

CLASSIFICAÇÃO DA CAFEICULTURA FAMILIAR PELO USO DE CLUSTERIZAÇÃO

Ana Maria Freitas FERREIRA¹, E-mail: terradoce@hotmail.com; Sônia Milagres TEIXEIRA², E-mail: smilagres@cnpaf.embrapa.br

¹SEMARH-GO, FANAP; 2 Embrapa - Café

Resumo:

Este estudo teve o objetivo de contribuir para a caracterização da cafeicultura familiar, utilizando-se ferramentas de análise de informação, especificamente de clusterização – *SimpleKmeans*. Os dados coletados sob a coordenação da Embrapa Café, dentro do projeto de pesquisa “Custos de Produção da Cafeicultura Brasileira” foram tratados e submetidos ao processo de clusterização. Foram feitos estudos comparativos entre os diversos critérios atualmente utilizados para a caracterização da agricultura familiar (PRONAF, CONTAG, FAO/INCRA), além de uma caracterização resultante da pesquisa e outra complementada considerando uma tipologia referenciada na lógica familiar. Os dados organizados possibilitaram análises *inter* e *intra* grupos familiares e patronais resultantes dos critérios classificatórios. Utilizando metodologia de *data mining*, buscou-se identificar similaridades e dissimilaridades *inter* e entre os grupos e inferir quais atributos se relacionam com os processos de produção de tipo familiar na cafeicultura. Conclui que há variação significativa no conceito de agricultura familiar no café para diferentes critérios de classificação, e na amostra estudada a inexistência de um padrão de comportamento produtivo que possa ser associado a um ou outro grupo.

Palavras-chave: agricultura familiar, cafeicultura, clusterização, *SimpleKmeans*

FAMILY COFFEE FARM CHARACTERIZATION WITH THE USE OF CLUSTERIZATION TOOLS

Abstract :

This study had the objective to contribute for the family farm's characterization of coffee farm business by using analytical information tools, specifically clustering - *SimpleKmeans*. The data were collected from a research project entitled “Costs of Production of the Brazilian Coffee Crop”, coordinated by Embrapa Café. They were analyzed and submitted to the clustering process. Comparative studies were undertaken among several criteria, lately used for family farm characterization (PRONAF, CONTAG, FAO/INCRA). Besides a characterization resulting from the field research complemented by a typology referred to as the family context. The data provided *inter* and *intra* family versus non-family groups resulting from various classification criteria. Using *data mining* methodology, from consistent signs, we sought to identify similarities and non similarities *inter* and *intra* groups and to infer which attributes are related to the family farm production processes in coffee cropping. The results showed a significant variation in the concept of family farm in coffee as the criteria were changed, for the sample the non existence of a productive behavior pattern that could be related to one or another group.

Key Words: family farm, coffee culture, clustering, *SimpleKmeans*.

Introdução

Agricultura Familiar

O debate sobre agricultura familiar no Brasil ainda é recente sem contornos definidos. Entre os estudos publicados merecem citação os trabalhos de Veiga (2000), Abramovay (1998) e de Lamarche (1998), que com seus estudos revelam ser a agricultura familiar uma forma social reconhecida e legitimada nos países desenvolvidos, onde a estrutura agrária é composta, em sua maior parte, por explorações nas quais o trabalho da família é preponderante. A expressão “agricultura familiar” ganhou projeção somente a partir do final dos anos oitenta, e somente na primeira metade da década de 90 o Estado concebe, pela primeira vez na história da política agrícola brasileira, um programa de políticas específicas destinado a fortalecer os estabelecimentos agrícolas de base familiar, conhecido como PRONAF.

A FAO, em estudo baseado no Censo Agropecuário de 1985 e desenvolvido em parceria com o INCRA, associou três características ao conceito de agricultura familiar: a) a gestão da unidade produtiva e os investimentos nela realizados são feitos por indivíduos que mantém entre si laços de sangue ou casamento; b) a maior parte do trabalho é igualmente fornecida pelos membros da família; e c) a propriedade dos meios de produção (embora nem sempre a terra)

pertence à família e é em seu interior que se realiza sua transmissão em caso de falecimento ou de aposentadoria dos responsáveis pela unidade produtiva.

Segundo Schneider (2003), as unidades familiares funcionam, predominantemente, com base na utilização da força de trabalho da família, podendo contratar outros trabalhadores temporariamente. Considera que a utilização de critérios de quantificação da unidade familiar que usa trabalho contratado deixa de ser considerada como tal constituí-se operação heurística que, isoladamente, não é suficiente para se entender e caracterizar sua natureza.

O Ministério de Desenvolvimento Agrário, Secretaria de Agricultura Familiar, segundo o Grupo de Trabalho Ater, para elaboração de sua Política Nacional de Assistência Técnica e Extensão Rural, considera: “[...] agricultura familiar”, aquela em que trabalhos em nível de unidade de produção são exercidos predominantemente pela família, mantendo ela a iniciativa, o domínio e o controle do *que* e do *como* produzir, havendo uma relação estreita entre o que é produzido e o que é consumido (ou seja, são unidades de produção e de consumo), mantendo também um alto grau de diversificação produtiva, tendo alguns produtos relacionados com o mercado.

A CONTAG (Confederação Nacional dos Trabalhadores Rurais) considera como familiares todos os agricultores que trabalham em menos de quatro módulos fiscais e que não contratem mão-de-obra permanente. A *área máxima regional* como limite superior para área total dos estabelecimentos familiares, foi incorporada ao conceito que não limitava tamanho, sendo a extensão máxima determinada pelo que a família pode explorar com seu próprio trabalho e tecnologia de que dispõe. O limite da área, nos estudos desenvolvidos pela FAO/INCRA teve a finalidade de evitar distorções resultantes da inclusão de grandes latifúndios no universo de unidades familiares,

O **objetivo** deste trabalho foi tipificar processos de produção de café no Brasil utilizando-se técnicas de agrupamento (*clustering*) que explorem semelhanças entre padrões e agrupam os padrões parecidos em categorias ou grupos. Com o uso de informação coletada junto a uma amostra de propriedades cafezeiras procurou-se identificar a cafeicultura familiar, com base nas definições mencionadas, os grupos e classificá-los por uma série de atributos.

Material e Métodos

Por intermédio de algoritmos baseados em redes neurais, estatísticas ou algoritmos de aprendizado, “mineram-se” informações coletadas em campo. A essência do *data mining* é descobrir uma informação, sem formular previamente uma hipótese. A mineração de dados pode tornar os resultados mais acessíveis, mas não, necessariamente, faz com que eles fiquem mais fáceis de interpretar. Cada classe de aplicação em *data mining* tem como base um conjunto de algoritmos que serão usados na extração de relações relevantes dentro de uma massa de dados: análise de seqüências, clusterização, classificação, estimativas e regras de associação.

A partir de um conjunto finito de objetos ou padrões X , cada um deles descritos por um conjunto de pares atributo-valor atributo , um algoritmo de clusterização deve atribuir rótulos aos objetos que identifiquem subgrupos naturais do conjunto. Isso é feito por meio da introdução de uma medida de distância (ou de similaridade) entre os objetos. Usando essa medida, determina-se um critério que expressa a idéia de clusterização, isto é, que particiona o conjunto de objetos em K *clusters* individuais e homogêneos, nos quais cada elemento de um *cluster* são tão similares entre si quanto possível e tão diferentes quanto possível, dos elementos dos outros *clusters*.

No método utilizado neste estudo, o número de classes K dever ser definido previamente. A técnica de clusterização utilizada neste trabalho é a denominada *K-mens*. O algoritmo consiste em estabelecer um número máximo de *cluster* que se deseja obter, busca a melhor maneira de separar o conjunto de dados. A idéia é que objetos de um mesmo *cluster* tenham um nível de similaridade maior do que entre objetos pertencentes a *clusters* distintos.

K-means é uma técnica que usa o algoritmo de agrupamento de dados por *K*-médias (*K-means clustering*). O objetivo é encontrar a melhor divisão de P dados em K grupos C_i , $i = 1, \dots, K$, de maneira que a distância total entre os dados de um grupo e o seu respectivo centro, somada por todos os grupos, seja minimizada. Segundo Jain citado por Luz (2004), *K-means* é um algoritmo de clusterização que particiona um grupo de dados em K clusters, representados por seus centros de massa, de acordo com sua distribuição. Iterativamente, divide os dados em conjuntos e estima os centros de massa destes conjuntos, de forma que a medida da equação abaixo, seja minimizada:

$$J = \sum_{j=1}^n \sum_{k=1}^K u_{kj} \|x_j - z_k\|^2$$

onde u_{kj} é a matriz de partição; z_k é o k -ésimo centro de massa ou centróide. Assim, as partições ocorrem seguindo o critério dos mínimos-quadrados. Atualiza-se cada centróide de forma que estes representem a média dos dados a eles pertencentes. Este processo de particionamento seguido de atualização é repetido até que os valores dos centróides se estabilizem, ou seja, que não haja mudanças significativas no valor de J em duas interações consecutivas.

Este método consiste em usar os valores dos primeiros n casos em um arquivo de dados, como estimativas temporárias das médias dos k clusters, onde k é o número de clusters especificado pelo usuário. Assim, o centro do *cluster* inicial é formado para cada caso em torno dos dados mais próximos e, então, comparados com os pontos mais

distantes e os outros clusters formados. A partir daí, dentro de um processo de atualização contínua e de um processo interativo, encontram-se os centros dos clusters finais (Silva *et al* 2001).

A ferramenta *Data Mining* utilizada, Weka é de fácil instação, tem sua implementação efetuada em Java e encontra-se disponível na Internet (<http://www.cs.waikata.ac.nz/~ml/weka>). Apresenta *interface* gráfica para o usuário. O formato do arquivo aceito é o “.arff” que, inicialmente apresenta os valores possíveis dos atributos da base e, em seguida, os dados propriamente ditos. Neste trabalho, foi utilizada a opção *Simplekmeans* para a clusterização.

Os Dados

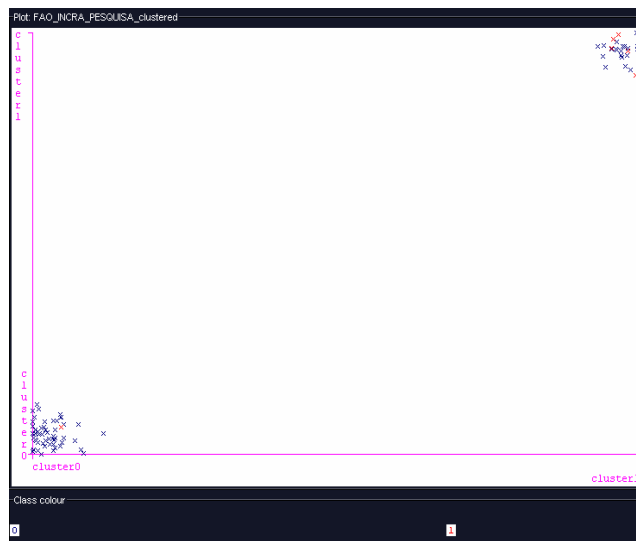
Os dados em *cross section*, foram coletados em 87 propriedades cafeeiras, nos Estados de Minas Gerais, Espírito Santo, Paraná, São Paulo e Bahia, em levantamentos realizados pelo Departamento de Economia Rural da Secretaria da Agricultura – Deral-PR; Instituto de Economia Agrícola - IEA-SP; Empresa de Assistência Técnica e Extensão Rural de Minas Gerais - Emater-MG; Instituto Capixava de Pesquisa, Assistência Técnica e Extensão Rural - Incaper-ES; e Centro Tecnológico de Café - Cetcaf-ES, sob coordenação da Embrapa Café, dentro do projeto de pesquisa Custos de Produção na Cafeicultura Brasileira.

Obedecendo às premissas estabelecidas para se caracterizar agricultura com caráter familiar e patronal, elaborou-se, em Excel, uma planilha, obedecendo aos critérios estabelecidos e procedeu-se a identificação em coluna específica das propriedades em familiar e patronal. A planilha foi complementada com informações úteis para se inferir quanto ao grau de dependência/independência do mercado e se a lógica se aproxima ou se afasta da família. Foram utilizadas informações sobre a participação efetiva da mulher e dos filhos no processo produtivo, tendo em vista sua permanência caracterizar uma opção pela continuidade do negócio e sua transmissão para futuras gerações. A utilização de administradores e encarregados pode ser entendida como um afastamento da família do processo de produção e que possivelmente o proprietário não encontre um herdeiro. A idade do proprietário *vis a vis* a idade dos talhões pode ser entendida como um sinal de perpetuação da propriedade na família. Proprietários jovens, com talhões antigos podem indicar que se tratam de herdeiros, o que não significa que a geração atual seguirá a mesma lógica das passadas. Para inferência da maior ou menor dependência do mercado (considera-se que a cafeicultura não oferece condições de ser independente do mercado, a montante ou a jusante) foram utilizadas as informações sobre processos produtivos mais ou menos dependentes de insumos externos como máquinas, equipamentos, fertilizantes e herbicidas; e dados sobre tratos culturais forneceram subsídios para inferências na dimensão sustentabilidade ambiental.

O cerne da metodologia interpretativa foi a análise tipológica. Os dados dos questionários foram classificados em função de critérios claramente explicitados. Cada grupo é representativo de um “tipo” relativo a uma temática particular. Posteriormente buscou-se estudar as propriedades desses grupos para entender melhor as especificidades do conceito “agricultura familiar na cafeicultura” a partir do conjunto de variáveis que se relaciona com esse conceito.

Resultados e Discussão

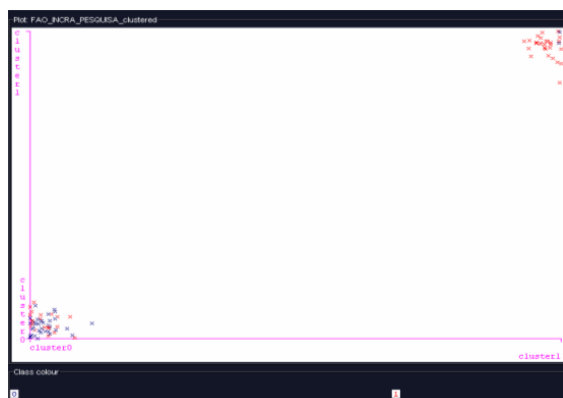
As Figuras 1 e 2 representam a visualização dos clusters e dos atributos determinantes da classificação das propriedades familiares. A exemplo, Administração pelo proprietário e trabalho na propriedade, pelos padrões determinados no estudo FAO-INCRA e trabalho familiar superior ao trabalho contratado .



Legenda: 0 = S(Administrado pelo proprietário) 1 =N (Não administrado pelo proprietário)

Figura 1 – Visualização gráfica dos Clusters formados considerando atributo: Dimensão Administração

Pela visualização da Figura 1 pode-se inferir que tanto as propriedades familiares, como as patronais são administradas pelos proprietários. O cluster “0” apresenta uma concentração maior de propriedades administradas pelo proprietário comparativamente ao cluster “1”. Para o atributo “Dimensão Trabalho”, a tela de visualização do processo de clusterização mostra que no cluster “0” aparecem propriedades em que o trabalho externo supera o familiar e no cluster “1”, é predominante o trabalho externo superar o trabalho familiar no conjunto das propriedades do estudo. Os grupos formados pelas propriedades estudadas, tanto familiares, como patronais apresentam peculiaridades inerentes da agricultura familiar que é ter a maior parte do trabalho executado por membros da família. Fig. 2.



Legenda: 0 = S (Trabalho Familiar igual ou superior ao trabalho contratado)
 1 =N (Trabalho Contratado supera o trabalho familiar)

Figura 2 – Visualização gráfica dos Clusters formados considerando atributo: dimensão trabalho

A análise dos clusters e seus diversos atributos possibilita se inferir sobre a heterogeneidade das propriedades analisadas, pois os atributos são em maior ou menor grau, presentes ou ausentes nos dois clusters. Verifica-se a heterogeneidade dos clusters, apresentando-se os cluster com características comuns, o que impossibilita inferências significativas. Essa heterogeneidade visualizada em todos os atributos foi determinante para se direcionar a análise dos dados considerando os diversos critérios utilizados para a classificação das propriedades familiares: FAO/INCRA; PRONAF e CONTAG, FAO/INCRA PESQUISA e LÓGICA FAMILIAR. (Tabela 1)

Observa-se que a classificação da CONTAG, ao restringir parte dos atributos para sua classificação torna o grupo de familiares muito menor quando comparado com a classificação PRONAF e FAO-INCRA (menos restritiva).

Tabela 1 – Classificação das propriedades da amostra da cafeicultura, conforme definição de familiar x patronal

| CRITÉRIOS | FAMILIAR | % | PATRONAL | % |
|--------------------|----------|-------|----------|-------|
| CONTAG | 17 | 20,48 | 66 | 79,52 |
| PRONAF | 27 | 32,53 | 56 | 67,47 |
| FAO/INCRA | 34 | 40,96 | 49 | 59,04 |
| FAO/INCRA PESQUISA | 42 | 50,6 | 41 | 49,4 |
| LÓGICA FAMILIAR | 53 | 63,86 | 30 | 36,14 |

Fonte: Dados da Pesquisa

Uma quarta classe analisada considerou informações inferidas pelos pesquisadores em campo, denominada FAO-INCRA-PESQUISA, considerando como familiares propriedades que no atributo “Trabalho” não se restringiam aos critérios das classificações anteriores. Essa classe demonstrou ser a mais coerente, não apenas em termos da clusterização final, como também evidenciou correlações de atributos que coadunam com o referencial teórico utilizado. Uma última caracterização foi estabelecida considerando-se como familiares todas as propriedades que obedeciam a pelo menos um dos critérios: Trabalho Familiar maior ou igual ao Contratado, Administração própria, Residência na

Propriedade, Familiares Trabalhando – que foi denominada: LÓGICA FAMILIAR, para este estudo. Nessa caracterização ampliou-se o número das propriedades familiares. Tabela 1.

Com relação ao atributo “Área” distinguem-se claramente propriedades patronais em todos os grupos, todavia, observam-se diferentes concentrações de propriedades familiares e não familiares, no grupo que pelo critério em análise seria considerado familiar. O atributo trabalho nas classificações FAO/INCRA; PRONAF e CONTAG, por ser um atributo determinante dessas classes, determinam claramente o grupo empresarial na cafeicultura neste critério. Por outro lado, o grupo “familiar” por este atributo, contém propriedades patronais (uma vez que este não é o único critério para a classificação final). O atributo estratificação de áreas (CLASSE_AREA) demonstra nos três primeiros conceitos (FAO-INCRA, PRONAF, CONTAG), a nítida presença de cafeicultura empresarial no maior estrato, menores propriedades com predominância do tipo familiar e a partir de 40 ha de área de propriedade, uma predominância da cafeicultura patronal. Também nas duas últimas classes, são patronais as propriedades nos dois maiores estratos, as familiares são predominantes até 20ha de área, confundindo-se nos estratos intermediários. Verifica-se um pequeno número de propriedades com áreas totais acima de 400 ha, consideradas como patronais em todos os critérios.

Os atributos relativos ao proprietário, idade e instrução, também foram analisados utilizando-se a classificação dos cinco critérios estabelecidos neste trabalho. Verifica-se a predominância de cafeicultores mais jovens e mais instruídos na atividade empresarial. Há, portanto, uma classe de cafeicultores empresariais com idade acima de 70 anos (5 casos), assim como há empresários com nível mínimo de escolarização. Com relação ao atributo “Trabalho Familiar” na cafeicultura fica mais evidenciada a diversidade de situações. Encontram-se proprietários, esposas, filhos maiores e filhos menores trabalhando tanto nas propriedades empresariais, quanto familiares, embora sua presença se mostre proporcionalmente maior nas propriedades familiares. A administração da propriedade cafeeira foi analisada do ponto de vista da presença de um administrador não proprietário (em 17 propriedades) e do trabalhador encarregado, num total de 22 propriedades. Predominam administradores e encarregados em propriedades patronais. Verifica-se em um pequeno número de propriedades familiares a presença do administrador externo.

O atributo “Mora na propriedade” é discriminante para o PRONAF e CONTAG, com um total de 38 propriedades em que o proprietário não reside na propriedade, classificadas como patronais para esses critérios. Para os demais critérios, tem-se para propriedades classificadas como patronais o proprietário residindo na propriedade e, em propriedades familiares, tendo a família sua residência fora da propriedade.

Para o atributo “Média de Plantas por ha” (adensamento), que caracteriza a tecnologia atualmente recomendada, observa-se o maior adensamento em duas propriedades patronais, ao mesmo tempo, predominam propriedades patronais nos menores adensamentos. Outro atributo de tecnologia refere-se à fertilização orgânica e química. Pelo referencial teórico, esperava-se que pequenos cafeicultores utilizassem mais a fertilização orgânica e os patronais a fertilização química. A análise evidencia a utilização indiscriminada tanto da fertilização química, quanto da orgânica. A degomagem mecânica é um atributo predominante da cafeicultura empresarial, apresentando em sua totalidade, 11 propriedades utilizando a prática.

O atributo Produtividade evidencia a presença de uma cafeicultura empresarial no nível mais alto, pelos critérios FAO-INCRA, PRONAF e CONTAG, havendo uma predominância de propriedades na média de 25 sacas por ha (32 propriedades). O atributo Grau de Mecanização - caracterizado pelo número de equipamentos em uso na cafeicultura, evidencia não apenas uma relação inversa entre número de propriedades e grau de mecanização, como demonstra a predominância da propriedade patronal mais mecanizada. Constitui um atributo não determinante das classes, mas que caracteriza o tipo de cafeicultura mais dependente de recursos de capital.

Conclusões

O complexo sistema da cadeia produtiva do agronegócio de cafeicultura brasileira foi analisado desde a produção agrícola (familiar, patronal, agroecológica e orgânica). Foram considerados a lógica familiar e o grau de dependência do mercado, os aspectos do cultivo do café adotados pela agricultura familiar e patronal. Estudos comparativos dos diversos critérios atualmente utilizados para a caracterização da agricultura familiar (PRONAF, CONTAG, FAO/INCRA), além de uma caracterização resultante da pesquisa e outra considerando uma tipologia referenciada na lógica familiar.

A análise dos dados fornece uma visão heterogênea da cafeicultura para os dois grupos em estudo, mostrando pelas análises visuais que os atributos se mesclam nos dois tipos de produção o “familiar” e o “patronal”. Verificam-se dentro do grupo familiar atributos que apresentam centróides representativos do grupo patronal, e vice-versa.

Conclui-se pelos resultados das análises deste estudo que o conceito de propriedade familiar necessita de uma decodificação, respeitando-se as especificidades de cada cultura, a exemplo da cafeicultura na qual o envolvimento da família no processo produtivo pode ser considerado mandatário para essa classificação, mas o envolvimento de mão-de-obra externa não pode ser um atributo excludente.

O alto grau de dependência de insumos químicos utilizados tanto no grupo patronal, como familiar mostra que estudos sobre impactos ambientais dessa prática seriam meritórios, principalmente nas propriedades familiares buscando-se a sustentabilidade desses empreendimentos, o que poderia se dar pela implantação de processos produtivos menos agressivos ao meio ambiente e economicamente mais viáveis, como exemplo, poder-se-ia avaliar viabilidade de consórcio com oleaginosas para a produção de biodiesel, plantas medicinais ou especiarias. Outro aspecto que merece estudos mais aprofundados, ainda na vertente da sustentabilidade ambiental, seria o destino das águas residuais da lavagem do café e da degomagem nas propriedades, tendo em vista que este processo se verifica tanto nas propriedades classificadas como “familiar”, como nas patronais.

Referências Bibliográficas

ABRAMOVAY, R. **Capital social**: cinco proposições sobre desenvolvimento rural. – Trabalho apresentado no II Fórum CONTAG de Cooperação Técnica – S.Luis – MA – dez/1998.

LAMARCHE, H. (Coord.) **A agricultura familiar**: comparação internacional. Vol II: do mito à realidade. Campinas: Ed. da Unicamp, 1998.

NEVES, D. **Agricultura familiar**: questões metodológicas” . In: Revista Reforma Agrária, 25: 21-37, maio;dezembro, Campinas, 1995.

SCHNEIDER, S. **Teoria Social, Agricultura Familiar e Pluratividade**. In: Revista Brasileira de Ciências Sociais. São Paulo: v.18, n.51, p.99-121, 2003.

SILVA, O.H. da. **Agricultura Familiar**: diversidade e adaptabilidade. Revista de Sociologia e Política no. 12: 161-167 jun. 1999.

SILVA, A.B.M, PORTUGAL, M.S., CECHIN, A. L. **Redes Neurais Artificiais e Análise de Sensibilidade**: Uma Aplicação à Demanda de Importações Brasileiras. Revecap vol.5 n. 4, 2001.

VEIGA, J. E. da . **A face rural do desenvolvimento. Natureza, território e agricultura**. Porto Alegre, Universidade Federal do RGS, 2.000.