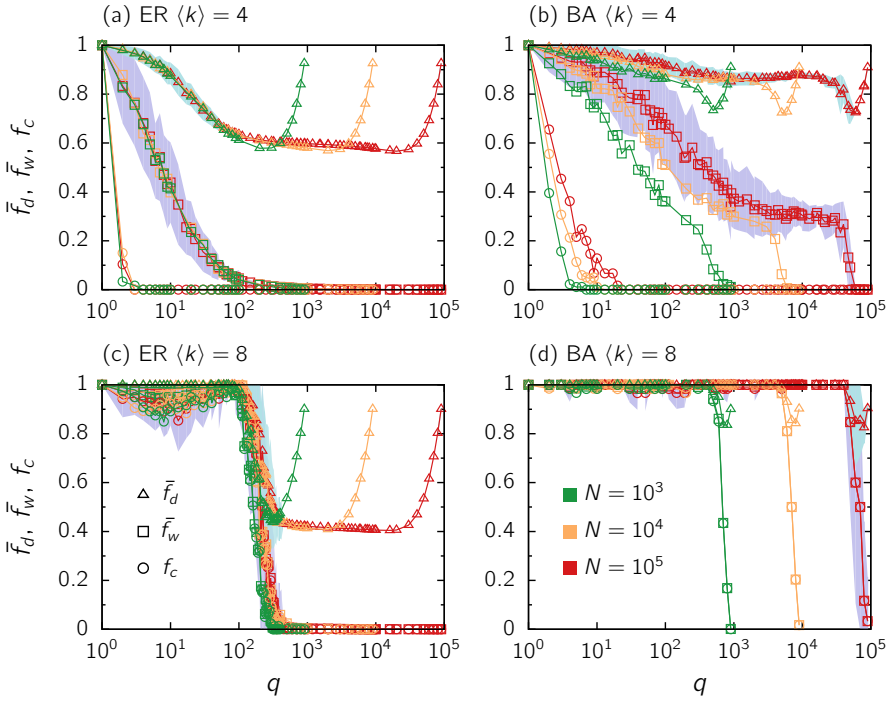


Figura 5.3: Fração média de pessoas decididas $\overline{f_d}$ (triângulos), fração média de sítios que escolheram a opinião vencedora $\overline{f_w}$ (quadrados) e fração de realizações que chegaram a consenso f_c (círculos), em função do número de opiniões diferentes q. Três tamanhos diferentes (indicados na figura) foram usados para redes de ER (esquerda) e BA (à direita), com conectividade média $\langle k \rangle = 4$ (painéis superiores) e 8 (painéis inferiores). As linhas sólidas são guias para os olhos, e as regiões sombreadas representam os erros padrão de f_d e f_w para o caso $N = 10^5$ (os erros são semelhantes para os outros tamanhos).

PUC-Rio - Certificação Digital Nº 1122074/CA

o crescimento de um grupo (sítios que compartilham a mesma opinião) é facilitada em redes heterogêneas devido à presença de *hubs*. Consequentemente, o grupo vencedor ganha mais adeptos. Note também que, em redes de ER, quando a fração f_d atinge o nível característico, uma opinião dominante está ausente, fato espelhado pelo pequeno valor $\overline{f_w}$ (quadrados na Fig. 5.3.a e 5.3.c). Entretanto, nas redes de BA o vencedor pode sempre conquistar uma fração expressiva da população (símbolos quadrados na Fig. 5.3.b e 5.3.d) expresso pelo fato que $\overline{f_w}$ permanece sempre finito, exceto no limite extremo $q \to N$.

Para ambos os tipos de redes, a probabilidade de ocorrência de consenso, f_c (círculos), tipicamente decai de 1 para 0 a medida em que q aumenta. Isto concorda com a intuição de que, quando existem mais opções disponíveis, é mais difícil se atingir o consenso. O decaimento para zero não é suave, mas apresenta um salto abrupto, tal que acima de um certo valor crítico q_c a fração de consenso f_c se torna desprezível. Observamos na Fig. 5.3 que quando a conectividade média da rede aumenta (veja os painéis de cima para baixo), assim como quando a rede se torna mais heterogênea (veja os painéis da esquerda para direita), o consenso se torna mais provável para um dado número de alternativas q, e q_c é maior. Isto é, alta conectividade e/ou heterogeneidade nos graus não somente evitam indecisão, mas também favorecem o consenso segundo a regra da pluralidade. A dependência precisa em $\langle k \rangle$ será estudada na sequência.

Efeito da conectividade média $\langle k \rangle$ As frações características $\overline{f_d}$, $\overline{f_w}$ e f_c como função da conectividade média da rede são mostradas na Fig. 5.4 para diversos valores de q, nas redes de ER e BA.

Analisaremos primeiro os resultados na rede de BA, onde as frações características exibem (veja a Fig. 5.4.b) um comportamento simples para $q \ll N$. A fração média de pessoas decididas $\overline{f_d}$, bem como a probabilidade de consenso f_c, aumentam com a conectividade. É interessante que o valor crítico da conectividade para a qual o consenso é atingido, $\langle k \rangle_c \simeq 7$, pareça ser aproximadamente independente do número de alternativas.

Em contrapartida, na rede de ER (see Fig. 5.4.a), $\langle k \rangle_c$ depende de q. Observe que a fração média de pessoas decididas $\overline{f_d}$ primeiro decresce com a conectividade até um valor mínimo localizado em $\langle k \rangle_{\min}$. Até esse ponto, a fração de simulações que chega ao consenso f_c é desprezível. No entanto, em $\langle k \rangle_{\min}$, acontece uma transição abrupta e ambas as frações f_c e $\overline{f_d}$ crescem rapidamente com $\langle k \rangle$, até 1. Poderíamos esperar mais sítios decididos quando a conectividade é grande, como na rede de BA, desde que, em princípio, mais conexões poderiam facilitar o espalhamento da informação. Porém, para uma

PUC-Rio - Certificação Digital Nº 1122074/CA

rede pouco conectada, os grupos que compartilham de uma dada opinião estão tipicamente isolados uns dos outros. Quando ligações são adicionadas na rede, grupos que estavam isolados se conectam e podem ocorrer empates. Isto é, por um lado, uma conectividade média alta implica que grupo com opiniões diferentes podem estar mais conectados entre si e em consequência acabam competindo. Por outro lado, um indivíduo terá contato com mais opiniões diferentes, dificultando o processo de decisão e fazendo com que haja mais sítios no estado indeciso. Portanto, não somente o excesso de alternativas (q grande) pode produzir a estagnação da dinâmica, esse congelamento também pode ser causado pelo "excesso" de ligações, ou seja, um valor grande de $\langle k \rangle$. Isto explica o decrescimento inicial de $\overline{f_d}$ com $\langle k \rangle$ que ocorre até $\overline{f_d}$ atingir o mínimo. Depois desse ponto, a inclusão de mais conexões permite que um grupo dominante imponha sua própria opinião sobre os demais, concomitantemente há um crescimento de f_c, até que seja atingido seu valor máximo igual a 1. Note também que em redes de ER, para cada valor de q, existe um intervalo da conectividade média para a qual a fração de pessoas decididas é minoria.

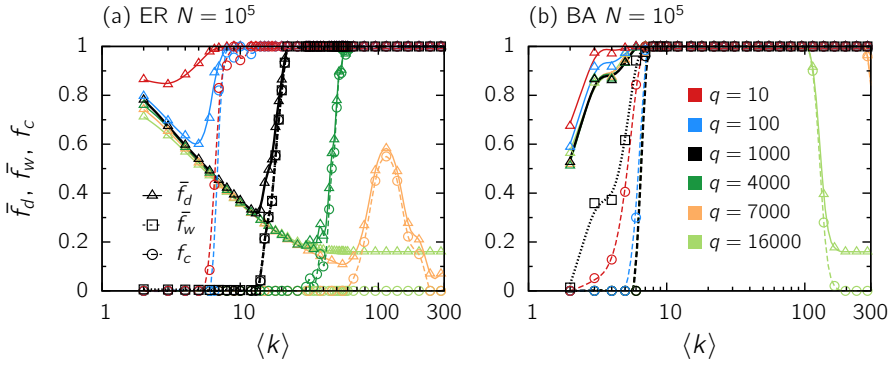


Figura 5.4: Frações $\overline{f_d}$ (linhas cheias) e f_c (linhas pontilhadas), como função da conectividade média $\langle k \rangle$, para vários valores de q (indicados por cores diferentes), em redes de ER (a) e BA (b) de tamanho $N = 10^5$. Por uma questão de clareza, o comportamento da fração vencedora $\overline{f_w}$ (linhas tracejadas) é mostrado apenas para o caso $q = 1000$ (realçado com linhas pretas). As médias foram calculadas usando pelo menos 50 amostras.

Para valores de q muito grandes, é provável que as frações características não atinjam seu valor máximo igual a 1, e quando a rede tende à saturação de opiniões diferentes ($q \to N$), a existência de pluralidade local de opniões que permita a evolução se torna improvável, daí $\overline{f_d} \to q/N$ e $f_c \to 0$. Algo similar ocorre quando nos aproximamos do limite de alta conectividade média. De fato, no limite de campo médio (grafo completamente conectado), onde todos os nós estão conectados com todos os outros nós, claramente o congelamento acontece para a condição inicial considerada aqui. Por outro lado, o limite de

PUC-Rio - Certificação Digital Nº 1122074/CA

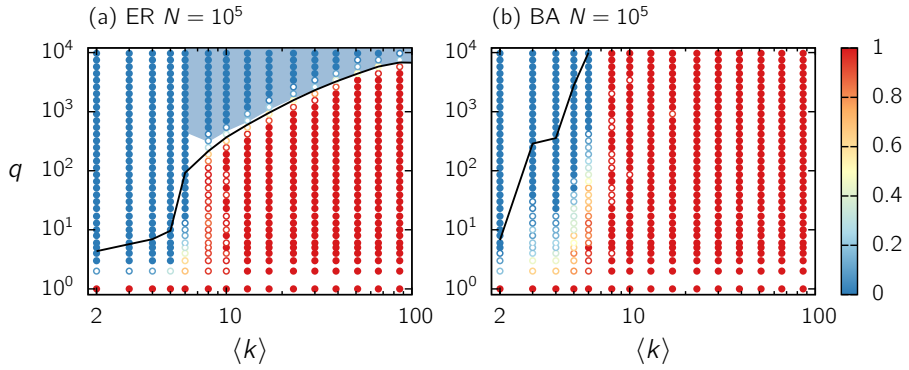


Figura 5.5: Diagrama de fases no plano $\langle k \rangle - q$ exibindo, em escala de cores, a fração f_c de realizações da dinâmica que atingiu o consenso. Os símbolos cheios enfatizam os pontos onde o consenso ocorre em todas as ($f_c = 1$, vermelho) ou nenhuma ($f_c = 0$, azul) das 30 realizações feitas para cada par $(\langle k \rangle, q)$. As linhas cheias representam os pontos em que $f_w = 0,5$ ($f_w > 0,5$ à direita dessas curvas). Áreas sombreadas representam os pontos para os quais $f_d < 0.5$. Os dois tipos de rede, ER e BA, têm tamanho $N = 10^5$.

campo médio não faz muito sentido no caso heterogêneo, porque perderíamos a heterogeneidade dos graus.

A existência de uma transição abrupta da situação onde muitas opiniões coexistem para o consenso indica que, pela adição de apenas algumas ligações ou algumas novas alternativas, a maior parte da população pode vir a adotar a mesma opinião. A transição para o consenso é mais abrupta em redes de BA e a largura do salto decresce com q. O maior grupo ganha adeptos adicionais mais facilmente com a vantagem cumulativa. Perto da conectividade crítica, a adição de algumas ligações ao acaso faz com que seja mais provável que grupos muito grandes se conectem com grupos menores que acabam sendo conquistados pela opinião dominante, rapidamente levando ao consenso. Em redes homogêneas esta transição é menos abrupta, porque um grupo dominante é menos provável, e a largura da transição aumenta com q.

Diagrama de fases A fim de resumir os resultados anteriores, apresentamos na Fig. 5.5 um diagrama de fases no plano de parâmetros $\langle k \rangle - q$. O diagrama mostra a transição entre regiões de consenso para as regiões com estado final fragmentado, como indicado pelas cores de vermelho para azul. Os símbolos cheios enfatizam os pontos onde o consenso acontece sempre ($f_c = 1$, vermelho) ou nunca ($f_c = 0$, azul).

As linhas sólidas representam a fronteira, onde $f_w = 0,5$ ($f_w > 0,5$ à direita das curvas), e a área sombreada destacam os pontos para os quais o maioria dos sítios permanece no estado indeciso ($f_d < 0,5$).

PUC-Rio - Certificação Digital Nº 1122074/CA

Fica claro que o domínio onde ocorre consenso ($f_c \simeq 1$, região vermelha) é maior para redes de BA, assinalando que a heterogeneidade da rede favorece o consenso. Em todos os casos, inserindo algumas poucas ligações, para um q fixo, pode conduzir ao consenso. Em redes de BA, para q suficientemente grande ($\gtrsim 200$) o valor crítico $\langle k \rangle_c \simeq 7$ torna-se independente de q.

Em contraste, para redes de ER, a dependência em q é mais forte. Os domínios de não consenso ($f_c \simeq 0$, região azul), para redes de Erdös e Rényi, cobrem a região de conectividade média alta, para q suficientemente grande. Nesse caso, também podemos atingir o consenso eliminando algumas alternativas. Concomitantemente, na área hachurada na Fig. 5.5.a., a fração de pessoas decididas se torna minoria ($f_d < 0,5$). Enquanto, no caso de BA, sítios decididos são maioria em todo o diagrama de fases. Além disso, mesmo na ausência de consenso, o grupo vencedor pode ser a maioria ($f_w > 0,5$), principalmente em redes de BA.

PUC-Rio - Certificação Digital Nº 1122074/CA

5.3.1
Distribuição de adoções simuladas e empíricas

Uma distribuição larga de opiniões embora com um forma diferente daquela observada no Cap. 4 pode manifestar-se nos estados estacionários em que não há consenso. Computamos os histogramas (normalizados) para estas realizações. As curvas típicas dessas distribuições são retratadas na Fig. 5.6. Podemos identificar comportamentos exponenciais, log-normais e leis de potência.

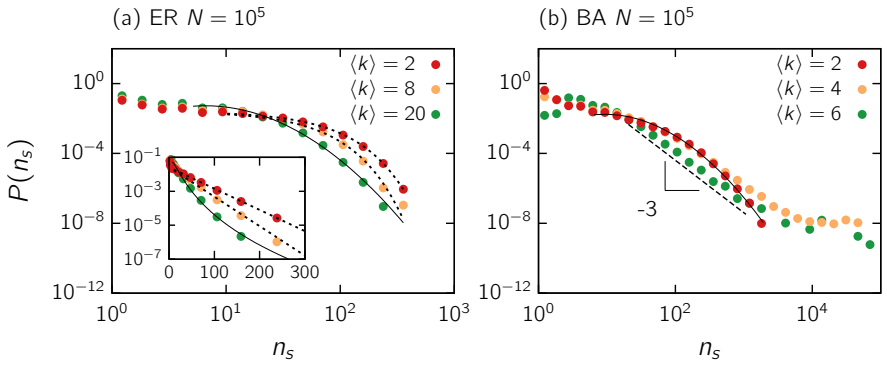


Figura 5.6: Distribuição do número de nós que compartilham uma determinada opinião n_s, $P(n_s)$, construído com 50 amostras, para $q = 2000$, em redes de ER (a) e BA (b) de tamanho $N = 10^5$, e diferentes valores de $\langle k \rangle$ (indicados na figura). As linhas sólidas são ajustes de log-normal, as linhas pontilhadas são ajustes exponenciais, e a linha tracejada com inclinação -3 foi desenhada como referência.

Em redes de ER, longe o suficiente da fronteira crítica do consenso,

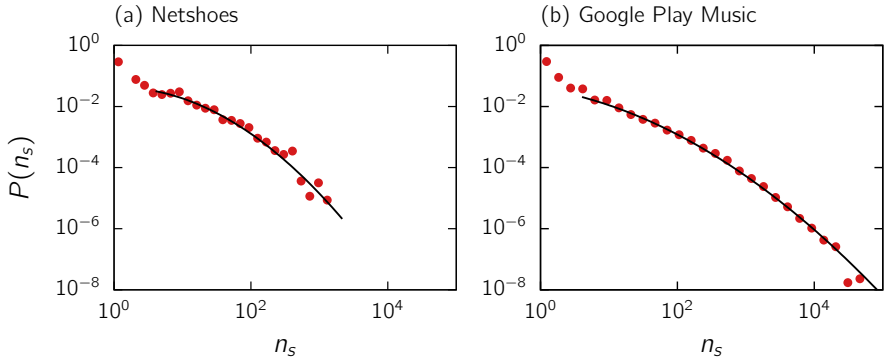


Figura 5.7: Distribuição $P(n_r)$ do número de avaliações n_r para os itens à venda no Netshoes [152] ($n_d \simeq 9 \times 10^4$; $q \simeq 1,2 \times 10^3$) e no GooglePlay [153] ($n_d \simeq 1,8 \times 10^6$; $q \simeq 7,5 \times 10^3$). Identificamos o número de avaliações com o número de pessoas decididas n_d e o número de itens com q. As linhas sólidas são ajustes de log-normal.

PUC-Rio - Certificação Digital Nº 1122074/CA

as preferências são distribuídas de maneira quase uniforme com um corte exponencial. Quando nos aproximamos da região de consenso (por exemplo, aumentando $\langle k \rangle$) a distribuição assume uma forma log-normal. Note que isso acontece na região do diagrama de fases onde a indecisão prevalece (área sombreada na Fig. 5.5).

Em redes de BA, a distribuição também se assemelha a uma log-normal, mas quando nos aproximamos da região de consenso a cauda aumenta devido à existência de uma opinião vencedora dominante. Além disso, nas situação onde o congelamento acontece nos estágios iniciais da dinâmica, $P(n_s)$ tende a refletir a distribuição de conectividade da rede com expoente -3.

A fim de comparar as distribuições simuladas com distribuições reais, analisamos dados nos quais as alternativas não são significantemente diferenciadas (por exemplo em preço e/ou qualidade), assim como é assumido pelo nosso modelo. Nós analisamos tênis masculinos dentro da mesma faixa de preços (R\$ 150 ± 50) do site Netshoes, uma empresa brasileira de artigos esportivos, e álbuns de música vendidos pelo Google Play (cujos preços são U\$ 9 ± 3). Identificamos q como o número de itens dentro de cada categoria (tênis ou álbuns) e consideramos a quantidade de resenhas recebidas por cada item como um indicador do número total de "adeptos". Para cada base de dados nós computamos o histograma do número de adoções (Fig. 5.7). A comparação entre as Figs. 5.6 e 5.7 coloca em evidência uma similaridade excepcional entre os distribuições reais e simuladas produzindo curvas log-normais. Também é notável a semelhança entre o formato das curvas para pequenos valores de s. A ausência de uma cauda como na rede de BA, indica que os dados empíricos são melhor modelados por redes de ER. De fato, uma vez que os usuários desses

serviços têm acesso às avaliações de qualquer outro usuário, a rede subjacente se assemelha a uma interação do tipo campo médio e portanto é melhor descrita por um grafo aleatório com uma conectividade relativamente alta.

5.4 Considerações Finais

Introduzimos um modelo baseado na regra da pluralidade que considera o processo de decisão como sendo governado pela influência da maioria relativa dos vizinhos. O modelo também incorpora a possibilidade de agentes indecisos. Apesar da sua simplicidade, o modelo oferece *insights* sobre sistemas reais. Pode-se aplicar essa modelagem não somente em situações que envolvem consumo de produtos ou serviços, também podemos aplicá-la em outros ambientes onde existe uma variedade de opções, desde que as alternativas sejam homogêneas com atratividade similar (em qualidade, custo, etc) e pessoas não têm uma escolha preferida *a priori*.

Diferentes estados estacionários são acessíveis dependendo do número de alternativas e da distribuição de graus da rede de contatos: podemos ter consenso, uma distribuição de adoções larga ou, para um número de alternativas suficientemente grande, situações onde a indecisão predomina. De fato, nos processos de tomada de decisão governados pela regra da pluralidade pode ocorrer empate (conflito e frustração), contribuindo para a estagnação da dinâmica pelo excesso de escolhas. Este efeito é mitigado pela heterogeneidade da rede. Adicionalmente, o modelo prevê que a estagnação também pode ser resultado do excesso de ligações.

Para qualquer rede o consenso é quase certo para um número suficientemente pequeno de opções ou para uma conectividade suficientemente grande. Para redes com distribuição de graus heterogênea, o consenso pode ocorrer mesmo se existem muitas opções disponíveis. Por outro lado, é improvável que ocorra consenso para um número grande de opções e quando a conectividade média da rede é baixa, especialmente se a rede é homogênea. Nessas situações de não consenso, uma distribuição de adoções larga pode surgir com um formato similar ao das distribuições reais onde a suposição de homogeneidade dos itens se mantém, como nos exemplos da Netshoes e álbuns no Google Music.

O modelo indica que o consenso pode surgir repentinamente através da introdução de algumas conexões ou através da eliminação de alguns itens. Além disso, o modelo também sugere que um item pode se tornar muito popular (com f_w relativamente grande) mesmo que a atratividade inicial de todos os itens sejam equivalentes. Estas observações fornecem outra explicação possível do

PUC-Rio - Certificação Digital Nº 1122074/CA

porquê alguns conteúdos amadores podem viralizar na internet, ou porque um serviço, um bem de consumo ou produto cultural pode se tornar um *best-seller* sem que tenha qualquer qualidade diferenciada aparente.

PUC-Rio - Certificação Digital Nº 1122074/CA

6
Modelo não consensual com opiniões contínuas

Investigamos o surgimento de pontos de vista extremos na sociedade através de uma modelagem via física estatística e de análise de pesquisas que abordam uma ampla gama de questões como religião, economia, política, aborto, sexo extraconjugal, livros, filmes, e voto eleitoral. As pesquisas apresentam um indicador claro do aumento de pontos de vista extremos. Seu precursor é uma relação não linear entre a fração de indivíduos que sustentam uma posição extrema e a fração de indivíduos que sustentam a mesma posição (seja extrema ou moderada). Por exemplo, na política, aqueles que estão "muito conservadores" contra aqueles que são "moderadamente conservadores ou muito conservadores". Propomos um modelo de dinâmica de ativação com regras de interação baseadas na existência da "teimosia" do indivíduo e que mimetiza as observações empíricas. De acordo com a nossa modelagem, o início da não linearidade pode ser associado com uma transição abrupta do tipo percolação de inicialização onde "cascatas de extremismo" acontecem na sociedade. Portanto, o início do comportamento não linear representa um alerta antecipado para prever a transição de visões moderadas para visões extremas. Além disso, por meio de um diagrama de fases, podemos classificar as sociedades de acordo com o regime percolativo ao qual elas pertencem em termos da fração crítica de extremistas e das conexões entre os indivíduos.

PUC-Rio - Certificação Digital Nº 1122074/CA

6.1
Introdução

As causas fundamentais do surgimento de opiniões extremas na sociedade constituem, nos dias de hoje, uma questão de intenso debate entre os acadêmicos [154, 3, 155, 63, 156, 157]. Ao longo das últimas décadas, parece haver uma tendência mundial de divisão das opiniões públicas sobre diversas questões como, por exemplo, controle de armas, aborto, imigração, aplicações biotecnológicas, aquecimento global, direitos LGBT, entre muitos outros.

Em muitos tópicos, um silenciamento notável de vozes mais moderadas pode ser observado, concomitantemente ao levante de pontos de vista extremos [158–160]. Não somente na política, mas também em tópicos cotidianos como filmes, livros, moda e outros temas culturais, posições extremas desabrocham e a opinião ou a atitude de grupos inicialmente pequenos podem acabar se tornando a regra.

Como essas tendências se estabelecem na sociedade ainda é um mistério. O grau de desenvolvimento econômico e social, as crenças religiosas, a história de vida dos indivíduos, e muitos outros fatores, sem dúvida, contribuem para moldar a distribuição das opiniões dos membros de uma sociedade. Porém, mesmo que características sociais sejam essenciais para moldar o comportamento coletivo, como viemos discutindo nos capítulos anteriores, as interações entre os indivíduos também desempenham um papel importante que muitas vezes é subestimado. Na rede social definida pelas conexões entre os membros da sociedade, se propagam informações, rumores e idéias. Neste processo, novas opiniões podem tomar forma e as opiniões existentes podem ser reforçadas ou enfraquecidas. Mas até que ponto essa interação ajuda a moldar a opinião pública? Pontos de vista extremos podem surgir apenas a partir das interações entre indivíduos? As respostas a estas perguntas podem nos ajudar a entender a dinâmica de polarização de opiniões públicas e podem também possibilitar a detecção de tendências à polarização.

Estudos longitudinais, que envolvem observações da evolução temporal da opinião de indivíduos em escalas de tempo longas, são difíceis de se conseguir. Felizmente, estão disponíveis para o público extensas bases de dados de estudos transversais (*cross-sectional studies*) da distribuição de opinião dos indivíduos sobre as mais diversas questões [161, 162]. Estes dados podem ser obtidos por meio de pesquisas, onde um grupo de pessoas é questionada sobre sua atitude em relação a um assunto específico, fornecendo assim uma evidência valiosa sobre a natureza complexa das opiniões públicas. As respostas são normalmente classificados em atitudes, e.g., muito favorável, moderadamente favorável, moderadamente desfavorável ou muito desfavorável. Para os nossos propósitos, pessoas com opiniões muito favoráveis ou muito desfavoráveis podem ser definidas como aquelas que detêm pontos de visto extremos. O cerne da questão é entender a dinâmica da opinião pública a partir dos dados transversais disponíveis, de tal forma a prever uma tendência à polarização antes que ela ocorra de fato [163]. A partir da análise desses dados estáticos, extraímos evidências da radicalização na população que se manifesta sob a forma de um comportamento não linear na região próxima a certos pontos críticos. Nessa região ondas de extremismo se espalham na sociedade através de uma dinâmica de avalanches que associamos a transições de fase críticas na mesma classe de universalidade da percolação de inicialização. Essas resultados elucidam a natureza da transição de grupos que passam a adotar posições mais extremas.

O achado central desse trabalho é a descoberta de um preditor estatístico do aumento das tendências de opinião extremas na sociedade em termos

PUC-Rio - Certificação Digital Nº 1122074/CA

de um comportamento não linear do número de indivíduos que sustentam uma determinada visão extrema, e o número de indivíduos que detêm uma opinião extrema ou moderada. Analisamos pesquisas que abrangem um vasto leque de questões, tais como religião, economia, política, aborto, sexo extraconjugal e voto eleitoral; além de analisar a opinião relacionada a bens de consumo cultural, como livros e filmes. Os levantamentos apresentam uma não linearidade que serve como um preditor do surgimento de pontos de vista extremos. Este preditor é ubíquo em relação à diversidade de pesquisas e países participantes, refletindo uma característica genérica da dinâmica de opinião humana. A metodologia desenvolvida a partir da não linearidade é assinalada por um ponto de inflexão no qual as sociedades se tornam extremas e, apesar disso, tanto quanto sabemos não foi usada antes para predizer tendências na opinião pública. O significado dessa não linearidade é explicada na sequência. Em geral, para um sistema físico não interagente, é esperado que haja isometria. Isto significa que o sistema é extensivo e os observáveis escalam com o tamanho do sistema [164]. Por exemplo, em um sistema físico não interagente, *e. g.*, um paramagneto, se dobrarmos o número de partículas, o valor da energia também dobra. Em termos dos nossos sistemas sociais de interesse, um sistema linear não interativo implica que o número de pessoas que sustentam uma posição extrema (por exemplo, muito favorável ao aborto) escala linearmente com o número de pessoas que compartilham da mesma opinião, sendo esta extrema ou moderada (seguindo o mesmo exemplo, muito favorável ou moderadamente favorável ao aborto). Portanto, linearidade é consequência da não interatividade entre os agentes. Por outro lado, é bem é conhecido em física estatística [164], que correlações entre as unidades do sistema levam a um comportamento não linear e à não extensividade. Este efeito é chamado de alometria no campo da sociofísica e vem sendo estudado nos últimos anos. Se investigou, por exemplo, de que maneira o tamanho das cidades escala com diferentes indicadores urbanos, como atividades tecnológicas [165] e indicadores de saúde [166, 167]. Nesses estudos, verifica-se que o número de homicídios escala superlinearmente com a população das cidades, enquanto o número de suicídios escala sublinearmente; ambos os casos são exemplos de comportamento alométrico não linear.

O aparecimento do comportamento não linear representa um sinal de alerta, que antecipa uma transição crítica abrupta de um cenário onde prevalecem opiniões moderadas para um regime dominado por opiniões extremas. O comportamento não linear, o qual antecipa a mudança abrupta, é facilmente detectado por meio dos levantamentos e quantifica o *status* da sociedade no caminho para o predomínio de atitudes extremas. Por meio

PUC-Rio - Certificação Digital Nº 1122074/CA

de uma modelagem de física estatística, encontramos que a não linearidade serve como prognóstico para o início de uma cascata de pontos de vista extremos. Mostramos que essas cascatas são uma consequência de uma transição de ativações que ocorre no ponto de inflexão quando a sociedade muda abruptamente de moderada para extrema.

6.2 Resultados

6.2.1 Resultados empíricos

Para ilustrar o resultados dos levantamentos, vamos considerar uma típica pesquisa realizada pelo Pew Research Center (ver Anexo D). Pergunta-se aos participantes de um determinado país se eles i) fortemente acreditam, ii) acreditam, iii) não acreditam, ou iv) fortemente não acreditam que a religião é uma parte importante de suas vidas. Usando esses dados, primeiro calculamos a fração f_e de pessoas que sustentam uma visão extrema. Calculamos essa fração em relação ao total da população pesquisada em cada país. Ou seja, nós computamos $f_e = N_e^+/N$ (ou N_e^-/N), onde N_e^+ (ou N_e^-) é o número de pessoas que expressam um ponto de vista extremo positivo (ou negativo), e N é o número total de pessoas que participam do levantamento. Em alguns levantamentos excluímos do total da população, a parte que não tem posição definida ou se recusou a responder. Calculamos então a fração de pessoas que sustentam um ponto de vista moderado ou extremo: $f = N^+/N$ (ou N^-/N), onde N^+ (ou N^-) é o número de indivíduos que acreditam ou acreditam fortemente em religião (ou que desacreditam e desacreditam fortemente). Veja exemplos na Fig. 6.1.

A figura 6.2a mostra o gráfico de f_e contra f, onde cada ponto representa o resultado da uma pesquisa realizada em um determinado país em um determinado ano. O conjunto de pontos embora espalhados, são nitidamente correlacionados e seguem uma tendência bem definida. Para extrair a tendência principal entre f_e e f sem usar uma forma funcional pre-determinada, usamos a regressão não local, conhecida como LOESS [111] com banda (parâmetro de suavização) $h = 0,8$, utilizamos também o estimador de Nadaraya-Watson [168] (veja apêndice C para detalhes). A regressão é representada pela linha sólida na Fig 6.2a. O resultado é paradigmático da dependência não trivial entre f_e e f. Para uma fração relativamente pequena de extremistas, f_e é aproximadamente proporcional a f (linha reta pontilhada na Fig. 6.2a). Este comportamento linear pode ser interpretado como originado

PUC-Rio - Certificação Digital Nº 1122074/CA

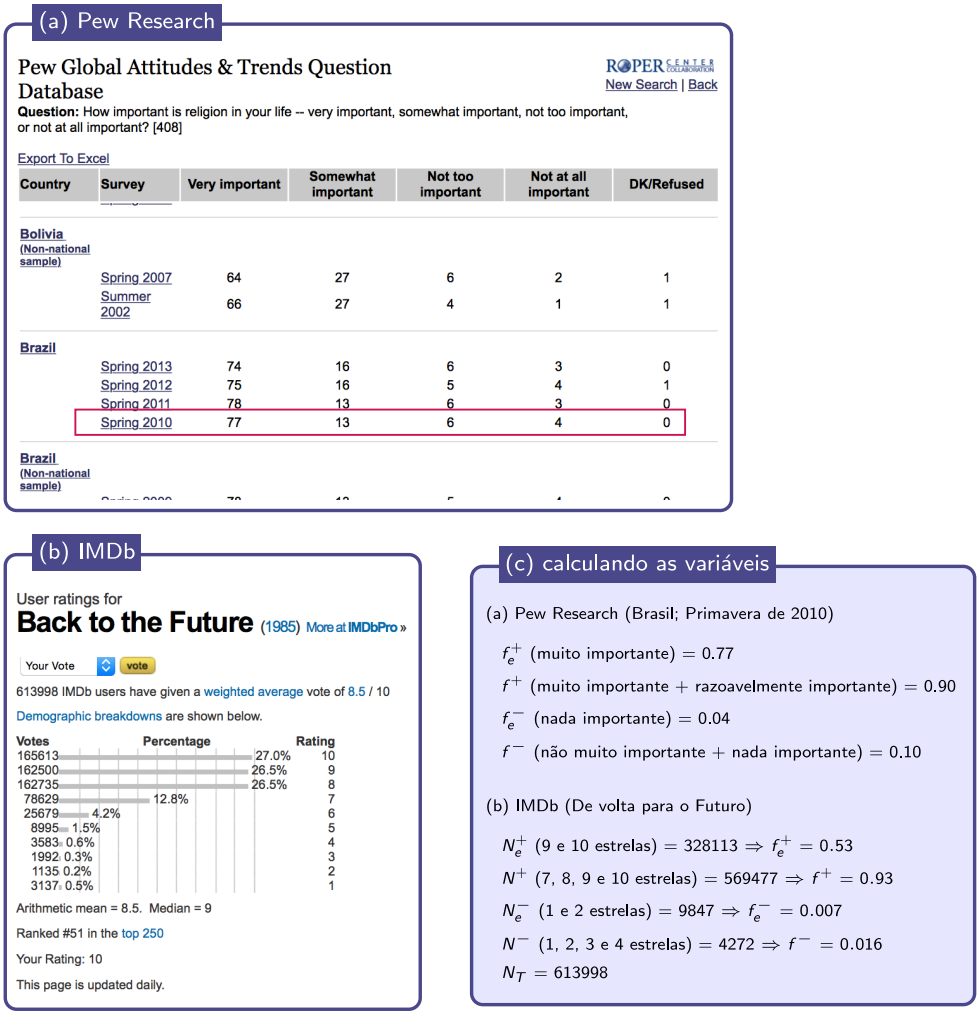


Figura 6.1: Exemplo de cálculos das frações: f_e^+, f^+, f_e^- e f^+. Usamos dois casos representativos (a) uma das pesquisas do Pew Research mencionada no texto, a saber, sobre o importância da religião na vida das pessoas. Usamos o levantamento feito no Brasil na primavera de 2010 (destacado em vermelho na figura) para fazer os cálculos e (b) um item do IMDb, usamos o mesmo item escolhido como exemplo no Cap. 4: o filme *De volta para o futuro*. O quadro (c) mostra como as frações foram computadas.

de um sistema de indivíduos não interagentes, i.e., que formam as suas opiniões de forma independente um do outro. Na ausência de interações entre as pessoas, o regime linear se estenderia até $f = 1$. No entanto, em $f_e \approx 0,20$, um desvio notável da linearidade é observado. Um comportamento não linear segue a partir desse ponto, o que marca o início de um "excesso" de extremistas em comparação com o número esperado no caso linear (não interactivo e extensivo). Portanto, quando uma maioria significativa de pessoas que compartilham do mesmo ponto de vista (extremo ou moderado) é atingida,

PUC-Rio - Certificação Digital Nº 1122074/CA

a interação entre os indivíduos com opinião semelhante leva a uma mudança do predomínio de opiniões moderadas para opiniões extremas. Este fenômeno está em bom acordo com o cenário observado para o caso de grupos pequenos em pesquisas de ciências do comportamento [169].

Um estudo de caso típico da transição em direção ao extremismo são as opiniões em relação a situação econômica depois da crise da dívida pública europeia em 2009. A evolução temporal de (f, f_e) para a França, Itália, Grécia e Espanha na figura Fig. 6.2b mostra que o comportamento não linear emerge após 2009, indicando que um negativismo extremo prevaleceu nestes países durante o período. Este resultado suporta a hipótese de que o desvio da linearidade marca o surgimento de pontos de vista extremos.

A não linearidade observada não é uma prerrogativa de questões religiosas ou econômicas onde as opiniões frequentemente parecem ser mais polarizadas, mas estende-se a vários outros tipos de pesquisas em todo o mundo. Pesquisas que variam de aborto à imigração (ver detalhes no Apêndice D) são apresentados na Fig. 6.2, todas exibindo características semelhantes. É um comportamento surpreendentemente onipresente, encontrado também em tópicos mais simples, como opiniões sobre livros e filmes (Fig. 6.2 n-o). Embora a forma precisa da curva (f, f_e) mude de um levantamento para outro, parece haver uma tendência universal muito diferente daquela encontrada, por exemplo, em dados embaralhados (Fig. 6.2p).

Na Fig. 6.2q, mostramos os resultados para votos em deputados estaduais brasileiros. Cada ponto corresponde a uma cidade para a qual nós computamos as frações de votos dentro de cada orientação política, como foi feito para as outras pesquisas. Este é um contra-exemplo notável: encontramos um padrão de dispersão em (f, f_e) semelhante ao que apareceria se as pessoas tivessem escolhido a orientação política (de extrema esquerda à extrema direita) associada ao voto de forma aleatória. Com efeito, os dados de eleição parecem ser não correlacionados, de uma forma similar à obtida com os dados embaralhados sobre a pesquisa com os livros apresentados na Fig. 6.2p. Mais pesquisas são necessárias para revelar se a ausência de uma tendência na votação eleitoral brasileira é uma característica genérica de eleições em geral.

Na sequência, interpretamos o comportamento não linear em termos de uma transição crítica de um regime de visões moderadas para visões extremas. Notavelmente, o desvio do comportamento linear, que aparece para f_e moderado, antecipa o aparecimento de um ponto crítico que marca a transição precisa entre esses dois tipos de regime. Sendo assim, as curvas (f, f_e), que podem ser facilmente obtidas a partir de pesquisas de opinião, anunciam prontamente o início do "comportamento" extremo antes que a transição de

PUC-Rio - Certificação Digital Nº 1122074/CA

PUC-Rio - Certificação Digital Nº 1122074/CA

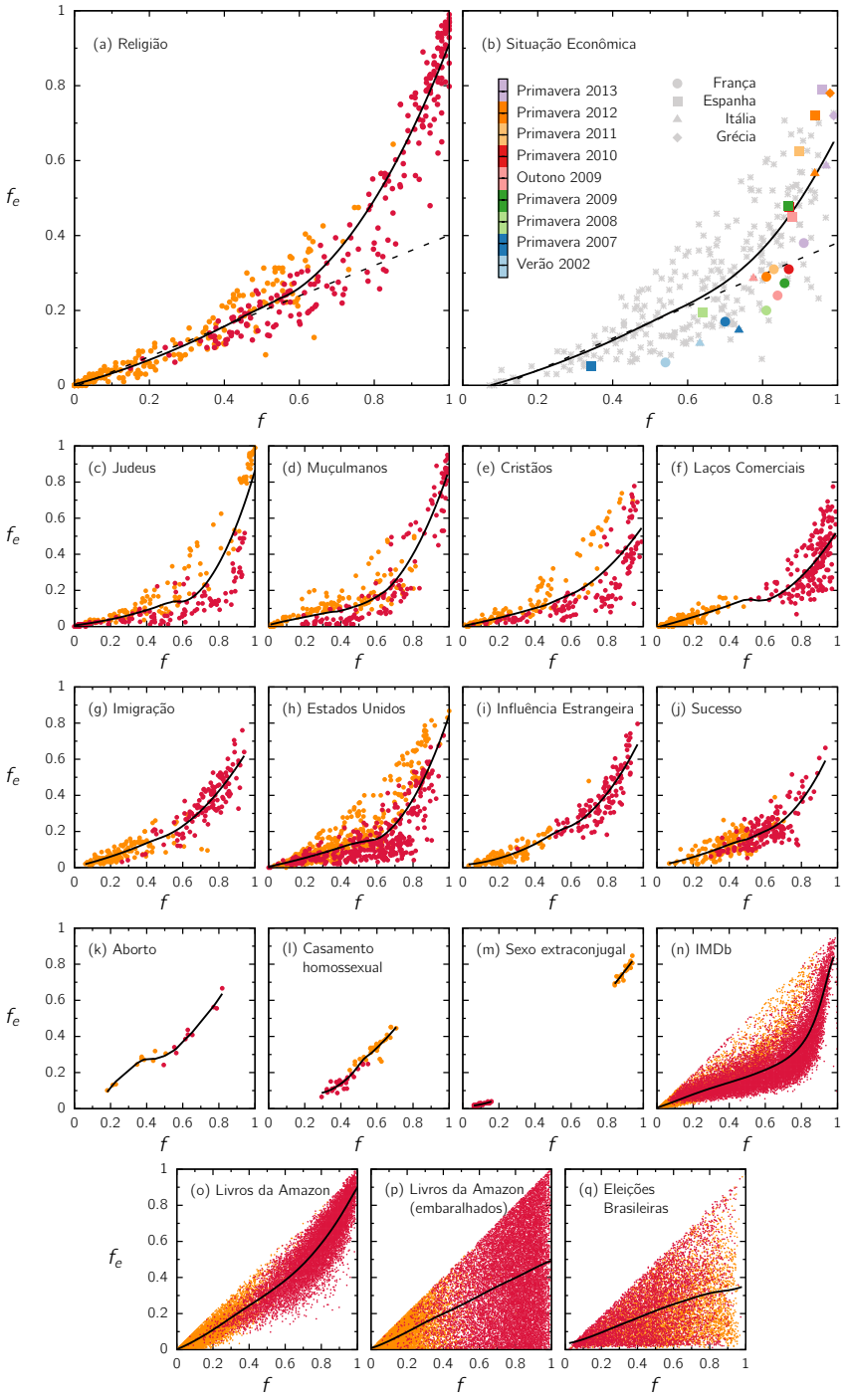


Figura 6.2: (Subtítulo na página seguinte)

Figura 6.2: **Observações Empíricas**. Dependência da fração de extremistas f_e como função de f, fração da população que compartilha da mesma posição em relação a um determinado tema. Os pontos vermelhos correspondem às opiniões positivas/favoráveis (f_e^+, f^+) e os pontos laranja correspondem às opiniões negativas/desfavoráveis (f_e^-, f^-). Os resultados foram obtidos a partir dos resultados dos levantamentos descritos no Apêndice D. As linhas sólidas são regressões não paramétrica (ver Apêndice C). Em (**a**) e (**b**), a linha tracejada representa o comportamento linear esperado para um grupo não interativo. No painel (**b**), estão representadas as respostas desfavoráveis em relação ao sentimento sobre a situação econômica. A evolução temporal (em escala de cores) de (f, f_e) é apresentada para França, Itália, Grécia e Espanha. Depois da crise da dívida pública europeia em 2009 surge um comportamento não linear nesses países.

fato tenha ocorrido.

6.2.2 Modelando a dinâmica de opinião extrema

As características observadas aqui, tanto quanto sabemos, não podem ser explicadas pelos modelos de opinião existentes. Como discutido no capítulo 5, boa parte dos modelos leva ao consenso, onde uma única opinião predomina, ou à uma situação onde há um equilíbrio, *i.e.*, temos frações iguais de opiniões distintas no estado estacionário [150, 143, 144]. Outros modelos permitem a coexistência de opiniões minoritárias e majoritárias [117, 170, 146, 171], mas não são adequados para descrever os dados empíricos, em que temos de distinguir as opiniões extremas das moderadas. Estes modelos são suficientes para abordar determinados problemas, mas não são adequados para estudar a emergência do extremismo. Há também os modelos de confiança delimitada (em inglês *bounded confidence models*), que assumem que apenas pessoas com atitudes suficientemente próximas interagem. Estes modelos foram considerados para estudar a dinâmica de opinião extrema, mas esses produzem uma distribuições descontínuas de opiniões, que não refletem os padrões empíricos que nós observamos.

Essas observações demandam um modelo simples e abrangente que capture a origem microscópica subjacente da formação de opiniões extremas. Propomos um modelo, aplicado à uma rede complexa, onde a opinião do indivíduo, q, assume valores reais entre -1 e $+1$. O sinal de q corresponde ao posicionamento do indivíduo (*e.g.*, contra ou a favor), enquanto o valor absoluto de q reflete o nível de "convicção" do indivíduo (*e.g.*, moderadamente a favor ou muito a favor) . São consideramos opiniões extremas estados onde $|q| > q_e$ e opiniões positivas (negativas) estados com $q > 0$ ($q < 0$). Motivados

PUC-Rio - Certificação Digital Nº 1122074/CA

pelas quatro opções da maioria das pesquisas, pode-se considerar $q_e = 0,50$. Chamamos de sítios "moderados" aqueles em que $|q| < q_e$.

Introduzimos um parâmetro a ($0 \leqslant a \leqslant 1$) que quantifica a teimosia dos indivíduos, esse é um ingrediente realista que mostraremos ser importante para entender o comportamento não linear da dinâmica de opinião. A dinâmica considera a opinião do indivíduo, bem como a opinião média $\bar{q}$ dos vizinhos na rede de contatos, segundo as seguintes regras (veja também a Fig. 6.3):

(i) $q \to \bar{q}$, se $|\bar{q}| > |q|$ e q tem o mesmo sinal de $\bar{q}$.

(ii) $q \to q$, se $(1-a)|q| \leqslant |\bar{q}| \leqslant |q|$ e q tem o mesmo sinal de $\bar{q}$.

(iii) $q \to \bar{q} + aq$, se $[\bar{q} < (1-a)q$ e $q > 0]$ ou $[\bar{q} > (1-a)q$ e $q < 0]$.

A regra *(i)* determina que um nó adotará a opinião média dos seus vizinhos, se essa média é mais "extrema" que a opinião do próprio nó. De fato, é razoável que as pessoas com um nível de convicção mais fraca têm mais chance de serem influenciadas por pessoas com uma opinião mais forte. Observe que, mesmo que o parâmetro a não esteja presente de forma explícita nesta regra, um indivíduo que tem uma opinião mais forte do que os seus contatos acaba sendo mais inflexível, uma vez que é mais difícil de mudar sua opinião. De acordo com a regra *(ii)*, nenhuma mudança no estado do indivíduo acontece para um certo intervalo de opiniões intermediárias de seus vizinhos, sendo este intervalo tanto maior quanto maior for a teimosia e mais forte a opinião do nó. Por fim, a regra *(iii)* determina que, quando a opinião média dos vizinhos é oposta ou muito menos radical do que a opinião do próprio nó, então a nova opinião do sítio será $\bar{q} + aq$.

Note que o novo estado de um sítio é influenciado não só pelos seus contatos, mas também é determinado parcialmente por sua própria opinião, ponderada pela teimosia a. Assim, o papel de teimosia a é duplo: se a é grande, não somente $\bar{q}$ deve estar distante o suficiente de q para que haja uma mudança na opinião do sítio, mas a também reduz o efeito relativo à sua vizinhança. No caso limite $a = 0$, o intervalo de inflexibilidade colapsa, e o nó se comporta como os indivíduos mais flexíveis, facilmente induzidos pelo ambiente próximo e assumindo a opinião média dos vizinhos, similarmente ao que acontece nos modelos com regra da maioria [143].

A teimosia é um ingrediente fundamental para termos uma população heterogênea, com opiniões diferentes. Na ausência da teimosia ($a = 0$), todas as três regras se reduzem à única prescrição de adotar o valor médio dos vizinhos, dando origem ao consenso de uma única opinião como no modelo da regra da maioria de Galam [143]. Isto é, a distribuição de opiniões colapsa

PUC-Rio - Certificação Digital Nº 1122074/CA

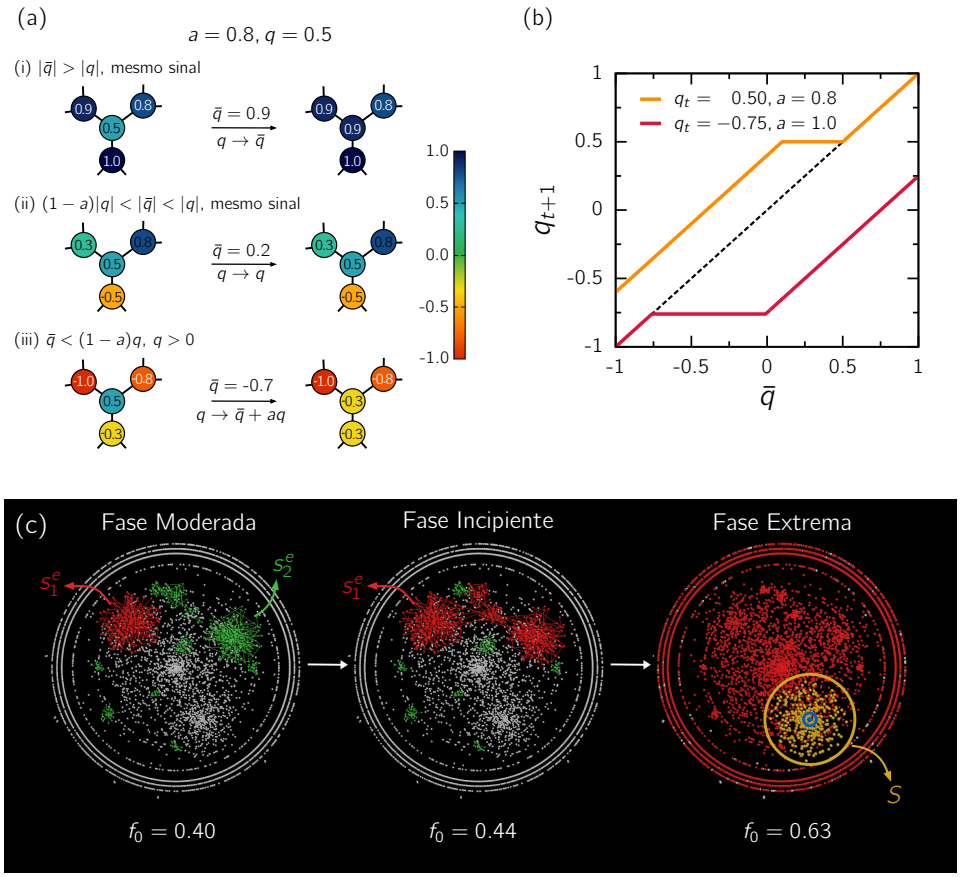


Figura 6.3: (Subtítulo na página seguinte)

numa função delta; ou seja, todos os sítios teriam o mesmo posicionamento e mesma convicção. Por outro lado, ao definir $a > 0$, pessoas com opiniões inicialmente diferentes, não serão facilmente convencidas e a heterogeneidade de opiniões irá persistir no estado final. Nesse caso, obtemos uma distribuição contínua de opiniões. Uma distribuição contínua de q é também um resultado realístico: mesmo que haja grupos que compartilhem da mesma opinião, há uma variabilidade no nível de convicção de cada pessoa. Um exemplo de distribuição contínua de opiniões para $a > 0$ é ilustrado na Fig. 6.4 para diferentes valores de f_0. Conforme o valor da fração f_0 aumenta, muda também a forma da distribuição $P(q)$; seu valor máximo h_p para valores positivos de q se torna mais acentuado e é deslocado na direção de valores cada vez mais extremos.

Note que aqui a teimosia é conceitualmente diferente de definições anteriores [136] que utilizam campos locais. Também é diferente da idéia de indivíduos intransigentes, onde nós inflexíveis não mudam de opinião [175, 176].

Simulamos o modelo em redes de Erdös-Rényi (ER) com conectividade

PUC-Rio - Certificação Digital Nº 1122074/CA

Figura 6.3: **Modelagem.** (**a**) Considere um nó com conectividade 3 e que detém uma opinião $q = 0,5$ com teimosia $a = 0,8$. Há três situações possíveis de acordo com as regras de modelo dependendo do valor de $\bar{q}$: *(i)* Se $\bar{q}(= 0,9) > q$ e tem o mesmo sinal de q, a opinião do nó se torna mais extrema, $q \to \bar{q}$. *(ii)* Se o $\bar{q} < q$, mas $\bar{q}$ é maior do que uma fração de q dada por $1-a$, então os vizinhos não conseguem mudar a opinião do nó devido à sua teimosia. Por conseguinte, a opinião do sítio permanece a mesma. *(iii)* Quando a opinião média dos vizinhos é mais moderada ou de sinal contrário, eles pode influenciar a opinião do sítio. Neste caso $\bar{q} = -0,7$, a opinião positiva do sítio muda, tornando-se $q = -0,3$. (**b**) Diagrama mostrando a nova opinião de um sítio no passo $t + 1$, q_{t+1}, como função da opinião média de seus primeiros vizinhos no passo t, $\bar{q}$. Dois casos típicos são retratados. Curva vermelha: $q_t = 0,5$ e $a = 0,8$ (este caso corresponde a situação ilustrada na parte (**a**)). Curva laranja: sítio com opinião negativa extrema $q_t = -0,75$ e teimosia $a = 1$. Quanto maior o valor de a mais amplo o intervalo de inflexibilidade da regra *(ii)* em que a opinião do sítio não muda. (**c**) Ilustração das diferentes fases do modelo para diferentes valores de f_0 em uma rede ER de tamanho $N = 10^4$ e $\langle k \rangle = 4$. Na fase moderada, aglomerados extremistas ficam fragmentados. O maior e-cluster está destacado em vermelho e os outros 10 maiores e-clusters estão destacados em verde, os nós brancos são sítios moderados [172–174]. Na fase incipiente, um e-cluster finito aparece pela primeira vez (em vermelho). O sistema é cada vez susceptível à perturbações; o cluster amarelo na fase extrema representa uma cascata resultante da mudança da opinião de um sítio que foi alterado externamente, passando de moderado para extremo. Para um estado final já bem imerso na fase III, a maioria absoluta dos nós (em vermelho) são extremistas.

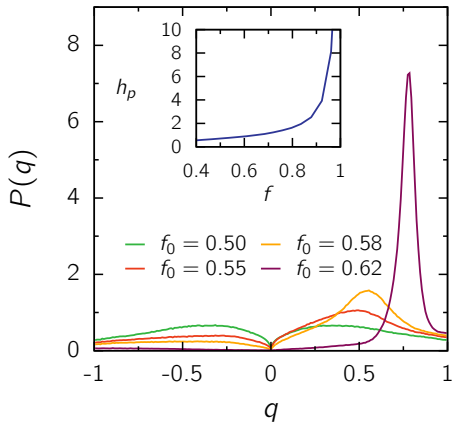


Figura 6.4: **Distribuição de opiniões** Função densidade de probabilidade $P(q)$ no estado final de uma simulação feita para $a = 1$ em uma rede de ER de tamanho 10^5 com $\langle k \rangle = 4,2$. As diferentes distribuições correspondem a diferentes valores de f_0 como indicado na figura. O gráfico inserido mostra a dependência da altura do pico h_p com f_0.

PUC-Rio - Certificação Digital Nº 1122074/CA

média $\langle k \rangle$, começando com uma fração f_0 dos sítios com opinião positiva. Para definir o estado inicial da dinâmica, simulada em uma rede de tamanho N, selecionamos um valor de f_0 que representa a fração inicial de nós com opinião positiva. Depois disso, selecionamos $f_0 N$ vértices e atribuímos a cada um deles um valor aleatório de opinião q uniformemente distribuída entre 0 e +1. Aos $(1 - f_0)N$ nós restantes, atribuímos um valor aleatório para q uniformemente distribuído entre -1 e 0. A cada passo de tempo t, as opiniões de todos os nós da rede são atualizadas de forma síncrona de acordo com as regras definidas acima. Extremistas positivos são minoria para qualquer condição inicial. No estado final, controlado por f_0, calculamos as frações f^+ (f^-), que corresponde a fração de sítios no estado final com $q > 0$ ($q < 0$), e f_e^+ (f_e^-), que é a fração de sítios no estado final com $q > q_e$ ($q < q_e$).

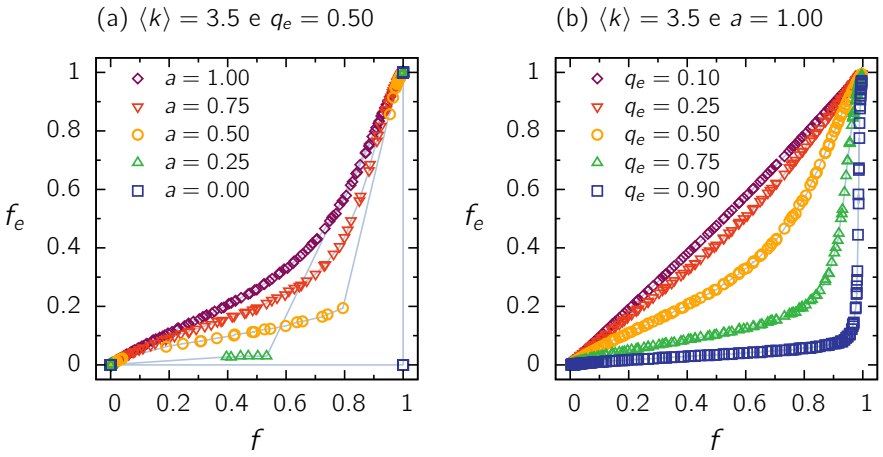


Figura 6.5: **Efeitos de a e q_e.** (**a**) Fração de extremistas f_e como função de f para diferentes valores de a e $q_e = 0,50$. Quando diminuímos o valor de a para zero a situação final se reduz ao regime de consenso. (**b**) f_e como função de f para $a = 1$ e diferentes valores de q_e (indicados na figura). Para q_e próximo de 0 ou 1, as curvas (f_e, f) apresentam comportamentos distintos daqueles observados nos dados empíricos: com q_e próximo de 0 encontramos uma relação de 1 para 1 entre f_e e f, e com q_e próximo de 1, a maioria dos sítios extremistas só emerge quando $f \to 1$. Portanto a escolha de valores intermediários de q_e é mais realística.

Como é ilustrado na Fig. 6.5, o modelo apresenta uma rica variedade de comportamentos para a dependência de f_e com f. Figura 6.5a mostra a redução do modelo para o regime de consenso. Para um valor de a grande o suficiente ($a = 1$ e $0,75$), encontramos um espectro contínuo de f_e como função de f, variando de $f = 0$ a $f = 1$. Diminuindo o valor de a, uma descontinuidade emerge, e a fração f_e colapsa e se torna dominante apenas para valores grandes de f. Para $a = 0$, o modelo atinge o regime de consenso e cada nó adota a

PUC-Rio - Certificação Digital Nº 1122074/CA

mesma opinião dada pela média da distribuição inicial de q. Neste regime, considerando que usamos uma distribuição uniforme de q, se $f_0 < 0,5$, f e f_e são nulos. No entanto, quando $f_0 > 0,5$ os valores positivos de q dominam, e $f = 1$. Como representado na figura Fig. 6.5a, para $a = 0$ e $f = 1$, f_e assume dois valores. Neste caso, $f_e = 0$ se a média da distribuição inicial de q é menor do que $q_e = 0,5$, e $f_e = 1$ se a média da distribuição inicial de q é maior do que $q_e = 0,5$. A Figura 6.5b ilustra como o modelo se comporta variando q_e para valores fixos de $\langle k \rangle$ e a. Se q_e é pequeno, por exemplo $q_e = 0,1$, nota-se uma relação linear entre f e f_e, uma vez que quase todos os nós com q positivo são classificados como extremos. Aumentando q_e diminuímos a inclinação do trecho linear (para o qual f_e é pequeno). Para valores ainda maiores de q_e uma emergencia súbita e não realística do extremismo é encontrada. Portanto, com base nos resultados empíricos apresentados na Fig. 6.2, e uma vez que estamos interessados em regimes não consensuais, consideramos, sem perda de generalidade, o conjunto representativo de parâmetros $a = 1$ e $q_e = 0,5$. Como mostrado na Fig. 6.6a o modelo com $a = 1$, $q_e = 0,5$, e $\langle k \rangle = 4,2$ mimetiza muito bem a curva (f, f_e) do levantamento sobre religião.

6.2.3
Fases de opinião extrema

Na sequência, discutimos como a fenomenologia do modelo nos permite interpretar a não linearidade em termos de mudanças na dinâmica microscópica do espalhamento de opiniões. Estas mudanças se manifestam em transições bem definidas entre as diferentes fases finais retratadas na Fig. 6.3c. As transições de uma fase para outra são caracterizadas pelo comportamento percolativo dos sítios extremistas relativo à rede de contatos. O comportamento dos *clusters*, formados pelos sítios extremistas (que chamaremos de *e-clusters* como no trabalho original) e ilustrados na Fig. 6.3c revela a origem da não linearidade. Variando f_0, o sistema passa por três fases distintas, separadas por dois pontos críticos como exemplificado na supracitada Fig. 6.3c. A fenomenologia das transições está intimamente relacionada a modelos de ativação como a percolação de inicialização [177, 103, 64, 102], o modelo de opinião de Watts [63, 178] e o modelo de competição de inovações de Helbing *et al.* [179]. De fato, há uma correspondência entre a dinâmica de ativação da percolação de inicialização e a mudança de visões predominante moderadas para extremas (*e-activation* em inglês) em nosso modelo; podemos considerar, no nosso modelo, a passagem de um sítio do estado moderado para o estado extremo como uma ativação.

O propósito do modelo é interpretar o comportamento não linear em

PUC-Rio - Certificação Digital Nº 1122074/CA

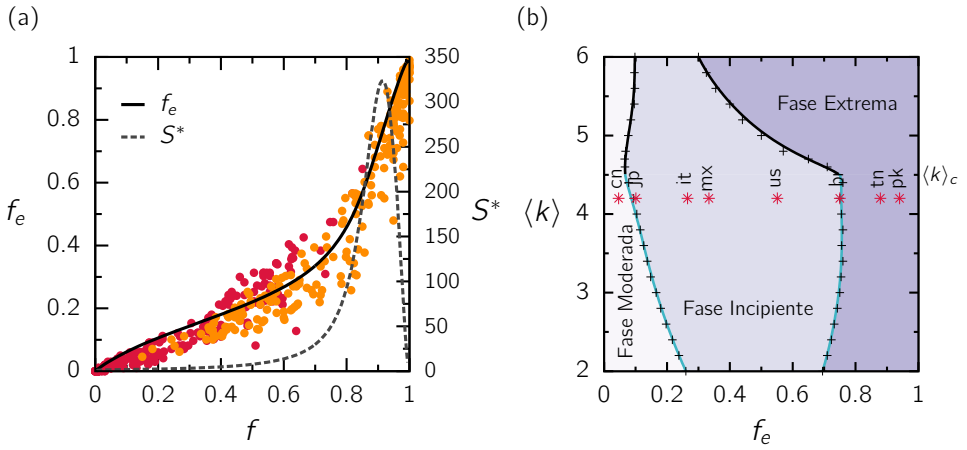


Figura 6.6: **Validação do modelo.** (**a**) f_e vs f para o levantamento sobre religião e o ajuste usando o modelo com os parâmetros $N = 10^4$, $a = 1$ e $\langle k \rangle = 4,2$. A curva obtida com o modelo é bem próxima do resultado empírico. Plotamos também o tamanho da maior avalanche S^*, obtida através de uma perturbação no estado final da dinâmica como explicado no texto. O aparecimento de comportamento não linear acompanha o aumento no tamanho das cascatas. (**b**) Diagrama de fases do modelo em termos de f_e e $\langle k \rangle$. As linhas pretas ($\langle k \rangle > k_c$) correspondem a transições de primeira ordem e as linhas azuis ($\langle k \rangle < k_c$) correspondem às transições contínuas. Fase moderada: não há e-cluster gigante. Fase incipiente: um e-cluster gigante aparece, e há um aumento dos efeitos em cascata. Fase extrema: caracterizada pelo consenso dos extremistas para uma conectividade média suficientemente alta. Os símbolos representam países selecionados a partir do levantamento sobre religião do painel (**a**) e abrangem todo o espectro de fases (os nomes dos países estão abreviados no formato FIPS). $\langle k \rangle$ deve ser interpretado como uma conectividade média efetiva da rede através da qual há o espalhamento da opinião e não como o número real de vínculos entre os indivíduos, podendo ser este último muito maior. A conectividade média efetiva é obtido a partir do ajuste feito no painel (**a**).

termos de transições de fase críticas que não podem ser medidas diretamente a partir dos dados reais, uma vez que normalmente a rede de contatos é desconhecida. O modelo identifica as seguintes fases:

Fase I ou fase moderada Para valores de f_0 pequenos, observamos e-clusters pequenos e isolados. O tamanho do maior e-cluster, s_1^e, como função de f_0 vai a zero (Figs. 6.7a e 6.7e) e o comportamento da curva (f, f_e) permanece aproximadamente linear.

PUC-Rio - Certificação Digital Nº 1122074/CA

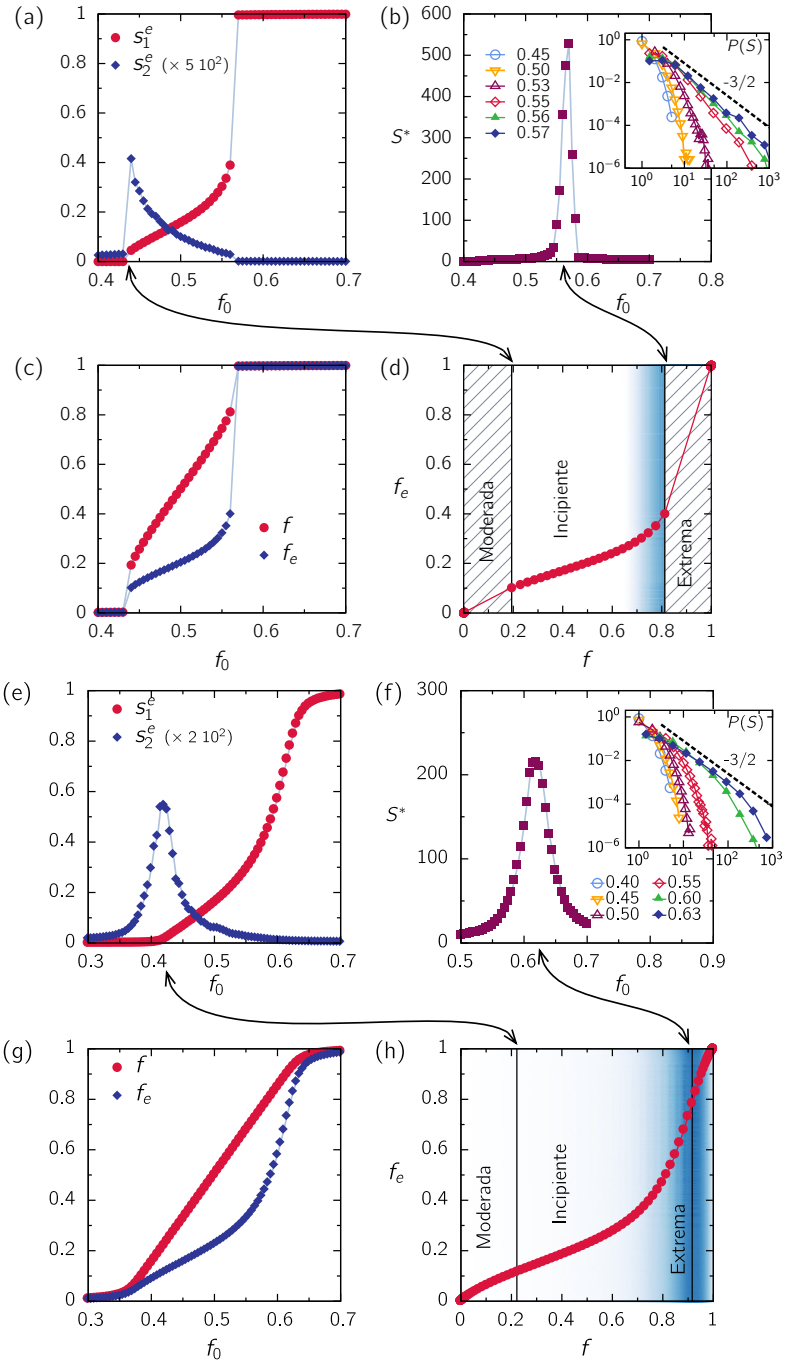


Figura 6.7: (Subtítulo na página seguinte)

Fase II ou fase incipiente Acima de um valor crítico, f_{0c_1}, surge um e-cluster de tamanho s_1^e que ocupa uma fração finita da rede. (Figs. 6.7a e 6.7e). O ponto crítico f_{0c_1} também é assinalado pelo pico no tamanho do segundo maior e-cluster, s_2^e. A ordem desta transição é determinada por $\langle k \rangle$, em comparação

PUC-Rio - Certificação Digital Nº 1122074/CA

Figura 6.7: **Transições críticas.** (**a**)-(**d**) Gráficos para $\langle k \rangle = 5 > k_c$. (**e**)-(**h**) Gráficos para $\langle k \rangle = 4 < k_c$. Os resultados apresentados são médias sobre 50 realizações da dinâmica em redes (diferentes) de ER (exceto nos painéis (**b**) e (**f**) onde as médias foram feitas sobre 300 realizações), em todos os painéis usamos $a = 1$. (**a**) e (**e**) s_1^e e s_2^e vs f_0. O tamanho dos clusters foram normalizados pelo tamanho da maior componente da rede ($N = 10^5$). ((**b**) e (**f**) Tamanho da maior cascata S^* vs f_0. O gráfico inserido mostra a distribuição de tamanhos de cascata para diferentes valores de f_0, exibindo uma lei de escala para f_0 próximo de f_{0c_2} ($N = 10^4$). (**c**) e (**g**) f_e e f vs f_0 ($N = 10^5$). (**d**) e (**h**) Curvas (f, f_e) exibem novamente o comportamento não linear. As regiões hachuradas em (**d**) correspondem aos saltos nas transições de primeira ordem, portanto, inacessível no limite de tamanho infinito. As áreas coloridas em tom azulado nos painéis (**d**) e (**h**) representam a região onde ocorrem as cascatas S^* nos painéis (**b**) e (**f**), mostrando que a não linearidade está associada com a ocorrência de cascatas progressivamente maiores (representadas por tons de azul gradualmente mais escuros) à medida que f aumenta. Por clareza, consideramos apenas a fração de vértices positivos ($q > 0$) e os positivos extremos ($q_e > 0$) no cálculo de f e f_e.

PUC-Rio - Certificação Digital Nº 1122074/CA

com um valor crítico $k_c = 4,5 \pm 0,1$. Para $\langle k \rangle > k_c$, s_1^e (Fig. 6.7a), f_e e f (Fig. 6.7c), apresentam uma descontinuidade em f_{0c_1}; uma assinatura de uma transição abrupta de primeira ordem. Para $\langle k \rangle < k_c$, a transição é de segunda ordem como na percolação tradicional: o tamanho s_1^e cresce continuamente em f_{0c_1} e s_2^e apresenta um pico (Fig. 6.7e). As frações f_e e f também aumentam de maneira suave (Fig. 6.7g).

Depois do aparecimento do e-cluster gigante, acontece um fenômeno coletivo de "avalanches" de opiniões extremas, por meio do qual o extremismo se propaga. Quantificamos a dinâmica de avalanches inspirados por uma dinâmica semelhante a que aparece na percolação de inicialização [64, 102].

Na percolação de inicialização [64, 102] os sítios da rede podem assumir dois valores: ativos ou inativos. No começo da dinâmica, uma fração f_a dos nós são escolhidos ao acaso e atribuídos ao estado ativo, ficando o restante dos nós no estado inativo. Já no nosso modelo partimos de um configuração inicial onde os nós estão ativos com probabilidade $f_0^e = (1 - q_e) f_0$. Um nó inativo na percolação de inicialização se torna ativo somente se ele tem pelo menos k vizinhos ativos, onde k é um parâmetro fixo do modelo. Sítios ativados permanecem nesse estado para sempre. Essa regra de ativação é iterativamente aplicada até que o sistema atinja um estado estacionário, sem mudanças adicionais. Uma variante foi introduzida por Watts [63] no qual a condição de ativação é dada por uma fração mínima de vizinhos ativos, ao invés de se considerar um número k fixo de vizinhos.

Na percolação de inicialização, quando temos a presença de um cluster

gigante de nós ativos, uma mudança infinitesimal da fração de nós ativos pode desencadear uma avalanche de ativações. Este processo em cascata está relacionado com a existência de clusters subcríticos de nós ativáveis. Um nó pertence a um cluster subcrítico se está conectado a outros $1 - k$ sítios fora do cluster gigante, onde k é a conectividade limiar necessária para ativação. Quando um vizinho j do sítio subcrítico i se torna ativo, então i também é ativado. Com isso, os nós do cluster subcrítico conectados ao sítio i passam a ter um vizinho ativo a mais, e por consequência também são ativados e acontece uma cascata (veja a Fig. 6.8). Em contraste, no nosso caso, as regras de ativação são muito mais complexas para permitir uma definição clara de sítios sub-críticos. De fato, a ativação de um nó "ativável" não garante a ativação dos seus vizinhos ativáveis mais próximos . Além disso, uma ativação indireta também é possível: um nó i pode ser transitivamente ativado através de algum nó intermediário já no estado ativo, desde que o nó i receba uma contribuição adicional em $\bar{q}$, devido à alteração de uma dos seus vizinhos.

PUC-Rio - Certificação Digital Nº 1122074/CA

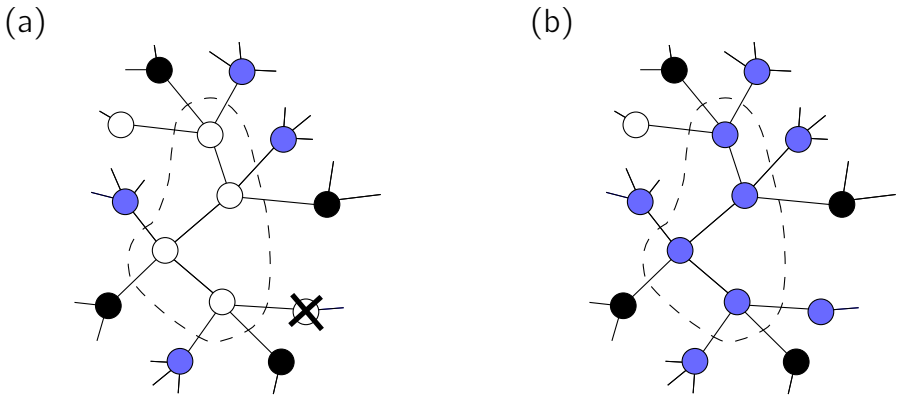


Figura 6.8: Ilustração do processo de ativação de um cluster subcrítico na percolação de inicialização. Em (a) temos um cluster subcrítico com limiar de ativação $k = 3$. Os vértices pretos representam as sementes, os vértices sombreados são os sítios ativos e os vazios são os sítios inativos. A área pontilhada delimita o cluster subcrítico. Se um sítio na vizinhança do cluster subcrítico se torna ativo, todos os pertencentes a esse cluster serão ativados, como ilustrado em (b). Essa figura foi adaptada da Ref. [64]

A fim de detectar e caracterizar as possíveis avalanches do presente modelo, nós contornamos as dificuldades mencionadas acima, fazendo uma pequena perturbação no estado estacionário do sistema. Consideramos os nós com opinião moderada positiva, (*i.e.* com $0 < q < q_e = 0,5$) como sendo os nós vulneráveis (ou subcríticos na nomenclatura da Ref. [64]). Escolhemos, então, um nó i com opinião $0 < q < q_e = 0,5$ e alteramos seu estado para

$q = 1$, dessa forma "acionamos" o sítio i. Aplicamos iterativamente as regras do modelo até que o sistema atinja novamente o equilíbrio. Por fim, medimos o número S de nós vulneráveis que se tornaram extremistas no novo estado final. Aqui S representa o tamanho da avalanche. A figura 6.3c mostra o resultado do processo descrito acima.

Para uma dada configuração final do sistema (*i.e.*, para uma dada amostra) acionamos, um por vez, todos os nós vulneráveis, retornando sempre para o estado não perturbado cada vez que acionamos um sítio novo. Acumulamos os dados para todos os nós vulneráveis que conseguiram desencadear uma avalanche, ou seja, para os quais $S \neq 0$. Para cada amostra computamos o tamanho da maior avalanche S^* gerada. O tamanho médio das avalanches $\langle S \rangle$ e tamanho da maior avalanche S^*, foram calculados como função de f_0. Encontramos que S é pequeno em torno de f_{0c_1} mas aumenta rapidamente com f_0. O tamanho da maior avalanche S^* como função de f é apresentado na figura Fig. 6.6a. A figura indica que a tendência não linear nas curvas (f, f_e) do modelo é acompanhada pelo aumento de tamanhos das avalanches. Assim, nós associamos o início do regime não linear com a fase extrema incipiente onde o sistema começa a ser mais sensível à perturbações; pequenas perturbações podem gerar avalanches de tamanho considerável e, por consequência, o espalhamento de opiniões extremas.

Fase III ou fase extrema S^* atinge o valor máximo no segundo ponto de crítico f_{0c_2} (Figs. 6.7b e 6.7f) sinalizando a transição para uma fase em que sociedade como um todo tornou-se extrema. Esta transição pode ser suave ou abrupta dependendo do valor de $\langle k \rangle$. Se $\langle k \rangle > k_c$, a transição é abrupta e de primeira ordem. A distribuição dos tamanho de avalanche desenvolve uma cauda em lei de potência com expoente 3/2 (veja o gráfico inserido na Fig. 6.7b). O valor deste expoente crítico sugere que o modelo está na classe de universalidade da percolação de inicialização [64, 102, 179]. Além disso, a dinâmica de ativação na percolação de inicialização [64] e no modelo de opinião Watts [63] exibem transições híbridas como no nosso modelo: uma combinação de um salto (como em transições de primeira ordem) e uma lei de potência (como em transições de segunda ordem) próximo ao ponto crítico. Perto do ponto crítico, o tamanho do maior e-cluster comporta-se como

$$|s_1^e - s_{1c}^e| \sim |f_0 - f_{0c}|^{\zeta}, \tag{6-1}$$

onde f_{0c} refere-se a f_{0c_1} ou f_{0c_2}, e onde o expoente $\zeta \approx 1/2$ (veja Fig. 6.9) , como na percolação de inicialização [64, 102].

Essas são transições híbridas e a aproximação de f_0 em direção ao

PUC-Rio - Certificação Digital Nº 1122074/CA

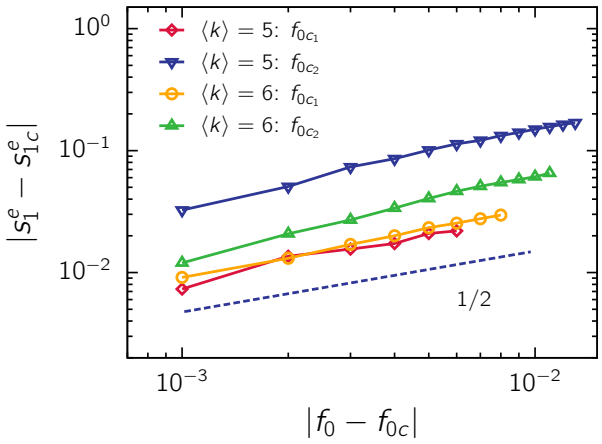


Figura 6.9: **Lei de escala da transição híbrida**, para dois casos de conectividade alta: $\langle k\rangle = 5$ e $\langle k\rangle = 6$, nos dois pontos críticos. O expoente da forma de escala é próximo de 1/2, como na percolação de inicialização. Usamos redes ER de tamanho 10^5.

PUC-Rio - Certificação Digital Nº 1122074/CA

ponto crítico em termos de leis de potência é dada acima e abaixo por f_{0c_1} e f_{0c_2}, respectivamente. Este resultado é mais um indicativo que nosso modelo, embora tenha várias diferenças em relação a percolação de inicialização, pode estar na mesma classe de universalidade.

O pico agudo de S^* (Fig. 6.7b) reflete a descontinuidade de s_1^e em f_{0c_2} (Fig. 6.7a), esta descontinuidade também pode ser notada em f e f_e (Fig. 6.7c). Após este salto abrupto, quase todos os nós pertencem ao e-cluster gigante. Quando $\langle k\rangle < k_c$, S^* apresenta um pico mais largo em f_{0c_2} (Fig. 6.7f). O expoente 3/2 do decaimento em lei de potência encontrado para $\langle k\rangle > k_c$ aplica-se aproximadamente ao envelope das distribuições de tamanho das avalanches (gráfico inserido da Fig. 6.7f). Neste caso, a transição para a fase extrema é progressiva em termos de f e f_e (Figs. 6.7g e 6.7h).

O impacto das transições críticas nas curvas (f, f_e) é ilustrado nas Figs. 6.6a, 6.7d e 6.7h. Elas mostram que o aparecimento da não linearidade na fase incipiente está associada ao aumento do tamanho das avalanches. A origem de não linearidade é a presença de cascatas de extremistas na fase II e o aparecimento da não linearidade é um preditor de mudanças mais drásticas que ocorrem quando o tamanho das avalanches é máximo. A perda da linearidade nos gráficos $f_e(f)$ é refletida na mudança drástica da forma de $P(q)$. Desde que f_e é diretamente relacionado à área sob o pico dessa distribuição, a mudança de $P(q)$ se manifesta no comportamento de f_e vs f. De fato, note a semelhança entre a forma de $h_p(f)$ mostrada no gráfico inserido da Fig. 6.4 com os gráficos

empíricos na Fig. 6.2.

6.2.4 Diagrama de fases

As diferentes fases previstas pelo modelo estão representadas na Fig. 6.6b em um diagrama de fases definido em termos de valores críticos de f_e e $\langle k \rangle$. O diagrama exibe a linha de transições de percolação que separa as fases moderada e incipiente, onde as características principais são a ausência ou a presença de um e-cluster gigante, respectivamente. O diagrama também exibe a linha que separa a fase extrema das demais. Para o caso da pesquisa sobre religião, usamos, na Fig. 6.6b, o valor $\langle k \rangle = 4,2$, que oferece o melhor ajuste entre os dados empíricos e os resultados da simulação do modelo. Para fazer esse ajuste usamos os dados da Fig. 6.6a. Uma vez que o valor de $\langle k \rangle$ é estimado, podemos plotar países particulares no diagrama de fases.

Por meio da interpretação fornecida pelo modelo, classificamos as sociedades de acordo com seu nível de extremismo; o diagrama de fases mensura a posição das sociedades em direção à predominância de atitudes extremas. Valores de f_e^+, selecionados do levantamento sobre religião da Fig. 6.6a para alguns países, são projetados no diagrama de fases, Fig. 6.6b. A maior parte dos países está localizada na fase II e alguns estão em fase III, onde a maioria da população assume uma posição extrema. Observamos, por exemplo, que em termos do posicionamento sobre a importância da religião, Paquistão e Tunísia estão na fase III positiva, quer dizer, consideram a religião muito importante, enquanto o Brasil está no ponto de transição entre a fase II e III. Os EUA também estão muito próximos do ponto de transição se aproximando bastante da fase III. A posição de um país no diagrama de fases poderia ser alterada por um aumento ou diminuição de f_e ou, mais importante, $\langle k \rangle$. A conectividade efetiva poder ser aumentada através da utilização de meios de comunicação social, por exemplo. Assim, os EUA pode entrar fase III apenas aumentando a conectividade média da população, do valor atual de $\langle k \rangle = 4,2$ para $\langle k \rangle = 5$. Isso resultaria em uma transição abrupta de primeira ordem para a fase III. Outros países, como o México, Itália e Japão estão na fase incipiente e a China está na fase moderada.

Esta classificação pode ter implicações importantes, já que poderia detectar se um país está à beira de uma mudança abrupta para a fase extrema, produzida tanto por um aumento de f_e quanto por um aumento no $\langle k \rangle$ (por exemplo, pelo aumento da conectividade média através do uso crescente de mídias sociais)

PUC-Rio - Certificação Digital Nº 1122074/CA

6.3 Considerações finais

A análise de pesquisas de opinião, como os levantamentos explorados nesse trabalho, servem como laboratório para estudo do surgimento de extremismo nas sociedades. A consistência entre os dados reais e as previsões do modelo sugere um escopo até mais amplo da análise estatística feita aqui. Os bons resultados apresentados tornam esse estudo um possível candidato para predizer outros aspectos do comportamento coletivo humano que envolvem crenças e tomada de decisão e onde prevalecem cascatas de opinião [63], tais como a competição no mercado de inovações [179, 180]. Por exemplo, a não linearidade pode ser um sinal capaz de antecipar a ampla adoção de certos bens de consumo; dado que o comportamento não linear aparece em avaliações de itens como livros e filmes. Mais pesquisas estão planejadas para investigar a aplicabilidade da análise não linear no comportamento humano coletivo em geral.

PUC-Rio - Certificação Digital Nº 1122074/CA

7
Conclusão e considerações finais

Através de uma abordagem de física estatística, exploramos mecanismos de interação em sistemas compostos de indivíduos ou "agentes sociais". Estes agentes interagem segundo regras simples e redes de contatos com topologia aleatória. Essas interações dão origem a padrões emergentes e auto-organização. Em todos os casos a modelagem foi validada através da comparação com dados empíricos. Este tipo de abordagem permite tratar problemas diversos, como a disseminação de adoções de filmes, a distribuição de preferências de produtos com múltiplas escolhas e o surgimento do extremismo.

As redes complexas em que esses processos tipicamente ocorrem são caracterizadas primariamente pela distribuição de graus. Entretanto, as correlações entre graus de nós vizinhos podem ser importantes. Estudamos como a correlação linear r se relaciona com outras características estruturais de grafos. Enquanto o impacto de r sobre o coeficiente de aglomeração não é muito marcante, o seu impacto sobre as distâncias características mostra como as correlações podem modificar de modo significativo os processos de propagação nessas redes. Como primeiro passo, nos trabalhos subsequentes, utilizamos redes aleatórias sem controlar de maneira sistemática o valor de r. Resta pesquisar o impacto das correlações nos processos estudados.

O IMDb (Internet Movie Database) é uma das páginas mais visitadas mundialmente e ao mesmo tempo fornece uma grande quantidade de dados sobre as preferências dos usuários. Assim, investigamos as distribuições desses dados procurando regularidades. Encontramos que a distribuição de votos apresenta um comportamento livre de escala, com um expoente muito próximo de 3/2. Este é um padrão robusto, independente de atributos dos filmes como nota média, idade ou gênero. O valor do expoente, típico de situações de campo médio, aponta para um mecanismo de propagação de adoções simples, onde a rede de contatos é irrelevante. Entretanto, algumas audiências (curtas e documentários) oferecem contraexemplos, possivelmente porque nesses casos a estrutura particular da rede de contatos ou o tipo de disseminação de adoções não seja tão aleatório como no caso mais geral.

Como perspectiva vale a pena pesquisar mais profundamente o que pode levar a esse comportamento excepcional. Seria interessante também procurar quais outros produtos seguem um comportamento semelhante ao observado para os filmes. Gostaríamos de salientar que além das implicações

PUC-Rio - Certificação Digital Nº 1122074/CA

dos resultados do trabalho, este trabalho mostra uma possibilidade de expansão das fontes de informação empíricas relacionados ao campo da dinâmica social, através do uso das páginas de recomendação.

Introduzimos também um modelo de opiniões para abordar o problema de múltiplas escolhas, incluindo a possibilidade de indecisão e onde as escolhas dos indivíduos evoluem segundo uma regra de pluralidade. Mostramos que essa dinâmica em redes com a propriedade de mundo pequeno produz diferentes estados estacionários realísticos, que dependem do número de alternativas e da distribuição de graus da rede de contatos: consenso, distribuição de adoções larga similar às reais e situações onde a indecisão predomina, quando o número de alternativas é suficientemente grande. Observamos que nos processos governados por uma regra da pluralidade pode ocorrer conflito e frustração que levam a uma estagnação da dinâmica por excesso de escolhas. Um efeito que pode ser mitigado pela heterogeneidade de graus da rede de contatos. Nestas redes pode ocorrer consenso mesmo havendo um número muito grande de escolhas. Entretanto o aumento de conexões também contribui para a estagnação. Os resultados do modelo são consistentes com observações empíricas para situações de escolha múltipla em que os itens apresentam uma atratividade semelhante. O modelo mostra como um item pode se tornar muito popular mesmo nessa situação em que não existe um diferencial importante de um item para outro. Além disso, o modelo reproduz a predominância de indecisão quando o número de alternativas excede certo limiar crítico.

Na última parte da tese, investigamos o surgimento de posições extremas na sociedade. Para isso foram usados os resultados de pesquisas de opinião em uma ampla gama de questões. Estas pesquisas fornecem outra importante fonte empírica para o estudo de fenômenos sociais. No trabalho desenvolvido, foi detectado, como precursor do surgimento de atitudes extremas, uma relação não linear entre a fração de extremistas e a de moderados. O modelo de opiniões proposto, com regras de ativação baseadas na teimosia dos indivíduos, permite interpretar o início dessa não linearidade em termos de uma transição abrupta do tipo percolação de inicialização onde acontecem cascatas de extremismo. O início do crescimento súbito da fração de extremistas numa população pode ser entendido com um indicador capaz de antecipar também a ampla adoção de certos bens de consumo, dado que o padrão linear/não linear também é observado em pesquisas sobre livros, filmes, etc.

Em todos estes casos estudados vemos como os modelos de opinião, aliados às enormes bases de dados, fornecem resultados com poder de interpretação e predição dos padrões empíricos.

PUC-Rio - Certificação Digital Nº 1122074/CA

A
Correlação em redes complexas - Cálculos

A.1
Coeficiente de correlação de Pearson

A covariância de x_i e x_j sobre ligações, Eq. 2-26 utilizada na Sec. 2.1, pode ser obtida da seguinte forma:

$$\begin{aligned}
\mathrm{cov}(x_i, x_j) &= \frac{\sum_{ij} A_{ij}(x_i - \langle x\rangle_e)(x_j - \langle x\rangle_e)}{\sum_{ij} A_{ij}} \\
&= \frac{1}{2M}\sum_{ij} A_{ij}(x_i x_j - \langle x\rangle_e x_i - \langle x\rangle_e x_j + \langle x\rangle_e^2) \\
&= \frac{1}{2M}\sum_{ij}(A_{ij}x_i x_j - \langle x\rangle_e^2) \\
&= \frac{1}{2M}\sum_{ij} A_{ij}x_i x_j \frac{1}{(2M)^2}\sum_{ij} k_i k_j x_i x_j \\
&= \frac{1}{2M}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2M}\right) x_i x_j, \qquad \text{(A-1)}
\end{aligned}$$

onde, nos passos acima, usamos a Eq 2-3 e a Eq. 2-24.

Nessa mesma seção apresentamos a expressão para o coeficiente de assortatividade (Eq. 2-31). Partindo da Eq. 2-29,

$$r = \frac{\sum_{ij}\left(A_{ij} - k_i k_j/2M\right) k_i k_j}{\sum_{ij}\left(k_i\delta_{ij} - k_i k_j/2M,\right) k_i k_j} \qquad \text{(A-2)}$$

podemos rearranjar os termos, fazendo

$$\begin{aligned}
r &= \frac{\sum_{ij}\left(A_{ij} - k_i k_j/2M\right) k_i k_j}{\sum_{ij}\left(k_i\delta_{ij} - k_i k_j/2M\right) k_i k_j} \\
&= \frac{\sum_{ij} A_{ij}k_i k_j - \frac{\sum_{ij}(k_i k_j)^2}{2M}}{\sum_{ij} k_i^2 k_j\delta_{ij} - \frac{\sum_{ij}(k_i k_j)^2}{2M}}. \qquad \text{(A-3)}
\end{aligned}$$

Note agora que

$$\frac{\sum_{ij}(k_i k_j)^2}{2M} = \frac{\sum_i k_i^2 \sum_j k_j^2}{N\langle k\rangle} \quad \text{e} \quad \sum_{ij} k_i^2 k_j\delta_{ij} = \sum_i k_i^3. \qquad \text{(A-4)}$$

Substituindo a Eq. A-4 na Eq. A-3 e multiplicando o numerador e o denominador por $\langle k\rangle/N$, temos

PUC-Rio - Certificação Digital Nº 1122074/CA

$$r = \frac{\langle k\rangle \sum_{ij} A_{ij}k_ik_j/N - \langle k\rangle \left(\sum_i k_i^2/N\right)^2}{\langle k\rangle \sum_i k_i^3/N - \langle k\rangle \left(\sum_i k_i^2/N\right)^2}. \tag{A-5}$$

Mas

$$\frac{\langle k\rangle \sum_{ij} A_{ij}k_ik_j}{N} = \frac{\sum_{ij} A_{ij}k_ik_j}{2M} = \langle kk'\rangle_e \quad \text{e}$$
$$\frac{\sum_i k_i^n}{N} = \langle k^n\rangle . \tag{A-6}$$

Substituindo a Eq. A-7 na Eq. A-5, temos

$$r = \frac{\langle k\rangle^2\langle kk'\rangle_e - \langle k^2\rangle^2}{\langle k\rangle\langle k^3\rangle - \langle k^2\rangle^2} . \tag{A-7}$$

por fim, usando que $\langle k\rangle_e = \langle k^{n+1}\rangle / \langle k\rangle$, temos

$$r = \frac{\langle kk'\rangle_e - \langle k\rangle_e^2}{\langle k^2\rangle_e - \langle k\rangle_e^2}, \tag{A-8}$$

que é a Eq. 2-31 apresentada na tese para a assortatividade.

A.2
Decomposição de r

A expressão para a assortatividade apresentada no Cap. 3 é

$$r = \frac{\langle kk'\rangle_e - \langle k\rangle_e^2}{\langle k^2\rangle_e - \langle k\rangle_e^2}, \tag{A-9}$$

onde $\langle\cdots\rangle_e$ denota média sobre ligações e k e k' são as conectividades dos vértices em cada extremidade de uma ligação. Lembrando que

$$\langle k^n\rangle_e = \frac{\langle k^{n+1}\rangle}{\langle k\rangle} \quad \text{e} \quad \langle k\rangle = \frac{2M}{N}, \tag{A-10}$$

onde $\langle\cdots\rangle$ (sem subíndice) significa média sobre a distribuição de graus, temos

$$r = \frac{\langle k\rangle^2\langle kk'\rangle_e - \langle k^2\rangle^2}{\langle k\rangle\langle k^3\rangle - \langle k^2\rangle^2} . \tag{A-11}$$

Os fragmentos (ou subgrafos) envolvidos no cálculo são os caminhos abertos (que envolvem $n+1$ nós), cujo número é P_n. Assim temos

$$\begin{aligned}
P_1 &= \frac{1}{2}\sum_i^N k_i \,, \\
P_2 &= \sum_i^N \binom{k_i}{2} = \frac{1}{2}\sum_i k_i(k_i-1) = \frac{N}{2}\left(\langle k^2\rangle - \langle k\rangle\right) \,, \\
P_3 &= \sum_{i,j}(k_i-1)(k_j-1) - 3n_\triangle \equiv \tilde{P}_3 - 3n_\triangle \,,
\end{aligned} \tag{A-12}$$

PUC-Rio - Certificação Digital Nº 1122074/CA

onde P_1 é o número de ligações do grafo, P_2 é o número de caminhos abertos de tamanho 2 (cunhas) e P_3 é o número de caminhos não triangulares de tamanho 3 (envolvendo 4 vértices). A expressão para P_3 na Eq. A-12 pode ser entendida da seguinte maneira (veja a Fig. A.1): considere a ligação entre os vértices i e j. De cada um desses dois vértices saem respectivamente outras $k_i - 1$ e $k_j - 1$ ligações, excluindo-se a ligação entre i e j. Somando sobre todas as ligações temos então o número de caminhos de tamanho 3, $\tilde{P}_3$. Entretanto, para contabilizar somente os caminhos abertos, envolvendo 4 vértices, devemos excluir os casos em que os vértices i' e j' representam o mesmo nó, situação que corresponde aos caminhos triangulares. Portanto devemos subtrair $n_\triangle$ vezes 3, dado que na soma sobre todos os k_i e k_j o número de triângulos é contabilizado 3 vezes.

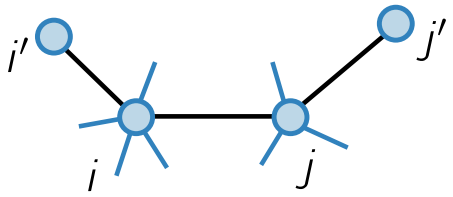


Figura A.1: Ilustração do cálculo de P_3.

Será útil ter em conta também a quantidade de caminhos de comprimento 3 (abertos e fechados) $\tilde{P}_3$ definida na Eq. A-12 e relembrar da definição do coeficiente de aglomeração

$$C = \frac{3n_\triangle}{P_2}\,. \tag{A-13}$$

Assim como foi feito na Eq. (A-13) para $3n_\triangle$, também vamos normalizar P_3 pelo número de caminhos de tamanho 2 (cunhas). A grandeza resultante,

$$B \equiv \frac{P_3}{P_2} \tag{A-14}$$

é chamada de ramificação escalada B (*scaled branching* em inglês).

Podemos reescrever $\tilde{P}_3$ de uma maneira mais conveniente para nossos cálculos como

$$\begin{aligned}
\tilde{P}_3 &= \sum_e (k-1)(k'-1) \\
&= \sum_e kk' - 2\sum_e k + \sum_e 1 \\
&= M\left(\langle kk'\rangle_e - 2\langle k\rangle_e + 1\right) \\
&= \langle k\rangle \frac{N}{2}\left(\langle kk'\rangle_e - 2\frac{\langle k^2\rangle}{\langle k\rangle} + 1\right)\,,
\end{aligned} \tag{A-15}$$

PUC-Rio - Certificação Digital Nº 1122074/CA

onde o subíndice "e" denota soma sobre ligações, ou seja sobre todos os pares (k, k') de vértices vizinhos.

Note também, a partir da Eq. A-12, que

$$\tilde{P}_3 = |P_2|(B+C) = \frac{N}{2}\left(\left\langle k^2\right\rangle - \langle k\rangle\right)(B+C)\,. \tag{A-16}$$

Igualando as Eq. A-15 e Eq. A-16, temos

$$\frac{N}{2}\left(\left\langle k^2\right\rangle - \langle k\rangle\right)(B+C) = \langle k\rangle\frac{N}{2}\left(\langle kk'\rangle_e - 2\frac{\langle k^2\rangle}{\langle k\rangle} + 1\right)\,. \tag{A-17}$$

Multiplicando os dois membros da igualdade por $\langle k\rangle$, obtemos

$$\langle k\rangle\left(\left\langle k^2\right\rangle - \langle k\rangle\right)(B+C) = \langle k\rangle^2\langle kk'\rangle_e - 2\left\langle k^2\right\rangle\langle k\rangle^2 + \langle k\rangle^2 \tag{A-18}$$

e somando $\left\langle k^2\right\rangle^2$ nos dois lados da equação acima, temos

$$\begin{aligned}\langle k\rangle^2\langle kk'\rangle_e - \left\langle k^2\right\rangle^2 &= \langle k\rangle\left(\left\langle k^2\right\rangle - \langle k\rangle\right)(B+C) + \\ &+ \langle k\rangle\left(\langle k\rangle^2 - \langle k\rangle\right) + \langle k\rangle\langle k\rangle^2 - \left\langle k^2\right\rangle^2\,,\end{aligned} \tag{A-19}$$

cujo primeiro membro é o numerador da Eq. A-11. Fazendo a manipulação

$$\langle k\rangle\langle k\rangle^2 - \left\langle k^2\right\rangle^2 = \langle k\rangle\left(\langle k\rangle^2 - \langle k\rangle\right)\left(-\frac{\langle k^2\rangle}{\langle k\rangle}\right)$$

e colocando $\langle k\rangle\left(\langle k\rangle^2 - \langle k\rangle\right)$ em evidência no lado direito da Eq. A-19, temos

$$\langle k\rangle^2\langle kk'\rangle_e - \left\langle k^2\right\rangle^2 = \langle k\rangle\left(\langle k\rangle^2 - \langle k\rangle\right)\left(B + 1 - \frac{\langle k^2\rangle}{\langle k\rangle} + C\right)\,. \tag{A-20}$$

Finalmente, substituindo o lado direito da equação acima no numerador da Eq. A-9, chegamos à equação

$$r = \frac{\langle k\rangle(\langle k^2\rangle - \langle k\rangle)\left(B + 1 - \frac{\langle k^2\rangle}{\langle k\rangle} + C\right)}{\langle k\rangle\langle k^3\rangle - \langle k^2\rangle^2}\,. \tag{A-21}$$

PUC-Rio - Certificação Digital Nº 1122074/CA

B
Base de dados do IMDb

Descrição dos dados do IMDb. Foram coletados votos (de 1 a 10 estrelas) para todos os itens do IMDb, exceto os episódios de TV (num total de 336.090.882 votos para 300.723 filmes). Os dados foram coletados de 19 a 28 de fevereiro de 2013 (conjunto # 1). Usando a mesma lista de filmes, coletamos novamente o número de votos de cada item no período de 08 a 18 de dezembro de 2014 (conjunto # 2: 465.292.451 votos) e mais uma vez de 05 a 10 de janeiro de 2015 (conjunto # 3: 471.222.420), conforme mostrado na (Fig. 4.11). Os itens com menos de 5 votos não são exibidos. Número de itens por tipo: 33.941 (Documentário); 133.775 (Longa-metragem); 3.172 (Minisséries); 50.408 (Curta); 1.071 (Especial de TV); 25.168 (Filme para TV); 33.165 (Série de TV); 2.450 (Especial de TV); 12.120 (Vídeo); 5.453 (Vídeo Game).

Por gênero: 24.911 (Ação); 93 (Adulto); 15.651 (Aventura); 18.918 (Animação); 5385 (Biografia); 74.393 (Comédia); 18.693 (Crime); 37.250 (Documentário); 97.087 (Drama); 16.022 (Família); 8677 (Fantasia); 567 (Filme *Noir*); 1575 (Game Show); 5525 (História); 15.072 (Horror); 10.212 (Música); 5840 (Musical); 8170 (Mistério); 1036 (Notícias); 3605 (Realidade Show); 21.165 (Romance); 8239 (Sci-Fi); 61.538 (Curta); 4360 (Esporte); 1467 (Talk Show); 16.246 (Suspense); 5080 (Guerra); 4549 (Western). Um item pode ser classificado em mais de um gênero. Como observação final, é possível para um usuário remover o seu voto; como conseqüência, uma pequena fração de filmes têm uma diminuição do número de votos. No entanto, isto representa uma fração desprezível dos filmes. Nós usamos a seguinte lista: `http://www.imdb.com/search/title?title_type=feature,tv_movie,tv_series,tv_special,mini_series,documentary,game,short,video&user_rating=1.0,10`.

PUC-Rio - Certificação Digital Nº 1122074/CA

C
Regressão não paramétrica

Consideramos procedimentos de regressão não paramétrica para obter uma curva suave estatisticamente representativa a partir dos conjuntos de dados dispersos (x_i, y_i), $i = 1, ..., n$ como aqueles da Fig. 6.2, 4.3 e 4.6: a regressão ponderada localmente (LOESS) e o estimador de Nadaraya-Watson (NW). Nós usamos o LOESS, com banda $h = 0.8$ para extrair a tendência das curvas (f, f_e) na Fig. 6.2, bem como o estimador de NW.

LOESS: Os valores estimados $\hat{y}_i$ para cada ponto x_i são obtidos através de um ajuste de mínimos quadrados com peso [111]. A função peso W depende da distância h_i ao i-ésimo vizinho mais próximo do ponto i. Os $k = 1, ..., n$, (com $k \neq i$) pesos para cada ponto x_i são dados por

$$w_k(x_i) = W\left(\frac{x_k - x_i}{h_i}\right), \tag{C-1}$$

onde W é a função peso cúbica

$$W = \begin{cases} (1 - |x|^3)^3 & , \text{se} \quad |x| < 1 \\ 0 & , \text{se} \quad |x| \geqslant 1\,. \end{cases}$$

A equação (C-1) determina o valor estimado de $\hat{y}_i$ como na referência [111].

Nadaraya-Watson: construímos o kernel suavizador [168]

$$\hat{m}_h(x) = \frac{\sum_i^n K_h(x - x_i)Y_i}{\sum_i^n K_h(x - y_i)}\,, \tag{C-2}$$

onde $K_h(x - x_i)$ é um kernel gaussiano com a forma,

$$K_h(x - x_i) = \exp\left[\frac{(x - x_i)^2}{2h^2}\right], \tag{C-3}$$

com largura de banda h estimada pelo método de validação cruzada por mínimos quadrados.

PUC-Rio - Certificação Digital Nº 1122074/CA

D
Descrição dos levantamentos usados na Fig.6.2

Nós fornecemos informações sobre os dados utilizados em cada painel exibido na Fig. 6.2. Para os levantamentos de dados, obtidos por exemplo do Pew Research Center, apresentamos explicitamente, quando disponível, *(i)* a pergunta utilizada em cada levantamento, *(ii)* a URL original, onde os dados podem ser encontrados, *(iii)* o número de países onde foram realizadas as pesquisas, *(iv)* o número de levantamentos realizados, o qual é maior que o número de países fornecidos em *(iii)* uma vez que as pesquisas são realizadas ao longo de muitos anos em determinados países, e *(v)* as datas em que as pesquisas foram realizadas.

(a) Religião:

Pergunta: Quão importante é a religião em sua vida – muito importante, pouco importante, não muito importante ou nada importante?

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=408&cntIDs=&stdIDs=`

Número total de países: 59

Número total de levantamentos: 231

Datas dos levantamentos: Verão de 2002, Primavera de 2005, Primavera de 2006, Primavera de 2007, Primavera de 2008, Primavera de 2009, Outono de 2009, Primavera de 2010, Primavera de 2011, Final da Primavera de 2011, e Primavera de 2012.

(b) Situação Econômica:

Pergunta: Agora pensando sobre nossa situação econômica, como você descreveria a atual situação econômica em (*país do levantamento*) – é muito boa, razoavelmente boa, um pouco ruim ou muito ruim?

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=753&cntIDs=&stdIDs=`

Número total de países: 59

Número total de levantamentos: 260

PUC-Rio - Certificação Digital Nº 1122074/CA

Datas dos levantamentos: Verão de 2002, Primavera de 2007, Primavera de 2008, Primavera de 2009, Outono de 2009, Primavera de 2010, Primavera de 2011, Final da Primavera de 2011, Primavera de 2012, e Primavera de 2013.

A evolução temporal é mostrada nos seguintes casos:

Países: França, Itália, Grécia, e Espanha.

Número total de levantamentos: 24

Datas dos levantamentos: Verão de 2002 (França e Itália), Primavera de 2007 (França, Itália, e Espanha), Primavera de 2008 (França e Espanha), Outono de 2009 (França, Itália, e Espanha), Primavera de 2009 (França e Espanha), Primavera de 2010 (França e Espanha), Primavera de 2011 (França e Espanha), Primavera de 2012 (França, Itália, Grécia, e Espanha), e Primavera de 2013 (França, Itália, Grécia, e Espanha).

(c) Judeus:

Pergunta: Por favor, me diga se você tem uma opinião muito favorável, moderadamente favorável, um moderadamente desfavorável, ou muito desfavorável em relação aos judeus

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=834&cntIDs=&stdIDs=`

Número total de países: 32

Número total de levantamentos: 131

Datas dos levantamentos: Primavera de 2004, Primavera de 2005, Primavera de 2006, Primavera de 2008, Primavera de 2009, Primavera de 2010, Primavera de 2011, e Final da Primavera de 2011.

(d) Muçulmanos:

Pergunta: Por favor, me diga se você tem uma opinião muito favorável, moderadamente favorável, um moderadamente desfavorável, ou muito desfavorável em relação aos muçulmanos

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=836&cntIDs=&stdIDs=`

Número total de países: 32

Número total de levantamentos: 135

Datas dos levantamentos: Primavera de 2004, Primavera de 2005, Primavera de 2006, Primavera de 2008, Primavera de 2009, Primavera de 2010, Primavera de 2011, e Final da Primavera de 2011.

PUC-Rio - Certificação Digital Nº 1122074/CA

(e) Cristãos:

Pergunta: Por favor, me diga se você tem uma opinião muito favorável, moderadamente favorável, um moderadamente desfavorável, ou muito desfavorável em relação aos cristãos

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=828&cntIDs=&stdIDs=`

Número total de países: 32

Número total de levantamentos: 133

Datas dos levantamentos: Primavera de 2004, Primavera de 2005, Primavera de 2006, Primavera de 2008, Primavera de 2009, Primavera de 2010, Primavera de 2011, Final da Primavera de 2011.

(f) Laços comerciais:

Pergunta: O que você pensa sobre os crescentes laços comerciais e empresariais entre (*país do levantamento*) e outros países - você acha que isso é uma coisa muito boa, razoavelmente boa, razoavelmente ruim ou muito ruim para o nosso país?

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=1011&cntIDs=&stdIDs=`

Número total de países: 55

Número total de levantamentos: 184

Datas dos levantamentos: Verão de 2002, Primavera de 2007, Primavera de 2008, Primavera de 2009, Primavera de 2010, Primavera de 2011, Final da Primavera de 2011.

(g) Imigração:

Pergunta: Conforme eu leio outra lista de declarações, para cada uma delas, por favor me diga se concorda totalmente, concorda em boa parte, discorda em boa parte ou discorda totalmente com... Nós devemos restringir e controlar a entrada de pessoas no nosso país mais do que fazemos agora.

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=54&cntIDs=&stdIDs=`

Número total de países: 54

Número total de levantamentos: 128

PUC-Rio - Certificação Digital Nº 1122074/CA

Datas dos levantamentos: Verão de 2002, Primavera de 2007, Primavera de 2009, Final do Outono de 2009.

(h) Estados Unidos:

Pergunta: Por favor, me diga se você tem uma opinião muito favorável, moderadamente favorável, um moderadamente desfavorável, ou muito desfavorável em relação aos Estados Unidos

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=844&cntIDs=&stdIDs=`

Número total de países: 59

Número total de levantamentos: 351

Datas dos levantamentos: Verão de 2002, March 2003, May 2003, Primavera de 2004, Primavera de 2005, Primavera de 2006, Primavera de 2007, Primavera de 2008, Primavera de 2009, Primavera de 2010, Primavera de 2011, Final da Primavera de 2011, Primavera de 2012, Final da Primavera de 2013.

(i) Influência estrangeira (proteção contra):

Pergunta: Conforme eu leio outra lista de declarações, para cada uma delas, por favor me diga se concorda totalmente, concorda em boa parte, discorda em boa parte ou discorda totalmente com... Nosso modo de vida precisa ser protegido contra a influência estrangeira.

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=51&cntIDs=&stdIDs=`

Número total de países: 52

Número total de levantamentos: 119

Datas dos levantamentos: Verão de 2002, Primavera de 2006, Primavera de 2007, Primavera de 2009, Final da Primavera de 2012.

(j) Sucesso (determinado por forças externas):

Pergunta: Por favor, me diga se concorda totalmente, concorda em boa parte, discorda em boa parte ou discorda totalmente com a seguinte afirmação... O sucesso na vida é praticamente determinado por forças fora do nosso controle.

Fonte: Pew Research Center

URL: `http://www.pewglobal.org/question-search/?qid=908&cntIDs=&stdIDs=`

PUC-Rio - Certificação Digital Nº 1122074/CA

Número total de países: 55

Número total de levantamentos: 155

Datas dos levantamentos: Verão de 2002, Primavera de 2007, Primavera de 2008, Primavera de 2009, Outono de 2009, Primavera de 2011, Final da Primavera de 2011.

(k) Aborto:

Pergunta: Você concorda muito, um pouco, não muito, ou nenhum pouco com a afirmação... Se uma mulher não quer ter filhos, ela deve poder fazer um aborto.

Fonte: Euro RSCG/TNS Sofres

URL: `http://en.wikipedia.org/wiki/Societal_attitudes_towards_abortion`

Número total de países: 10 (Somente Europa)

Número total de levantamentos: 10

Datas dos levantamentos: Maio de 2005

(l) Casamento entre pessoas do mesmo sexo:

Pergunta: Por favor, me diga se você é fortemente favorável, favorável, desfavorável ou fortemente desfavorável à ... Permitir que casais de gays e lésbicas se casem legalmente?

Fonte: Pew Research Center

URL: `http://pt.scribd.com/doc/131666438/Polls-on-Attitudes-on-Homosexuality-Gay-Ma`

Número total de países: 1

Número total de levantamentos: 28

Datas dos levantamentos: Maio de 1996-Outubro de 2012

(m) Sexo Extraconjugal:

Pergunta: O que você pensa sobre uma pessoa casada ter relações sexuais com alguém que não seja seu cônjuge, é sempre errado, quase sempre errado, errado só às vezes, ou nunca é errado?

Fonte: NORC/GSS

URL: `http://pt.scribd.com/doc/131666438/Polls-on-Attitudes-on-Homosexuality-Gay-Ma`

Número total de países: 1 (USA)

Número total de levantamentos: 23

Datas dos levantamentos: 1973-2010

PUC-Rio - Certificação Digital Nº 1122074/CA

(n) IMDb:

Usamos nesse painel os dados descritos no Apêndice B, especificamente usamos o conjunto # 1, coletado em fevereiro de 2013, mas excluímos itens nos quais o número de votos é menor que 1000. Com esse corte, consideramos 19.405 no total. Convertemos a avaliação dada em estrelas nas variáveis do nosso problema da seguinte maneira:

Número de opiniões extremas positivas (N_e^+): 9 e 10 estrelas

Número de opiniões positivas (N^+): 7, 8, 9 e 10 estrelas

Número de opiniões negativas (N^-): 1, 2, 3 e 4 estrelas

Número de opiniões extremas negativas (N_e^-): 1 e 2 estrelas

(o) Livros da Amazon:

Coletamos as avaliações de (de 1 a 5 estrelas), dos livros vendidos em `www.amazon.com`. Os pontos da Fig. 6.2o correspondem aos livros com mais de 50 votos. Convertemos as avaliações nas variáveis do modelo de maneira similar ao que fizemos com os dados do IMDb:

Número de opiniões extremas positivas (N_e^+): 5 estrelas

Número de opiniões positivas (N^+): 4 e 5 estrelas

Número de opiniões negativas (N^-): 1 e 2 estrelas

Número de opiniões extremas negativas (N_e^-): 1 estrela

URL: `http://www.amazon.com/`

Número total de livros coletados: 291.428

Livros com mais de 50 avaliações: 16.390

(p) Livros da Amazon (embaralhados):

Para cada livro da Amazon apresentado na Fig 6.2o redistribuídos aleatoriamente os votos positivos (4 e 5 estrelas) e negativos (1 e 2 estrelas), separadamente.

(q) Eleições brasileiras (para deputados estaduais em 2010):

Na disputa eleitoral de 2010, existiam 27 partidos elegíveis no Brasil: PMDB, PT, PP, PSDB, PDT, PTB, PTdoB, DEM, PR, PSB, PPS, PSC, PCdoB, PV, PRB, PRP, PMN, PSL, PTC, PSDC, PHS, PTN, PRTB, PSOL, PSTU, PCB, e PCO. Obtemos a orientação potítica de cada um desses partidos de `http://en.wikipedia.org/wiki/List_of_political_parties_in_Brazil` (acessado em 22/11/2013):

PUC-Rio - Certificação Digital Nº 1122074/CA

- Extrema esquerda: PSTU, PCB, PCO (Total: 3)
- Esquerda: PT, PSB, PCdoB, PSOL (Total: 4)
- Centro esquerda: PSDB, PDT, PTB, PPS, PV, PMN (Total: 6)
- Centro: PMDB, PTdoB, PRB, PRP, PSL, PHS, PTN, PRTB (Total: 8)
- Centro direita: PTC, PSDC (Total: 2)
- Direita: PP, DEM, PR, PSC (Total: 4)

Analisamos a eleição brasileira de 2010 para deputados estaduais, que corresponde a indicar representantes para as assembléias legislativas das unidades federativas do Brasil (Fig. 6.2q). Estes dados estão disponíveis em `http://agencia.tse.jus.br/estatistica/sead/odsele/votacao_partido_munzona/votacao_partido_munzona_2010.zip`. Para cada cidade do Brasil, calculamos o número de votos recebidos pelos partidos, incluindo votos nominais e na legenda, associados a cada uma das seis orientações políticas. Observe que não há nenhum partido de extrema-direita no Brasil. Arbitrariamente, tomamos votos nos partidos de extrema-esquerda, esquerda, e centro-esquerda como opiniões negativas, N^-. Os votos em partidos de centro, centro-direita e direita são considerados como opiniões positivas, N^+. Consideramos opiniões extremas os votos em partidos da extrema-esquerda e esquerda N_e^- (extremo negativo), e os votos em partidos de direita e centro-direita N_e^+ (extremo positivo). Esta escolha é motivada pelo fato de que frações muito pequenas do eleitorado correspondem a orientações de extrema-esquerda e de centro-direita. Atualmente, existem 32 partidos no Brasil, cinco novos partidos foram criados no país desde a eleição de 2010. Nenhuma dos partido existente em 2010, considerados no presente trabalho, foi dissolvido.

PUC-Rio - Certificação Digital Nº 1122074/CA

Referências Bibliográficas

[1] ASIMOV, I. **Fundação: trilogia: fundação-fundação e imperio-segunda fundação**. HEMUS, 1975.

[2] http://www.technovelgy.com/ct/Science-Fiction-News.asp?NewsNum=1925. Acessado em 05/06/2015.

[3] CASTELLANO, C.; FORTUNATO, S. ; LORETO, V. **Reviews of Modern Physics**. Statistical physics of social dynamics, journal, v.81, p. 591–646, 2009.

[4] BALL, P. **Critical Mass: How One Thing Leads to Another : Being an Enquiry Into the Interplay of Chance and Necessity in the Way that Human Culture, Customs, Institutions, Cooperation and Conflict Arise**. Farrar, Straus and Giroux, 2004.

[5] COHEN, R. **The Natural Sciences and the Social Sciences: Some Critical and Historical Perspectives**. Springer Netherlands, 2013.

[6] QUÉTELET, A. **Physique sociale: ou, Essai sur le développement des facultés de l'homme**.

[7] BUCKLE, H. **History of Civilization in England**. Read Books, 2008.

[8] CAMPBELL, L.; GARNETT, W. 2001, 294–295p.

[9] BOLTZMANN, L.; HASENÖHRL, F. **Weitere studien über das wärmegleichgewicht unter gas-molekülen**. In: Hasenöhrl, F., editor, WISSENSCHAFTLICHE ABHANDLUNGEN, volume 1, p. 316–402. J. A. Barth, Leipzig, 1909.

[10] RAMOS, M.; ANTENEODO, C. **J. Stat. Mech. Theory Exp.** Random degree–degree correlated networks, journal, n.2, p. P02024, 2013.

[11] RAMOS, M.; SHAO, J.; REIS, S. D. S.; ANTENEODO, C.; ANDRADE JR, J. S.; HAVLIN, S. ; MAKSE, H. A. **Scientific reports**. How does public opinion become extreme?, journal, v.5, 2015.

[12] DE OLIVEIRA, S. M. M.; DE OLIVEIRA, P. **Física Em Computadores**. Livraria da Física, 2010.

PUC-Rio - Certificação Digital Nº 1122074/CA

[13] WATTS, D. J.; STROGATZ, S. H. **Nature**. Collective dynamics of 'small-world' networks, journal, v.393, p. 440–442, 1998.

[14] BARABASI, A.-L.; ALBERT, R. **Science (80-.).** Emergence of scaling in random networks, journal, v.286, n.October, p. 11, 1999.

[15] NEWMAN, M. **Networks: An Introduction**. New York, NY, USA: Oxford University Press, 2010.

[16] ALBERT, R.; BARABÁSI, A.-L. **Rev. Mod. Phys.** Statistical mechanics of complex networks, journal, v.74, p. 47–97, Jan 2002.

[17] DOROGOVTSEV, S.; MENDES, J. ; SAMUKHIN, A. **Phys. Rev. Lett.** Structure of Growing Networks with Preferential Linking, journal, v.85, n.21, p. 4633–4636, 2000.

[18] KRAPIVSKY, P.; REDNER, S. ; LEYVRAZ, F. **Phys. Rev. Lett.** Connectivity of Growing Random Networks, journal, v.85, n.21, p. 4629–4632, Nov. 2000.

[19] SIMON, H. A. **Phys. Rev. Lett.** On a class of skew distribution functions, journal, v.42, n.21, p. 425–440, 1955.

[20] PRICE, D. J. D. S. **J. Amer. Soc. Inform. Sci.** A general theory of bibliometric and other cumulative advantage processe, journal, v.27, p. 292–306, 1976.

[21] AMARAL, L. A. N.; SCALA, A.; BARTHÉLÉMY, M. ; STANLEY, H. E. **Proc. Natl. Acad. Sci. USA**. Classes of small-world networks, journal, v.97, p. 11149–11152, 2000.

[22] DAVIS, G. F.; YOO, M. ; BAKER, W. E. **Strateg. Organ.** The small world of the american corporate elite, 1982-2001, journal, v.1, p. 301–326, 2003.

[23] NEWMAN, M. E. J.; STROGATZ, S. H. ; WATTS, D. J. **Phys. Rev. E**. Random graphs with arbitrary degree distributions and their applications, journal, v.64, n.2, p. 026118, 2001.

[24] DE CASTRO, R.; GROSSMAN, J. W. **Math. Intell.** Famous trails to paul erd´os, journal, v.21, p. 51–63, 1999.

[25] GROSSMAN, J. W.; ION, P. D. F. **Congressus Numerantium**. On a portion of the well-known collaboration graph, journal, v.108, p. 129–131, 1995.

PUC-Rio - Certificação Digital Nº 1122074/CA

[26] NEWMAN, M. E. J. **Phys. Rev. E**. Scientific collaboration networks: I. network construction and fundamental results, journal, v.64, p. 016131, 2001.

[27] NEWMAN, M. E. J. **PNAS**. The structure of scientific collaboration networks, journal, v.98, p. 404–409, 2001.

[28] Dados podem ser encontrados em `http://www-personal.umich.edu/~mejn/netdata`.

[29] AIELLO, W.; CHUNG, F. ; LU, L. **A random graph model for massive graphs**. In: PROCEEDINGS OF THE 32ND ANNUAL ACM SYMPOSIUM ON THEORY OF COMPUTING, p. 171–180, New York, NY, USA, 2000. Association of Computing Machinery.

[30] AIELLO, W.; CHUNG, F. ; LU, L. **Random evolution of massive graphs**. In: Abello, J.; Pardalos, P. M. ; Resende, M. G. C., editors, HANDBOOK OF MASSIVE DATA SETS.

[31] EBEL, H.; MIELSCH, L.-I. ; BORNHOLDT, S. **Phys. Rev. E**. Scale-free topology of e-mail networks, journal, v.66, p. 035103, 2002.

[32] NEWMAN, M. E. J.; FORREST, S. ; BALTHROP, J. **Phys. Rev. E**. Email networks and the spread of computer viruses, journal, v.66, p. 035101, 2002.

[33] BEARMAN, P.; MOODY, J. ; STOVEL, K. **Am. J. Sociol.** Chains of affection: The structure of adolescent romantic and sexual networks, journal, v.110, p. 44–91, 2004.

[34] LILJEROS, F.; EDLING, C. R. ; AMARAL, L. A. N. **Microbes Infec.** Sexual networks: Implication for the transmission of sexually transmitted infection, journal, v.5, p. 189–196, 2003.

[35] LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E. ; ABERG, Y. **Nature**. The web of human sexual contacts, journal, v.411, p. 907–908, 2001.

[36] ALBERT, R.; JEONG, H. ; BARABÁSI, A.-L. **Nature**. Diameter of the world-wide web, journal, v.401, p. 130–131, 1999.

[37] BARABÁSI, A.-L., A. R.; JEONG, H. **Physica A**. Scale-free characteristics of random networks: The topology of the world wide web, journal, v.281, p. 69–77, 2000.

PUC-Rio - Certificação Digital Nº 1122074/CA

[38] BRODER, A.; KUMAR, R.; MAGHOUL, F.; RAGHAVAN, P.; RAJAGOPALAN, S.; STATA, R.; TOMKINS, A. ; WIENER, J. **Comput. Netw.** Graph structure in the web, journal, v.33, p. 309–320, 2000.

[39] REDNER, S. **Eur. Phys. J. B**. How popular is your paper? an empirical study of the citation distribution, journal, v.4, p. 131–134, 1998.

[40] BOGUNÁ, M.; PASTOR-SATORRAS, R.; DÍAZ-GUILERA, A. ; ARENAS, A. **Phys. Rev. E**. Models of social networks based on social distance attachment, journal, v.70, p. 056122, 2004.

[41] KNUTH, D. E. **The Stanford GraphBase: A Platform for Combinatorial Computing**. Addison-Wesley, Reading, MA, 1993.

[42] DOROGOVTSEV, S. N.; MENDES, J. F. F. **Proc. R. Soc. London B**. Language as an evolving word web, journal, v.268, p. 2603–2606, 2001.

[43] FERRER I CANCHO, R.; SOLÉ, R. V. **Proc. R. Soc. London B**. The small world of human language, journal, v.268, p. 2261–2265, 2001.

[44] CHEN, Q.; CHANG, H.; GOVINDAN, R.; JAMIN, S.; SHENKER, S. J. ; WILLINGER, W. **The origin of power laws in internet topologies revisited**. In: PROCEEDINGS OF THE 21ST ANNUAL JOINT CONFERENCE OF THE IEEE COMPUTER AND COMMUNICATIONS SOCIETIES, p. 171–180, IEEE Computer Society, New York, 2002. Association of Computing Machinery.

[45] FALOUTSOS, M., F. P.; FALOUTSOS, C. **Comput. Commun. Rev.** On power-law relationships of the internet topology, journal, v.29, p. 251–262, 1999.

[46] SEN, P.; DASGUPTA, S.; CHATTERJEE, A.; SREERAM, P. A.; MUKHERJEE, G. ; MANNA, S. S. **Phys. Rev. E**. Small-world properties of the indian railway network, journal, v.67, p. 036106, 2003.

[47] NEWMAN, M. E. J. **Phys. Rev. E**. Mixing patterns in networks, journal, v.67, p. 026126, 2003.

[48] VALVERDE, S.; CANCHO, R. F. ; SOL'E, R. V. **Europhys. Lett.** Scale-free networks from optimal design, journal, v.60, p. 512–517, 2002.

[49] FERRER I CANCHO, R.; JANSSEN, C. ; SOLÉ, R. V. **Phys. Rev. E**. Topology of technology graphs: Small world patterns in electronic circuits, journal, v.64, p. 046119, 2001.

PUC-Rio - Certificação Digital Nº 1122074/CA

[50] Dados podem ser encontrados em `http://snap.stanford.edu/data/p2p-Gnutella04.html`.

[51] JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z. N. ; BARABÁSI, A.-L. **Nature**. The large-scale organization of metabolic networks, journal, v.407, p. 651–654, 2000.

[52] JEONG, H.; MASON, S.; BARABÁSI, A.-L. ; OLTVAI, Z. N. **Nature**. Lethality and centrality in protein networks, journal, v.411, p. 41–42, 2001.

[53] HUXHAM, M.; BEANEY, S. ; RAFFAELLI, D. **Oikos**. Do parasites reduce the chances of triangulation in a real food web?, journal, v.76, p. 284–300, 1996.

[54] MARTINEZ, N. D. **Ecol. Monographs**. Artifacts or attributes? effects of resolution on the little rock lake food web, journal, v.61, p. 367–392, 1991.

[55] WHITE, J. G.; SOUTHGATE, E.; THOMPSON, J. N. ; BRENNER, S. **Phil. Trans. R. Soc. London**. The structure of the nervous system of the nematode caenorhabditis elegans, journal, v.314, p. 1–340, 1986.

[56] BACKSTROM, L.; BOLDI, P.; ROSA, M.; UGANDER, J. ; VIGNA, S. **Four degrees of separation**. In: PROCEEDINGS OF THE 4TH ANNUAL ACM WEB SCIENCE CONFERENCE, WebSci '12, 2012.

[57] SPIEGEL, M.; STEPHENS, L. **Schaum's Outline of Statistics**. Schaum's Outline Series. McGraw-Hill Education, 2007.

[58] FESTINGER, L.; SCHACHTER, S. ; BACK, K. **Social pressures in informal groups: A study of human factors in housing**. Stanford University Press, 1950.

[59] KELMAN, H. **Journal of Conflict Resolution**. Compliance, identification, and internalization: Three processes of attitude change, journal, v.2, p. 51–60, 1958.

[60] WEIDLICH, W. **British Journal of Mathematical and Statistical Psychology**. The statistical description of polarization phenomena in society, journal, v.24, n.2.

[61] GALAM, S.; GEFEN, Y. ; SHAPIR, Y. **J. Math. Psychol.** Sociophysics: A mean behavior model for the process of strike, journal, v.9, p. 1, 1982.

PUC-Rio - Certificação Digital Nº 1122074/CA

[62] GALAM, S.; MOSCOVICI, S. **European Journal of Social Psychology**. Towards a theory of collective phenomena: Consensus and attitude changes in groups, journal, v.21, n.1.

[63] WATTS, D. J. **PNAS**. A simple model of global cascades on random networks., journal, v.99, p. 5766–71, 2002.

[64] BAXTER, G. J.; DOROGOVTSEV, S. N.; GOLTSEV, A. V. ; MENDES, J. F. F. **Phys. Rev. E**. Bootstrap percolation on complex networks, journal, v.82, p. 011103, 2010.

[65] YEOMANS, J. **Statistical Mechanics of Phase Transitions**. Clarendon Press, 1992.

[66] PATHRIA, R.; BEALE, P. **Statistical Mechanics**. Elsevier Science, 1996.

[67] NEWMAN, M. E. J.; BARKEMA, G. T. **Monte Carlo Methods in Statistical Physics**. Clarendon Press; Oxford University Press, 1999.

[68] DE OLIVEIRA, P. **Computing Boolean Statistical Models**. World Scientific, 1991.

[69] BONABEAU, E. **Pnas**. Agent-based modeling: methods and techniques for simulating human systems., journal, v.99, n.suppl. 3, p. 7280–7287, 2002.

[70] GARDNER, M. **Scientific American**. The fantastic combinations of john conway's new solitaire game "life", journal, v.223, p. 120–123, 1970.

[71] `https://pt.wikipedia.org/wiki/Web_crawler`. Acessado em 10/06/2015.

[72] CHEONG, F. **Internet agents: spiders, wanderers, brokers, and bots**. New Riders, 1996.

[73] VAZQUEZ, A. **Phys. Rev. E**. Spreading dynamics on small-world networks with connectivity fluctuations and correlations, journal, v.74, p. 056101, 2006.

[74] SMALL, M.; XU, X.; ZHOU, J.; ZHANG, J.; SUN, J. ; LU, J. **Phys. Rev. E**. Scale-free networks which are highly assortative but not small world, journal, v.77, p. 066112, 2008.

[75] XUE, Y.; WANG, J.; LI, L.; HE, D. ; HU, B. **Phys. Rev. E**. Optimizing transport efficiency on scale-free networks through assortative or dissortative topology, journal, v.81, p. 037101, 2010.

PUC-Rio - Certificação Digital Nº 1122074/CA

[76] MORENO, Y.; GÓMEZ, J. B. ; PACHECO, A. F. **Phys. Rev. E**. Epidemic incidence in correlated complex networks, journal, v.68, p. 035103, 2003.

[77] GALLOS, L. K.; SONG, C. ; MAKSE, H. A. **Phys. Rev. Lett.** Scaling of degree correlations and its influence on diffusion in scale-free networks, journal, v.100, p. 248701, 2008.

[78] BOGUNÁ, M.; PASTOR-SATORRAS, R. **Phys. Rev. E**. Epidemic spreading in correlated complex networks, journal, v.66, p. 047104, 2002.

[79] NEWMAN, M. E. J. **Phys. Rev. Lett.** Assortative mixing in networks, journal, v.89, p. 208701, 2002.

[80] NEWMAN, M.; GIRVAN, M. **Mixing patterns and community structure in networks**. In: Pastor-Satorras, R.; Rubi, M. ; Diaz-Guilera, A., editors, STATISTICAL MECHANICS OF COMPLEX NETWORKS, volume 625 de **Lecture Notes in Physics**, p. 66–87. Springer Berlin Heidelberg, 2003.

[81] PARK, J.; NEWMAN, M. E. J. **Phys. Rev. E**. Statistical mechanics of networks, journal, v.70, n.6, p. 066117, 2004.

[82] FOSTER, D. V.; FOSTER, J. G.; GRASSBERGER, P. ; PACZUSKI, M. **Phys. Rev. E**. Clustering drives assortativity and community structure in ensembles of networks, journal, v.84, n.6, p. 066117, 2011.

[83] NOH, J. **Phys. Rev. E**. Percolation transition in networks with degree-degree correlation, journal, v.76, n.2, p. 026116, 2007.

[84] ESTRADA, E. **Phys. Rev. E**. Combinatorial study of degree assortativity in networks, journal, v.84, n.4, p. 047101, 2011.

[85] SERRANO, M. A.; BOGUNÁ, M.; PASTOR-SATORRAS, R.; VESPIGNANI, A. ; CALDARELLI, G. **Structure and dynamics of complex networks**. In: INFORMATION TECHNOLOGY TO FINANCE AND NATURAL SCIENCE. 2006.

[86] DOROGOVTSEV, S. **Phys. Rev. E**. Clustering of correlated networks, journal, v.69, n.2, p. 027104, 2004.

[87] E REIS, S. D. S. **Local and global effects on navigation in small-world networks and explosive percolation**. 2012. Tese de Doutorado - Universidade Federal do Ceará.

PUC-Rio - Certificação Digital Nº 1122074/CA

[88] BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M. ; HWANG, D. **Phys. Rep.** Complex networks: Structure and dynamics, journal, v.424, n.4-5, p. 175–308, 2006.

[89] LATORA, V.; MARCHIORI, M. **Phys. Rev. Lett.** Efficient Behavior of Small-World Networks, journal, v.87, n.19, p. 198701, 2001.

[90] BARRAT, A.; WEIGT, M. **Eur. Phys. J. B**. On the properties of small-world network models, journal, v.13, n.3, p. 547–560, 2000.

[91] ERDŐS, P.; RÉNYI, A. **Publ. Math. Debrecen**. On random graphs I., journal, v.6, p. 290–297, 1959.

[92] MOLLOY, M.; REED, B. **Random Struct. Algorithms**. A critical point for random graphs with a given degree sequence, journal, v.6, n.2-3, p. 161–180, 1995.

[93] ROSS, S. M. **Simulation, Fourth Edition**. Orlando, FL, USA: Academic Press, Inc., 2006.

[94] DOROGOVTSEV, S. N.; MENDES, J. F. F. **Evolution of networks**. In: ADV. PHYS, volume 51, p. 1079–1187, 2002.

[95] MASLOV, S.; SNEPPEN, K. **Science**. Specificity and stability in topology of protein networks, journal, v.296, n.5569, p. 910–3, Mai 2002.

[96] FOSTER, D.; FOSTER, J.; PACZUSKI, M. ; GRASSBERGER, P. **Phys. Rev. E**. Communities, clustering phase transitions, and hysteresis: Pitfalls in constructing network ensembles, journal, v.81, n.4, p. 046115, Abr. 2010.

[97] PARK, J.; NEWMAN, M. **Phys. Rev. E**. Origin of degree correlations in the Internet and other networks, journal, v.68, n.2, p. 026112, 2003.

[98] MASLOV, S.; SNEPPEN, K. ; ZALIZNYAK, A. **Physica A**. Detection of topological patterns in complex networks: Correlation profile of the internet, journal, v.333, p. 529–540, 2004.

[99] NEWMAN, M. E. J.; PARK, J. **Phys. Rev. E**. Why social networks are different from other types of networks, journal, v.68, p. 036122, 2003.

[100] MENCHE, J.; VALLERIANI, A. ; LIPOWSKY, R. **Phys. Rev. E**. Asymptotic properties of degree-correlated scale-free networks, journal, v.81, p. 046103, 2010.

PUC-Rio - Certificação Digital Nº 1122074/CA

[101] GALAM, S. **Sociophysics: A Physicist's Modeling of Psycho-political Phenomena**. Springer complexity. Springer, 2012.

[102] BAXTER, G. J.; DOROGOVTSEV, S. N.; GOLTSEV, A. V. ; MENDES, J. F. F. **Phys. Rev. E**. Heterogeneous k-core versus bootstrap percolation on complex networks, journal, v.83, p. 051134, 2011.

[103] GOLTSEV, A.; DOROGOVTSEV, S. ; MENDES, J. **Phys. Rev. E**. k-core (bootstrap) percolation on complex networks: Critical phenomena and nonlocal effects, journal, v.73, p. 056101, 2006.

[104] `http://www.similarweb.com/website/imdb.com`; `http://www.alexa.com/topsites`.

[105] `http://www.netflixprize.com`. Em Outubro de 2006, a Netflix lançou uma competição para melhora da perfomance de seu sistema de recomendação. Como parte da competição, eles liberaram uma base de dados composta de aproximadamente 100 milhões de avaliações de usuários sobre 17 mil filmes.

[106] **Contemporary Physics**. NEWMAN, M. E. J., journal, v.79, 2000.

[107] MITZENMACHER, M. **Internet Mathematics**. A brief history of generative models for power law and lognormal distributions, journal, v.1, p. 226–251, 2004.

[108] DROR, G.; KOREN, Y. **RecSys '11 Proceedings of the fifth ACM conference on Recommender systems**. Yahoo! Music Recommendations: Modeling Music Ratings with Temporal Dynamics and Item Taxonomy, journal.

[109] LORENZ, J. **Eur. Phys. J. B**. Universality in movie rating distributions, journal, v.71, p. 251–258, 2009.

[110] `http://www.imdb.com/features/15thanniversary/history`.

[111] CLEVELAND, W. S. **Journal of the American Statistical Association**. Robust locally weighted regression and smoothing scatterplots, journal, v.74, p. 829, 1979.

[112] SALGANIK, M. J.; WATTS, D. J. **Top. Cogn. Sci.** Web-based experiments for the study of collective social dynamics in cultural markets, journal, v.1, p. 439–468, 2009.

PUC-Rio - Certificação Digital Nº 1122074/CA

[113] PRZYBYLSKI, A. K.; MURAYAMA, K.; DEHAAN, C. R. ; GLADWELL, V. **Comput. Human Behav.** Motivational, emotional, and behavioral correlates of fear of missing out, journal, v.29, n.4, p. 1841–1848, Jul 2013.

[114] DE OLIVEIRA, M. J.; MENDES, J. F. F. ; SANTOS, M. A. **Journal of Physics A: Mathematical and General**. Nonequilibrium spin models with ising universal behaviour, journal, v.26, n.10, p. 2317, 1993.

[115] GALAM, S. **J. Math. Psychol.** Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach, journal, v.30, p. 426–434, 1986.

[116] CHEN, P.; REDNER, S. **Journal of Physics A: Mathematical and General**. Consensus formation in multi-state majority and plurality models, journal, v.38, n.33, p. 7239, 2005.

[117] SHAO, J.; HAVLIN, S. ; STANLEY, H. **Phys. Rev. Lett.** Dynamic opinion model and invasion percolation, journal, v.103, p. 018701, 2009.

[118] ALSTRØM, P. **Phys. Rev. A**. Mean-field exponents for self-organized critical phenomena, journal, v.38, p. 4905, 1988.

[119] CHRISTENSEN, K. **Self-Organization in Models of Sandpiles, Earthquakes, and Flashing Fireflies**. Institute of Physics and Astronomy, 1992.

[120] BAK, P. **How nature works: the science of self-organized criticality**. Springer Verlag, New York, 1996.

[121] CENTOLA, D.; EGUÍLUZ, V. M. ; MACY, M. W. **Physica A**. Cascade dynamics of complex propagation, journal, v.374, p. 449–456, 2007.

[122] HARRIS, T. E. **The theory of branching processes**. Springer-Verlag, Berlin, 1963.

[123] GOH, K. I.; LEE, D. S.; KAHNG, B. ; KIM, D. **Phys. Rev. Lett.** Branching Process Approach to Avalanche Dynamics on Complex Networks, journal, v.91, p. 148701, 2003.

[124] LEE, D. S.; GOH, K. I.; KAHNG, B. ; KIM, D. **J. Korean Phys. Soc.** Branching Process Approach to Avalanche Dynamics on Complex Networks, journal, v.44, p. 633, 2004.

PUC-Rio - Certificação Digital Nº 1122074/CA

[125] BONABEAU, E. **J. Phys. Soc. Japan**. Sandpile Dynamics on Random Graphs, journal, v.64, p. 327, 1995.

[126] OLAMI, Z.; CHRISTENSEN, K. **Phys. Rev. A**. Temporal correlations, universality, and multifractality in a spring-block model of earthquakes, journal, v.46, p. 1720–1723, 1992.

[127] FORTUNATO, S.; CASTELLANO, C. **Phys. Rev. Lett.** Scaling and Universality in Proportional Elections, journal, v.99, p. 138701, 2007.

[128] COSTA FILHO, R. N.; ALMEIDA, M. P.; ANDRADE, J. S. ; MOREIRA, J. E. **Physical Review E**. Scaling behavior in a proportional voting process., journal, v.60, p. 1067, 1999.

[129] LYRA, M. L.; COSTA, U. M. S.; FILHO, R. N. C. ; ANDRADE JR, J. S. **EPL (Europhysics Letters)**. Generalized zipf's law in proportional voting processes, journal, v.62, p. 131, 2003.

[130] LU, L.; MEDO, M.; YEUNG, C. H.; ZHANG, Y.-C.; ZHANG, Z.-K. ; ZHOU, T. **Phys. Rep.** Recommender systems, journal, v.519, p. 1–49, 2012.

[131] `http://www.ers.usda.gov/media/287629/new_product_d.html`. Acessado em 20/03/2015.

[132] TOFFLER, A. **Future Shock**. Random House, 1970.

[133] IYENGAR, S. S.; LEPPER, M. R. **Journal of personality and social psychology**. When choice is demotivating: Can one desire too much of a good thing?, journal, v.79, n.6, p. 995–1006, 2000.

[134] BIKHCHANDANI, S.; HIRSHLEIFER, D. ; WELCH, I. **A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades**, 1992.

[135] KIRMAN, A. **Q. J. Econ.** Ants, Rationality, and Recruitment, journal, v.108, p. 137–156, 1993.

[136] GALAM, S. **Physica A**. Rational group decision making: A random field Ising model at $T = 0$, journal, v.238, p. 66–80, 1997.

[137] DEFFUANT, G.; NEAU, D.; AMBLARD, F. ; WEISBUCH, G. **Adv. Complex Syst.** Mixing beliefs among interacting agents, journal, v.3, p. 87–98, 2000.

PUC-Rio - Certificação Digital Nº 1122074/CA

[138] CONT, R.; BOUCHAUD, J.-P. **Macroecon. Dyn.** Herd behavior and aggregate fluctuations in financial markets, journal, v.100, p. 170–196, 2000.

[139] CHALLET, D.; MARSILI, M. ; ZHANG, Y. **Minority Games**. Oxford University Press, 2005.

[140] CURTY, P.; MARSILI, M. **J. Stat. Mech.** Phase coexistence in a forecasting game, journal, p. P03013, 2000.

[141] DE MARTINO, A.; MARSILI, M. **Journal of Physics A: Mathematical and General**. Statistical mechanics of socio-economic systems with heterogeneous agents, journal, v.39, n.43, p. R465–R540, 2006.

[142] HOLME, P.; NEWMAN, M. E. J. **Phys. Rev. E**. Nonequilibrium phase transition in the coevolution of networks and opinions, journal, v.74, p. 056108, 2006.

[143] GALAM, S. **Eur. Phys. J. B**. Minority opinion spreading in random geometry., journal, v.25, p. 403–406, 2002.

[144] KRAPIVSKY, P. L.; REDNER, S. **Phys. Rev. Lett.** Dynamics of majority rule in two-state interacting spin systems, journal, v.90, p. 238701, 2003.

[145] CROKIDAKIS, N.; BLANCO, V. H. ; ANTENEODO, C. **Phys. Rev. E**. Impact of contrarians and intransigents in a kinetic model of opinion dynamics, journal, v.89, p. 013310, 2014.

[146] CROKIDAKIS, N.; ANTENEODO, C. **Phys. Rev. E**. Role of conviction in nonequilibrium models of opinion formation, journal, v.86, p. 061127, 2012.

[147] HE, J.; CHU, W. W. **A social network-based recommender system (SNRS)**. Springer, 2010.

[148] BALBÁS GAMBRA, M.; FREY, E. **Eur. Phys. J. B**. Social dynamics with peer support on heterogeneous networks : the "mafia model", journal, v.83, p. 507–518, 2011.

[149] TRAVIESO, G.; DA FONTOURA COSTA, L. **Phys. Rev. E**. Spread of opinions and proportional voting, journal, v.74, n.3, p. 036112, 2006.

[150] SZNAJD-WERON, K. & SZNAJD, J. **Int. J. Mod. Phys. C**. Opinion evolution in closed community, journal, v.11, p. 1157–1165, 2000.

PUC-Rio - Certificação Digital Nº 1122074/CA

[151] SLANINA, F.; LAVICKA, H. **Eur. Phys. J. B**. Analytical results for the Sznajd model of opinion formation, journal, v.35, p. 279–288, 2003.

[152] **NetShoes**. selected. Foram selecionados apenas tênis masculinos com o intervalo de preço entre R$ 100 e R$200. Número de items: 1.216. Número de avaliações: 91.816. URL: `http://www.netshoes.com.br`. Data: Abril de 2015.

[153] **Google Music Albums**. Número de items: 7.496. Número de avaliações: 1.778.226. URL: `https://play.google.com/store/music` Data: Dezembro de 2014.

[154] HOFFER, E. **The True Believer: Thoughts On The Nature Of Mass Movements**. Harper & Row, Publishers, Inc., New York, 1951.

[155] SIMON, B.; KLANDERMANS, B. **American Psychologist**. Politicized collective identity: A social psychological analysis, journal, v.56, p. 319–331, 2001.

[156] LAZER, D.; PENTLAND, A.; ADAMIC, L.; ARAL, S.; BARABASI, A.-L.; BREWER, D.; CHRISTAKIS, N.; CONTRACTOR, N.; FOWLER, J.; GUTMANN, M.; JEBARA, T.; KING, G.; MACY, M.; ROY, D. ; VAN ALSTYNE, M. **Science**. Computational social science, journal, v.323, p. 721–723, 2009.

[157] HELBING, D.; BALIETTI, S. **Eur. Phys. J. Special Topics**. From social data mining to forecasting socio-economic crisis, journal, v.195, p. 3–68, 2011.

[158] ABRAMOWITZ, A. I.; SAUNDERS, K. L. **The Journal of Politics**. Is polarization a myth?, journal, v.70, p. 542–555, 2008.

[159] LAYMAN, G. C.; CARSEY, T. M. ; HOROWITZ, J. M. **Annual Review of Political Science**. Party Polarization in American politics: Characteristics, causes, and consequences., journal, v.9, p. 83–110, 2006.

[160] MERKL, P. H.; WEINBERG, L. **The Revival of Right-Wing Extremism in the Nineties**. Frank Class, London, 1997.

[161] `https://en.wikipedia.org/wiki/Longitudinal_study`. Acessado em: 17/06/2015.

[162] `https://en.wikipedia.org/wiki/Cross-sectional_study`. Acessado em: 17/06/2015.

PUC-Rio - Certificação Digital Nº 1122074/CA

[163] SCHEFFER, M.; BASCOMPTE, J.; BROCK, W. A.; BROVKIN, V.; CARPENTER, S. R.; DAKOS, V.; HELD, H.; VAN NES, E. H.; RIETKERK, M. ; SUGIHARA, G. **Nature**. Early-warning signals for critical transitions, journal, v.461, p. 53–59, 2009.

[164] STANLEY, H. E. **Introduction to Phase Transitions and Critical Phenomena**. Oxford University Press, Oxford, 1988.

[165] BETTENCOURT, L. M. A.; LOBO, J.; HELBING, D.; KÜHNERT, C. ; WEST, G. B. **Proc. Natl. Acad. Sci.** Growth, innovation, scaling, and the pace of life in cities., journal, v.104, p. 7301–7306, 2007.

[166] MELO, H. P. M.; MOREIRA, A. A.; MAKSE, H. A. ; ANDRADE, J. S. **Sci. Rep.** Statistical Signs of Social Influence on Suicides, journal, v.4, p. 6239, 2014.

[167] OLIVEIRA, E. A.; ANDRADE, J. S. ; MAKSE, H. A. **Sci. Rep.** Large cities are less green, journal, v.4, p. 4235, 2014.

[168] NADARAYA, E. **Theor. Probab. Appl.** On estimating regression., journal, v.9, p. 141–142, 1964.

[169] SUNSTEIN, C. R.; HASTIE, R. **Wiser: Getting Beyond Groupthink to Make Groups Smarter**. Harvard Business Review Press, Boston, 2015.

[170] BISWAS, S.; CHATTERJEE, A. ; SEN, P. **Physica A**. Disorder induced phase transition in kinetic models of opinion dynamics, journal, v.391, p. 3257–3265, 2012.

[171] SINGH, P.; SREENIVASAN, S.; SZYMANSKI, B. K. ; KORNISS, G. **Sci. Rep.** Threshold-limited spreading in social networks with multiple initiators, journal, v.3, p. 2330, 2013.

[172] KITSAK, M.; GALLOS, L. K.; HAVLIN, S.; LILJEROS, F.; MUCHNIK, L.; STANLEY, H. E. ; MAKSE, H. A. **Nat. Phys.** Identification of influential spreaders in complex networks, journal, v.6, p. 888–=893, 2010.

[173] PEI, S.; MUCHNIK, L.; ANDRADE JR, J. S.; ZHENG, Z. ; MAKSE, H. A. **Sci. Rep.** Searching for superspreaders of information in real-world social media, journal, v.4, p. 5547, 2014.

[174] PEI, S.; MAKSE, H. A. **J. Stat. Mech.** Spreading dynamics in complex networks, journal, p. P12002, 2013.

PUC-Rio - Certificação Digital Nº 1122074/CA

[175] GALAM, S.; JACOBS, F. **Physica A**. The role of inflexible minorities in the breaking of democratic opinion dynamics, journal, v.381, p. 366–376, 2007.

[176] GALAM, S. **Physica A**. Public debates driven by incomplete scientific data: The cases of evolution theory, global warming and H1N1 pandemic influenza, journal, v.389, p. 3619–3631, 2010.

[177] CHALUPA, J.; LEATH, P. L. ; REICH, G. R. **J. Phys. C**. Bootstrap percolation on a Bethe lattice, journal, v.12, p. L31–L35, 1979.

[178] GRANOVETTER, M. S. **Am. J. Sociol.** Threshold models of collective behavior, journal, v.83, p. 1420–1443, 1978.

[179] HELBING, D. **Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior**. Springer-Verlag, Berlin, 2012, 169–184p.

[180] SHILLER, R. J. **Irrational Exuberance**. Princeton University Press, Princeton, 2002.

PUC-Rio - Certificação Digital Nº 1122074/CA