

Antonio Andre Monteiro Manoel

***Mecânica Estatística de Sistemas de Reputação
em Redes Autônomas***

Dissertação de mestrado apresentada ao Instituto de Física para a obtenção do título de Mestre em Ciências

Orientador:

Prof. Dr. Renato Vicente

Comissão Examinadora:

Prof. Dr. Renato Vicente (IME/USP)

Prof. Dr. Silvio Roberto de Azevedo Salinas (IF/USP)

Prof. Dr. Fernando Fagundes Ferreira (EACH/USP)

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

São Paulo

2012

Agradecimentos

Gostaria de agradecer aqueles que me apoiaram ao longo dos dois últimos anos, tempo em que estive preparando este trabalho, e em especial: aos professores Renato Vicente e Nestor Caticha, pela orientação e motivação dada; à minha família, à Cristine e a meus pais, Edson e Elizabeth; aos amigos que me acompanham desde a graduação — Bruno, Gabriel, Henrique, João Pedro, Lucas, Michel, Paulo e Petre; e aos amigos com quem convivi durante o mestrado — Alex, Diogo, Jonatas, Leonardo e Rafael.

Devo a estes sincero reconhecimento por todos os estímulos e incentivos, tanto intelectuais como emocionais, cada um deles de extrema importância na conclusão deste trabalho. Agradeço ainda aos demais com quem cruzei pelo Instituto de Física da Universidade de São Paulo desde 2006, proporcionando um ambiente de grande estímulo intelectual.

Por fim, agradeço ao Conselho Nacional de Pesquisa (CNPq) pelo apoio e financiamento que viabilizaram a realização desta tese.

A todos, muito obrigado!

Resumo

Dá-se o nome de sistemas de reputação a mecanismos em que membros de uma comunidade emitem avaliações sobre os demais e a partir destas se inferem quais dos membros podem ou não ser considerados confiáveis. Apresentamos, nesta dissertação de mestrado, um estudo sobre estes sistemas. Seguindo idéias recentes na literatura, modela-se o problema de calcular reputações a partir de avaliações não-confiáveis como um problema de inferência estatística, que é então analisado com o uso de uma técnica conhecida como propagação de crenças, permitindo que obtenhamos estimativas. Em seguida, utilizamo-nos da relação existente entre problemas de inferência e mecânica estatística para realizar um estudo analítico mais profundo, inédito, por meio de uma generalização do método de cavidade. São traçados diagramas de fase, em que se observam regiões de parâmetros para as quais o problema torna-se mais difícil de resolver; esta análise nos dá alguma intuição sobre o problema, possibilitando que sejam propostas melhorias aos métodos existentes para tratá-lo.

Abstract

It's given the name of reputation system to mechanisms in which members of a community issue each other ratings and from these it is inferred which can be trusted and which can't. We present, in this master's dissertation, a study on these systems. Following ideas from recent works on the engineering literature, the problem of calculating reputations from unreliable ratings is modeled as one of statistical inference, and then analyzed with the use of a technique known as belief propagation, allowing us to obtain estimatives. Next, we use the existing relation between inference problems and statistical mechanics to motivate a novel deeper study, by means of a generalization of the cavity method. Phase diagrams are drawn, making possible to identify regions of parameters for which the problem is harder to solve; this analysis brings insight to the problem, allowing one to propose improvements to the methods available for its treatment.

Sumário

1	Introdução	p. 9
1.1	Inferência e mecânica estatística	p. 10
1.2	Organizando a dissertação	p. 11
2	Estimando Reputações	p. 13
2.1	Sistemas de reputação — definição e exemplos	p. 15
2.2	Modelo para inferência	p. 18
2.3	Técnica de propagação de crenças	p. 22
2.4	Formulação do algoritmo	p. 28
3	Análise Teórica	p. 36
3.1	Método de cavidade	p. 37
3.2	Vidro de spin de campo aleatório	p. 47
3.3	Estudo do algoritmo	p. 51
4	Testes de Robustez	p. 55
4.1	Diferentes topologias	p. 55
4.2	Ataques ao sistema	p. 60
5	Conclusão	p. 63
	Referências Bibliográficas	p. 65
	Índice Remissivo	p. 69

1 Introdução

“Where is the knowledge we have lost in information?”

— T.S. Eliot

Nas mais diversas áreas do conhecimento, é bastante comum e por vezes até necessário tirar conclusões sobre algo a partir de observações incertas; a tal processo, dá-se o nome de *inferência*.

Raciocínios lógicos indutivos, que efetuamos a todo instante, são um exemplo de inferência; ocorrem ainda nas ciências, em que se deseja estudar a partir de medições como diferentes quantidades se relacionam entre si ou como variam no tempo.

Neste contexto, costuma-se descrever a tarefa em termos de modelos estatísticos; a relação entre duas quantidades a e b pode assim ser estudada a partir da distribuição de probabilidade conjunta $P(a, b)$; das distribuições condicionais $P(a|b)$ e $P(b|a)$; e das distribuições marginais $P(a)$ e $P(b)$. Analogamente, uma quantidade $x^{(t)}$ pode ser estudada a partir de distribuições do tipo $P(x^{(t)}|x^{(t-1)}, x^{(t-2)}, \dots)$.

É de fundamental importância, nestes estudos, a utilização do chamado teorema de Bayes, que permite que calculemos distribuições de probabilidade *a posteriori* — incorporando novas informações. Dada uma hipótese H e um novo conjunto de dados D , o teorema de Bayes toma a forma $P(H|D) \propto P(D|H)P(H)$.

Na mecânica estatística, o problema central de calcular a termodinâmica de um sistema a partir de sua descrição microscópica pode ser também visto como um problema de inferência. Desenvolveu-se na área, ao longo dos dois últimos séculos, um extenso ferramental matemático capaz de modelar com sucesso uma grande variedade de sistemas físicos.

Em trabalhos desenvolvidos na década de 1950, sobretudo por Edwin Jaynes (1), percebeu-se que este ferramental pode ser utilizado em problemas de inferência análogos — que se referem não só a sistemas físicos, mas também à outros sistemas complexos diversos. Modelos clássicos para sistemas magnéticos, assim, podem ser vistos mais geralmente como modelos para problemas de inferência em alta dimensionalidade, em que as variáveis assumem valores binários; e vice-versa.

1.1 Inferência e mecânica estatística

Outra área cujos problemas típicos são de inferência é a teoria de informação. Em 1948, Claude Shannon introduziu um *framework* (2) que permite analisar em termos estatísticos o problema de transmissão de dados; um modelo consiste, basicamente, em uma fonte que codifica a mensagem inicial, um canal (ruído) que a transmite, e um receptor que a decodifica (ver figura).

Utilizaremos a teoria da informação nesta seção para exemplificar como problemas de inferência pertencentes a outras áreas podem ser analisados por meio de técnicas da mecânica estatística.

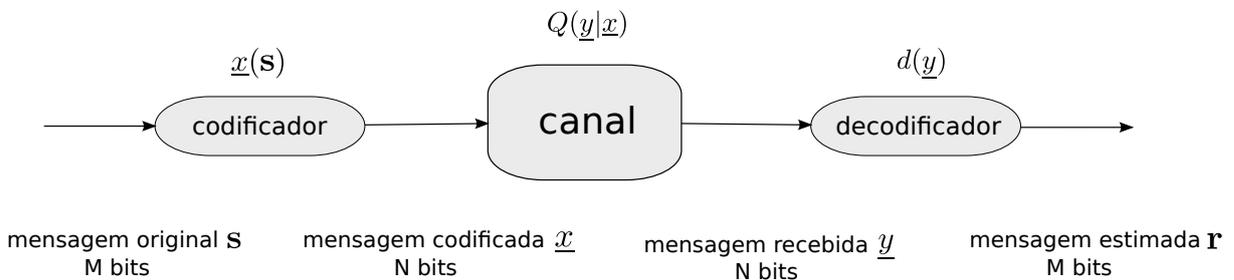


Figura 1.1: Modelo típico para transmissão de dados; figura baseada em (3)

Deseja-se que a mensagem recebida, após decodificada, seja idêntica ou ao menos bastante semelhante a mensagem enviada. Para este fim, são utilizados os chamados códigos de correção de erros, que se aproveitam de alguma redundância na mensagem transmitida para inferir a mensagem original. Matematicamente (4, 5), uma mensagem \mathbf{s} de M bits é codificada em uma mensagem \underline{x} de N bits, que é transmitida; recebe-se uma mensagem \underline{y} de N bits, que é decodificada em uma mensagem \mathbf{r} de, novamente, M bits. Um código de correção de erros consiste em uma forma para a codificação $\underline{x} = \underline{x}(\mathbf{s})$, e uma para a decodificação $\mathbf{r} = d(\underline{y})$.

Trabalha-se tipicamente com a distribuição de probabilidades $Q(\underline{y}|\underline{x})$, relacionada ao ruído no canal. São objetos de estudo quantidades como a probabilidade de erro em blocos, $p_B = P(\mathbf{r} \neq \mathbf{s})$, e a probabilidade de erro em bits, $p_b = \frac{1}{M} \sum_{i=1}^M P(r_i \neq s_i)$; procura-se, usualmente, minimizar uma destas duas quantidades.

Uma análise detalhada vem da distribuição a posteriori $P(\mathbf{s}|\mathbf{r})$, que pelo teorema de Bayes pode ser escrita como

$$P(\mathbf{s}|\mathbf{r}) = \frac{1}{Z} \exp[\log P(\mathbf{r}|\mathbf{s}) + \log P(\mathbf{s})], \quad (1.1)$$

cujas forma é denominada medida de Gibbs, $P(\underline{s}) = \frac{1}{Z} \exp[-\beta H(\underline{s})]$, com

$$H(\underline{s}) = -\log P(\underline{s}|\mathbf{r}) - \log P(\underline{s}). \quad (1.2)$$

É útil, ainda, trabalhar com $s_i \in \{-1, 1\}$; somas binárias entre variáveis pertencentes a $\{0, 1\}$, assim, passam a ser representadas por produtos.

Dada a analogia, as tarefas de minimizar p_B e p_b relacionam-se agora a outros cálculos, típicos da mecânica estatística (6, 7). A primeira está associada a determinar as mensagens que maximizam $P(\mathbf{s}|\mathbf{r})$ — ou os estados fundamentais, que minimizam $H(\underline{s})$; já minimizar p_b está relacionado a determinar as mensagens que maximizam as distribuições marginais $\{P(s_i)\}$ — ou os estados que, para $\beta = 1$, respeitam o sinal das magnetizações locais ($\beta = 1$ é dito o valor do parâmetro na *linha de Nishimori*).

Trata-se, mais geralmente, de dois problemas de inferência distintos: aquele que visa maximizar a distribuição a posteriori é comumente denominado MAP (do inglês, *maximum a posteriori*); o que vem das magnetizações locais, MPM (*maximum posterior marginal*). Técnicas usuais para determinação dos estados fundamentais e das magnetizações locais podem assim ser úteis na execução destas tarefas, no estudo de problemas quaisquer.

Outras técnicas permitem ainda que estudemos o problema mais a fundo, de forma analítica; é possível traçar diagramas de fase que são, por exemplo, capazes de explicar a dificuldade de se alcançar os estados fundamentais em alguns casos. Estas técnicas provêm intuição sobre o problema, permitindo que proponhamos novas maneiras de resolvê-lo.

1.2 Organizando a dissertação

Introduzimos, em seguida, um problema específico de inferência. Agentes, desempenhando alguma atividade qualquer, interagem entre si e avaliam uns aos outros; a partir destas avaliações, queremos determinar quais destes agentes podem ser considerados, de alguma forma, confiáveis. Mecanismos responsáveis por esta tarefa recebem o nome de **sistemas de reputação**.

No capítulo 2, são dados uma definição e exemplos de tais sistemas; discutem-se modelos gerais para os mesmos, para então formularmos um modelo específico baseados na literatura. Métodos para inferência são então introduzidos e analisados com o uso de simulações computacionais.

Uma análise teórica do processo de inferência a partir de técnicas de mecânica estatística — nossa principal contribuição — ocupa o capítulo 3. São, primeiramente, apresentadas tais

técnicas; nos utilizamos sobretudo do **método de cavidade**, um método de campo médio para sistemas desordenados.

No capítulo 4, finalmente, aproveitamo-nos do discernimento fornecido pela análise para estudar outros aspectos do problema, tal como a robustez do método proposto para inferência.

2 *Estimando Reputações*

“Uncertainty is the only certainty there is, and knowing how to live with insecurity is the only security.”

— John Allen Paulos

Avanços tecnológicos das últimas décadas deram origem a uma série de situações em que pessoas ou máquinas interagem entre si sem saber ao certo o quão confiável são as entidades com que estão lidando. Com o advento da internet, por exemplo, passamos a trocar informações, compartilhar arquivos e fazer negociações com pessoas que não conhecemos a fundo. Em realizações presenciais destas tarefas, uma certa empatia permite que avaliemos melhor o risco a que estamos nos expondo; o contato virtual, no entanto, não permite tal tipo de avaliação, de forma que se faz necessária a introdução de ferramentas capazes de fazê-la em nosso lugar.

Nas engenharias também, avanços em áreas como inteligência artificial e robótica tornaram possível a troca de informações entre dispositivos eletrônicos que são, por vezes, autônomos — não necessitam ser explicitamente controlados por alguém, sendo capazes de lidar com novas situações por si próprios através de, por exemplo, métodos de aprendizado estatístico. Estes dispositivos são, no entanto, susceptíveis à ataques maliciosos e mal-funcionamento; é importante assim que disponhamos de técnicas que permitam a um dispositivo avaliar se outro com o qual interage está em pleno funcionamento.

Pode-se tentar resolver o problema a partir do estabelecimento de algum tipo de **confiança** entre as entidades — sejam humanas ou não. Não há consenso na literatura sobre a definição deste conceito; duas definições comumente utilizadas (8) são a de confiança para segurança (*reliability trust*) — que é dita a probabilidade subjetiva que um agente atribui à execução por outro agente de uma certa ação de que seu bem-estar depende; e de confiança para decisão (*decision trust*), que mede a disposição de uma entidade em depender de outra em uma dada situação com relativa segurança. São ambas definições vagas, mas também gerais.

Distinguem-se, usualmente (9), alguns tipos ou classes de confiança:

- confiança de provisão (*provision trust*): descreve a confiança de uma entidade em um serviço ou provedor de recursos.

- confiança de acesso (*access trust*): descreve a confiança do agente em outros que acessarão recursos seus ou sob sua supervisão; se faz importante em esquemas de controle de acesso, bastante utilizados em computação.
- confiança de delegação (*delegation trust*): descreve a confiança de uma entidade em outras que agirão e tomarão decisões em seu lugar.
- confiança de identidade (*identity trust*): a confiança em um agente ser de fato quem afirma ser; para este fim são geralmente utilizados esquemas de autenticação
- confiança de contexto (*context trust*): dá uma medida de quanto uma entidade acredita estarem em ordem os sistemas e instituições necessários para que a transação seja realizada com sucesso.

Em um dado cenário, vários dos tipos de confiança são por vezes utilizados. Por exemplo, em processos em que se faz necessária a confiança de provisão, também se costuma utilizar a confiança de identidade. A esta combinação, dá-se o nome de propósito de confiança (*trust purpose*).

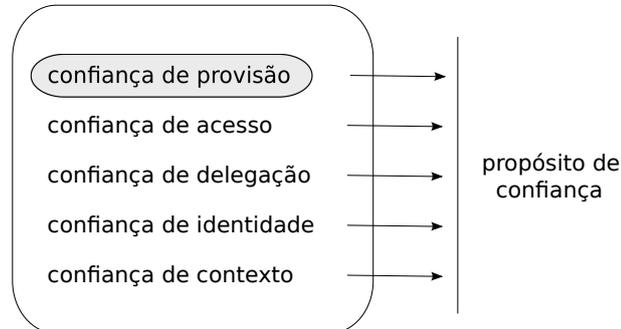


Figura 2.1: Classes de confiança de acordo com (9); a reputação, objeto de nosso estudo, está ligada à confiança de provisão. Figura baseada em (8).

Um conceito que por vezes se confunde ao de confiança é o de **reputação**. De acordo com o dicionário Houaiss, reputação é “o conceito de que alguém ou algo goza num grupo”. Reputação e confiança estão de certo associados, mas um não implica em outro: um indivíduo pode confiar em outro *devido* à sua boa reputação, ou *apesar* de sua má reputação.

A reputação é, no entanto, um conceito menos vago e mais mensurável do que a confiança. Uma boa maneira de inferir que entidades num certo grupo são confiáveis — sobretudo em relação à confiança de provisão — parece ser, assim, identificar aquelas que possuem uma boa reputação. Mecanismos responsáveis por tal tarefa são comumente denominados **sistemas de reputação**.

2.1 Sistemas de reputação — definição e exemplos

Identifica-se em geral um sistema de reputação como composto por cinco diferentes componentes (10): uma para a captação de informações, uma para a emissão de pontuações e classificações, uma para a seleção de com que entidades interagir, uma para que se efetuem de fato as transações e outra ainda para recompensar e punir os agentes dependendo de sua classificação.

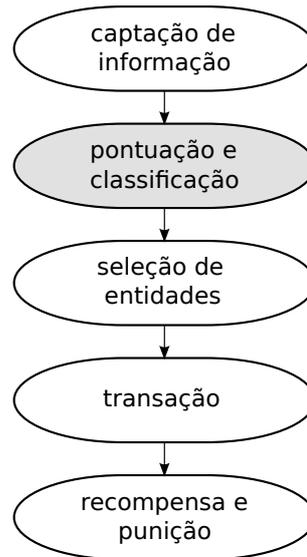


Figura 2.2: Componentes de um sistema de reputação, de acordo com (10); o foco de nosso estudo é a 2ª componente, responsável pela emissão de pontuações e classificações.

É importante diferenciar entre dois tipos de sistemas (8): os centralizados e os distribuídos. Nos centralizados, há uma autoridade central que recebe as avaliações emitidas pelos agentes, computa as pontuações e classificações e as transmite para os agentes; nos distribuídos, por outro lado, tal autoridade não está presente, e são os próprios agentes que, trocando informações entre si, obtêm amostras de avaliações emitidas e a partir destas calculam os índices. Sistemas distribuídos são importantes em contextos como o de redes de sensores sem fio (11, 12, 13).

Pode-se perceber que para ambos, no entanto, é de extrema importância a 2ª componente mencionada — responsável pelo cálculo de pontuações e classificações. O estudo desta componente é o foco de nossa dissertação; gostaríamos de, a partir das informações obtidas, calcular índices que apontem quais dos agentes são de fato bem ou mal-intencionados — um problema que é, basicamente, de **inferência**.

Desde o início da década de 90, quando os primeiros sistemas de reputações foram propostos (14), muitos outros surgiram em diferentes contextos. Na academia, passou a haver uma grande quantidade de propostas a partir dos anos 2000 que procuram em geral aplicar o método a problemas específicos, delineando o processo a ser executado de maneira *ad-hoc*. Só mais re-

centemente, no entanto, passou a haver tentativas de impor alguma padronização às propostas. (10).

Na prática, estes esquemas passaram a ser aplicados a uma grande gama de situações. Na internet, são utilizados por *websites* de *e-commerce* e mídia social; em foruns de discussão; na melhoria da segurança de redes *peer-to-peer*, como aquelas de compartilhamento de arquivos; e até mesmo por mecanismos de busca. São também muito aplicados à redes de computadores e dispositivos eletrônicos como sensores sem fio (12, 11). Os esquemas utilizados costumam ser, no entanto, mais simples do que aqueles propostos na literatura, talvez para que se mantenham compreensíveis ao usuário comum.

De acordo com as diferentes características que os sistemas propostos e utilizados apresentam, é possível classificá-los (15) em termos de

- modelo conceitual: cognitivo (relacionados à confiança para decisão) ou ligado à teoria de jogos (relacionados à confiança para segurança).
- fontes de informação: proveniente de contato direto, de terceiros, deduzidas ou preconcebidas.
- visibilidade: os índices de reputação são globais, isto é, cada agente tem acesso aos mesmos índices; ou subjetivos, e cada agente atribui índices diferentes para uma mesma entidade
- granularidade: os índices são atribuídos pensando em um contexto específico, ou em vários contextos diferentes
- hipóteses de comportamento: considera ou não que os agentes podem ser tendenciosos e/ou trapacear
- tipo de informação trocada: valores binários ou contínuos

Há ainda algumas técnicas típicas para o cálculo dos índices de reputação (8), tal como: de soma ou média simples; bayesianos, que empregam modelagem estatística; baseados em crenças, métodos similares aos probabilísticos mas em que não há normalização; *fuzzy*, que emprega a lógica de mesmo nome; e modelos de fluxo, em que índices são computados através de iterações transitivas em grafos.

Um dos primeiros modelos para sistema de reputação foi introduzido por Marsh em 1994 (14). O modelo diferencia entre uma confiança *geral*, $T_x^t(y)$, e outra *situacional*, $T_x^t(y, \alpha)$, que

o agente x possui no agente y num dado tempo t (e para um dada situação α). Para o cálculo da confiança situacional, utiliza-se a relação

$$T_x^t(y, \alpha) = U_x^t(\alpha) I_x^t(\alpha) \widehat{T}_x^t(y), \quad (2.1)$$

na qual $U_x^t(\alpha)$ é a utilidade de x com relação a α , $I_x^t(\alpha)$ a importância de α para x , e $\widehat{T}_x^t(y)$ uma estimativa para a confiança geral, calculada levando-se em conta $T_x^\tau(y, \sigma)$ para situações semelhantes σ e tempos passados $\tau < t$; é proposto tomar-se a média, o máximo ou o mínimo de tais valores.

Utiliza-se posteriormente o valor de $T_x^t(y, \alpha)$ para que x decida, de acordo com um dado critério, se deve ou não cooperar com y . De acordo com a classificação mencionada, este modelo é conceitualmente ligado à teoria de jogos; leva em conta informações provenientes de contato direto, apenas; e fornece índices de reputação subjetivos, atribuídos pensando-se em um contexto específico.

Das propostas mais recentes, vale mencionar aquela por Schillo et al. (2000) (16), que abordam o problema como um de inferência — de maneira semelhante a que faremos em seguida; e também o sistema de reputação Regret, introduzido por Sabater e Sierra em 2001 (17) e bastante representativo do nicho de sistemas multi-agentes.

Um sistema que obteve grande sucesso em aplicações é o PageRank (18). O modelo é formulado pensando-se em páginas da *web* como entidades: sejam u e v páginas contidas num conjunto de páginas P , e $N^-(u)$ o número de páginas que apontam para u e $N^+(v)$ o número de páginas para que v aponta; calcula-se o índice de reputação para u (seu *PageRank*) a partir de

$$R(u) \cong E(u) + \sum_{v \in N^-(u)} \frac{R(v)}{|N^+(v)|}, \quad (2.2)$$

onde E é um vetor sobre P que fornece um valor de reputação intrínseco às páginas (o símbolo \cong indica aqui igualdade a menos de um fator de normalização). Neste contexto, o esquema foi e continua sendo muito utilizado pela Google em seu mecanismo de buscas; pode-se considerar ainda outras entidades — o método já foi utilizado por exemplo em estudos cientométricos, para avaliar a reputação de artigos científicos (19).

São apontados alguns problemas de que costumam sofrer as propostas usuais para sistemas de reputação, tais como a importância indevida dada a avaliações emitidas de maneira injusta e, sobretudo, a possibilidade de que ocorram fraudes no esquema de votação — por exemplo, através de trocas de identidade, permitindo que uma mesma entidade se passe por outras e emita

avaliações em seu lugar (comumente denominados de ataques Sybil na literatura); ou por meio de conluio, havendo um grupo de agentes que emitem opiniões de forma a propositalmente impedir o cálculo correto de reputações.

2.2 Modelo para inferência

A fim de calcular reputações de maneira ótima, baseados em (20, 21), formularemos agora um modelo que procura reproduzir a relação existente entre as opiniões formadas pelos diversos agentes e suas reputações.

Na discussão que segue, consideraremos que um conjunto de opiniões $\{J_{ij}\}$ foi amostrado, seja por uma autoridade central ou por um agente qualquer; aqui J_{ij} representa a avaliação emitida por i sobre j tal como recebida, constituindo a matriz J de dimensões $n \times n$. Um vetor $\mathbf{r} = (r_1, r_2, \dots, r_n)$ dá as reputações de fato dos agentes, de acordo com alguma medida a ser definida.

Podemos afirmar sem perda de generalidade que

$$J = f(\mathbf{r}, \{\xi\}), \quad (2.3)$$

onde $\{\xi\}$ é um conjunto de variáveis aleatórias que representam externalidades diversas. Um modelo específico dependerá de uma forma para f , da especificação dos domínios de J_{ij} e r_i , e das probabilidades de ξ assumirem diferentes valores. Queremos que estes parâmetros sejam determinados de forma a modelar situações reais da maneira mais geral possível, mas também que sejam simples o bastante a ponto de permitir que estudemos a distribuição posterior $P(\mathbf{r}|J)$ analiticamente em detalhes. Levando isto em conta, uma boa escolha parece ser

$$J_{ij} = \begin{cases} \xi_{ij} r_i r_j & \text{se } (i, j) \in \Omega, \\ 0 & \text{do contrário,} \end{cases} \quad (2.4)$$

com $r_i = \pm 1$ representando reputações boas (+1) ou ruins (-1), $J_{ij} = \pm 1$ avaliações positivas ou negativas e $\xi_{ij} \sim \text{Bimod}_{\pm 1}(p)$ um ruído na formação ou transmissão de opiniões, onde

$$\text{Bimod}_{\pm 1}(p) = \begin{cases} +1 & \text{com prob. } p, \\ -1 & \text{com prob. } 1 - p, \end{cases} \quad (2.5)$$

e o conjunto Ω é dado pelos pares de agentes cujas opiniões foram amostradas.

Em princípio, um agente i a julgar um outro agente j de reputação boa ($r_j = +1$) deve se sentir inclinado a formar uma opinião boa do mesmo ($J_{ij} = +1$). Em situações reais, no entanto, isto não é o que acontece; como descrito na seção 2.1, por vezes há conluio: um grupo de agentes de reputação ruim se infiltra no sistema emitindo avaliações positivas uns sobre os outros, de forma a tentar alavancar suas reputações. A forma proposta para f busca evitar este tipo de ataque, admitindo que isto ocorra; o produto $r_i r_j$ leva a uma avaliação boa mesmo no caso em que $r_i = r_j = -1$. Admite-se ainda que, com uma probabilidade $1 - p$, agentes podem se enganar ao emitir opiniões ou falhas podem ocorrer no mecanismo de transmissão.

Para estudar o modelo proposto, determinemos $P(\mathbf{r}|J)$. Admitiremos que as opiniões J_{ij} são emitidas de maneira independente umas em relação às outras — o que pode não ocorrer em muitos dos casos de interesse prático, mas parece funcionar bem em primeira aproximação — e que as reputações r_i são também independentes entre si. Temos então do teorema de Bayes que

$$\begin{aligned} P(\mathbf{r}|J) &\propto P(J|\mathbf{r}) P_0(\mathbf{r}) \\ &\propto \prod_{(i,j) \in \Omega} P(\xi_{ij}) \prod_i P_0(r_i). \end{aligned} \quad (2.6)$$

Vamos supor uma distribuição *a priori* para \mathbf{r} de forma a ter $P_0(r_i = +1) = q$, ou seja, $r_i \sim \text{Bimod}_{\pm 1}(q)$. Levemos em conta ainda que se $P(x) = \text{Bimod}_{\pm 1}(p)$, pode-se reescrever $P(x)$ como

$$P(x) = p^{\frac{1+x}{2}} (1-p)^{\frac{1-x}{2}} = c_p \exp(\alpha_p x), \quad (2.7)$$

com $c_p = \sqrt{p(1-p)}$ e $\alpha_p = \frac{1}{2} \log \frac{p}{1-p}$. Prosseguindo com o cálculo, obtemos

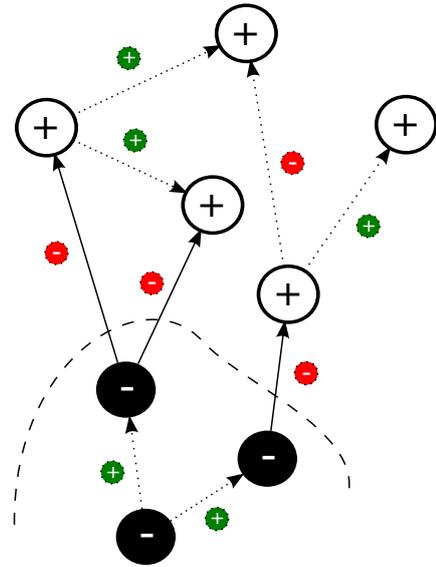


Figura 2.3: Um dos cenários possíveis que procurou-se modelar; agentes de má reputação emitem avaliações positivas uns aos outros, de forma a tentar enganar o mecanismo. Agentes de boa reputação são aqui indicados por círculos brancos, e os de má reputação por círculos pretos; as setas indicam avaliações emitidas — acompanhadas de um círculo vermelho, caso negativas, e de um círculo verde, caso positivas.

$$\begin{aligned}
P(\mathbf{r}|J) &\propto \prod_{(i,j) \in \Omega} \exp(\alpha_p \xi_{ij}) \prod_i \exp(\alpha_q r_i) \\
&\propto \exp\left(\alpha_p \sum_{(i,j) \in \Omega} \xi_{ij} + \alpha_q \sum_i r_i\right) \tag{2.8}
\end{aligned}$$

$$\propto \exp\left[\alpha_p \left(\sum_{(i,j) \in \Omega} J_{ij} r_i r_j + \frac{\alpha_q}{\alpha_p} \sum_i r_i\right)\right]. \tag{2.9}$$

Trabalharemos sobretudo com valores de $p > 0.5$, situação em que o ruído não é prevalente; e de $q > 0.5$, assumindo que a maioria das entidades é confiável (pode-se mostrar, no entanto, que o caso de $q < 0.5$ é simétrico a este).

Queremos fazer uma estimativa $\hat{\mathbf{r}}$ para \mathbf{r} sendo consistentes com a amostra de opiniões obtida e de maneira a minimizar o erro cometido¹ $\varepsilon = h(\mathbf{r}, \hat{\mathbf{r}}) = \frac{1}{2} \left(1 - \frac{\mathbf{r} \cdot \hat{\mathbf{r}}}{n}\right)$; delineado um algoritmo para este cálculo, pode-se utilizá-lo como base para a formulação de sistemas de reputação.

A fim de exemplificar o tipo de análise a ser feita, utilizemo-nos primeiro de um método simples para a obtenção de estimativas; mais precisamente, calculemos as componentes de $\hat{\mathbf{r}}$ a partir de

$$\hat{r}_i = \operatorname{sgn}\left(\sum_{k \in \partial i} J_{ki}\right), \tag{2.10}$$

onde ∂i indica o conjunto de vizinhos de i no grafo — mais precisamente, neste caso, os agentes que avaliaram ou foram avaliados por i . Podemos simular computacionalmente um cenário em que n agentes dispostos sobre um grafo da topologia desejada emitem opiniões uns sobre os outros de acordo com o modelo formulado (eq. 2.4); calculamos, em seguida, as estimativas da maneira proposta e comparamos os valores reais e estimados das reputações, dado que na simulação dispomos de ambos. Geramos em seguida um grande número de cenários — variando em cada um deles as reputações, conexões formadas e avaliações emitidas — e calculamos médias das grandezas de interesse.

Por outro lado, é possível estudar mais a fundo, de forma analítica, as implicações de utilizar tal método. Uma vez que, pelo modelo, $J_{ki} = \xi_{ki} r_k r_i$, temos que $\hat{r}_i = \operatorname{sgn}[(\sum_{k \in \partial i} \xi_{ki} r_k) r_i]$; como se admitem conhecidas as distribuições de ξ e (a priori) de r , e que tais variáveis são

¹ $h(\mathbf{r}, \hat{\mathbf{r}})$ aqui é a distância de Hamming entre \mathbf{r} e $\hat{\mathbf{r}}$.

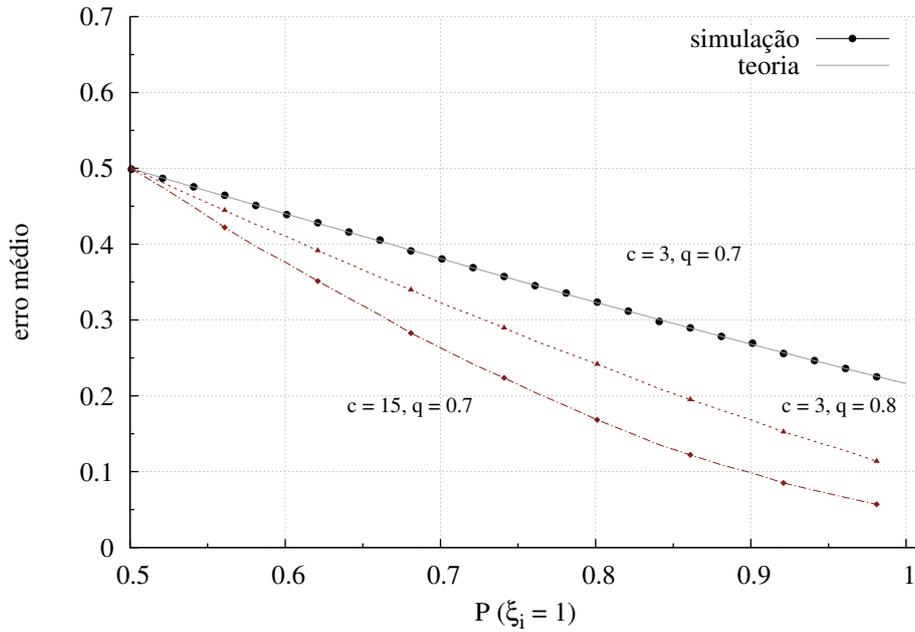


Figura 2.4: Erro médio na obtenção de estimativas com o uso do método simples delineado, para $q \equiv P_0(r_i = 1) = 0.7$ e $c = 3$ (as curvas pontilhadas dão o erro para outros valores dos parâmetros); o resultado da simulação com $n = 100$ é indicado por pontos, e aquele fornecido pela teoria por uma linha contínua. Pode-se observar que o erro é, de um modo geral, grande. Para a simulação, foram gerados cerca de 3000 cenários para cada valor de p , sobre os quais calculou-se o erro médio; o tamanho das barras de incerteza é desprezível.

independentes e identicamente distribuídas, obtemos²

$$\hat{r}_i = \text{sgn} \left[\left(\sum_{k=1}^{c_i} \xi_k r_k \right) r_i \right], \quad (2.11)$$

onde c_i é amostrado da distribuição de graus do grafo em questão; assumimos, na análise que segue, $c_i = 3 \forall i$ — mais precisamente, que os agentes estão dispostos numa rede aleatória regular³ de grau 3. A probabilidade de que \hat{r}_i seja igual a r_i é, portanto, $P(\sum_{k=1}^c \xi_k r_k > 0)$. Como $P(\xi_i = 1) = p$ e $P_0(r_i = 1) = q$, temos que $P(\xi_i r_i = 1) = pq + (1-p)(1-q) \equiv \varpi$, e assim

$$P(\hat{r}_i = r_i) = P\left(\sum_{k=1}^c \xi_k r_k > 0\right) = \sum_{n=0}^{\lfloor \frac{c}{2} \rfloor} \binom{c}{n} \varpi^{c-n} (1-\varpi)^n. \quad (2.12)$$

Pode-se ver na figura 2.4 o erro médio $\bar{\varepsilon} = 1 - P(\hat{r}_i = r_i)$ em função de p para $q = 0.7$, que é de um modo geral grande. Para $p = 1$, não há ruído algum na transmissão ou formação de opiniões, e há no entanto uma probabilidade de aproximadamente 20% de se obter uma

²Tratamos aqui ξ_k e r_k como cópias de variáveis aleatórias ξ e r , amostradas respectivamente de $P(\xi)$ e $P_0(r)$.

³Sobre grafos aleatórios regulares, ver por exemplo (22, 23).

estimativa incorreta; vale notar ainda que, para $q = 0.7$, estimativas atribuída com probabilidade uniforme trariam uma probabilidade de erro de 30%. Isto mostra que tal método é um tanto quanto inadequado para a obtenção de estimativas, ao menos no caso em que avaliações são emitidas de forma semelhante a que admitimos em nosso modelo.

Talvez por sua simplicidade, este é um método adotado com alguma frequência, por exemplo por sistemas de *e-commerce*; um dos *websites* mais influentes do meio, o eBay (24) utiliza sistema bastante semelhante — calcula não um índice binário de reputação, mas apenas a soma de todas as opiniões emitidas sobre um certo usuário.

Um dos motivos pelos quais se obteve um erro grande é que, apesar de possuímos um modelo de como as avaliações são formadas, não o levamos em conta. Introduziremos, na próxima seção, a técnica de **propagação de crenças**, que nos auxiliará a obter estimativas de modo mais compatível com o desejado. Por ser um método mais elaborado, uma análise teórica não se dá de maneira trivial como a feita acima; seremos capazes de fazê-lo com o uso de ferramentas da mecânica estatística (cap. 3).

2.3 Técnica de propagação de crenças

Seja $P(\underline{x})$ uma distribuição de probabilidades definida sobre um conjunto de variáveis aleatórias discretas $\underline{x} = \{x_1, x_2, \dots, x_N\}$. Em problemas de inferência estatística, é por vezes de interesse, dado um subconjunto $\underline{x}_S \in \underline{x}$, calcular somas do tipo⁴

$$\mu(\underline{x}_S) = \sum_{x_i \in \underline{x}/\underline{x}_S} P(\underline{x}), \quad (2.13)$$

levando a distribuições marginais ou fatores de normalização (no caso em que $\underline{x}_S = \{\}$). Esta tarefa é especialmente demorada para valores grandes de $M = N - |\underline{x}_S|$, dado que se executa em $O(|\mathcal{X}|^M)$ — \mathcal{X} aqui é o alfabeto de \underline{x} , isto é, o conjunto de possíveis valores que as variáveis x_i assumem.

Em alguns casos é possível, no entanto, fatorar esta soma, facilitando o seu cálculo. Se, por exemplo, uma distribuição $P(x_1, x_2, x_3)$ puder ser escrita como $P(x_1, x_2, x_3) = f(x_1, x_3)g(x_2, x_3)$, pode-se fazer o cálculo a partir de

$$\mu(x_3) = \sum_{x_1, x_2} P(x_1, x_2, x_3) = \left(\sum_{x_1} f(x_1, x_3) \right) \cdot \left(\sum_{x_2} g(x_2, x_3) \right), \quad (2.14)$$

⁴Aqui, $\underline{x}/\underline{x}_S$ denota o complemento de \underline{x}_S em \underline{x} .

ou seja, a partir de $2|\chi| + 1$ operações, em vez de $|\chi|^2$ — de um modo geral, a partir de $O(M)$ operações em vez de $O(|\chi|^M)$, caso a soma possa ser fatorada em somas sobre apenas uma variável.

Esta fatoração é de certa forma trivial; a percepção de que é utilizada por uma série de cálculos é todavia recente e fornece um maior entendimento (25). Formalizaremos, a seguir, esta técnica sob o nome de **propagação de crenças**.

2.3.1 Campos de Markov e grafos de fatores

Para que seja possível fatorar as somas, as expressões para as distribuições devem ser da forma

$$P(\underline{x}) = \frac{1}{Z} \prod_{a=1}^M \psi_a(\underline{x}_{\partial a}), \quad (2.15)$$

dizendo-se, neste caso, que as variáveis pertencentes a \underline{x} formam um **campo de Markov**. Neste contexto, ψ_a são denominadas *funções de compatibilidade*.

Uma boa representação para tal estrutura é fornecida por **grafos de fatores**; trata-se de um multigrafo composto por dois tipos de nós: nós-variável (\circ) $i \in V, |V| = N$, relacionados as x_i , e nós-função (\blacksquare) $a \in F, |F| = M$, relacionados as ψ_a . Nós-função se ligam apenas a nós-variável e vice-versa, de forma que o nó-função referente a ψ_a se liga aos nós-variável referentes à seus argumentos, $\underline{x}_{\partial a}$. Um grafo de fatores G é completamente definido pelos conjuntos de nós, V e F , e pelo conjunto de arestas, E .

Uma distribuição do tipo $P(x_1, x_2, x_3, x_4) = \psi_1(x_1, x_2) \psi_2(x_1, x_3) \psi_3(x_2, x_4) \psi_4(x_3, x_4)$, por exemplo, é representada pelo grafo de fatores da fig. 2.5.

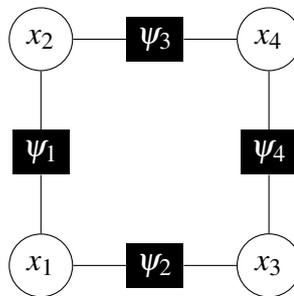


Figura 2.5: Grafo de fatores para uma distribuição fatorável em pares, $P(x_1, x_2, x_3, x_4) = \psi_1(x_1, x_2) \psi_2(x_1, x_3) \psi_3(x_2, x_4) \psi_4(x_3, x_4)$

2.3.2 Equações de propagação de crenças

Definiremos, sobre o grafo de fatores de um campo de Markov qualquer, um esquema de transmissão de mensagens: um nó-variável i envia, no tempo t , mensagens $v_{i \rightarrow a}^{(t)}(x_i)$ para os nós-função em sua vizinhança, $a \in \partial i$; e, analogamente, um nó-função envia mensagens $\hat{v}_{a \rightarrow i}^{(t)}(x_i)$ para $i \in \partial a$.

Queremos utilizar uma regra para a transmissão que faça com que as mensagens convirjam após um dado número de iterações para $\{v_{i \rightarrow a}^*\}$, $\{\hat{v}_{a \rightarrow i}^*\}$, e que permita o cálculo das distribuições marginais após a convergência, por exemplo a partir de ⁵

$$\mu_i(x_i) \cong \prod_{a \in \partial i} \hat{v}_{a \rightarrow i}^*(x_i), \quad (2.16)$$

ou seja, o produto de todas as mensagens enviadas ao nó-variável pelos nós-função vizinhos. Se considerarmos que o grafo de fatores com que estamos trabalhando é uma árvore, isto é, não possui ciclos, isto pode ser feito através do seguinte conjunto de equações

$$\begin{aligned} v_{i \rightarrow a}^{(t+1)}(x_i) &\cong \prod_{b \in \partial i/a} \hat{v}_{b \rightarrow i}^{(t)}(x_i) \\ \hat{v}_{a \rightarrow i}^{(t)}(x_i) &\cong \sum_{\underline{x}_{\partial a/i}} \psi_a(\underline{x}_{\partial a}) \prod_{j \in \partial a/i} v_{j \rightarrow a}^{(t)}(x_j), \end{aligned} \quad (2.17)$$

o qual denominaremos de *equações de propagação de crenças*. Percebe-se que as mensagens enviadas combinam as mensagens recebidas de todos os vizinhos exceto o destinatário; os nós-função fazem um tipo de processamento, combinando as mensagens ao valor da função de compatibilidade naquele ponto.

É possível obter boas aproximações para as marginais mesmo no caso em que o grafo não é uma árvore, mas possui ciclos de longo comprimento médio.

Este esquema permite ainda que computemos as marginais locais — distribuições conjuntas de duas ou mais variáveis, que representaremos por nós-variável V_R ligados entre si por nós-fator F_R — através de

$$\mu(\underline{x}_R) \cong \prod_{a \in F_R} \psi_a(\underline{x}_a) \prod_{a \in \partial R} \hat{v}_{a \rightarrow i(a)}^*(x_{i(a)}), \quad (2.18)$$

⁵O termo \cong é utilizado ao longo desta dissertação para indicar a igualdade a menos de uma constante de normalização.

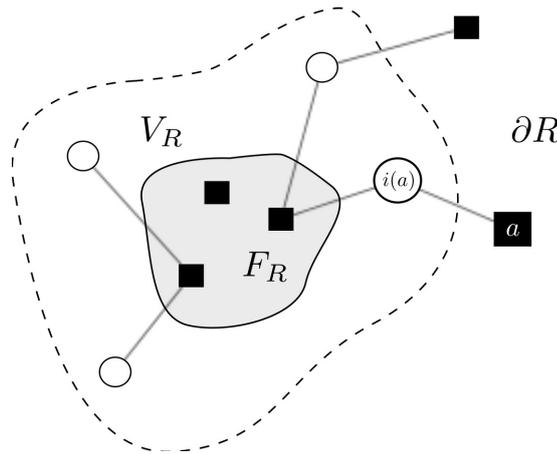


Figura 2.6: Indicação das variáveis de que depende o cálculo das marginais locais (eq. 2.18). Aqui, ∂R é o conjunto de nós-fator que se conectam à V_R mas não estão em F_R , e $i(a)$ é o nó-variável que liga o nó-fator a a um nó-fator em F_R (há apenas um destes, dado que se houvessem dois ou mais, a estaria em F_R)

onde ∂R é o conjunto de nós-fator que se conectam à V_R mas não estão em F_R , e $i(a)$ é o nó-variável que liga o nó-fator a a um nó-fator em F_R (há apenas um destes, dado que se houvessem dois ou mais, a estaria em F_R).

2.3.3 Aplicações em mecânica estatística

Um bom exemplo de aplicação do esquema citado acima vem da física estatística. Consideremos uma hamiltoniana para um sistema magnético do tipo

$$H(\underline{s}) = - \sum_{(i,j) \in E} J_{ij} s_i s_j - \sum_i B_i s_i, \quad (2.19)$$

com $s_i = \pm 1$ e $E, \{J_{ij}\}, \{B_i\}$ quaisquer⁶ — uma generalização do modelo de Ising. Os estados de equilíbrio, a uma dada temperatura β^{-1} , são distribuídos de acordo com

$$P(\underline{s}) = \frac{1}{Z} \exp(-\beta H(\underline{s})) = \exp \left[\beta \left(\sum_{(i,j) \in E} J_{ij} s_i s_j + \sum_i B_i s_i \right) \right], \quad (2.20)$$

que pode ser fatorada da seguinte maneira,

$$P(\underline{s}) = \frac{1}{Z} \prod_{(i,j) \in E} \exp(\beta J_{ij} s_i s_j) \prod_i \exp(\beta B_i s_i), \quad (2.21)$$

⁶Em princípio supõe-se, contudo, que a matriz J é simétrica.

com estrutura semelhante à mencionada anteriormente. De fato, pode-se neste caso construir um grafo de fatores com apenas dois tipos de nós-função: um referente às funções $\psi_{ij}(s_i, s_j) = \exp(\beta J_{ij} s_i s_j)$ (■), e outro às $\phi_i(s_i) = \exp(\beta B_i s_i)$ (□). Exemplos de construções para uma rede quadrada e uma rede totalmente conexa estão na figura 2.7

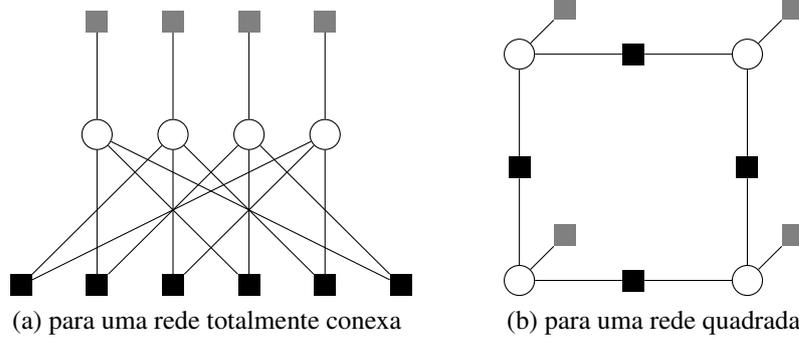


Figura 2.7: Grafos de fatores para a distribuição em questão, sobre diferentes redes; os nós-variável são indicados por \circ , e os dois tipos de nós função por ■ e □.

Indexemos o nó-fator ■ que liga os nós-variável i e j por (ij) , e a mensagem enviada pelo mesmo para i por $\hat{v}_{(ij) \rightarrow i}$; analogamente, o nó-fator ■ que se conecta apenas ao nó-variável i é indexado por i , e a mensagem enviada pelo mesmo por \hat{v}_i . O esquema de transmissão de mensagens para esta distribuição fica assim

$$\begin{aligned}
 \mathbf{v}_{i \rightarrow (ij)}^{(t+1)}(s_i) &\cong \hat{v}_i(s_i) \prod_{k \in \partial i / j} \hat{v}_{(ki) \rightarrow i}^{(t)}(s_i) \\
 \hat{v}_{(ji) \rightarrow i}^{(t)}(s_i) &\cong \sum_{s_j} \exp(\beta J_{ij} s_i s_j) \mathbf{v}_{j \rightarrow (ji)}^{(t)}(s_j) \\
 \hat{v}_i(s_i) &\cong \exp(\beta B_i s_i),
 \end{aligned} \tag{2.22}$$

ou, substituindo os termos devidamente e abreviando a notação de $i \rightarrow (ij)$ para $i \rightarrow j$

$$\mathbf{v}_{i \rightarrow j}^{(t+1)}(s_i) \cong \exp(\beta B_i s_i) \prod_{k \in \partial i / j} \sum_{s_k} \exp(\beta J_{ik} s_i s_k) \mathbf{v}_{k \rightarrow i}^{(t)}(s_k), \tag{2.23}$$

que é a regra para a atualização de mensagens neste caso. Num período de tempo, propagar-se-ão $|\mathcal{X}| = 2$ mensagens sobre cada aresta $(i, j) \in E$ e em cada sentido, contendo informações sobre a probabilidade dos nós assumirem cada um dos valores do alfabeto. Caso as mensagens converjam após um número de iterações da regra, uma aproximação para a distribuição marginal da variável s_i pode ser obtida através da eq. 2.16, e ficamos com

$$\mu_i(s_i) \cong \exp(\beta B_i s_i) \prod_j \sum_{s_j} \exp(\beta J_{ij} s_i s_j) v_{j \rightarrow i}^*(s_j), \quad (2.24)$$

onde $\{v_{i \rightarrow j}^*\}$ é o conjunto de mensagens obtidas após a convergência. Vale notar que a convergência do mapa não é garantida para uma rede de topologia qualquer, e sim tão somente para árvores — caso também em que as aproximações obtidas são exatas. A partir da distribuição marginal, grandezas termodinâmicas podem ser calculadas com facilidade, como veremos adiante; a magnetização local, por exemplo, é dada por $m_i = \mu_i(1) - \mu_i(-1)$.

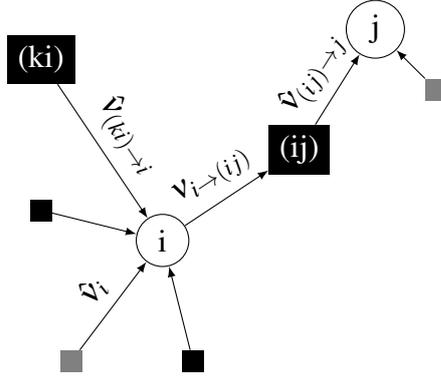


Figura 2.8: Esquema de transmissão de mensagens: o nó-variável i envia ao nó-fator (ij) mensagem gerada a partir do processamento de mensagens recebidas dos demais nós-fatores que o cercam; em seguida, o nó-fator (ij) processa esta mensagem e transmite a mensagem resultante ao nó-variável j . Posteriormente, todo o processo se repete no sentido inverso.

Sendo s_i uma variável binária e $v_{i \rightarrow j}(s_i)$ definida a menos de uma constante de normalização, pode-se fazer uma parametrização de forma a haver apenas 1 mensagem se propagando em cada sentido de aresta. Introduzamos a razão de log-verossimilhança,

$$h_{i \rightarrow j} \equiv \frac{1}{2\beta} \log \frac{v_{i \rightarrow j}(1)}{v_{i \rightarrow j}(-1)}, \quad (2.25)$$

definindo inversamente $v_{i \rightarrow j}(s_i) \cong \exp(\beta h_{i \rightarrow j} s_i) = \frac{1}{2} [1 + s_i \tanh(\beta h_{i \rightarrow j})]$; substituindo na eq. 2.23

$$v_{i \rightarrow j}^{(t+1)}(s_i) \cong \exp(\beta B_i s_i) \prod_{k \in \partial i/j} \cosh(\beta J_{ik} s_i) \left[1 + \tanh(\beta J_{ik} s_i) \tanh(\beta h_{k \rightarrow i}^{(t)}) \right], \quad (2.26)$$

e, usando a definição de $h_{i \rightarrow j}$

$$h_{i \rightarrow j}^{(t+1)} = B_i + \frac{1}{2\beta} \sum_{k \in \partial i/j} \log \left[\frac{1 + \tanh(\beta J_{ik}) \tanh(\beta h_{k \rightarrow i}^{(t)})}{1 - \tanh(\beta J_{ik}) \tanh(\beta h_{k \rightarrow i}^{(t)})} \right], \quad (2.27)$$

onde levou-se em conta que que a função $\cosh(x)$ é par em x e $\tanh(x)$, ímpar. Utilizando a identidade $\tanh^{-1}(x) = \frac{1}{2} \log \frac{1+x}{1-x}$, para $|x| < 1$, temos finalmente

$$\boxed{h_{i \rightarrow j}^{(t+1)} = B_i + \frac{1}{\beta} \sum_{k \in \partial i / j} \tanh^{-1} \left[\tanh(\beta J_{ik}) \tanh(\beta h_{k \rightarrow i}^{(t)}) \right]}, \quad (2.28)$$

A equação 2.28 fornece uma regra para a transmissão do novo conjunto de mensagens, $\{h_{i \rightarrow j}^{(t)}\}$. Iterando-a até a convergência, é possível obter as marginais a partir de $\mu_i(s_i) \cong \exp(\beta \hat{h}_i s_i)$, onde \hat{h}_i é dado por expressão semelhante a anterior, mas com soma efetuada sobre todos os vizinhos

$$\hat{h}_i = B_i + \frac{1}{\beta} \sum_{k \in \partial i} \tanh^{-1} [\tanh(\beta J_{ik}) \tanh(\beta h_{k \rightarrow i}^*)], \quad (2.29)$$

e a magnetização local vem de $m_i = \mu_i(1) - \mu_i(-1) = \tanh(\beta \hat{h}_i)$.

Iterando este mapa, obtemos $\{h_{i \rightarrow j}^*\}$ e $\{\hat{h}_i\}$ e, a partir destes, grandezas termodinâmicas podem ser calculadas, como descreveremos no capítulo 3. De fato, esta é uma forma alternativa e mais geral para a formulação de técnicas clássicas da mecânica estatística — mais precisamente, a aproximação de Bethe-Peierls e o método de cavidade; um estudo analítico mais aprofundado da equação 2.28 leva a uma *prescrição moderna* do último (seção 3.1).

2.4 Formulação do algoritmo

Queremos estudar a distribuição a posteriori encontrada (eq. 2.9) utilizando-nos do método descrito na seção anterior. A distribuição conjunta em questão possui formato bastante semelhante à do exemplo dado (eq. 2.20); fazendo $\beta_N = \alpha_p = \frac{1}{2} \log \frac{p}{1-p}$ e $B_N = \frac{\alpha_q}{\alpha_p} = \frac{\log \frac{q}{1-q}}{\log \frac{p}{1-p}}$, podemos escrevê-la como

$$P(\underline{s}|J) = \frac{1}{Z} \exp \left[\beta_N \left(\sum_{(i,j) \in \Omega} J_{ij} s_i s_j + B_N \sum_i s_i \right) \right], \quad (2.30)$$

em que passou-se a rotular o argumento como \underline{s} , a fim de explicitar que trabalhamos agora com um conjunto de variáveis dinâmicas que assumem valores quaisquer. Temos então na prática a mesma distribuição nos dois casos, com $B_i = B_N$ para todo i ; isto permite que utilizemos as equações derivadas no exemplo (2.28) para o cálculo das distribuições marginais.

Obtidas aproximações para as distribuições marginais, $\{\mu(r_i)\}$, adotaremos as estimativas seguindo

$$\hat{r}_i = \arg \max_{r_i} \mu(r_i). \quad (2.31)$$

Como mencionado anteriormente, a tal tarefa dá-se o nome de inferência MPM (do inglês, *maximum posterior marginal*). Pode-se resumir o esquema proposto da seguinte forma:

1. Numa rede com n agentes, é amostrado um conjunto de avaliações $\{J_{ij}\}$, com $J_{ij} = \pm 1$ representando uma opinião positiva ou negativa do agente i em relação ao agente j .
2. Inicializa-se um conjunto $\{h_{i \rightarrow j}\}$ amostrando valores de uma determinada distribuição $P^{(0)}(h)$; itera-se o conjunto de equações

$$h_{i \rightarrow j}^{(t+1)} = B_N + \frac{1}{\beta_N} \sum_{k \in \partial i/j} \tanh^{-1} \left[\tanh(\beta_N J_{ik}) \tanh(\beta_N h_{k \rightarrow i}^{(t)}) \right] \quad \forall i, j, \quad (2.32)$$

em um determinado número de passos t_{\max} ou até que se atinja a convergência, com $\max |h_{i \rightarrow j}^{(t+1)} - h_{i \rightarrow j}^{(t)}| < \varepsilon$. β_N e B_N são calculados a partir de estimativas obtidas para os parâmetros p e q — $\{\hat{p}, \hat{q}\}$ — através das expressões $\beta_N = \frac{1}{2} \log \frac{\hat{p}}{1-\hat{p}}$ e $B_N = \frac{\log \frac{\hat{q}}{1-\hat{q}}}{\log \frac{\hat{p}}{1-\hat{p}}}$.

3. Atingida a convergência, calcula-se os valores de $\{\hat{h}_i\}$ a partir de

$$\hat{h}_i = B_N + \frac{1}{\beta_N} \sum_{k \in \partial i} \tanh^{-1} \left[\tanh(\beta_N J_{ik}) \tanh(\beta_N h_{k \rightarrow i}^*) \right] \quad \forall i. \quad (2.33)$$

4. Como $\mu(r_i) \cong \exp(\beta_N \hat{h}_i r_i)$, seguindo 2.31 as estimativas vem de

$$\hat{r}_i = \operatorname{sgn}(\hat{h}_i) \quad \forall i. \quad (2.34)$$

Faz-se necessário, assim, especificar uma distribuição $P^{(0)}(h)$ e valores para t_{\max} e ε , além de obter estimativas para os parâmetros p e q .

Dado que trabalhamos com valores de $q > 0.5$, é razoável tomar $P^{(0)}(h)$ sobre valores de $h \geq 0$; a parte deste requisito, não há grande diferença em adotar uma ou outra distribuição — exceto em alguns casos atípicos, que serão melhor analisados no cap. 3. Valores para t_{\max} e ε dependem da situação específica e podem ser tomados com base no bom senso; valores de $t_{\max} \sim 100$ e $\varepsilon \sim 10^{-7}$ geram tipicamente bons resultados.

Já os parâmetros p e q estão relacionados, respectivamente, ao grau de ruído na formação e transmissão de opiniões, e na esperança depositada a priori sobre os agentes estarem em seu

pleno funcionamento; com base nesta descrição, técnicas para estimá-los podem ser utilizadas. Estudaremos, mais a frente, o comportamento do algoritmo para valores estimados, $\{\hat{p}, \hat{q}\}$, tanto iguais como diferentes dos reais, $\{p, q\}$.

Algoritmo 1 *Iteração numérica de mensagens* (distribuição $P^{(0)}(h)$, número máximo de iterações t_{\max} , precisão mínima ε , parâmetros estimados \hat{p} e \hat{q})

inicializa H a partir de amostras de $P^{(0)}(h)$

$\beta \leftarrow \beta_N(\hat{p}), B \leftarrow B_N(\hat{p}, \hat{q})$

while $\Delta \geq \varepsilon$ **and** $t < t_{\max}$ **do**

for $i = 1 \rightarrow n$ **do**

for $j \in \partial i$ **do**

$H'_{ij} \leftarrow f(H)$

end for

end for

$\Delta \leftarrow \max(|H' - H|)$

$t \leftarrow t + 1, H \leftarrow H'$

end while

if convergiu **then**

 calcula \hat{h}_i

$\hat{r}_i \leftarrow \text{sgn}(\hat{h}_i)$

end if

Podemos delinear o método, ainda, de forma algorítmica, como se pode ver sob o rótulo de algoritmo 1. Admite-se, na listagem, que H é uma matriz $n \times n$ cujo elemento H_{ij} dá o valor de $h_{i \rightarrow j}$ e que f é a forma para a iteração, dada pela equação 2.32.

Como mencionado, o algoritmo pode ser utilizado tanto por sistemas centralizados quanto distribuídos; podemos considerar que as amostras foram obtidas por uma autoridade central, ou por um agente qualquer. Caso venha a ser utilizado de forma distribuída, torna-se provavelmente necessário o uso de informações de terceiros; admitimos também esta possibilidade na formulação do algoritmo, de forma que nesta situação ξ pode vir a representar ruídos na transmissão de avaliações.

Um sistema de reputação que se utilize deste algoritmo pode ser classificado de acordo com as denominações sugeridas na seção 2.1; o mesmo seria de visibilidade global, levando em conta que agentes são tendenciosos e podem trapacear, e processando informações binárias. Além disso, podemos considerar o método um *modelo de fluxo*.

É importante observar que, ao longo da derivação da técnica de propagação de crenças, assume-se que a matriz J é simétrica, isto é, que $J_{ij} = J_{ji}$. Não há razão a priori para que isto ocorra nas amostras de opiniões obtidas; podemos no entanto, dada uma amostra $\{J_{ij}\}$, tomar $\tilde{J}_{ij} = \frac{J_{ij} + J_{ji}}{2}$, de modo a obter uma matriz \tilde{J} simétrica. Isto equivale em nosso caso a ignorar

opiniões dissonantes, o que pode levar a maus resultados caso estas sejam muitas; é sabido todavia que são poucas em vários dos casos de interesse, como por exemplo em sistemas de *e-commerce* (8).

2.4.1 Análise empírica

Analisemos o método proposto, primeiramente, de maneira empírica: através de simulações, de forma análoga à feita com o método mais simples delineado anteriormente.

Novamente, geramos um grande número de cenários (~ 5000), instanciando para cada um deles um novo grafo de interações e diferentes amostras de reputações; as avaliações eram em seguida geradas de acordo com o modelo proposto. Simulou-se a execução do algoritmo nos diferentes cenários, possibilitando que calculássemos médias sobre as quantidades de interesse.

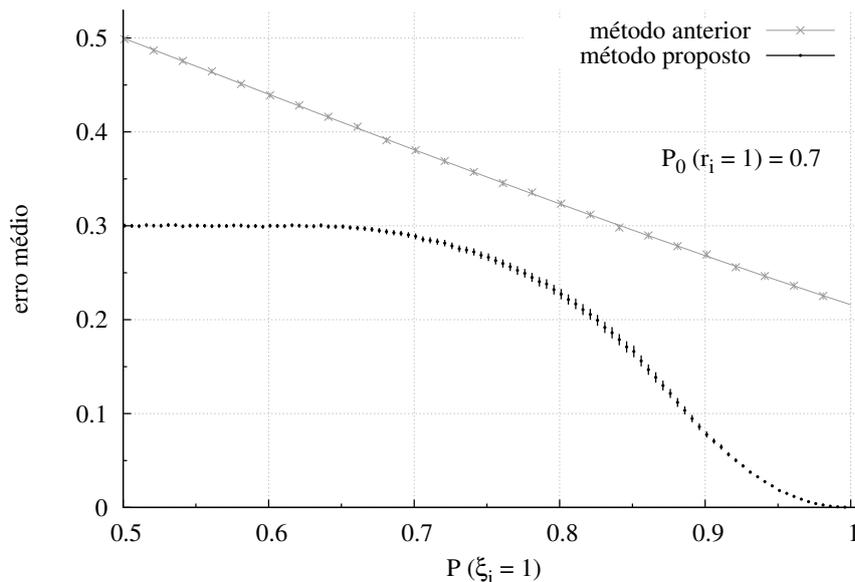


Figura 2.9: Erro médio obtido para estimativas usando o método mais simples delineado anteriormente (via simulação e cálculos analíticos), e o método proposto (via simulação, apenas), para $c = 3$ e $q = 0.7$; o segundo alcança erros menores.

Para uma primeira análise, tomemos como no caso anterior $n = 100$ agentes dispostos sob uma rede aleatória regular com $c = 3$. Como dito, o método de propagação de crenças fornece marginais exatas no caso em que o grafo de fatores em questão é uma árvore⁷; as aproximações para as marginais são também, no entanto, bastante boas para grafos de fatores com ciclos de longo comprimento médio, tal como grafos aleatórios regulares.

⁷Por só haverem nós-função de grau 1 e 2 no grafo de fatores, pode-se pensar no esquema de transmissão de mensagens como atuando no grafo de interações; a topologia deste e do grafo de fatores serão portanto a mesma.

Pode-se ver na figura 2.9 os resultados desta simulação, comparados àqueles obtidos com o método mencionado anteriormente, para valores diversos de $p \equiv P(\xi_{ij} = 1)$ e $q \equiv P_0(r_i = 1) = 0.7$. O erro médio obtido pelo novo método é, de um modo geral, menor; se mantém sempre abaixo de 30%, e para $p = 1$ é basicamente nulo.

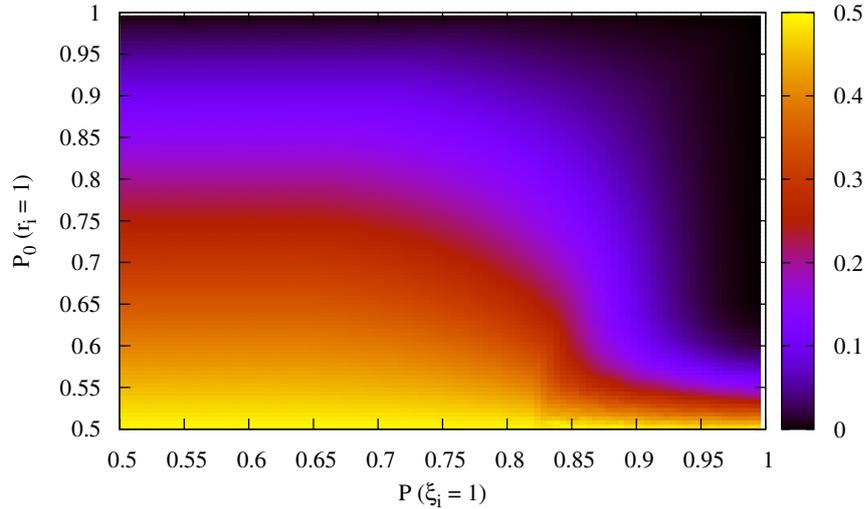


Figura 2.10: Erro médio $\bar{\epsilon} = \frac{1}{n} \langle |\hat{r}_i \neq r_i| \rangle$ no cálculo de estimativas pelo método proposto, obtido via simulação para vários valores de $p \equiv P(\xi_i = 1)$ e $q \equiv P_0(r_i = 1)$. As médias são calculadas a partir de cenários simulados diversos, cada qual contendo amostra de reputações e grafo distintos

Estudou-se o comportamento do erro médio, nesta situação, para valores diversos de p e q , tal como visualizado na fig. 2.10. Com a finalidade de analisar o algoritmo, podemos avaliar ainda outras quantidades, como a fração de vezes em que o esquema converge ou o número médio de iterações necessário para que isto ocorra (ver figura 2.11).

O erro é, como seria de se esperar, baixo para valores de p e q próximos de 1 — situação em que a maioria das entidades é confiável e as avaliações emitidas são fidedignas — aumentando a medida que os parâmetros se aproximam de 0.5. Já o estudo da convergência do algoritmo mostra que há uma região de parâmetros em que esta se dá de maneira lenta, por vezes não ocorrendo.

É interessante ainda que analisemos a sensibilidade às condições iniciais por parte do algoritmo; seriam, para diferentes distribuições $P^{(0)}(h)$, alcançados diferentes resultados? Utilizou-se, para este teste⁸, $h^{(0)} \sim \frac{1}{\beta_N} (\text{Beta}(a, b) - \frac{1}{2})$ com $b = \{1, 2, \dots, 7\}$ e $a = 2b$, de modo a manter a distribuição inclinada para a direita. Em cada cenário gerado, o esquema foi executado para os diferentes valores de b , gerando diferentes estimativas $\{s_i^{(b)}\}$; calculava-se então o *overlap* entre as estimativas, dado por $\Upsilon(\{s_i^{(b)}\}) \cong \sum_{m,n=1}^{b_{\max}} \sum_{i=1}^n s_i^{(m)} s_i^{(n)}$. Na figura 2.12, pode-se observar a

⁸A família de distribuições beta é dada por $\text{Beta}(a, b) \cong x^{a-1}(1-x)^{b-1}$.

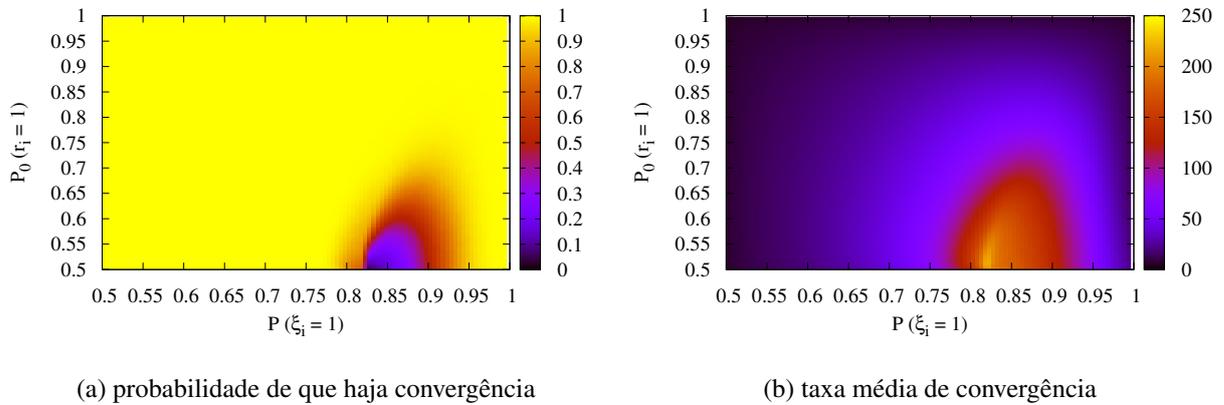


Figura 2.11: Análise de outras quantidades na execução do algoritmo, em função de p e q ; identifica-se uma região parâmetros em que o desempenho é baixo. O cálculo das grandezas é feito de maneira empírica, aproximando-as respectivamente (a) pela fração de cenários em que há convergência e (b) pelo número médio de iterações necessário para que a mesma ocorra.

fração de cenários para os quais $\Upsilon(\{s_i^{(b)}\}) = 1$, isto é, em que se obteve as mesmas estimativas para todos os valores de b .

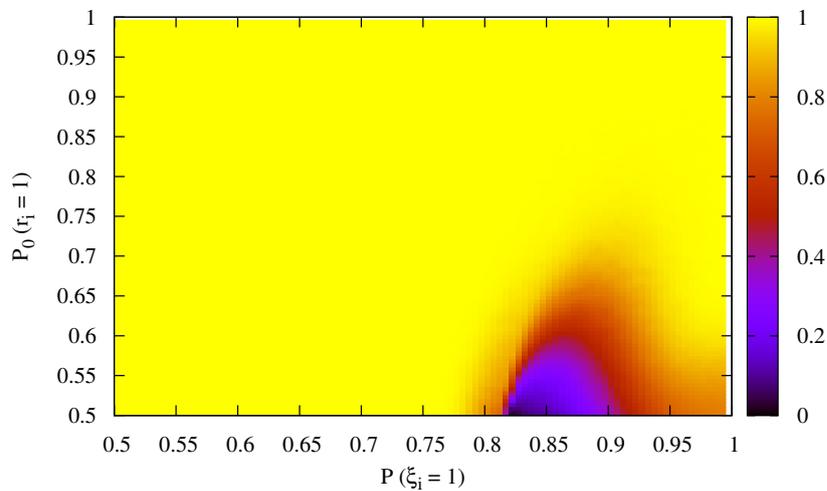


Figura 2.12: Fração de cenários em que obteve-se resultados similares dadas diferentes condições iniciais ($\Upsilon(\{s_i^{(b)}\}) = 1$); aqui também há uma região de parâmetros para a qual os resultados da execução são ruins, apresentando alta sensibilidade às condições iniciais.

Percebe-se assim que existe uma região de parâmetros na qual o desempenho do algoritmo é baixo — há dificuldades na convergência e alta sensibilidade às condições iniciais. Parece portanto interessante que investiguemos o porquê disto ocorrer nesta região específica, e não em uma outra qualquer; a análise do capítulo 3 permitirá que discorramos sobre o assunto.

Vale notar ainda que até agora em nossa análise assumimos fornecidos ao algoritmo os valores reais dos parâmetros p e q — de acordo com os quais as avaliações são geradas. Estas não são, no entanto, as condições sob as quais a execução se dará tipicamente; de um ponto de vista mais geral, são quatro os parâmetros presentes: $\{p, q\}$, especificados pela instância do problema; e $\{\hat{p}, \hat{q}\}$, estimados e fornecidos ao algoritmo.

Exemplos de como a análise pode ser estendida ao caso mais geral são dados pelas figuras 2.13 e 2.14: na primeira, fixamos $q = 0.7$ e analisamos erro e tempo de convergência médios para diferentes valores de p e \hat{p} ; em seguida, fixam-se $q = 0.7$ e $p = 0.7$, e estudamos o comportamento das mesmas quantidades em função de \hat{p} e \hat{q} .

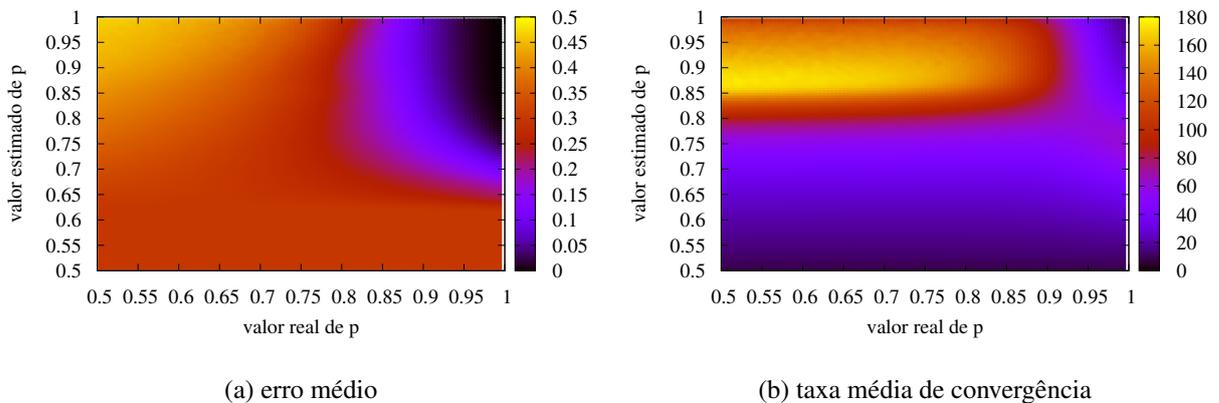


Figura 2.13: Análise para parâmetros estimados $\{\hat{p}, \hat{q}\}$ diferentes dos reais, $\{p, q\}$; fixou-se aqui $\hat{q} = q = 0.7$, e estudou-se as quantidades em função de p e \hat{p} . As quantidades em questão — o erro médio $\bar{\epsilon}$ e a taxa média de convergência (o número médio de iterações necessário para que a convergência ocorra) — servem como medidores para o desempenho do algoritmo sob as diferentes condições. Para cada valor de p , o erro mantém-se mínimo em $\hat{p} = p$; o estudo da taxa média de convergência indica região de baixo desempenho no canto superior esquerdo do gráfico.

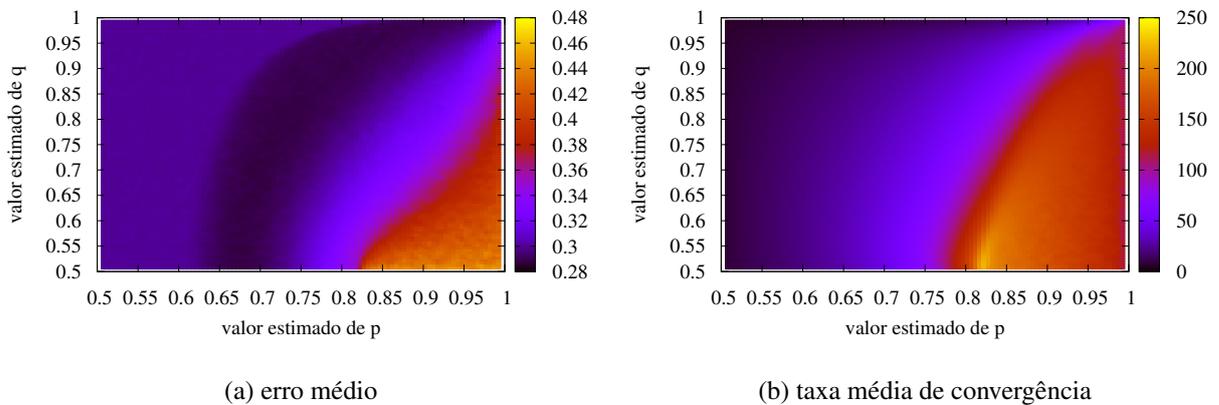


Figura 2.14: Fixando os parâmetros reais $p = q = 0.7$, pode-se estudar como as mesmas quantidades variam em função dos parâmetros estimados \hat{p} e \hat{q} . Como esperado, erros baixos ocorrem para valores de $\hat{p} \sim p$ e $\hat{q} \sim q$; pode-se, a partir da análise da taxa média de convergência, identificar uma região de baixo desempenho no canto direito do gráfico.

Após modelar o problema de estimar reputações como um de inferência, formulamos neste capítulo um método para a obtenção de estimativas e analisamos o mesmo de maneira empírica, por meio de simulações — executou-se o algoritmo para um grande número de cenários, estudando seu desempenho típico por meio de indicadores como o erro e a taxa de convergência.

Esta análise, enquanto útil, apresenta alguns problemas: é bastante custosa, e aumenta em pouco nossa compreensão do problema — não procura explicar, por exemplo, porque o desempenho é maior ou menor para uma dada região de parâmetros.

Técnicas de mecânica estatística permitem que analisemos o problema de forma analítica — e portanto, menos custosa — e que investiguemos mais a fundo sua estrutura. Este é o objeto de estudo de nosso próximo capítulo.

3 *Análise Teórica*

“You don’t understand anything until you learn it more than one way”

— Marvin Minsky

Consideremos um sistema físico a temperatura fixa β^{-1} , que assume diferentes configurações microscópicas $\underline{x} = (x_1, x_2, \dots, x_N)$ de respectivas energias $E(\underline{x})$. É sabido da mecânica estatística (26) que as configurações de equilíbrio deste sistema distribuem-se de acordo com $P(\underline{x}) = \frac{1}{Z} \exp[-\beta E(\underline{x})]$, onde $Z = \sum_{\underline{x}} \exp(-\beta E(\underline{x}))$ é dita sua *função de partição*. O cálculo da termodinâmica do sistema pode em seguida ser efetuado a partir de sua energia livre, que é dada por $\frac{F}{N} = \frac{1}{\beta} \lim_{N \rightarrow \infty} \frac{1}{N} \log Z$.

Em modelos para sistemas desordenados em equilíbrio, a energia dependerá por vezes não só da configuração das variáveis microscópicas \underline{x} , mas também de uma certa aleatoriedade introduzida no sistema - seja na posição das partículas ou, mais comumente, nas direções e intensidades dos campos e ligações. Temos assim $E = E(\underline{x}, \{\xi_i\})$, onde $\{\xi_i\}$ é um conjunto de variáveis aleatórias que assumem cada qual valores num determinado intervalo de acordo com uma distribuição de probabilidades $P(\xi_i)$.

Queremos, em modelos como estes, calcular médias das grandezas termodinâmicas com relação a $\{\xi\}$. A energia livre, por exemplo, fica

$$\langle\langle F \rangle\rangle_{\{\xi\}} = \frac{1}{\beta} \lim_{N \rightarrow \infty} \langle\langle \log Z \rangle\rangle_{\{\xi\}}. \quad (3.1)$$

onde $\langle\langle \cdot \rangle\rangle_{\{\xi\}}$ denota média sobre $\{\xi\}$, isto é, $\langle\langle \cdot \rangle\rangle_{\{\xi\}} = \int (\cdot) \prod_i P(\xi_i) d\xi_i$.

Neste contexto, distingue-se usualmente entre dois tipos de modelo para desordem (27): temperada (*quenched*) e recozida (*annealed*). No primeiro, admite-se que a desordem é estática — servindo como aproximação para situações em que as variáveis microscópicas evoluem muito mais rapidamente que a desordem; modelos de desordem recozida, por outro lado, assumem escala de evolução temporal semelhante para ambas. Trabalharemos aqui apenas com a desordem temperada.

A abordagem usual para modelos de desordem temperada se dá a partir do *truque de répli-*

cas (28), que é como se denomina a identidade $\log Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n}$. Pode-se prosseguir com o cálculo da energia livre, obtendo

$$\langle\langle F \rangle\rangle_{\{\xi\}} = \frac{1}{\beta} \lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \frac{1 - \langle\langle Z^n \rangle\rangle_{\{\xi\}}}{n}. \quad (3.2)$$

Dependeríamos agora tão somente do cálculo de $\langle\langle Z^n \rangle\rangle_{\{\xi\}}$, que é feito a partir do produto de n réplicas do sistema — daí o nome do método. Este cálculo é em geral deveras complicado, havendo na maior parte das vezes a necessidade de se fazer a hipótese de que as réplicas são simétricas.

O método de réplicas é sem dúvida o de maior sucesso neste tipo de estudos. Desde sua introdução por Sam Edwards e Phillip Anderson em 1975 (29), foi aplicado incansavelmente no estudo de diversos modelos — que, por sua vez, mostraram-se capazes de descrever uma série de fenômenos dentro e fora do âmbito da física.

Uma alternativa, no entanto, surge em torno de 1986. Neste ano, Marc Mézard e Giorgio Parisi, seguindo propostas semelhantes por outros autores (30, 31), formalizam técnica semelhante à aproximação de Bethe-Peierls (32, 33) a ser aplicada no estudo de sistemas desordenados, dando origem ao *método de cavidade* (34). Trata-se, como a aproximação, de uma teoria de *campo médio*: admite-se que a rede em que se dão as interações é uma árvore.

Mais recentemente, por volta do ano 2000, mostrou-se que há grande semelhança entre o método de cavidade e o que denominamos neste trabalho de técnica de propagação de crenças (35, 36, 7). Esta percepção possibilitou uma formulação moderna do método, que tem sido aplicada com sucesso nos últimos anos na análise de problemas de inferência e otimização (37, 38, 39)

Nas próximas seções, serão mencionadas as principais características do método de cavidade, e em seguida vamos utilizá-lo para a análise de nosso problema.

3.1 Método de cavidade

Na seção 2.3, foi discutido um esquema de transmissão de mensagens que possibilita o cálculo de marginais dadas distribuições de probabilidade do tipo $P(\underline{s}) = \frac{1}{Z} \exp[-\beta H(\underline{s})]$, típicas da mecânica estatística — fornecem a distribuição dos estados de equilíbrio a temperatura fixa β^{-1} . Para uma hamiltoniana da forma $H(\underline{s}) = -\sum_{(i,j) \in E} J_{ij} s_i s_j - \sum_i B_i s_i$, a regra para a transmissão fica

$$h_{i \rightarrow j}^{(t+1)} = B_i + \frac{1}{\beta} \sum_{k \in \partial i / j} \tanh^{-1} \left[\tanh(\beta J_{ik}) \tanh(\beta h_{k \rightarrow i}^{(t)}) \right]. \quad (3.3)$$

Iterando esta regra até que uma possível convergência seja alcançada, é obtido um conjunto de mensagens de ponto fixo, $\{h_{i \rightarrow j}^*\}$. Aproximações para as marginais são então calculadas a partir de $\mu_i(s_i) \cong \exp(\beta \hat{h}_i s_i)$, onde $\{\hat{h}_i\}$ são fornecidos por expressão semelhante à anterior, porém com soma efetuada sobre todos os vizinhos (eq. 2.29).

O cálculo de grandezas termodinâmicas pode então ser feito, dado que existe relação entre estas e $\{h_{i \rightarrow j}^*\}$ e $\{\hat{h}_i\}$, como veremos a seguir. Neste contexto, denominaremos $\{h_{i \rightarrow j}^*\}$ de *campos de cavidade*, e $\{\hat{h}_i\}$ de *campos efetivos*.

3.1.1 Cálculo de grandezas termodinâmicas

A partir do cálculo da energia livre do sistema, é possível obter uma descrição termodinâmica completa do mesmo. Outras grandezas, no entanto, como energia interna e entropia, possuem uma conexão mais direta com os campos de cavidade; determinaremos expressões para estas inicialmente, e a energia livre pode ser facilmente calculada em seguida a partir de $F = U - \frac{1}{\beta} S$.

Energia interna e entropia

As funções de compatibilidade, em nosso caso, são da forma $\psi_a(\underline{x}_{\partial a}) = \exp[-\beta E_a(\underline{x}_{\partial a})]$, pelo que $E_a(\underline{x}_{\partial a}) = -\frac{1}{\beta} \log \psi_a(\underline{x}_{\partial a})$, e a energia total é $E = -\frac{1}{\beta} \sum_{a=1}^M \log \psi_a(\underline{x}_{\partial a})$. Mas ora, a energia interna nada mais é que o valor médio da energia total, e assim

$$U = -\frac{1}{\beta} \sum_{\underline{x}} P(\underline{x}) \sum_{a=1}^M \log \psi_a(\underline{x}_{\partial a}). \quad (3.4)$$

Nos utilizando em seguida da equação 2.18, temos que $P(\underline{x}) = \prod_a \frac{1}{Z_a} \psi_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} v_{i \rightarrow a}^*(x_j)$, e U pode ser reescrita como

$$\begin{aligned} U[V] &= -\frac{1}{\beta} \sum_{\underline{x}} \prod_a \frac{1}{Z_a} \psi_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} v_{i \rightarrow a}^*(x_j) \sum_a \log \psi_a(\underline{x}_{\partial a}) \\ &= -\frac{1}{\beta} \sum_a \frac{1}{Z_a} \sum_{\underline{x}_{\partial a}} \left\{ \psi_a(\underline{x}_{\partial a}) \log \psi_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} v_{i \rightarrow a}^*(x_j) \right\}. \end{aligned} \quad (3.5)$$

Substituindo as funções de compatibilidade e a partir da eq. 2.25, podemos reescrever a expressão para nosso caso específico em termos de $\{h_{i \rightarrow j}^*\}$ e $\{\hat{h}_i\}$ como

$$U[h] = - \sum_{i,j} \xi_{ij} \frac{\tanh(\beta J_{ij}) \tanh(\beta h_{i \rightarrow j}^*) \tanh(\beta h_{j \rightarrow i}^*)}{1 + \tanh(\beta J_{ij}) \tanh(\beta h_{i \rightarrow j}^*) \tanh(\beta h_{j \rightarrow i}^*)} - \sum_i B_i \tanh(\beta \hat{h}_i). \quad (3.6)$$

A entropia, por sua vez, é dada por $S = - \sum_{\underline{x}} P(\underline{x}) \log P(\underline{x})$. Dado que o grafo de fatores com que estamos trabalhando é uma árvore, pode-se escrever que $P(\underline{x}) \cong \prod_a \mu_a(\underline{x}_{\partial a}) \prod_i \mu_i(x_i)^{1-|\partial i|}$ (para uma demonstração, ver por exemplo (3)), e a entropia fica

$$S[\mu] = - \sum_a \sum_{\underline{x}_{\partial a}} \mu_a(\underline{x}_{\partial a}) \log \mu_a(\underline{x}_{\partial a}) - \sum_i (1 - |\partial i|) \sum_{x_i} \mu_i(x_i) \log \mu_i(x_i). \quad (3.7)$$

Energia livre

Re-escrevendo $\prod_{i \in \partial a} v_{i \rightarrow a}^*$ em U (eq. 3.5) como $\prod_{b \in \partial R} \hat{v}_{b \rightarrow i}^*$ — onde ∂R é o conjunto de nós-função diferentes de a que se conectam a i — temos pela eq. 2.18 que

$$U[\mu] = - \frac{1}{\beta} \sum_a \sum_{\underline{x}_{\partial a}} \mu_a(\underline{x}_{\partial a}) \log \psi_a(\underline{x}_{\partial a}). \quad (3.8)$$

A partir desta expressão e utilizando-nos da relação $F[\mu] = U[\mu] - \frac{1}{\beta} S[\mu]$, temos para a energia livre

$$- \beta F[\mu] = \sum_a \sum_{\underline{x}_{\partial a}} \log \left[\frac{\mu_a(\underline{x}_{\partial a})}{\psi_a(\underline{x}_{\partial a})} \right] + \sum_i (1 - |\partial i|) \sum_{x_i} \mu_i(x_i) \log \mu_i(x_i), \quad (3.9)$$

ou, em termos de $\{v_{i \rightarrow a}, \hat{v}_{a \rightarrow i}\}$, novamente através de 2.18

$$\begin{aligned} - \beta F[\underline{v}] &= \sum_a \log \left[\sum_{\underline{x}_{\partial a}} \psi_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} v_{i \rightarrow a}^*(x_i) \right] + \sum_i \log \left[\sum_{x_i} \prod_{b \in \partial i} \hat{v}_{b \rightarrow i}^*(x_i) \right] \\ &\quad - \sum_{ia} \log \left[\sum_{x_i} v_{i \rightarrow a}^*(x_i) \hat{v}_{a \rightarrow i}^*(x_i) \right]. \end{aligned} \quad (3.10)$$

Para nosso caso, substituindo as funções de compatibilidade e pondo a expressão em termos dos campos de cavidade, ficamos com

$$\begin{aligned}
& -\beta F[\underline{h}] = \sum_{ij} \log \cosh \beta J_{ij} - \sum_{ij} \log [1 + \tanh(\beta J_{ij}) \tanh(\beta h_{i \rightarrow j}^*) \tanh(\beta h_{j \rightarrow i}^*)] \quad (3.11) \\
& + \sum_i \log \left\{ e^{\beta B_i} \prod_{j \in \partial i} [1 + \tanh(\beta J_{ij}) \tanh(\beta h_{j \rightarrow i}^*)] + e^{-\beta B_i} \prod_{j \in \partial i} [1 - \tanh(\beta J_{ij}) \tanh(\beta h_{i \rightarrow j}^*)] \right\}.
\end{aligned}$$

Magnetização e outros parâmetros de ordem

Como já discutido na seção 2.3, a magnetização local vem de $m_i = \tanh(\beta \hat{h}_i)$; uma maneira de calcular a magnetização média é a partir da média das magnetizações locais, isto é

$$m[\underline{h}] = \frac{1}{n} \sum_i \tanh(\beta \hat{h}_i). \quad (3.12)$$

Podemos também calculá-la supondo a presença de um campo externo B_0 , fazendo $H(\underline{s}) \rightarrow H(\underline{s}) - B_0 \sum_i s_i$; e posteriormente tomando $m = \left. \frac{\partial F}{\partial B_0} \right|_{B_0=0}$, o que fornece o mesmo resultado da eq. 3.12.

Um caso de interesse é aquele em que $B_i = Br_i$, com $r_i = \pm 1$; estudaremos este caso em detalhes mais a frente. Neste contexto, outra grandeza relevante é a *magnetização alternada* — dada por $\hat{m} = \frac{\partial F}{\partial B} = \frac{1}{n} \sum_i r_i s_i$, ou

$$\hat{m}[\underline{h}] = \frac{1}{n} \sum_i r_i \tanh(\beta \hat{h}_i). \quad (3.13)$$

É útil ainda introduzirmos um parâmetro de ordem capaz de caracterizar a fase vítrea. Adota-se usualmente a susceptibilidade de vidro de spin

$$\chi_{vs} = \frac{1}{n} \sum_{i,j} (\langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle)^2, \quad (3.14)$$

cuja divergência indica o aparecimento de tal fase. Classicamente, relaciona-se esta quantidade ao espectro da matriz hessiana, tal como proposto por Almeida e Thouless (40); no formalismo utilizado, no entanto, é mais natural relacioná-la aos campos de cavidade.

Como estamos trabalhando sobre uma árvore, podemos escrever que $\chi_{vs} = \sum_{r=1}^{\infty} C(r) (\langle s_0 s_r \rangle - \langle s_0 \rangle \langle s_r \rangle)^2$ em que $C(r)$ dá o número de nós a uma distância r de s_0 ; para que χ_{vs} não divirja devemos portanto ter $\lim_{R \rightarrow \infty} (\langle s_0 s_R \rangle - \langle s_0 \rangle \langle s_R \rangle)^2 = 0$.

Do teorema de flutuação-dissipação, vem que $\langle s_0 s_r \rangle - \langle s_0 \rangle \langle s_r \rangle \propto \frac{\partial \langle s_0 \rangle}{\partial h_{\rightarrow r}} = \frac{\partial \langle s_0 \rangle}{\partial h_{0 \rightarrow}} \frac{\partial h_{0 \rightarrow}}{\partial h_{\rightarrow r}}$, e deve-

se analisar assim

$$\rho = \lim_{R \rightarrow \infty} \frac{\partial h_{0 \rightarrow \cdot}}{\partial h_{\cdot \rightarrow R}}, \quad (3.15)$$

onde denotamos por $h_{0 \rightarrow \cdot}$ e $h_{\cdot \rightarrow R}$ mensagens respectivamente enviadas e recebidas pelo caminho que leva de 0 a R na árvore. Dependendo da maneira como a análise é realizada, há diferentes maneiras de efetuar o cálculo desta quantidade (41); exporemos em seguida como se dará a análise em nosso caso.

3.1.2 Equação de cavidade

É possível desta forma, a partir da iteração numérica das eqs. 3.3 sobre grafos da topologia desejada, simular sistemas físicos de tamanho finito e efetuar o cálculo de sua termodinâmica. É de nosso interesse, no entanto, realizar um estudo analítico mais aprofundado destas equações, sobretudo no limite $n \rightarrow \infty$.

Consideremos agora a variável aleatória $h^{(t)}$, cuja evolução é dada por

$$h^{(t+1)} = \left\langle \left\langle B + \frac{1}{\beta} \sum_{i=1}^{c-1} \tanh^{-1} \left[\tanh(\beta J_i) \tanh(\beta h_i^{(t)}) \right] \right\rangle \right\rangle_{B, J}, \quad (3.16)$$

em que $h_i^{(t)}$ são cópias i.i.d.¹ de $h^{(t)}$, distribuídas de acordo com $P(h^{(t)})$, e B e J_i são também aleatórias, distribuídas de acordo com dadas $P(B)$ e $P(J)$; continuamos pressupondo que os nós possuem cada um número fixo de vizinhos c . Afirmaremos sem provar² que, no limite $n \rightarrow \infty$, uma mensagem $h_{i \rightarrow j}^{(t)}$ amostrada de maneira uniforme converge em distribuição para $h^{(t)}$. À esta equação, dá-se o nome de *equação de evolução de densidade*.

Faremos algumas hipóteses; supondo a aplicação do esquema de propagação de crenças à um grafo de fatores aleatório arbitrariamente grande descrevendo $P(\underline{s})$, admitimos que: (a) há um conjunto de mensagens de ponto fixo, que (b) fornece boas aproximações para as marginais da distribuição conjunta, e (c) cujas mensagens se distribuem de acordo com o ponto fixo de 3.16.

A este conjunto de assunções, dá-se o nome de *hipótese de réplica simétrica*, em analogia àquela feita na utilização do método de réplicas; admite-se que, na presença de fases vítreas, a quebra de ergodicidade ocorrerá de forma a fracionar o espaço de fases em regiões não-ergódicas que relacionam-se entre si por meio de simetrias (42). Caso estas regiões não ocupem

¹ Isto é, independentes e identicamente distribuídas.

² Para uma prova, ver (3).

todo o volume do espaço de fases, é dito que há quebra na simetria de réplicas; faz-se necessário o uso de técnicas mais avançadas para o tratamento de tais situações (43, 44).

Foquemo-nos então na equação acima em seu ponto fixo, isto é

$$h = \left\langle \left\langle B + \frac{1}{\beta} \sum_{i=1}^{c-1} \tanh^{-1} [\tanh(\beta J_i) \tanh(\beta h_i)] \right\rangle \right\rangle_{B,J}, \quad (3.17)$$

denominada **equação de cavidade**. Esta mesma equação faz-se presente, por exemplo, nos trabalhos pioneiros de Mezard e Parisi; sob este *framework*, no entanto, podemos compreender que se trata em última instância de um estudo probabilístico sobre as marginais — mais precisamente, efetuado em árvores. A utilização deste mesmo esquema em grafos de topologia qualquer é um tipo de aproximação de campo médio; de fato, como mostrado por Yedidia (35), a minimização da energia livre³ $F[\mu]$ em uma árvore (eq. 3.9), dadas as marginais $\mu(x_i)$ de uma distribuição $P(\underline{x})$ qualquer, equivale à obtenção das soluções da eq. 3.17.

Nosso objetivo é determinar $P(h)$. Denotando $u(h_i, J_i) \equiv \frac{1}{\beta} \tanh^{-1} [\tanh(\beta J_i) \tanh(\beta h_i)]$, temos

$$P(h) = \left\langle \left\langle \int \delta \left[h - \left(B + \sum_{i=1}^{c-1} u(J_i, h_i) \right) \right] \prod_{i=1}^{c-1} P(h_i) dh_i \right\rangle \right\rangle_{B,J}. \quad (3.18)$$

Em casos mais simples, este estudo pode ser feito analiticamente; mais geralmente, no entanto, se faz necessária a obtenção numérica de uma estimativa para $P(h)$, através de um algoritmo de *dinâmica de populações*.

Parte-se, no algoritmo, de um conjunto grande de variáveis (uma *população*) que representam cópias i.i.d. de h . A cada período de tempo, cada uma destas variáveis é atualizada de acordo com a eq. 3.17, sendo os h_i amostrados da população de maneira uniforme, e B e J sorteados de acordo com as respectivas distribuições. Espera-se que, após um número de períodos, a população esteja distribuída de maneira similar a distribuição real $P(h)$, e podemos desta forma obter uma estimativa para a mesma. Vemos abaixo uma descrição algorítmica deste processo (algoritmo 2).

A inicialização de $\{h_i\}$ se dá a partir da amostragem de valores de alguma dada distribuição. São utilizadas, tipicamente, duas condições iniciais distintas (45): uma *livre*, com $h_i \sim 0 \forall i$, e outra *ferromagnética*, com $h_i \rightarrow \infty \forall i$. Havendo transições de fase de primeira ordem, a primeira condição inicial costuma levar a soluções metaestáveis. Dadas soluções para as duas condições

³Alternativamente, poder-se-ia minimizar a divergência de Kullback-Leibler $KL(P|Q)$ entre uma distribuição $P(\underline{x})$ qualquer e $Q(\underline{x}) \cong \prod_a \mu_a(\underline{x}_{\partial a}) \prod_i \mu_i(x_i)^{1-|\partial i|}$ em uma árvore.

Algoritmo 2 *Dinâmica de populações* (tamanho da população n_{pop} , número de coordenação c , passos no tempo t_{max})

```

inicializa  $\{h_i\}$ 
for  $t = 1 \rightarrow t_{\text{max}}$  do
  for  $i = 1 \rightarrow n_{\text{pop}}$  do
    amostra  $B; \{J_j\}, j = 1, c - 1$ 
    sorteia conjunto de variáveis  $\{h_j\}, j = 1, c - 1$ 
     $h_i \leftarrow B + \sum_{j=1}^{c-1} u(J_j, h_j)$ 
  end for
end for
calcula distribuição  $P(h)$ 
return  $P(h)$ 

```

iniciais, identifica-se a estável como aquela de menor energia livre.

Vale notar que os algoritmos 1 e 2, apesar de algo semelhantes, são responsáveis por tarefas bastante distintas. Enquanto o primeiro supõe situações reais em que uma amostra de avaliações foi obtida, o foco aqui é o estudo numérico das soluções de uma equação estocástica.

Determinada uma estimativa para $P(h)$ (e, conseqüentemente, para $P(\hat{h})$), podemos calcular agora valores médios para as grandezas termodinâmicas a partir das expressões deduzidas acima. A magnetização fica

$$m = \left\langle\left\langle \tanh(\beta \hat{h}) \right\rangle\right\rangle_{B,J,\hat{h}}, \quad (3.19)$$

e a (densidade de) energia livre, por sua vez

$$\begin{aligned}
-\beta f = & -\frac{c}{2} \log \cosh \beta + \frac{c}{2} \left\langle\left\langle \log [1 + \tanh(\beta J) \tanh(\beta h_1) \tanh(\beta h_2)] \right\rangle\right\rangle_{B,J,h} \quad (3.20) \\
& + \left\langle\left\langle \log \left\{ e^{\beta B} \prod_{i=1}^{c-1} [1 + \tanh(\beta J_i) \tanh(\beta h_i)] + e^{-\beta B} \prod_{i=1}^{c-1} [1 - \tanh(\beta J_i) \tanh(\beta h_i)] \right\} \right\rangle\right\rangle_{B,J,h}.
\end{aligned}$$

Para o cálculo computacional das médias, para por exemplo uma grandeza $\phi(h_1, h_2, \dots, h_k)$ que depende de k cópias de h , tomamos

$$\left\langle\left\langle \phi \right\rangle\right\rangle_h = \frac{1}{l} \sum_{i=1}^l \phi(h_1^{(i)}, h_2^{(i)}, \dots, h_k^{(i)}), \quad (3.21)$$

com $\underline{h}^{(i)}$ amostrados uniformemente da população, e $l \sim n$.

Expansão para baixas temperaturas

Podemos ainda determinar uma forma para a equação de cavidade no limite $\beta \rightarrow \infty$, por vezes de interesse em situações físicas. Para uma variável x qualquer, podemos escrever

$$\tanh(\beta x) = \operatorname{sgn}(x) \frac{1 - \exp(-2\beta|x|)}{1 + \exp(-2\beta|x|)} \sim \operatorname{sgn}(x) (1 - e^{-2\beta|x|}). \quad (3.22)$$

Utilizando-nos desta aproximação na eq. 3.17, ficamos com

$$h = B + \frac{1}{\beta} \sum_{i=1}^{c-1} \operatorname{sgn}(J_i h_i) \tanh^{-1} \left[1 - 2(e^{-2\beta|h_i|} + e^{-2\beta|J_i|}) \right]. \quad (3.23)$$

Levando em conta que $\beta|x| \sim \tanh^{-1}(1 - 2e^{-2\beta|x|})$, temos

$$h = B + \sum_{i=1}^{c-1} \operatorname{sgn}(J_i h_i) \min(|J_i|, |h_i|). \quad (3.24)$$

Podemos aplicar a aproximação, de forma análoga, no cálculo das grandezas termodinâmicas; a magnetização, por exemplo, fica $m \sim \left\langle \left\langle \operatorname{sgn}(\hat{h}) \right\rangle \right\rangle_{B,J,\hat{h}}$.

3.1.3 Exemplo: modelo S-K

Em 1975, Sherrington e Kirkpatrick propuseram (46) uma generalização do modelo de Ising, dada por um hamiltoniano $H = -\frac{1}{N} \sum_{(i,j)} J_{ij} s_i s_j$ — com $\{J_{ij}\}$ amostrados de certa distribuição (geralmente gaussiana ou bimodal) e soma efetuada sobre todos os pares (i, j) . Cada spin interage assim com todos os demais mas fracamente, devido a presença do termo $\frac{1}{N}$. Denomina-se geralmente este modelo de S-K, em alusão ao nome de seus autores; no artigo original, procede-se com o cálculo da termodinâmica a partir do método de réplicas.

Tomemos, nas equações que derivamos, $B_i = 0 \forall i$ e $J_{ij} = \pm 1$, com $P(J_{ij} = 1) = \frac{1}{2}$, continuando a trabalhar com spins dispostos sobre uma rede regular aleatória de grau $c = 3$. Passa-se a descrever assim um modelo bastante similar ao S-K; a equação de cavidade fica

$$h = \left\langle \left\langle \frac{1}{\beta} \sum_{i=1}^{c-1} \tanh^{-1} [\tanh(\beta J_i) \tanh(\beta h_i)] \right\rangle \right\rangle_{J,h}. \quad (3.25)$$

Obtêm-se em seguida estimativas numéricas para a distribuição $P(h)$, através do algoritmo de dinâmica de populações; na figura 3.1, pode-se ver as estimativas obtidas para alguns va-

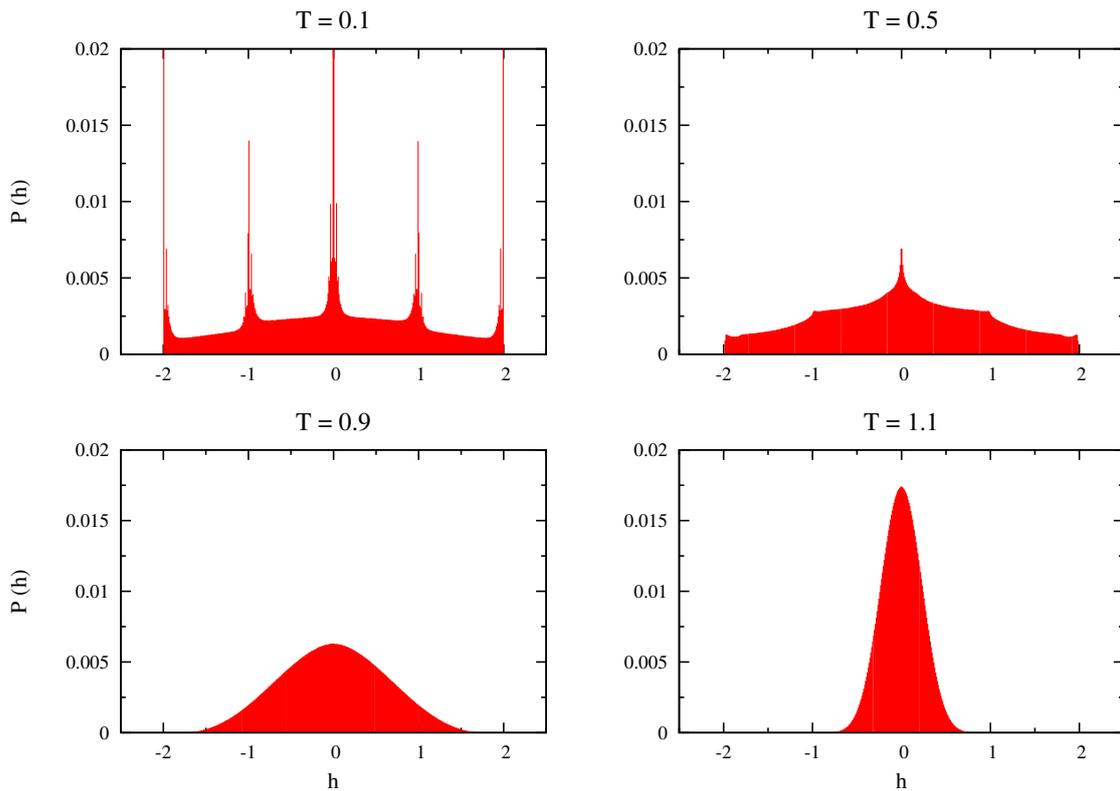


Figura 3.1: Estimativas obtidas para a distribuição $P(h)$ através do algoritmo de dinâmica de populações, para o caso em que $B_i = 0 \forall i$ e $J_{ij} = \pm 1$, $P(J_{ij} = 1) = \frac{1}{2}$.

lores de $T = \beta^{-1}$. Calculamos em seguida médias das grandezas termodinâmicas a partir, por exemplo, das eqs. 3.19 e 3.20.

A magnetização a campo nulo permite que identifiquemos se a fase termodinâmica em que o sistema se encontra é para- ($m = 0$) ou ferromagneticamente ($m \neq 0$) ordenada. Distribuições simétricas em torno de zero, como as que obtivemos, implicam em $m \sim 0$.

Pode-se observar na fig. 3.1 que a distribuição parece concentrar-se num pico em torno de 0 a medida que T aumenta. De fato, a partir de uma determinada temperatura, obtemos na dinâmica de populações $P(h) = \delta(h)$; denomina-se tal solução de *paramagnética*, por implicar em $m_i = 0 \forall i$. Analogamente, distribuições do tipo $P(h) = \delta(h - h_0)$ para algum dado h_0 denominam-se *ferromagnéticas* — aparecem por exemplo no estudo do modelo de Ising. Presupondo que a distribuição assume alguma destas formas, obtemos aproximações analíticas por vezes consistentes com alguma região de parâmetros. Na figura 3.2 calculamos a média da energia livre a partir das estimativas obtidas para $P(h)$ pelo algoritmo de dinâmica de populações, e também a partir de $P(h) = \delta(h)$; a altas temperaturas, as curvas coincidem.

Na análise de Sherrington e Kirkpatrick, foi identificada ainda uma fase de propriedades incomuns; estudos iniciando-se à época tratavam-a pelo nome de fase de *vidro de spin*. Tal

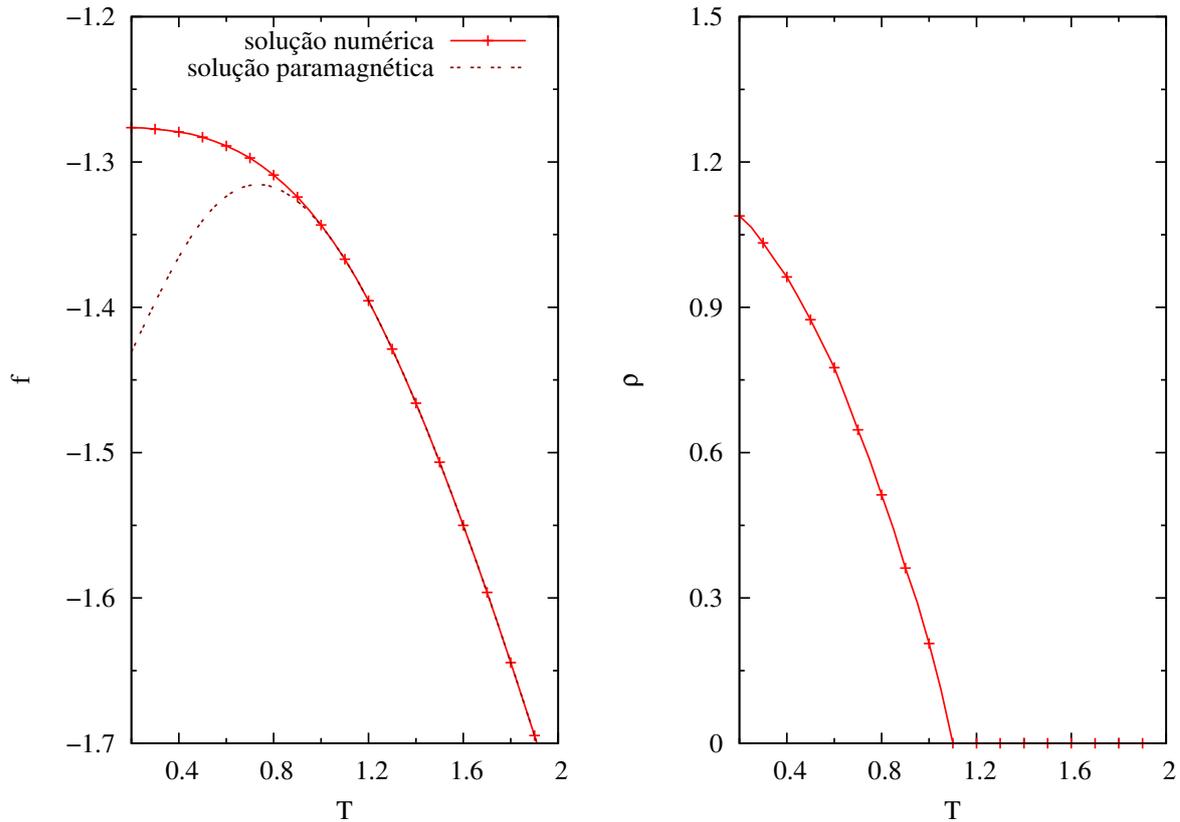


Figura 3.2: Grandezas termodinâmicas no modelo S-K em função da temperatura. À esquerda: densidade de energia livre, obtida através da dinâmica de populações (linha cheia) e de hipótese para $P(h) = \delta(h)$ (linha pontilhada) — ambas coincidem no limite de altas temperaturas; à direita: parâmetro ρ ; $\rho \neq 0$ indica divergência de χ_{vs} , e por conseguinte, a existência de uma fase vítrea

fase, típica de sistemas desordenados, apresenta características como (47) estados fundamentais frustrados — incapazes de minimizar globalmente termos locais de interação; superfícies de energia livre rugosas⁴; longos tempos de decorrelação; quebra de ergodicidade, etc.

Como mencionado acima, no formalismo de réplicas o aparecimento da fase é indicado pela divergência de 3.14, posteriormente relacionada ao espectro da matriz hessiana. No uso do método de cavidade, no entanto, queremos analisar o parâmetro ρ , dado pela eq. 3.15. Há várias formas de fazê-lo (41); optamos aqui pelo estudo de desvios entre duas réplicas da população. Calculamos a derivada

$$\frac{\partial h_0}{\partial h_R} \sim \frac{h_0(h_R + \delta) - h_0(h_R)}{\delta}, \quad (3.26)$$

através de um truque numérico: são introduzidas duas réplicas da população, atribuindo-se a

⁴Entende-se aqui as superfícies como função de um conjunto de valores médios que parametrizam $P(\underline{s})$ (ver por exemplo (48, 49)). A energia livre mantém-se convexa em função dos parâmetros extensivos.

uma valores iniciais $h_i^{(0)}$, e a outra valores $h_i'^{(0)} = h_i^{(0)} + \delta$, $\delta \sim 10^{-4}$. Após um número grande τ de passos no tempo, aproxima-se ρ por

$$\rho = \lim_{R \rightarrow \infty} \left(\frac{\partial h_0}{\partial h_R} \right)^2 \sim \sum_i \left[h_i^{(\tau)} - h_i^{(0)} \right]^2. \quad (3.27)$$

Os passos de tempo na dinâmica de população simulam assim a troca de mensagens entre nós cada vez mais distantes, devido à estrutura de árvore sobre a qual trabalhamos; a partir da eq. 3.27, é possível obter uma estimativa de como mensagens similares divergem em sua propagação — fornecendo uma aproximação para $\frac{\partial h_0}{\partial h_R}$.

Tomaremos, portanto, como parâmetros de ordem m e ρ ; pode-se daí diferenciar entre as seguintes fases termodinâmicas

- fase paramagnética, com $m = 0$ e $\rho = 0$
- fase ferromagnética, com $m \neq 0$ e $\rho = 0$
- fase de vidro de spin, com $m = 0$ e $\rho \neq 0$
- uma fase de vidro de spin ferromagneticamente ordenada, que denominaremos *mista*, com $m \neq 0$ e $\rho \neq 0$

O calculo da magnetização média nos dá neste caso $m = 0$, como argumentado acima; já pela figura 3.2, podemos ver que ρ se anula a partir de uma certa temperatura $T_c \sim 1.13$. Identificamos portanto duas fases distintas: uma fase vítrea para $T < T_c$, e uma paramagnética para $T > T_c$.

3.2 Vidro de spin de campo aleatório

Utilizando o método descrito na seção anterior, estudemos a hamiltoniana

$$H(\underline{s}) = - \sum_{(i,j) \in E} J_{ij} s_i s_j - B \sum_i r_i s_i, \quad (3.28)$$

com as variáveis de spin $s_i = \pm 1$ dispostas sobre um grafo aleatório regular de grau c , e as variáveis temperadas J_{ij} e r_i distribuídas de acordo com $\text{Bimod}_{\pm 1}(p)$ e $\text{Bimod}_{\pm 1}(q)$, respectivamente.

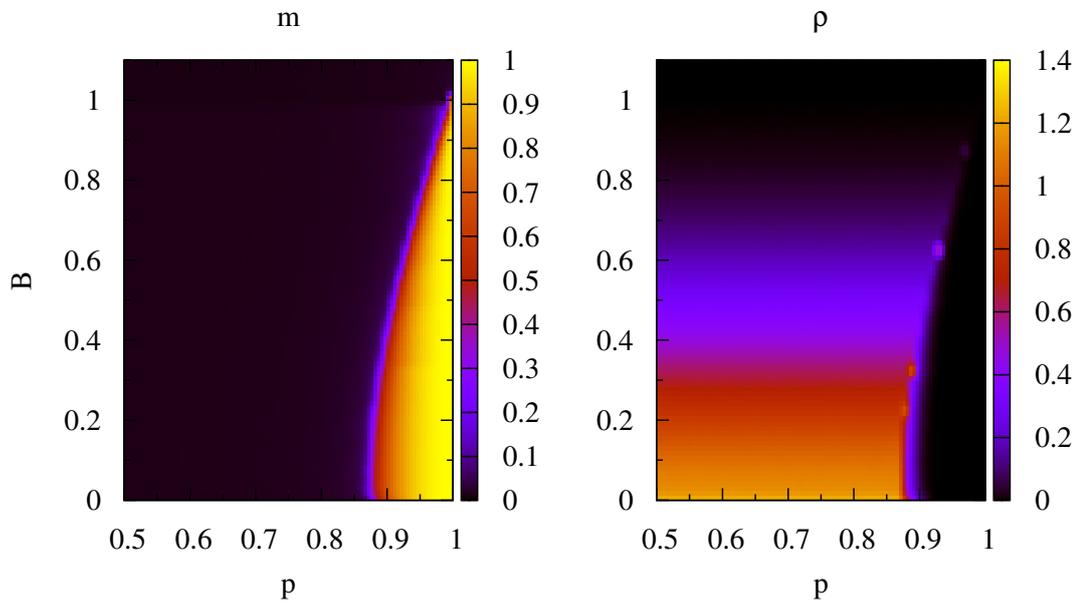


Figura 3.3: Parâmetros de ordem em função de p e B , para $T = 0$; m dá a magnetização à campo nulo, e ρ indica a divergência da susceptibilidade vítrea χ_{vs} . A região com $m = 0$ coincide em grande parte com aquela em que $\rho = 0$, dando origem à uma fase vítrea; há no entanto uma pequena região em que não coincidem, de forma que se faz também presente uma fase mista.

Este estudo abrirá caminho para a análise da distribuição a posteriori determinada na seção 2.2, $P(\mathbf{r}|J)$.

Iniciemos com a análise do limite para baixas temperaturas ($\beta \rightarrow \infty$), utilizando a equação 3.24. Através do algoritmo de dinâmica de populações, efetuamos o cálculo de m e ρ para valores diversos de p e B (fig. 3.3); traçamos em seguida o diagrama de fases da figura 3.4, seguindo a denominação sugerida na seção anterior.

Este procedimento pode ser repetido para o desenho dos demais diagramas. Temos a princípio quatro parâmetros de controle: β , B , p e q ; escolhemos dois destes, fixando os demais, e os variamos, executando para cada par de valores o algoritmo de dinâmica de populações. Geramos assim mapas para os parâmetros de ordem m e ρ , como os da figura 3.3; a partir destes mapas, determinamos as linhas do diagrama, dadas pelas linhas de contorno para $m = 0$ e $\rho = 0$ (na prática, utilizam-se as linhas de contorno para m e ρ infinitesimalmente pequenos, a fim de evitar problemas numéricos).

Estudemos assim o comportamento dos parâmetros de ordem, de maneira análoga, a temperaturas finitas; na fig. 3.5, vemos diagramas B-T para valores diversos de p . Os diagramas das figuras 3.4 e 3.5 foram feitos a fim de que comparássemos nossos resultados aos de (45), que

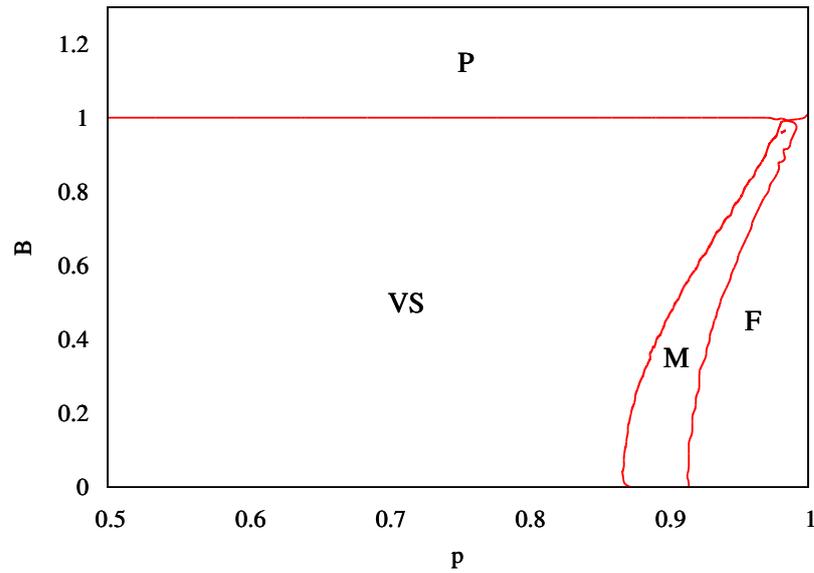


Figura 3.4: Diagrama de fases p-B para $T = 0$. Quatro fases termodinâmicas distintas fazem-se presentes: paramagnética (P), com $m = 0$, $\rho = 0$; ferromagnética (F), com $m \neq 0$, $\rho = 0$; vítrea (VS), com $m = 0$, $\rho \neq 0$; e mista (M), com $m \neq 0$, $\rho \neq 0$.

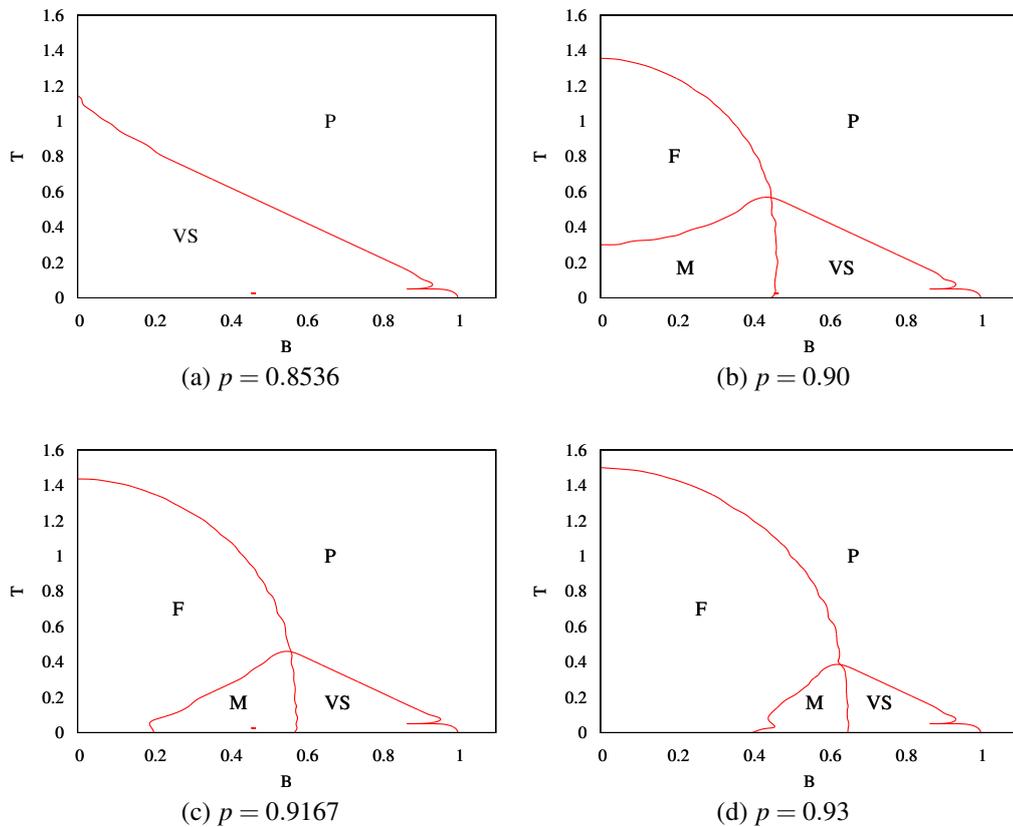


Figura 3.5: Diagramas de fases B-T para $q = \frac{1}{2}$ e valores diversos de p , escolhidos de forma a replicar os diagramas de (45). As irregularidades presentes em $T \sim 0.1$ se devem à problemas numéricos com a aproximação para baixas temperaturas.

realiza análise similar a esta; uma vez replicados com sucesso, validou-se nossa implementação do método.

As irregularidades presentes nos diagramas devem-se ao uso de pouca precisão no cálculo dos mapas de m e ρ , e um número pequeno de passos no tempo utilizados na execução da dinâmica de populações. É possível evitá-las impondo maiores precisão e número de passos, o que incorreria no entanto num maior custo computacional. Optando por um custo computacional baixo, as linhas dos diagramas apresentam por vezes pequenas falhas, que implicam tão somente um pequeno desvio em relação ao ponto em que de fato ocorre a transição de fases.

3.2.1 Linha de Nishimori

A distribuição a posteriori determinada anteriormente (eq. 2.30) sugere que trabalhemos com a hamiltoniana

$$H(\underline{s}) = - \sum_{(i,j) \in \Omega} J_{ij} s_i s_j - B \sum_i s_i. \quad (3.29)$$

Rotacionemos esta hamiltoniana fazendo $s_i \rightarrow r_i s_i$, a fim de que estimativas corretas, $s_i = r_i$, passem a ser representadas por $s_i = 1$. Usando que $J_{ij} = \xi_{ij} r_i r_j$, ficamos com

$$H(\underline{s}) = - \sum_{(i,j) \in \Omega} \xi_{ij} s_i s_j - B \sum_i r_i s_i. \quad (3.30)$$

Os valores de β e B são, à princípio, livres; tomando $\beta = \beta_N = \frac{1}{2} \log \frac{p}{1-p}$ e $B = B_N = \frac{\log \frac{q}{1-q}}{\log \frac{p}{1-p}}$, no entanto, passamos a estudar a distribuição de fato, $P(\underline{s}|J)$; Estes valores dos parâmetros definem a *linha de Nishimori* (50).

Utilizemos nossa análise do vidro de spin de campo aleatório para traçar o diagrama de fases sobre a linha de Nishimori (3.6), seguindo o procedimento delineado. Ao compará-lo com os mapas das figuras 2.11 e 2.12, percebemos que a região de parâmetros correspondente aqui à fase mista coincide com aquela na qual o algoritmo para inferência apresenta maus resultados.

Tal semelhança motiva uma investigação mais profunda sobre a relação entre as fases termodinâmicas deste sistema e os diferentes comportamentos do algoritmo.

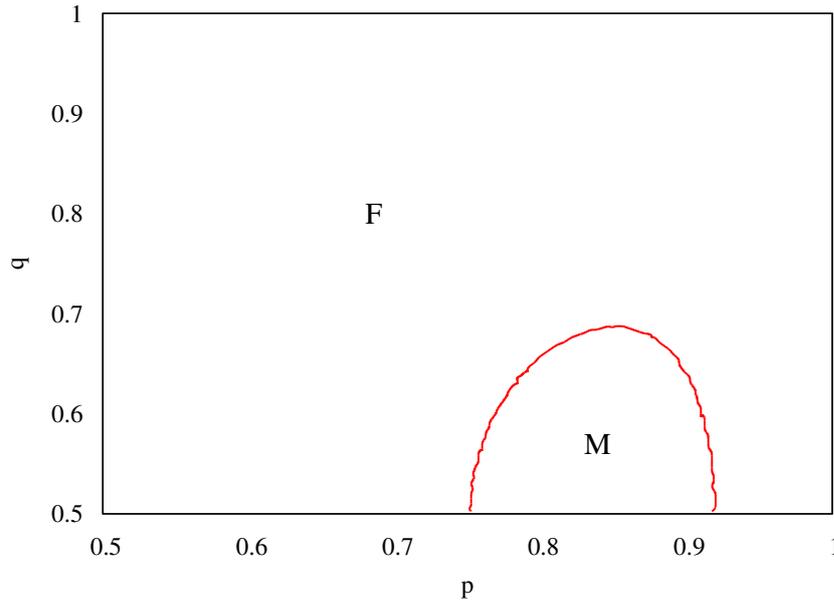


Figura 3.6: Diagrama de fases p - q , com β e B na linha de Nishimori. A região de parâmetros relacionada aqui à fase mista coincide com aquela em que o algoritmo apresenta baixo desempenho, na análise empírica

3.3 Estudo do algoritmo

No estudo empírico feito na seção 2.4, preocupamo-nos em analisar o erro médio do algoritmo, dado por $\bar{\varepsilon} = \left\langle \frac{1}{2} \left(1 + \frac{\mathbf{r} \cdot \hat{\mathbf{r}}}{n} \right) \right\rangle$. Lembrando que para a estimativa fizemos $\hat{r}_i = \text{sgn}(\hat{h}_i) \forall i$ e que, para o estudo analítico, efetuamos a rotação $s_i \rightarrow r_i s_i$, notamos que a quantidade $\frac{\mathbf{r} \cdot \hat{\mathbf{r}}}{n}$ pode ser analisada neste caso a partir do cálculo de $\left\langle \left\langle \text{sgn}(\hat{h}) \right\rangle \right\rangle_{\xi, r, h}$.

Esta grandeza pode ser facilmente calculada com o uso da dinâmica de populações; na figura 3.7, vemos o valor de $\bar{\varepsilon} = \frac{1}{2} \left(1 - \left\langle \left\langle \text{sgn}(\hat{h}) \right\rangle \right\rangle_{\xi, r, h} \right)$ em função de p e q , para β e B na linha de Nishimori.

Tal gráfico é, como se pode observar, bastante semelhante ao da figura 2.10; a parte de diferenças no canto inferior direito (região próxima a $p = 1$, $q = 0.5$), os valores obtidos para o erro são praticamente idênticos. Esta diferença parece existir por assumirmos estar aqui no limite termodinâmico ($n \rightarrow \infty$), enquanto nas simulações trabalhamos com um número finito de agentes; espera-se assim que, na análise empírica, sejam obtidos nesta região de parâmetros erros cada vez menores para valores maiores de n (figura 3.8).

Podemos, de forma análoga, estudar esta quantidade e traçar os diagramas de fase para os casos em que os parâmetros estimados $\{\hat{p}, \hat{q}\}$ são diferentes dos parâmetros reais $\{p, q\}$. Para isto, calculamos β_N e B_N na dinâmica de populações a partir de \hat{p} e \hat{q} , enquanto amostrando ξ

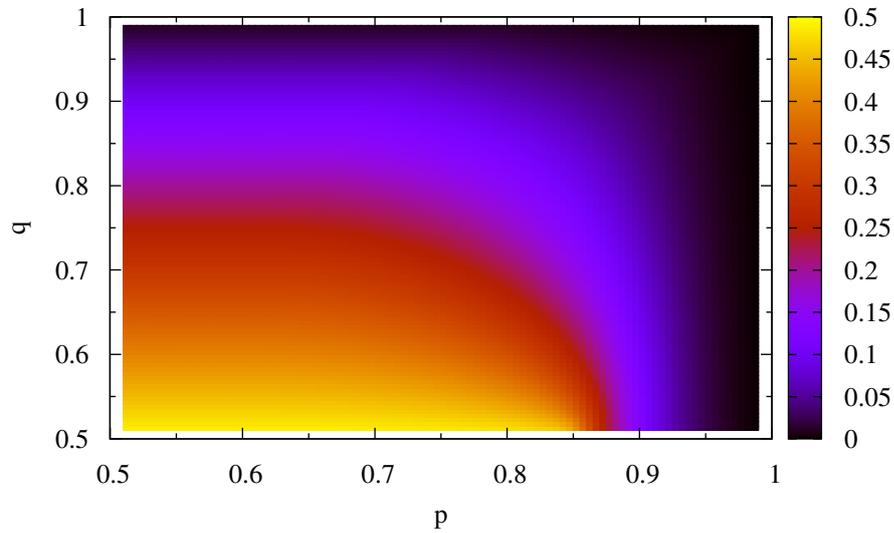


Figura 3.7: Erro teórico do algoritmo, calculado a partir de $\bar{\varepsilon} = \frac{1}{2} \left(1 - \left\langle \left\langle \text{sgn}(\hat{h}) \right\rangle \right\rangle_{\xi, r, h} \right)$, em função de $p \equiv P(\xi = 1)$ e $q \equiv P_0(r_i = 1)$

e r de $\text{Bimod}_{\pm 1}(p)$ e $\text{Bimod}_{\pm 1}(q)$, respectivamente.

Também nestes casos, a região de parâmetros em que a fase é mista coincide com aquela em que o desempenho é, de acordo com a análise anterior, baixo — em que a convergência não ocorre ou demora a ocorrer, e há alta sensibilidade às condições iniciais. A análise teórica do erro médio, por sua vez, fornece também resultados bastante semelhantes aos obtidos empiricamente.

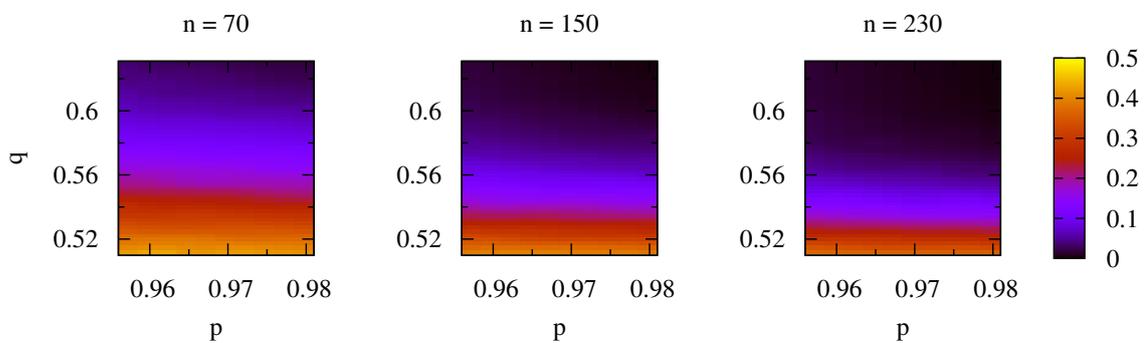


Figura 3.8: Erro médio obtido na análise empírica, na região próxima a $p = 1$, $q = 0.5$, para $n = 70$, 150 e 230; espera-se que, para n arbitrariamente grande, os valores sejam próximos aos da figura 3.7

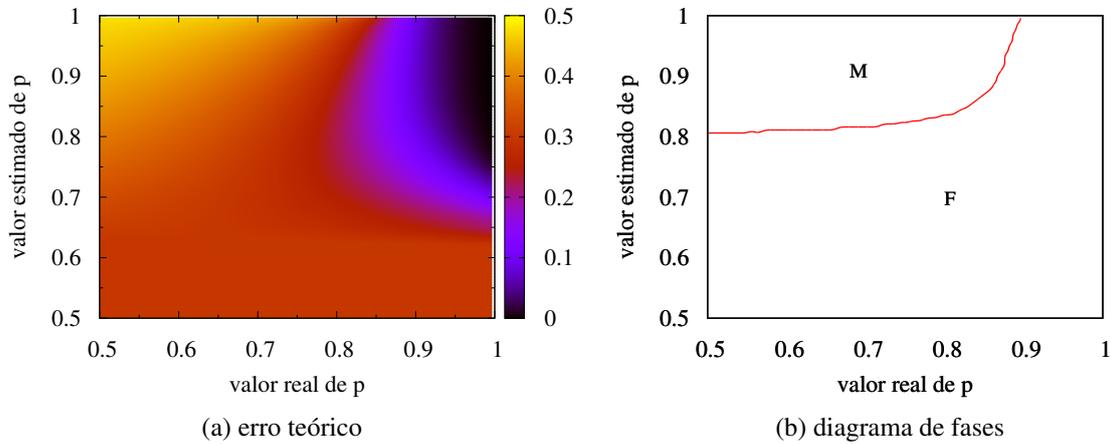


Figura 3.9: Análise teórica para o caso em que o valor estimado \hat{p} é diferente do real, p , com q fixo em 0.7. Também aqui a fase mista coincide com a de baixo desempenho, e o erro médio calculado possui grande semelhança ao obtido empiricamente (fig. 2.13).

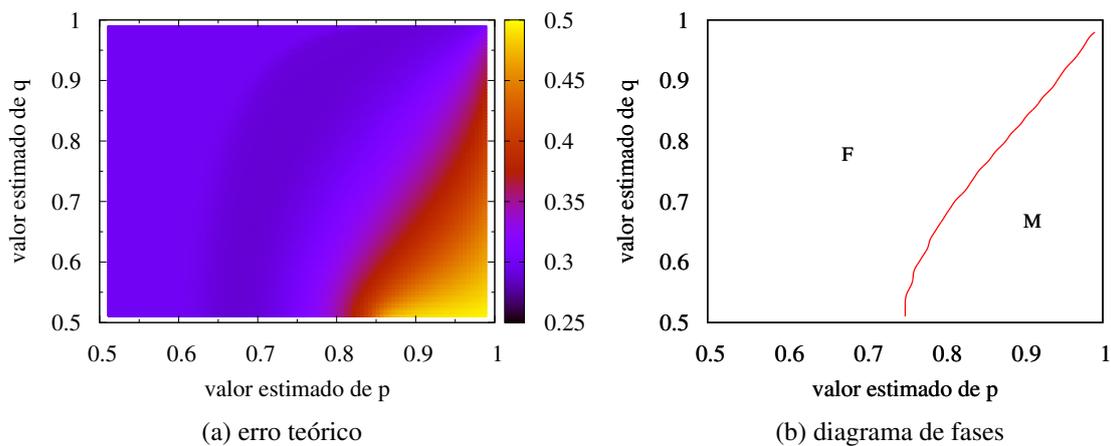


Figura 3.10: Análise teórica para valores estimados \hat{p} e \hat{q} diferentes dos reais, $p = q = 0.7$. Pode-se notar aqui também as semelhanças com a análise empírica (fig. 2.14).

Mostramos assim que se pode realizar, através do estudo do hamiltoniano, análise similar à empírica; a partir de considerações sobre o problema de inferência e com o uso de técnicas da mecânica estatística, determinamos de forma aproximada o erro médio do algoritmo e ainda regiões de baixo e alto desempenho. Há, portanto, o mérito de sermos capazes de calcular o desempenho esperado do algoritmo sem passar por uma custosa análise via simulações; ainda que dependamos da dinâmica de populações, seu custo tende a ser reduzido em comparação àquele das simulações realizadas no capítulo 2.

Ganhamos, ainda, intuição sobre o problema. A existência de fases vítreas (ou mistas) para determinadas regiões de parâmetros indica que a busca por mínimos da energia livre nestas regiões se dará com alguma dificuldade, devido às características já mencionadas de tal fase termodinâmica. Esta dificuldade deve-se sobretudo às propriedades intrínsecas do processo de inferência, e não àquelas de um ou outro algoritmo específico.

Tal análise aprofundada é de fato inédita na literatura; esta é, portanto, a principal contribuição desta dissertação.

4 *Testes de Robustez*

Até agora, analisamos o problema em questão sob duas perspectivas distintas: uma mais pragmática, na qual propomos uma maneira de obter estimativas; e outra teórica, em que aspectos intrínsecos do problema são estudados mais a fundo.

Em ambos os estudos, no entanto, limitamo-nos a amostras de opiniões distribuídas sobre um grafo aleatório regular de grau 3. Assumimos ainda que as avaliações eram emitidas de acordo com o modelo prescrito, ou ao menos de forma similar.

Ao admitir topologias diversas ou opiniões emitidas de maneiras distintas, podemos testar a robustez do algoritmo e também explorar os limites de nossa análise.

4.1 Diferentes topologias

Passemos a considerar agora amostras distribuídas sobre grafos de topologia qualquer. Tal tarefa faz-se trivial na análise empírica: basta gerarmos estruturas que representem estes grafos, por exemplo suas matrizes de adjacência, e executarmos o algoritmo para obtenção de estimativas sobre estas estruturas. Para gerá-las, podemos nos utilizar de bibliotecas especializadas (ver, por exemplo, (51)).

Na dedução do método de cavidade feita no seção 3.1, no entanto, admitimos estar trabalhando sobre uma rede regular aleatória de grau c fixo. Felizmente, o método é facilmente estendido; podemos reescrever as equações de modo à representar uma rede aleatória com um perfil de graus qualquer¹. Denominemos de Λ a distribuição de graus com que trabalharemos, e de $\bar{\Lambda}$ seu valor médio; $\Lambda(\gamma)$ dá assim a fração de nós com γ vizinhos.

A equação de cavidade fica

¹O perfil de graus a que nos referimos aqui é o de nós-variável; permanecemos trabalhando com o mesmo perfil para os nós-fator, de forma a haver sempre um nó-fator entre dois nós-vértice.

$$h = \left\langle \left\langle Br + \frac{1}{\beta} \sum_{i=1}^{\gamma-1} \tanh^{-1} [\tanh(\beta \xi_i) \tanh(\beta h_i)] \right\rangle \right\rangle_{r, \xi, \gamma}, \quad (4.1)$$

com γ amostrado de Λ ; a densidade de energia livre, por sua vez, é calculada a partir de

$$\begin{aligned} -\beta f &= -\frac{\bar{\Lambda}}{2} \log \cosh \beta + \frac{\bar{\Lambda}}{2} \left\langle \left\langle \log [1 + \tanh(\beta \xi) \tanh(\beta h_1) \tanh(\beta h_2)] \right\rangle \right\rangle_{r, \xi, \gamma, h} \quad (4.2) \\ &+ \left\langle \left\langle \log \left\{ e^{\beta Br} \prod_{i=1}^{\gamma-1} [1 + \tanh(\beta \xi_i) \tanh(\beta h_i)] + e^{-\beta Br} \prod_{i=1}^{\gamma-1} [1 - \tanh(\beta \xi_i) \tanh(\beta h_i)] \right\} \right\rangle \right\rangle_{r, \xi, \gamma, h}. \end{aligned}$$

Reescrevamos também o algoritmo de dinâmica de populações de forma a refletir tais mudanças (ver algoritmo 3).

Algoritmo 3 *Dinâmica de populações para rede aleatória com perfil de graus qualquer* (tamanho da população n_{pop} , perfil de graus $\Lambda(\gamma)$, passos no tempo t_{max})

```

inicializa  $\{h_i\}$ 
for  $t = 1 \rightarrow t_{\text{max}}$  do
  for  $i = 1 \rightarrow n_{\text{pop}}$  do
    amostra  $\gamma; B; \{J_j\}, j = 1, \gamma - 1$ 
    sorteia conjunto de variáveis  $\{h_j\}, j = 1, \gamma - 1$ 
     $h_i \leftarrow B + \sum_{j=1}^{\gamma-1} u(J_j, h_j)$ 
  end for
end for
calcula distribuição  $P(h)$ 
return  $P(h)$ 

```

Estamos agora em posição de estudar o algoritmo sobre redes de diferentes topologias, tanto de maneira empírica quanto teórica. Começemos analisando redes aleatórias "quase-regulares"; definamos a distribuição de graus $\Lambda_c(\gamma)$, dada por

$$\Lambda_c(\gamma) = \begin{cases} 1 - (c - \lfloor c \rfloor) & \text{p/ } \gamma = \lfloor c \rfloor \\ c - \lfloor c \rfloor & \text{p/ } \gamma = \lfloor c \rfloor + 1 \\ 0 & \text{p/ demais } \gamma \end{cases} \quad (4.3)$$

Para valores inteiros de c , temos que $\Lambda_c(\gamma) = \delta(\gamma - c)$, isto é, Λ_c dá o perfil de uma rede regular aleatória de grau c ; no caso mais geral, no entanto, por exemplo para $c = 2.3$, uma fração de 0.7 dos nós apresenta grau 2 e uma fração de 0.3, grau 3.

Para o estudo empírico, assim, gerou-se grafos com este perfil para valores diversos de c , permitindo que traçássemos os gráficos da fig. 4.1. Para o estudo teórico, utilizou-se a

variação sugerida da dinâmica de populações (alg. 3). Como feito anteriormente, calculamos o erro teórico e desenhamos o diagrama de fases. Há novamente notória similaridade entre os resultados obtidos, apesar da região de baixo desempenho não coincidir inteiramente com a fase mista.

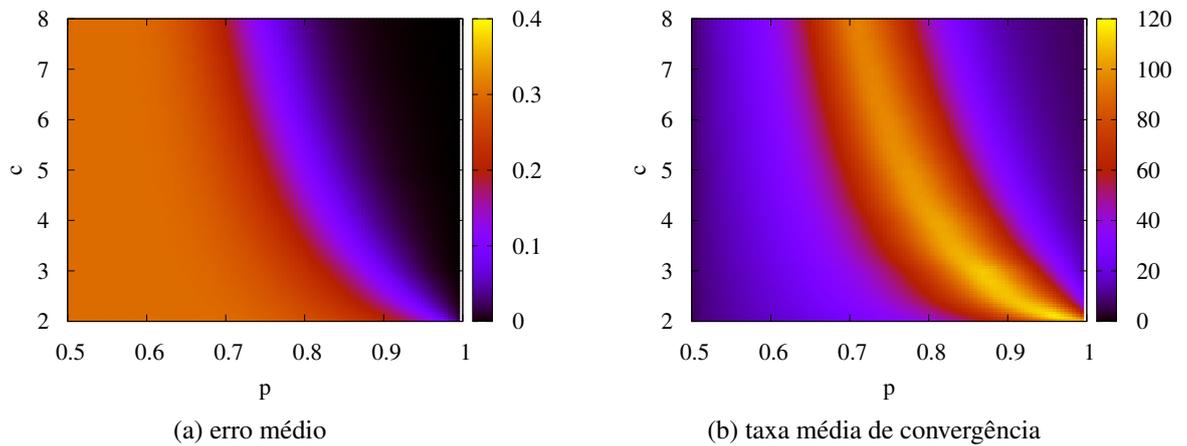


Figura 4.1: Análise empírica para redes aleatórias "quase-regulares", com perfis de grau dados pela eq. 4.3. Para valores inteiros de c , o perfil coincide com aquele de uma rede regular aleatória. Pode-se notar que o erro médio é menor para redes com um maior número médio de vizinhos.

Pode-se perceber ainda que o erro tende a ser cada vez menor para valores maiores de c ; tal resultado é intuitivo — trabalhar com amostras maiores de avaliações torna o ruído menos relevante.

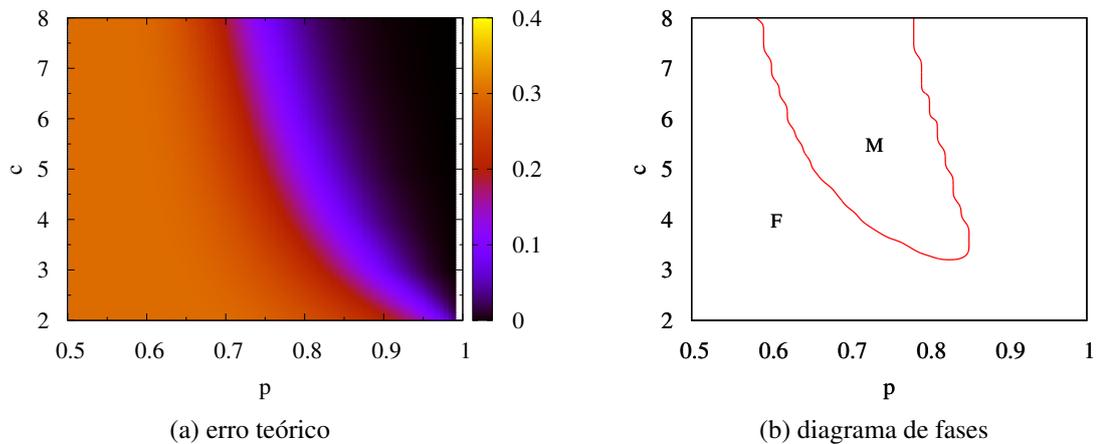


Figura 4.2: Análise teórica por meio da eq. 4.1, realizada numericamente com o uso do alg. 3 e tomando como distribuição de graus Λ_c . Pode-se notar aqui também a similaridade com o estudo empírico (fig. 4.1).

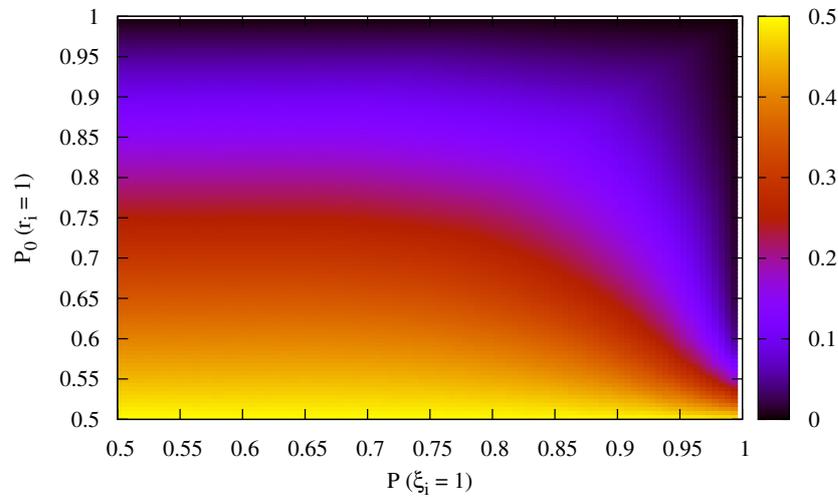


Figura 4.3: Erro empírico para execuções em árvores binárias; o perfil do gráfico se assemelha àquele obtido para grafos regulares aleatórios com $c = 3$. Como esperado, neste caso a convergência ocorre em 100% das execuções.

Derivamos o método de propagação de crenças, no capítulo 2, assumindo estar trabalhando em árvores — caso em que as marginais obtidas são exatas; repetindo a análise empírica para, por exemplo, uma árvore binária, obtemos portanto convergência em todas as execuções do algoritmo, e o número de passos necessários para que esta ocorra é o mesmo para valores quaisquer de p e q , dependendo apenas de n . Na figura 4.3, vemos o erro médio obtido na análise — há alguma semelhança com aquele obtido para grafos aleatórios regulares com $c = 3$.

Dado que trabalhamos na maior parte do tempo com grafos que contêm ciclos de longo comprimento médio (nos quais a técnica funciona com razoável precisão), é interessante que atestemos a validade do algoritmo em grafos nos quais ciclos pequenos ocorrem com frequência. Na figura 4.4 podemos ver o erro empírico para execuções numa rede quadrada, em função de p e com q fixo em 0.7. Nota-se assim que, apesar da convergência do algoritmo não ser garantida para grafos com ciclos, a mesma por vezes ocorre e leva à bons resultados.

Partindo de tal topologia e religando uma fração f de arestas de maneira aleatória, dá-se origem às chamadas redes de Watts-Strogatz. À medida que p aumenta, a distância média entre dois nós quaisquer da rede diminui drasticamente, pelo que tais redes são denominadas de redes de *mundo pequeno*. Há um grande interesse no estudo destas redes, pelo fato de apresentarem propriedades conhecidas de redes reais — por exemplo, redes sociais. Na figura 4.5, vemos que a fração de arestas religadas influi muito pouco no erro médio; a dependência maior parece ocorrer assim sobre o número médio de vizinhos no grafo.

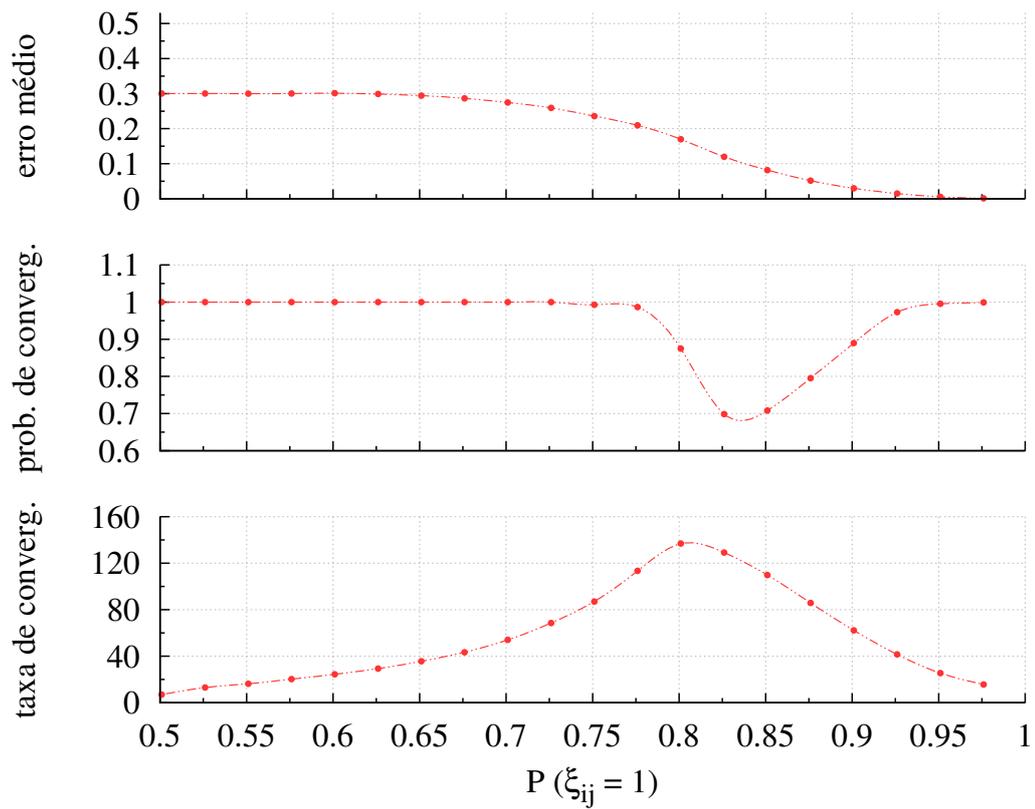


Figura 4.4: Resultados da análise via simulações para execuções numa rede quadrada com condições periódicas de contorno, com $q \equiv P_0(r_i = 1)$ fixo em 0.7. Apesar das execuções se darem num grafo com ciclos de pequeno comprimento médio, a convergência do mapa ocorre e fornece bons resultados.

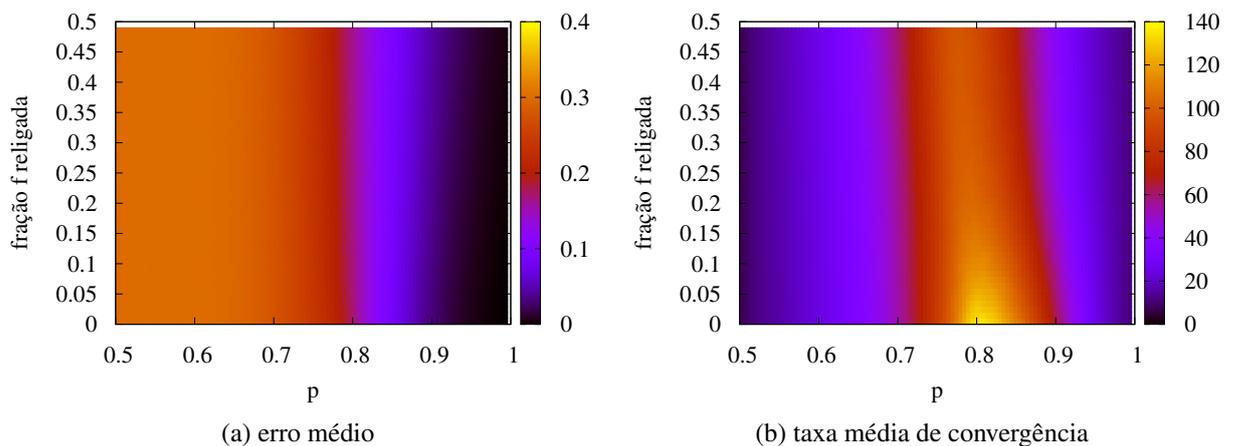


Figura 4.5: Análise empírica para rede de Watts-Strogatz com diferentes frações de arestas religadas, e q fixo em 0.7; a fração de arestas religadas parece não possuir grande influência sobre o erro.

4.2 Ataques ao sistema

Na seção 2.1, identificamos alguns dos problemas de que sofrem os sistemas de reputação; sua susceptibilidade a ataques mal-intencionados é possivelmente o mais grave deles.

Uma variedade comum de ataques se dá a partir do conluio entre agentes, como já mencionado anteriormente. O algoritmo que formulamos para a obtenção de estimativas procura evitar que este tipo de conluio influa no cálculo das reputações, invalidando-o.

Equipamentos eletrônicos que não estão em seu pleno funcionamento podem passar a emitir avaliações de maneira qualquer; este, também, pode ser considerado um tipo de ataque. Estudemos o que ocorre com nosso algoritmo em situações como esta.

Modificaremos a análise empírica de forma a fazer com que uma fração f dos agentes emita opiniões de maneira aleatória, isto é, um agente i que pertença a esta fração emitirá avaliações a partir de $J_{ij} = \eta_{ij}$, com $\eta_{ij} = \text{Bimod}_{\pm 1}(\frac{1}{2})$.

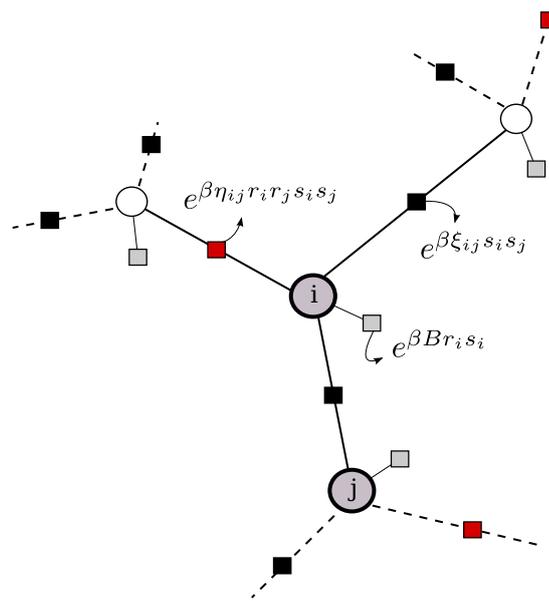


Figura 4.6: Grafo de fatores para a distribuição de Gibbs definida pela nova hamiltoniana, eq. 4.4. Além dos nós-função já presentes anteriormente, há agora um associado às funções de compatibilidade da forma $\exp(\beta \eta_{ij} r_i r_j s_i s_j)$.

Podemos modificar também a análise teórica de maneira a admitir que isto ocorra. Ao rotacionarmos a hamiltoniana sugerida por $P(\underline{s}|J)$ (eq. 3.30), eliminamos a dependência em J_{ij} fazendo $J_{ij} = \xi_{ij} r_i r_j$; assumamos agora que, para uma parte das interações, $J_{ij} = \eta_{ij}$. Passa-se então a trabalhar com a seguinte hamiltoniana

$$H(\underline{s}) = - \sum_{(i,j) \in \Omega/\Omega_f} \xi_{ij} s_i s_j - \sum_{(i,j) \in \Omega_f} \eta_{ij} r_i r_j s_i s_j - B \sum_i r_i s_i. \quad (4.4)$$

A fim de derivar uma nova equação de cavidade para esta hamiltoniana, devemos primeiro escrever as equações de propagação de crenças para o novo grafo de fatores.

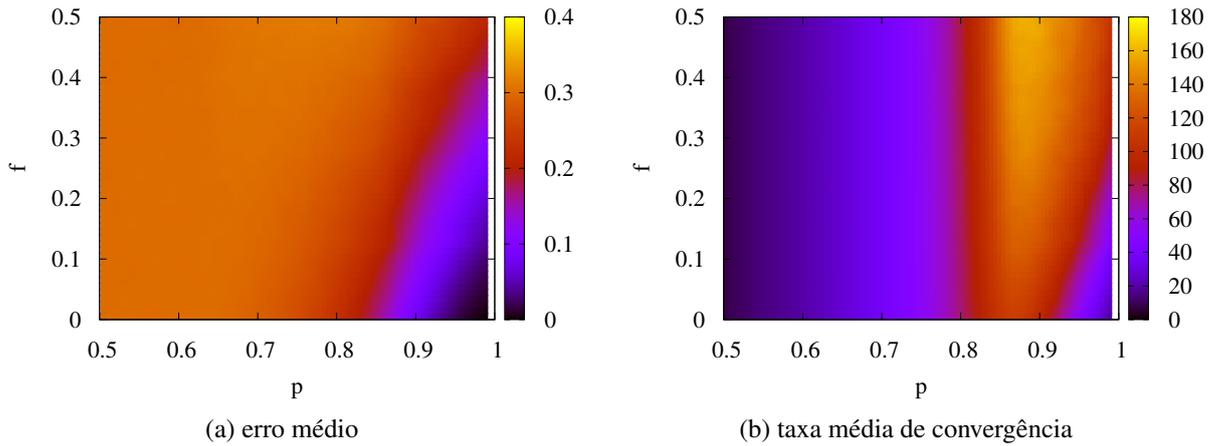


Figura 4.7: Resultados obtidos por meio de simulações para o desempenho do algoritmo sob ataque, em que uma fração f de agentes emite opiniões de maneira aleatória, $J_{ij} = \text{Bimod}_{\pm 1}(\frac{1}{2})$. Como pode-se observar, a medida que f aumenta, o erro cresce rapidamente.

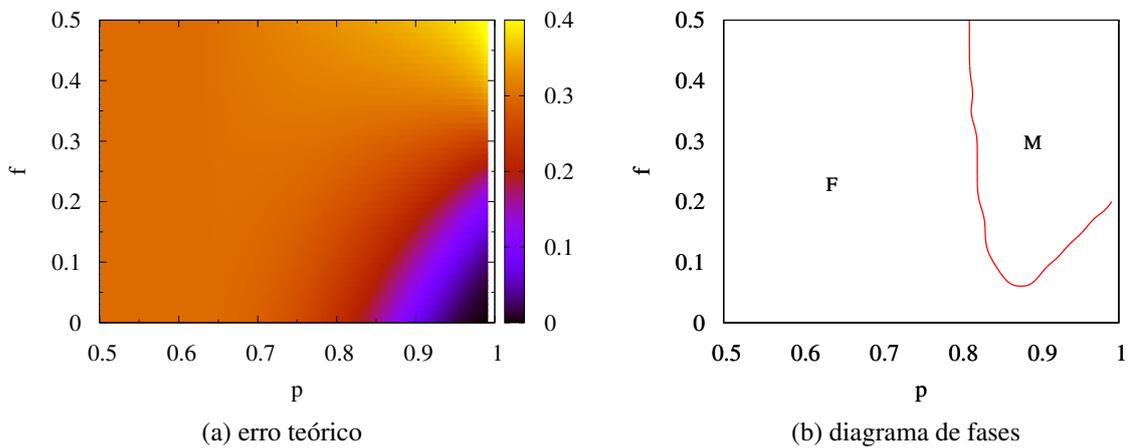


Figura 4.8: Análise teórica do algoritmo sob ataque, a partir da nova equação de cavidade 4.5; os resultados obtidos são novamente bastante similares aos empíricos (fig. 4.7), com a região de parâmetros em que a fase é mista coincidindo em grande parte com aquela de baixo desempenho.

Para tratar o novo termo, $\sum_{(i,j) \in \Omega_f} \eta_{ij} r_i r_j s_i s_j$, devemos incluir um novo tipo de nó-função no grafo de fatores, ligado às funções de compatibilidade da forma $\exp(\beta \eta_{ij} r_i r_j s_i s_j)$; cada nó-vértice se conecta a um destes nó-função com probabilidade f , e ao nó-função usual — ligado à função $\exp(\beta \xi_{ij} s_i s_j)$ — com probabilidade $(1 - f)$. Escrevemos as equações de propagação de crenças como na seção 2.3. Utilizando o mesmo raciocínio do capítulo 3, podemos escrever a equação de cavidade como

$$h = \left\langle \left\langle Br + \sum_{i=1}^{c-1} u_i(h) \right\rangle \right\rangle_{r, \xi, \eta}, \quad (4.5)$$

onde $u_i(h)$ é amostrado da distribuição de probabilidades em questão. Com probabilidade f , teremos $u_i(h) = \frac{1}{\beta} \tanh^{-1} [\tanh(\beta \eta r r') \tanh(\beta h)]$, e com probabilidade $(1 - f)$, o mesmo termo que utilizávamos anteriormente, $u(h) = \frac{1}{\beta} \tanh^{-1} [\tanh(\beta \xi) \tanh(\beta h)]$. Voltamos aqui a trabalhar sobre um grafo aleatório regular de grau c . De maneira análoga, escreve-se a energia livre como

$$\begin{aligned} -\beta f &= -\frac{c}{2} \log \cosh \beta + f \frac{c}{2} \left\langle \left\langle \log [1 + \tanh(\beta \eta r_1 r_2) \tanh(\beta h_1) \tanh(\beta h_2)] \right\rangle \right\rangle_{r, \xi, \eta, h} \\ &\quad + (1 - f) \frac{c}{2} \left\langle \left\langle \log [1 + \tanh(\beta \xi) \tanh(\beta h_1) \tanh(\beta h_2)] \right\rangle \right\rangle_{r, \xi, \eta, h} \\ &\quad + \left\langle \left\langle \log \left\{ e^{\beta B r} \prod_{i=1}^{c-1} [1 + \theta_i \tanh(\beta h_i)] + e^{-\beta B r} \prod_{i=1}^{c-1} [1 - \theta_i \tanh(\beta h_i)] \right\} \right\rangle \right\rangle_{r, \xi, \eta, h}, \end{aligned} \quad (4.6)$$

atribuindo $\theta_i = \tanh(\beta \eta r_1 r_2)$ com probabilidade f e $\theta_i = \tanh(\beta \xi)$ com probabilidade $(1 - f)$. As expressões que derivamos para as demais grandezas termodinâmicas permanecem inalteradas. A partir de pequenas modificações na dinâmica de populações (alg. 2), podemos obter soluções numéricas para a equação 4.5 e em seguida calcular a termodinâmica do sistema para os diferentes conjuntos de parâmetros.

Vemos na fig. 4.7 os resultados da análise empírica, e na fig. 4.8, os da análise teórica. O resultado de ambas é, como nos outros casos, bastante similar. Nota-se que, a medida que f aumenta, o erro também cresce rapidamente.

5 *Conclusão*

Ao tomar a tarefa realizada por sistemas de reputação como um cálculo de inferência estatística, abriu-se caminho para que estudássemos o problema de diferentes maneiras. Após dar base à nosso estudo com uma pequena revisão bibliográfica, formulamos um modelo que busca representar as situações desejadas de maneira bastante geral, mantendo-se contudo simples o bastante a ponto de conseguirmos atacá-lo com o uso de técnicas analíticas. Derivamos deste modelo uma distribuição *a posteriori*, que viria a se tornar nosso objeto de estudo.

Apresentamos em seguida uma técnica para a obtenção de estimativas em modelos gráficos denominada propagação de crenças. Abordamos a distribuição derivada com esta técnica — idéia presente na literatura recente da área — e mostramos que o cálculo de reputações se faz possível com seu uso; delineamos assim um algoritmo para a execução de tal tarefa. Analisamos este algoritmo de maneira empírica — simulamos sua execução sobre um grande número de cenários e avaliamos seu desempenho típico.

Passamos então a nos utilizar de técnicas da mecânica estatística para a análise do problema de inferência; a conhecida relação entre as duas áreas, mencionada na introdução desta dissertação, possibilita que estudemos o problema em detalhes, vindo a conhecer suas propriedades intrínsecas. Esta análise reduz-se ao estudo de uma hamiltoniana, que em nosso caso é aquela de um vidro de spin de campo aleatório; estudamos esta hamiltoniana com o uso do método de cavidade, um método moderno para o tratamento de sistemas desordenados. Apesar de introduzido nos anos 80, apenas há pouco tempo tal método veio a ser estudado mais a fundo, e é sabido hoje que se trata de um estudo probabilístico feito sobre a técnica de propagação de crenças.

A partir desta análise teórica, pudemos obter resultados similares aos da análise empírica com um menor esforço computacional — nossa principal contribuição. Apesar de ter tratado inicialmente apenas o caso em que avaliações distribuíam-se por grafos aleatórios regulares, estendemos em seguida este tratamento para redes de topologias diversas; atestou-se com isso que o algoritmo funciona com razoável acurácia mesmo em situações não previstas inicial-

mente. Estudamos também o que ocorre com o algoritmo quando, devido a ataques ao sistema, opiniões são emitidas de maneira diferente àquela modelada; apesar do desempenho ser ruim neste caso, mostramos de que forma pode-se modificar as análises a fim de incluir situações como esta.

Identificamos, ainda, relação entre regiões para as quais o desempenho do algoritmo é baixo e aquelas em que o sistema se encontra numa fase vítrea — havendo pois dificuldade na determinação dos estados de equilíbrio. Isto permite que formulemos novos algoritmos, capazes de levar em consideração as características típicas desta fase termodinâmica, tal como o fracionamento de seu espaço de fases.

Esta é, de fato, uma das direções futuras propostas para este trabalho. Formulada recentemente com base na técnica de propagação de crenças, a propagação de sondagens (44) (do inglês *survey propagation*) é uma técnica que procura atacar este tipo de dificuldades; há a expectativa de que aplicá-la ao nosso problema gere bons resultados. Aliada a esta técnica estão métodos analíticos mais avançados que permitem considerarmos a quebra da simetria de réplicas; seu uso levaria à uma análise ainda mais detalhada da estrutura do problema de inferência.

A abordagem de que nos utilizamos nesta dissertação, ligando problemas de inferência à mecânica estatística, é bastante geral. Tomamos modelo e problema específicos, mas análise similar pode ser realizada caso, por exemplo, não se julgue o modelo adequado. Processos diversos de inferência e otimização, típicos das áreas de computação e engenharia, podem vir a ser atacados com o uso das mesmas técnicas — gerando por vezes bons resultados, como tem sido mostrado por avanços das últimas décadas.

Referências Bibliográficas

- 1 JAYNES, E. T. *Probability Theory: The Logic of Science*. [S.l.]: Cambridge University Press, 2003. 758 p. ISBN 9780521592710.
- 2 SHANNON, C. A Mathematical Theory of Communication. *The Bell System Technical Journal*, v. 27, p. 379–423, 1948. ISSN 0724-6811.
- 3 MEZARD, M.; MONTANARI, A. *Information, Physics, and Computation*. [S.l.]: Oxford University Press, USA, 2010. ISBN 9780198570837.
- 4 COVER, T.; THOMAS, J. *Elements of Information Theory*. 2nd. ed. [S.l.]: Wiley-Interscience, 2006. ISBN 9780471241959.
- 5 MACKAY, D. *Information Theory, Inference and Learning Algorithms*. 1st. ed. [S.l.]: Cambridge University Press, 2003. ISBN 9780521642989.
- 6 SOURLAS, N. Spin Glasses, Error-Correcting Codes and Finite-Temperature Decoding. *Europhysics Letters (EPL)*, v. 25, n. 3, p. 159–164, jan. 1994. ISSN 0295-5075.
- 7 VICENTE, R.; SAAD, D.; KABASHIMA, Y. Low-density parity-check codes - A statistical physics perspective. *Advances in Imaging and Electron Physics*, Elsevier, v. 125, p. 231–353, 2003.
- 8 JOSANG, A.; ISMAIL, R.; BOYD, C. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, v. 43, n. 2, p. 618–644, mar. 2007. ISSN 01679236.
- 9 GRANDISON, T.; SLOMAN, M. A survey of trust in internet applications. *Communications Surveys & Tutorials, IEEE, IEEE*, v. 3, n. 4, p. 2–16, 2000.
- 10 MARMOL, F. G.; PÉREZ, G. M. Towards pre-standardization of trust and reputation models for distributed and heterogeneous systems. *Computer Standards & Interfaces*, Elsevier B.V., v. 32, n. 4, p. 185–196, jun. 2010. ISSN 09205489.
- 11 BUCHMANN, E. Trust Mechanisms and Reputation Systems. In: WAGNER, D.; WATTENHOFER, R. (Ed.). *Algorithms for Sensor and Ad Hoc Networks: Advanced Lectures*. Berlin: Springer, 2007. p. 325–337. ISBN 9783540749905.
- 12 ALZAID, H. Reputation-Based Trust System in Wireless Sensor Networks. In: *Security of Self-Organizing Networks: MANET, WSN, WMN, VANET*. [S.l.: s.n.], 2010. p. 493.
- 13 SEN, J. Reputation- and Trust-Based Systems for Wireless Self-Organizing Networks. In: *Security of Self-Organizing Networks: MANET, WSN, WMN, VANET*. [S.l.: s.n.], 2010. p. 91.
- 14 MARSH, S. *Formalising trust as a computational concept*. 170 p. Tese (Doutorado) — University of Stirling, 1994.

- 15 SABATER, J.; SIERRA, C. Review on Computational Trust and Reputation Models. *Artificial Intelligence Review*, v. 24, n. 1, p. 33–60, set. 2005. ISSN 0269-2821.
- 16 SCHILLO, M.; FUNK, P.; ROVATSOS, M. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence*, v. 14, n. 8, p. 825–848, 2000.
- 17 SABATER, J.; SIERRA, C. REGRET: reputation in gregarious societies. *Proceedings of the 5th International Conference on Autonomous Agents*, p. 194–195, 2001.
- 18 PAGE, L. et al. The PageRank citation ranking: Bringing order to the web. *World Wide Web Internet And Web Information Systems*, Stanford InfoLab, p. 1–17, 1999.
- 19 CHEN, P. et al. Finding scientific gems with Google’s PageRank algorithm. *Journal of Informetrics*, v. 1, n. 1, p. 8–15, jan. 2007. ISSN 17511577.
- 20 ERMON, S.; SCHENATO, L.; ZAMPIERI, S. Trust Estimation in autonomic networks: a statistical mechanics approach. In: *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*. [S.l.]: IEEE, 2009. p. 4790–4795. ISBN 978-1-4244-3871-6.
- 21 JIANG, T.; BARAS, J. S. Trust evaluation in anarchy: A case study on autonomous networks. In: *Proceedings of IEEE Infocom*. [S.l.: s.n.], 2006.
- 22 WORMALD, N. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, Citeseer, p. 239–298, 1999.
- 23 VIGER, F.; LATAPY, M. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. *Computing and Combinatorics*, Springer, p. 440–449, 2005.
- 24 eBay. Disponível em: <<http://www.ebay.com/>>.
- 25 AJI, S.; MCELIECE, R. The generalized distributive law. *IEEE Transactions on Information Theory*, v. 46, n. 2, p. 325–343, mar. 2000. ISSN 00189448.
- 26 SALINAS, S. *Introdução à Física Estatística*. 2. ed. [S.l.: s.n.], 1997. ISBN 9788531403866.
- 27 CATICHA, N. From quenched to annealed: a study of the intermediate dynamics of disorder. *Journal of Physics A: Mathematical and General*, IOP Publishing, v. 27, p. 5501, 1994.
- 28 NISHIMORI, H. *Statistical physics of spin glasses and information processing: an introduction*. [S.l.]: Oxford University Press, USA, 2001. ISBN 0198509405.
- 29 EDWARDS, S.; ANDERSON, P. Theory of spin glasses. *Journal of Physics F: Metal Physics*, v. 965, 1975.
- 30 SCHNEIDER, T.; PYTTE, E. Random-field instability of the ferromagnetic state. *Physical Review B*, APS, v. 15, n. 3, p. 1519–1522, 1977. ISSN 1550-235X.
- 31 NAKANISHI, K. Two- and three-spin cluster theory of spin-glasses. *Physical Review B*, v. 23, n. 7, p. 3514–3522, abr. 1981. ISSN 0163-1829.

- 32 WEISS, P. The application of the Bethe-Peierls method to ferromagnetism. *Physical Review*, v. 74, n. 10, p. 1493–1504, 1948.
- 33 STEIN-BARANA, A.; YOSHIDA, M.; LÍBERO, V. A aproximação de campo médio de Bethe-Peierls. *Revista Brasileira de Ensino de Física*, v. 26, n. 4, p. 385–393, 2004.
- 34 MEZARD, M.; PARISI, G.; VIRASORO, M. SK model: The replica solution without replicas. *Europhysics Letters (EPL)*, v. 77, 1986.
- 35 YEDIDIA, J.; FREEMAN, W.; WEISS, Y. Characterization of belief propagation and its generalizations. *IT-IEEE*, v. 51, p. 2282–2312, 2001.
- 36 MÉZARD, M.; PARISI, G. The Bethe lattice spin glass revisited. *The European Physical Journal B*, v. 20, p. 217–233, 2001.
- 37 MEZARD, M.; PARISI, G.; ZECCHINA, R. Analytic and algorithmic solution of random satisfiability problems. *Science*, AAAS, v. 297, n. 5582, p. 812, 2002. ISSN 1095-9203.
- 38 KRZAKALA, F.; RICCI-TERSENGHI, F.; ZDEBOROVÁ, L. Elusive Spin-Glass Phase in the Random Field Ising Model. *Physical Review Letters*, v. 104, n. 20, p. 1–4, maio 2010. ISSN 0031-9007.
- 39 YEDIDIA, J. S. Message-Passing Algorithms for Inference and Optimization. *Journal of Statistical Physics*, out. 2011. ISSN 0022-4715.
- 40 ALMEIDA, J.; THOULESS, D. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, IOP Publishing, v. 11, p. 983, 1978.
- 41 ZDEBOROVÁ, L. Statistical physics of hard optimization problems. *Acta Physica Slovaca. Reviews and Tutorials*, v. 59, n. 3, p. 169–303, jun. 2009. ISSN 0323-0465.
- 42 GOLDENFELD, N. *Lectures on Phase Transitions and the Renormalization Group*. [S.l.]: Perseus Book Publishing, 1992. ISBN 9780201554098.
- 43 PARISI, G. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, Institute of Physics Publishing, v. 13, p. L115–L121, 1980.
- 44 MEZARD, M. The cavity method at zero temperature. *Journal of Statistical Physics*, v. 111, n. April, 2003.
- 45 MATSUDA, Y. et al. Random-field p -spin-glass model on regular random graphs. *Journal of Physics A: Mathematical and Theoretical*, v. 44, n. 18, p. 185002, maio 2011. ISSN 1751-8113.
- 46 SHERRINGTON, D.; KIRKPATRICK, S. Solvable model of a spin-glass. *Physical Review Letters*, v. 35, n. 26, 1975.
- 47 CASTELLANI, T.; CAVAGNA, A. Spin-glass theory for pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2005, n. 05, p. P05012, maio 2005. ISSN 1742-5468.
- 48 AMIT, D. J. *Modeling Brain Function*. [S.l.]: Cambridge University Press, 1992. ISBN 9780521421241.

- 49 BINDER, K.; YOUNG, A. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern physics*, APS, v. 58, n. 4, p. 801, 1986.
- 50 IBA, Y. The Nishimori line and Bayesian statistics. *J. Phys. A: Math. Gen.*, n. 32, p. 3875–3888, 1999.
- 51 CSÁRDI, G.; NEPUSZ, T. *igraph*. 2009. Disponível em: <<http://igraph.sourceforge.net/>>.

Índice Remissivo

- propagação de crenças
 - equações de, 24
- Bethe-Peierls, 28
- campo de
 - cavidade, 38
 - Markov, 23
- confiança, 13
- desordem
 - recozida, temperada, 36
- dinâmica de populações, 42
- equação de cavidade, 42
- fases termodinâmicas, 47
- função de compatibilidade, 23
- grafo de fator, 23
- inferência
 - MAP, 11
 - MPM, 11, 29
- linha de Nishimori, 50
- método de
 - cavidade, 37
 - réplicas, 37
- magnetização, 43, 45
 - alternada, 40
 - local, 27
- magnetização local, 28
- propagação de crenças, 23
- réplicas simétricas, 42
 - hipótese de, 41
- rede
 - aleatória regular, 21
- reputação, 14
- sistemas de reputação, 11
- vidro de spin, 45
 - susceptibilidade de, 40