UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

UMA ANÁLISE DA DIVISÃO DIGITAL NO BRASIL ATRAVÉS DA
APLICAÇÃO DA APRENDIZAGEM DE
REDES BAYESIANAS

Luís Fernando Britto Pereira de Mello Barreto

Orientador: Prof. Dr. Hiroo Takaoka

SÃO PAULO
2012

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Reinaldo Guerreiro
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Adalberto Américo Fischmann
Chefe do Departamento de Administração

Prof. Dr. Lindolfo Galvão de Albuquerque
Coordenador do Programa de Pós-Graduação em Administração

LUÍS FERNANDO BRITTO PEREIRA DE MELLO BARRETO

**UMA ANÁLISE DA DIVISÃO DIGITAL NO BRASIL ATRAVÉS DA
APLICAÇÃO DA APRENDIZAGEM DE REDES BAYESIANAS**

Dissertação apresentada ao Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Hiroo Takaoka

Versão Corrigida

(versão original disponível na Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2012

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Barreto, Luis Fernando Britto Pereira de Mello

Uma análise da divisão digital no Brasil através da aplicação da aprendizagem de redes bayesianas / Luis Fernando Britto Pereira de Mello Barreto. – São Paulo, 2012.

148 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2012.

Orientador: Hiroo Takaoka.

1. Tecnologia da informação – Aspectos econômicos 2. Desigualdades sociais 3. Inclusão digital 4. Desenvolvimento econômico I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade II. Título.

CDD – 338.9

Aos meus filhos

Alice e Lipe.

AGRADECIMENTOS

Ao professor Dr. Ronaldo Zwicker, em memória, que me inspirou no caminho a seguir, mas infelizmente não pode me acompanhar ao longo dele.

Ao meu orientador, professor Dr. Hiroo Takaoka, que me acolheu sendo sempre muito solícito e paciente para me ensinar e auxiliar.

Aos professores da FEA que muito me ensinaram e contribuíram na realização deste trabalho, em especial ao professor Dr. Cesar Alexandre de Souza, à professora Dra. Adriana Backx Noronha e ao professor Dr. Nicolau Reinhard.

Aos meus colegas da FEA, especialmente Sergi Pauli, Cláudia Mendes Nogueira e Fernando Fonseca por proporcionarem um ambiente rico para troca de experiências e realizações conjuntas.

À Lícia pela boa vontade e disposição em ajudar a resolver todo tipo de problema.

À dona Roselene Neves pelo carinho e tempo dedicado aos meus filhos, principalmente nos momentos de maior demanda de esforço para elaboração deste trabalho.

Ao amigo e sócio Luiz Eduardo, companheiro de longas discussões sobre qualquer assunto que pareça interessante.

Aos meus pais, João Heitor e Maria Regina pela formação que me deram e o incentivo constante.

À minha esposa Glícia pelo apoio e incentivo constante, além da paciência e companheirismo sem os quais eu não seria capaz de correr atrás de sonhos e realizações.

Finalmente, à Alice e ao Lipe, por fazerem todo o resto valer à pena.

“I don’t pretend we have all the answers.

But the questions are certainly

worth thinking about.”

Arthur C. Clarke

RESUMO

Este trabalho buscou identificar como o Brasil está posicionado no contexto internacional da divisão digital assim como também os fatores que influenciam a divisão digital no país. Utilizando dados do cálculo do índice IDI disponibilizados pelo International Telecommunication Union (ITU) para 152 diferentes países obteve-se um modelo de clusterização a partir do aprendizado de Redes Bayesianas que resultou numa distribuição dos países em oito clusters. Este modelo foi alimentado com subconjuntos incompletos de indicadores correspondentes aos componentes do IDI para diversas subdivisões regionais brasileiras (regiões, estados, regiões metropolitanas/interiores dos estados e zonas rural/urbana), conforme a disponibilidade em diferentes fontes como a ANATEL, o IBGE e o CGI.BR. Desta forma foi possível inferir a que cluster mais provavelmente cada subdivisão regional deve pertencer. Enquanto a posição 64 no *ranking* do IDI coloca o Brasil em uma posição à frente de mais da metade dos 152 países incluídos, a concentração de diversas sub-regiões no sexto cluster coloca a maior parte do país apenas dois níveis acima dos países de pior classificação e a cinco níveis de distância dos melhores classificados. Já a análise dos fatores que influenciam a divisão digital no Brasil utilizou dados fornecidos pelo CGI.BR na obtenção de um modelo diferente de rede *Augmented Naive Bayes* para cada uma de quatro variáveis representativas de adoção tecnológica: uso da internet, uso do celular, adoção do comércio eletrônico e do internet banking. Como variáveis independentes, os modelos incluíram sexo, faixa etária, grau de instrução, situação de emprego, raça, classe social, zona rural/urbana, região e estado, sendo que as que apresentaram maior grau de influência foram a classe social e o grau de instrução. O conjunto de fatores selecionado demonstrou ser bastante preciso apenas na identificação da divisão digital em relação ao uso de internet, enquanto a utilização do celular, do comércio eletrônico e do internet banking não puderam ser totalmente explicadas, sugerindo a necessidade de mais estudos no sentido de identificar variáveis complementares que influenciem a adoção dessas tecnologias.

ABSTRACT

This study aimed to identify how Brazil is positioned in the international context of the digital divide as well as the factors influencing the digital divide within the country. A clustering model was obtained by applying Bayesian Networks learning on data for calculating the IDI index provided by the International Telecommunication Union (ITU) for 152 different countries, resulting in the distribution of the countries among eight clusters. This model was fed with incomplete subsets of indicators corresponding to the various components of the IDI for Brazilian regional subdivisions (regions, states, metropolitan areas/state country sides and rural/urban zones), as available from different sources such as ANATEL, IBGE and CGI.BR. This way it was possible to infer the most likely cluster to which each regional subdivision should belong. The 64 IDI ranking position puts Brazil ahead of more than half of the 152 countries included, but with the majority of sub-regions placed in the sixth cluster, most of the country appears just two levels above the lowest ranking countries and five levels away from the top ranked. The analysis of the factors influencing the digital divide in Brazil used data provided by CGI.BR to obtain a different Augmented Naive Bayes network model for each of four variables representing technology adoption: internet use, cell phone use, e-commerce adoption and internet banking adoption. Gender, age, education level, employment status, race, social class, rural / urban areas, region and state were included as independent variables, with social class and education level showing the most influence. The selected factors proved to be quite accurate in identifying the internet use digital divide, while cell phone use, e-commerce adoption and internet banking adoption could not be fully explained, suggesting the need for further studies to identify additional variables influencing the adoption of these technologies.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	3
LISTA DE GRÁFICOS	4
LISTA DE ILUSTRAÇÕES	5
1 INTRODUÇÃO	7
1.1 Problema de pesquisa e objetivos	10
1.2 Organização do trabalho	11
2 ICT4D: TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO PARA O DESENVOLVIMENTO	13
2.1 Base Teórica	15
2.2 Histórico	16
2.3 Desenvolvimento	20
2.4 Tecnologias de Informação e Comunicação	23
2.5 O papel das Tecnologias de Informação e Comunicação no Desenvolvimento	28
2.6 A pesquisa na área	31
3 DIVISÃO DIGITAL	35
3.1 Indicadores	38
3.2 Fatores de influência	41
3.3 A divisão digital no Brasil	46
3.4 Fontes de dados disponíveis	46
4 MÉTODO	49
4.1 Contextos Metodológicos	49
4.2 O processo de <i>Knowledge Discovery</i> e <i>Data Mining</i> (KDDM)	49
4.3 Redes Bayesianas	53
4.3.1 A aprendizagem de Redes Bayesianas	57
4.3.2 Algoritmos	61
4.3.3 Comparação com outras técnicas de modelagem	63
4.4 Aplicações de Redes Bayesianas	66
4.5 Passos para desenvolvimento do trabalho	67
4.5.1 Compreensão do domínio da aplicação	67
4.5.2 Compreensão dos dados	68
4.5.3 Preparação dos dados e identificação da tecnologia de <i>Data Mining</i>	68
4.5.4 <i>Data Mining</i>	69
4.5.5 Avaliação do conhecimento descoberto	71
4.5.6 Uso do conhecimento descoberto	71
4.6 Dados utilizados	73
5 MEDINDO A DIVISÃO DIGITAL BRASILEIRA	77
5.1 Visão Geral dos dados	79
5.2 Clusterização e inferência	83
5.3 Análise dos resultados	86
6 AVALIANDO OS FATORES DE INFLUÊNCIA NA DIVISÃO DIGITAL NO BRASIL	89
6.1 Adoção de tecnologias	91
6.1.1 Internet	93
6.1.2 Celular	98

6.2	Aplicações da internet.....	103
6.2.1	Comércio eletrônico.....	105
6.2.2	Internet banking	111
6.3	Análise dos resultados.....	117
7	CONCLUSÃO.....	121
	REFERÊNCIAS	123
	Apêndices	133
	Apêndice 1 – <i>Ranking</i> de países em função do IDI.....	133
	Apêndice 2 – Indicadores componentes do IDI para sub-regiões brasileiras (normalizados).....	136
	Apêndice 3 – PIB per capita dos países incluídos no <i>ranking</i> em função do IDI	137
	Apêndice 4 – PIB per capita dos estados e regiões do Brasil.....	141
	Apêndice 5 –Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso da internet	142
	Apêndice 6 – Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso do celular	144
	Apêndice 7 – Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso do comércio eletrônico	146
	Apêndice 8 – Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso do internet banking.....	148

LISTA DE ABREVIATURAS E SIGLAS

ANB: Augmented Naive Bayes
CETIC.br: Centro de Estudos sobre as Tecnologias da Informação e da Comunicação
CGI.br: Comitê Gestor da Internet no Brasil
DAG: Directed Acyclic Graph
DAI: Digital Access Index
DIDIX: Digital Divide Index
DM: Data Mining
DOI: Digital Opportunity Index
DOT Force: Digital Opportunity Task Force
ICI: Internet Connectedness Index
ICT4D: Information and Communication Technologies for Development
ICTD: Information and Communication Technologies and Development
IDI: ICT Development Index
ISI: Information Society Index
ITU: International Telecommunication Union
KD: Knowledge Discovery
KDD: Knowledge Discovery in Databases
KDDM: Knowledge Discovery and Data Mining
KL: Kullback-Leibler
MAP: Máxima a posteriori
MDG: Millenium Development Goals
MDL: Minimum Description Length
NRI: Networked Readiness Index
ODM: Objetivos de Desenvolvimento do Milênio
OECD: Organisation for Economic Co-operation and Development
PIB: Produto interno Bruto
PLS: Partial Least Squares
PNB: Produto Nacional Bruto
RM: Região metropolitana
SEM: Structural equation modeling
SIBIS: Statistical Indicators Benchmarking Information Society
TAI: Technology Achievement Index
TIC: Tecnologias da Informação e Comunicação
UNDP: United Nations Development Programme
WSIS: World Summit on the Information Society

LISTA DE GRÁFICOS

Gráfico 1 - Entropia do lançamento da moeda.....	60
Gráfico 2 - Análise de sensibilidade para o uso da Internet=Sim	95
Gráfico 3 - Probabilidades de uso da internet para os estados das variáveis grau de instrução, faixa etária, classe social e situação de emprego.....	96
Gráfico 4 - Análise de sensibilidade para o uso da Celular=Sim	100
Gráfico 5 - Probabilidades de uso do celular para os estados das variáveis grau de instrução, faixa etária, classe social e situação de emprego.....	101
Gráfico 6 - Análise de sensibilidade para o uso de comércio eletrônico=Sim.....	107
Gráfico 7 - Probabilidades de uso do comércio eletrônico para os estados das variáveis grau de instrução, faixa etária e classe social	108
Gráfico 8 - Probabilidades de uso do comércio eletrônico para os estados da variável estado brasileiro	109
Gráfico 9 - Análise de sensibilidade para o uso de internet banking=Sim	113
Gráfico 10 - Probabilidades de uso do internet banking para os estados das variáveis grau de instrução, faixa etária e classe social	114
Gráfico 11 - Probabilidades de uso do internet banking para os estados da variável estado brasileiro	115

LISTA DE FIGURAS

Figura 1 - Tipologia das TIC	24
Figura 2 - Gráfico acíclico direcionado	54
Figura 3 - Tipos de conexão	56
Figura 4 - Diagrama metodológico adaptado de Kurgan e Musilek (2006) e Lauría e Duchessi (2007).....	72
Figura 5 - Rede Bayesiana dos indicadores de cálculo do IDI (informação mútua e entropia relativa).....	80
Figura 6 - Rede Bayesiana dos indicadores de cálculo do IDI (correlações de Pearson).....	81
Figura 7 - Identificação por cores dos clusters de variáveis da Rede Bayesiana dos indicadores de cálculo do IDI.....	82
Figura 8 - Rede Bayesiana gerada a partir da clusterização dos indicadores de cálculo do IDI.....	83
Figura 9 - Probabilidades associadas aos estados possíveis dos nós da Rede Bayesiana obtida pela clusterização dos indicadores de cálculo do IDI.....	84
Figura 10 - Rede Bayesiana com a relação entre a adoção de tecnologias e variáveis independentes	91
Figura 11 - Rede Augmented Naive Bayes do uso de internet.....	93
Figura 12 - Rede Augmented Naive Bayes do uso de celular.....	98
Figura 13 - Rede Bayesiana com a relação entre uso do comércio eletrônico e do internet banking e variáveis independentes.....	103
Figura 14 - Rede Augmented Naive Bayes do uso de comércio eletrônico.....	105
Figura 15 - Rede Augmented Naive Bayes do uso de internet banking	111

1 INTRODUÇÃO

Ao longo das duas últimas décadas a indústria da tecnologia da informação vem testemunhando uma série de sucessos em um ritmo que parece só acelerar, com as notícias rotineiramente proclamando o acontecimento de novas vitórias empresariais e revoluções tecnológicas (TOYAMA, 2011). A expansão das tecnologias de informação e comunicação (TIC) tem estimulado a produtividade, impulsionado o crescimento econômico dos países, encurtado os ciclos de vida de produtos, diminuído a importância da distância, globalizado mercados e economias e alterado radicalmente a forma como as pessoas vivem e trabalham (PICK; AZARI, 2008).

De ferramentas fundamentais para o ambiente corporativo, os computadores migraram para o ambiente doméstico, aumentaram sua mobilidade e hoje estão em todo lugar na forma de smartphones e tablets. O e-mail, a navegação na internet e as redes sociais tornaram-se integrados na vida cotidiana. Da mesma forma, atividades como e-learning, governo eletrônico e comércio eletrônico são elementos cada vez mais importantes na vida de muitos indivíduos (VICENTE; LÓPEZ, 2011). A internet facilitou enormemente o acesso à informação, servindo também de plataforma para o surgimento de novas tecnologias que ganham rapidamente status de indispensáveis para a vida moderna como o Google, Youtube, Facebook, Skype, Twitter, Messenger, etc.

Apesar desta rápida evolução, a disseminação da tecnologia no cenário mundial não vem ocorrendo de forma uniforme, resultando em uma disparidade crescente no grau de utilização, despesas e disponibilidade de acesso às TIC que criou uma “divisão digital” percebida não apenas entre nações ricas e pobres, mas também entre as regiões mais ou menos favorecidas no interior das nações. Descrita tanto como causa e efeito das desigualdades sociais (PARAYIL, 2005), esta nova forma se junta ao rol de elementos associados às desigualdades sociais, juntamente com as disparidades no acesso à comida, saúde, educação, entre outras.

Esta associação entre o progresso social e econômico no mundo em desenvolvimento e a capacidade de criação, disseminação e utilização das TIC tem recebido uma atenção cada vez maior de governos e da comunidade internacional.

Uma melhor compreensão da divisão digital possibilita que um país avalie melhor o impacto das TIC em suas economias, compare o desempenho interno entre suas regiões e em relação a outros países e identifique áreas prioritárias para investimento, ficando em posição melhor para elaborar políticas públicas mais eficientes na difusão e aproveitamento das TIC.

No entanto, enquanto os governos, as empresas e a sociedade civil ao redor do mundo buscam a melhor forma de aproveitar o potencial das TIC através da criação de agendas políticas e ações proativas, nos meios acadêmicos ainda existe uma falta de consenso em relação à forma como a transição para sociedades de informação ocorre, sem que um quadro conceitual coerente que englobe todos os aspectos deste esforço multidisciplinar tenha sido estabelecido (HILBERT, 2010b).

Tal contexto demanda um esforço no sentido da realização de pesquisas que possibilitem avançar no entendimento do fenômeno, passando principalmente pela busca da mensuração do grau em que a divisão ocorre, assim como da identificação de quais fatores desempenham papel significativo na sua determinação.

Para isso, a obtenção de informações estatísticas sobre o uso das TIC tem demonstrado ser cada vez mais importante, refletindo-se em iniciativas cada vez mais frequentes neste sentido.

No entanto, um dos principais obstáculos para a análise da divisão digital está associado à falta de disponibilidade de dados em contextos regionais, o que faz com que as autoridades muitas vezes não tenham informações suficientes para acompanhar devidamente a adoção tecnológica e avaliar os efeitos das políticas de TIC. Talvez por este motivo a maior parte dos estudos existentes tenha foco nos países como a configuração geográfica relevante na análise da divisão digital, sendo muito escassa a investigação em contextos regionais (VICENTE; LÓPEZ, 2011).

No que diz respeito ao Brasil, o país aparece em diversos estudos que estabelecem a classificação de países em função de determinados índices de adoção tecnológica, mas pouco tem sido realizado em relação à avaliação da divisão interna, com os poucos estudos existentes usualmente se restringindo a contagem de usuários de uma determinada tecnologia. Da mesma forma são raros os estudos que tentam explicar os fatores que levam à ocorrência da divisão.

Este estudo busca contribuir para um melhor entendimento do fenômeno de adoção das TIC no Brasil, assim como permitir avaliar o posicionamento regional interno em um contexto internacional. Em função da constante evolução das TIC assim como do processo de sua disseminação no Brasil, o estudo não se propõe a apresentar um modelo definitivo, mas sim a estabelecer uma forma de análise que possa ser replicada ao longo do tempo, permitindo avaliações contínuas e adaptáveis a contextos diferentes.

A técnica utilizada levou em consideração principalmente dois aspectos associados à análise da divisão digital em seu atual estágio: a pouca disponibilidade de dados e a inexistência de um modelo teórico pré-estabelecido.

Em um estudo que utiliza o aprendizado de Redes Bayesianas associado à adoção de tecnologia, Nedeveschi et al. (2006) afirmam que a dificuldade em decifrar a complexidade da adoção de tecnologia em regiões emergentes utilizando uma metodologia quantitativa se deve em parte à incerteza da informação e em parte à complexa interação de muitas variáveis, sugerindo o aprendizado de Redes Bayesianas como adequado para tratar ambos os problemas.

A dificuldade na obtenção de dados leva Donner e Toyama (2009) a afirmar que nos próximos anos algumas das maiores contribuições do ramo quantitativo para o campo de pesquisa que associa as TIC ao desenvolvimento social e econômico será a partir de dados desorganizados e multivariados, utilizando-se de técnicas que consigam extrair o máximo possível de dados pobres e incompletos.

Em uma Rede Bayesiana, as dependências probabilísticas entre as variáveis permitem que sejam tratadas situações onde se desconheçam uma parte dos dados, uma vez que cada variável conhecida faz com que as probabilidades de ocorrência das demais variáveis sejam atualizadas, reduzindo as incertezas e aprimorando a visão da situação do caso a ser tratado.

Sob o aspecto de modelagem de fenômenos, uma característica interessante das Redes Bayesianas não presentes em outros métodos de modelagem é o fato de não ser necessário basear-se em uma teoria ou modelo teórico pré-existente (LAURÍA; DUCHESSI, 2007). Além disso, elas permitem identificar relações não lineares, o tratamento de variáveis multinomiais e a sua aplicação resulta em uma representação

gráfica do conhecimento obtido a partir dos dados, possibilitando uma análise mais intuitiva e de compreensão mais simples para o cérebro humano.

As motivações pessoais para elaboração deste estudo se deram em função do interesse pessoal deste autor pela área das TIC, principalmente em relação às novas tecnologias, juntamente com a possibilidade de contribuição em um território ainda relativamente pouco explorado e relevante como é o caso da divisão digital.

1.1 Problema de pesquisa e objetivos

De acordo com a proposta apresentada para este trabalho, duas perguntas podem ser destacadas na delimitação do problema de pesquisa:

- Como o Brasil está posicionado no contexto internacional da divisão digital?
- Quais são os fatores que influenciam a divisão digital no Brasil?

Desta forma os objetivos principais a serem alcançados com este trabalho foram:

- Identificar o grau de desigualdade entre as diferentes sub-regiões brasileiras no que se refere à utilização das TIC, situando-as em um contexto de comparação internacional;
- Identificar o grau de influência de diferentes fatores na determinação da divisão digital no Brasil.

Além disso, uma vez que os estudos utilizando aprendizado de Redes Bayesianas na área de ciências sociais são escassos, um objetivo secundário do trabalho é servir de referência para trabalhos futuros de natureza semelhante.

1.2 Organização do trabalho

Para responder a ambas as questões de pesquisa, a abordagem adotada foi a realização de uma revisão da literatura sobre o tema buscando identificar as formas existentes para mensuração da divisão digital, assim como quais os fatores usualmente são citados como seus determinantes. A partir desta identificação foi realizada uma análise buscando combinar os dados existentes e disponíveis com os elementos encontrados na literatura de forma a obter modelos adaptados ao contexto brasileiro.

Para a compreensão da divisão digital é necessário primeiramente entender melhor a relação existente entre o uso das TIC com o desenvolvimento econômico e social. Neste sentido o capítulo 2 faz uma revisão da literatura da área conhecida como ICT4D (*Information and Communication Technologies for Development*) enquanto o capítulo 3 apresenta uma revisão focada especificamente na área de divisão digital.

O capítulo 4 apresenta os aspectos metodológicos, incluindo uma revisão da literatura sobre Redes Bayesianas.

Os capítulos 5 e 6 apresentam os resultados obtidos e as análises visando responder às duas questões de pesquisa propostas, enquanto o capítulo 7 apresenta as conclusões.

2 ICT4D: TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO PARA O DESENVOLVIMENTO

Donner e Toyama (2009) definem a expressão em inglês “*Information and Communication Technologies for Development*” (“Tecnologias de Informação e Comunicação para o Desenvolvimento”) como o conjunto de atividades que considera como as tecnologias eletrônicas podem ser aplicadas no sentido de promover o desenvolvimento socioeconômico de comunidades em desenvolvimento ao redor do mundo. As tecnologias usualmente consideradas dentro deste contexto incluem os computadores pessoais, a internet e telefones celulares, com domínios de aplicação de abrangência análoga ao estudo do próprio desenvolvimento global, incluindo áreas como a saúde pública, a educação, a agricultura, o microcrédito, o saneamento e o ativismo político, entre outras.

A área de ICT4D trabalha na intersecção de dois domínios do conhecimento com definições abrangentes e muitas vezes controversas. Enquanto interpretações diferentes da abrangência das TIC podem considerar elementos tão distintos que vão da imprensa escrita até os tambores africanos (TOYAMA; DIAS, 2008), a especificação do que deve ser considerado na definição do conceito de “desenvolvimento” também tem sido tema de longos debates teóricos (AVGEROU, 2010). No entanto, independentemente da forma utilizada para definir os dois componentes, a pesquisa na área de ICT4D engloba de forma abrangente a consideração de relações humanas e sociais com o mundo tecnológico e especificamente o potencial efeito positivo da utilização de tecnologia no que se refere à mudança socioeconômica (BURREL; TOYAMA, 2009).

O resultado da combinação de uma ampla gama de assuntos faz com que o campo de pesquisas relacionadas com o potencial para o desenvolvimento e o impacto das TIC seja bastante diverso, com o envolvimento de disciplinas que vão desde análises econométricas e etnográficas até a engenharia, fazendo com que a comunidade trabalhando com ICT4D seja particularmente e necessariamente multidisciplinar (DONNER; TOYAMA, 2009).

Entre as disciplinas citadas como contribuintes para a área de ICT4D estão os sistemas de informação, o estudo da interação homem-máquina, estudos sobre comunicação e estudos sobre desenvolvimento (AVGEROU, 2010), assim como também a engenharia,

as políticas públicas, as ciências sociais e a teoria do desenvolvimento (BURREL; TOYAMA, 2009).

Como fator chave a ser alcançado para o progresso da área, Heeks (2008) identifica a necessidade de combinar os domínios intelectuais da ciência da computação, dos sistemas de informação e dos estudos sobre desenvolvimento de forma que cada um busque respostas para as seguintes perguntas:

- O que é possível com a tecnologia digital? (ciência da computação)
- O que é viável com a tecnologia digital? (sistemas de informação)
- O que é desejável com a tecnologia digital? (desenvolvimento)

Se por um lado as disciplinas envolvidas na pesquisa são diversas, o mesmo também é válido para as possíveis áreas de aplicação onde as TIC podem ser utilizadas como fatores geradores de impacto para o desenvolvimento. Em um estudo envolvendo 50 pesquisadores e praticantes com diversos anos de experiência na área de ICT4D, Patra et al. (2009) identificaram as seguintes áreas de atuação como sendo as mais citadas:

- Saúde
- Educação
- Negócios
- Agricultura
- Comunicação e infraestrutura de rede
- Design de interface com o usuário
- Governança

Além do termo ICT4D, o mais usualmente encontrado na literatura, também são bastante utilizadas as expressões e siglas "ICT4Dev" , ICTD ("*ICT and Development*") e "*Development Informatics*" para tratarem da mesma área.

2.1 Base Teórica

Em sua proposta buscando especificar um framework conceitual para a área de ICT4D, Hilbert (2010b) apresenta como base teórica o conceito de evolução socioeconômica que propõe que o progresso humano se dá em unidades cíclicas distintas separadas por períodos intermediários de equilíbrio (SCHUMPETER, 1939 e PEREZ, 2004). Cada uma destas unidades cíclicas se inicia através de um impulso gerado pelo surgimento de atividades empreendedoras que dão origem ao processo de afastamento de uma situação estabelecida de equilíbrio, seguindo-se posteriormente de uma fase de acomodação em torno de um novo ponto de equilíbrio a partir do qual o ciclo se reinicia (SCHUMPETER, 1939). O resultado é um processo incessante de “destruição criativa” que transforma e moderniza a sociedade como um todo, abrangendo aspectos econômicos, sociais, culturais e de organização política.

Cinco ciclos motivados pela força de destruição criativa associada a grandes mudanças tecnológicas podem ser identificados ao longo dos últimos séculos. Tendo como tecnologia motora a utilização da força da água principalmente na mecanização da indústria têxtil, o primeiro ciclo se deu na primeira Revolução Industrial (1770-1850), enquanto o segundo foi baseado na disseminação da tecnologia de mecanização a vapor (1850-1895). O terceiro ciclo se caracterizou pela eletrização das organizações sociais e produtivas (1895-1940), seguido pelo quarto ciclo caracterizado pela motorização baseada em petróleo (1940-1970). Já o ciclo mais recente associa-se com a disseminação das TIC dando início à “Era da Informação” (FREEMAN; LOUÇÃ, 2001). Cada um destes ciclos foi caracterizado por períodos prolongados de modernização social assim como também de aumento da produtividade econômica. De acordo com Perez (1983), este salto na produtividade pode ser visto como uma revolução tecnológica que se torna possível pelo surgimento de um fator chave qualquer que afeta a estrutura geral de custos anteriormente estabelecida e cumpre as seguintes condições:

- Custo relativo percebido claramente mais baixo e descendente;
- Suprimento ilimitado para todos os efeitos práticos;
- Potencial onipresença;

- Capacidade de reduzir os custos de capital, trabalho e produtos, assim como para altera-los qualitativamente.

De acordo com Hilbert (2010b), os requisitos estipulados acima são plenamente atendidos pelas TIC conforme pode ser ilustrado pelas altas taxas de crescimento anual alcançadas pela relação custo-desempenho de computadores e dispositivos de armazenamento e comunicação durante o período 1980-2005 (HILBERT et al., 2010), assim como por um processo de difusão tecnológica que não encontra outro precedente na história humana (ITU, 2009). Além disso, as características das TIC permitem sua aplicação de uma forma geral, afetando todos os aspectos da conduta humana indo desde a formação da vontade política nas pessoas a novas formas de relacionamento pessoal, trazendo também uma modernização da organização social a partir de sua incorporação (PERES; HILBERT, 2010) e um aumento da produtividade mensurável (CIMOLI et al., 2010).

Este processo de transformação social recebeu diferentes denominações na literatura, entre as quais podemos destacar a “Ascensão da Sociedade Pós-industrial” (BELL, 1976), a “Sociedade da Informação” (WEBSTER, 2002), a “Era Digital” (NEGROPONTE, 1995), a “Sociedade em Rede” (CASTELLS, 1996) e a “Era da Tecnologia da Informação e Comunicação” (FREEMAN; LOUÇÃ, 2001).

2.2 Histórico

O primeiro computador digital utilizado em um país em desenvolvimento foi instalado em Calcutá no Instituto Indiano de Estatística em 1956, sendo empregado no trabalho de cálculos científicos (HEEKS, 2008). Deste início até a década de 1990 a computação para o desenvolvimento teve suas aplicações focadas em duas áreas: governamental e corporativa. Inicialmente o governo era o ator principal utilizando as TIC principalmente para funções administrativas internas do setor público. Durante a década de 1980 as multinacionais e outras empresas vieram à tona passando a considerar as TIC como ferramentas úteis para garantir um crescimento econômico no setor privado. Heeks (2009) descreve através da expressão “ICT4D 0.0” este período anterior à criação do termo ICT4D e que durou de meados dos anos 1950 até o final dos anos 1990.

Gomez et al. (2012) descrevem a área de ICT4D como a sucessora das primeiras aplicações do uso de imprensa, rádio, fotografia e televisão como meio de promover o desenvolvimento comunitário, dentro do que era conhecido nos anos 1970 e 1980 como Comunicação para o Desenvolvimento, ou DevCom.

Já de acordo com Patra et al. (2009), o primeiro local de aplicação de ICT4D na forma como hoje é conhecida foi a Escandinávia, através da implantação de salas comunitárias de computador em suas vilas. Começando com a vila de Fjaltring na Dinamarca e depois se espalhando por outros locais, em 1985 já existiam várias pequenas “tele-cabanas” implantadas por toda a Suécia, Dinamarca e Noruega. Alguns anos mais tarde o Sri Lanka se tornou o primeiro país em desenvolvimento a implantar centros comunitários de acesso à tecnologia (QVORTRUP, 1989).

O termo ICT4D começou a aparecer no meio acadêmico e na indústria por volta dos meados dos anos 1990, com pequenas variações alfabéticas como ICTD, ICT4B (bilhões), IT4D e assim por diante. Os anos anteriores tiveram um caráter formativo no crescimento do interesse no assunto em todo o mundo. A abertura das economias do leste europeu coincidiu com períodos fenomenais de crescimento na China e na Índia. O boom da tecnologia nos Estados Unidos contou com um grande grupo de engenheiros oriundos de diversos países em desenvolvimento. Os efeitos econômicos colaterais desta força de trabalho internacional na área de tecnologia se deram a partir do regresso de engenheiros levando conhecimento tecnológico a seus países de origem (VERTOVEC, 2002), mas também através surgimento de empreendedores que criaram uma série de novas empresas de engenharia de pequeno e médio porte (SAXENIAN, 1999).

O discurso da tecnologia como sendo um motor para o crescimento macroeconômico ganhou força rapidamente assim que os impactos iniciais da globalização na indústria de tecnologia manifestaram-se, primeiramente no mundo desenvolvido ocidental, e em seguida em diversas partes do mundo em desenvolvimento. Os responsáveis internacionais por este fenômeno muitas vezes eram técnicos transnacionais expatriados de países em desenvolvimento. Por meados dos anos 1990, o papel do boom das TIC como motivador de uma mudança dramática global ganhou bastante destaque e uma primeira geração de literatura sobre ICT4D discutindo a natureza específica de projetos

de tecnologia no desenvolvimento despertou um interesse mundial para esta área (CASTELLS, 1996; NEGROPONTE, 1995).

Nesta época houve um aumento dramático no número de “Projetos de ICTD”, ou seja, projetos de tecnologia buscando especificamente gerar resultados para o desenvolvimento em seus locais de aplicação. Esta tendência foi parcialmente motivada por uma série de trabalhos de pesquisa e elaboração de documentos de políticas públicas realizados por agências internacionais. Como resultado, na virada do século existiam milhares de telecentros financiados através de diferentes fontes em todo o mundo. A partir do início dos anos 2000, engenheiros começaram a se interessar pela área de ICT4D não apenas buscando atender a uma agenda de desenvolvimento, mas também por apresentar problemas de pesquisa interessantes dentro de seus próprios campos de trabalho. Seguiu-se então a criação de inúmeros espaços para discussão acadêmica, tanto como parte de fóruns existentes anteriormente, assim como através de eventos independentes especificamente criados para o estudo e discussão de ICT4D (PATRA et al., 2009).

Heeks (2009) atribui este crescimento do interesse e das aplicações na área principalmente a dois eventos que aconteceram na década de 1990: a disseminação do uso da internet e a elaboração dos *Millennium Development Goals* (MDGs ou Objetivos de Desenvolvimento do Milênio/ODM). Segundo ele, a internet provocou uma onda generalizada de interesse nas TIC, incluindo um interesse revigorado na forma como elas podiam ser aplicadas nos países em desenvolvimento. Ao mesmo tempo, o desenvolvimento internacional começou ganhar espaço na agenda política, culminando com a formalização dos MDGs em Setembro de 2000 com o objetivo de reduzir a pobreza, melhorar a saúde e educação e promover a igualdade de gênero. O cruzamento destes dois domínios deu origem a ICT4D. Heeks (2009) descreve este segundo período indo do final dos anos 1990 até o final dos anos 2000 como “ICTD4 1.0” onde as aplicações mais comuns foram os telecentros, utilizados para trazer informações sobre questões de desenvolvimento, tais como saúde, educação e extensão rural em comunidades pobres.

De acordo com Heeks (2010) o foco de interesse entre a comunidade ICT4D mudou ao longo do tempo, sendo possível traçar este processo de mudança através de quatro

estágios relacionados com a realização dos eventos mais significativos para discussão do tema ao longo dos últimos anos:

- A *Digital Opportunity Taskforce* (DOTForce) surgiu a partir da cúpula do G8 de 2000, em Okinawa. Em 2001, ela produziu o relatório “*Digital Opportunities for All*” (“Oportunidades digitais para todos”) que tratou de quatro áreas focais (DOTForce, 2001): conectividade, prontidão (*readiness*), capacidade humana e participação em e-networks.
- Em 2003 foi realizado em Genebra o primeiro *World Summit on the Information Society*. O relatório principal foi intitulado “*Building the Information Society*” (“Construindo a Sociedade da Informação”) (WSIS, 2003) procurando principalmente descrever as tecnologias relacionadas ao uso de computadores, com foco no provimento de acesso à infraestrutura e reforço da dimensão técnica das TIC.
- O segundo *World Summit on the Information Society* foi realizado em Túnis, em 2005. O seu foco ainda foi bastante voltado para as áreas técnica e de infraestrutura, mas já puderam ser verificadas as primeiras tentativas de tratar novas aplicações e realizar discussões mais aprofundadas sobre o impacto causado pelo uso das TIC (WSIS, 2005).
- A reunião de maior importância seguinte foi o evento GK3 em Kuala Lumpur, no final de 2007, o último dos grandes eventos internacionais de ICT4D. Este evento ampliou o escopo de abordagem tratando de uma dúzia de diferentes subtemas, incluindo o acesso, a prontidão, a disponibilidade de financiamento, a governança, as aplicações das TIC, a absorção e o impacto (HEEKS, 2010).

Heeks (2009) usa a expressão “ICT4D 2.0” para descrever o período atual, iniciado no final dos anos 2000, sugerindo mudanças do enfoque nos telecentros para o telefone celular como a aplicação arquetípica, um maior interesse no impacto das TIC no desenvolvimento, e mais foco nos pobres como produtores e inovadores na utilização das TIC ao invés de terem apenas o papel de consumidores de informação nelas baseadas.

2.3 Desenvolvimento

Apesar de não ser uma área de estudo recente, o entendimento dado ao conceito de desenvolvimento é de natureza altamente complexa, contraditória e sujeita a diferentes debates e discursos. O termo significa coisas diferentes para pessoas diferentes, baseando-se em contextos econômicos, geográficos, políticos, sociais, culturais, religiosos e étnicos. Além disso, pode ser interpretado a partir das perspectivas de uma série de disciplinas acadêmicas, compondo um conjunto de características que fazem com que o desenvolvimento seja um termo difícil de definir e compreender (REDDI, 2011). A incerteza decorrente da falta de consenso dentro da comunidade ligada ao estudo do desenvolvimento com relação aos aspectos da vida que são ou não importantes como parte dos objetivos de desenvolvimento global é herdada também pela comunidade atuando na área de ICT4D (BURREL e TOYAMA, 2009).

As políticas e ações de desenvolvimento se misturam com interesses conflitantes e relações de poder na política contemporânea em níveis nacional e global. Nos países emergentes as políticas para o crescimento econômico e reformas institucionais adotadas por agências internacionais de desenvolvimento são amplamente contestadas (AVGEROU, 2010).

Enquanto alguns defendem o crescimento econômico e a criação de oportunidades de negócios como as métricas mais adequadas, outros enfatizam a participação e a capacitação das pessoas como condições essenciais para o desenvolvimento (GOMEZ et al., 2012).

Mowlana e Wilson (1988) afirmam que a disseminação do uso da expressão desenvolvimento como um framework conceitual para tratar de uma série de mudanças individuais, institucionais, nacionais, internacionais e do progresso foi um fenômeno iniciado após a Segunda Guerra Mundial. Começando na década de 1940 e ganhando força nas décadas de 1950 e 1960, o termo desenvolvimento tornou-se sinônimo de crescimento, modernização, mudança, democracia, produtividade e industrialização. Popularizado primeiramente entre os estudiosos e os decisores de políticas norte-americanos, o uso do termo se espalhou para a Europa e os países menos industrializados do mundo, transformando-se em uma questão importante nas

organizações internacionais, apesar de seu significado mal definido e sem um reconhecimento universalmente estabelecido.

Uma vez que no mundo ocidental a economia de desenvolvimento foi responsável por direcionar a mudança social planejada durante o pós-guerra, a maior parte das discussões iniciais teve como foco o crescimento econômico assumindo a premissa de uma natureza inerente aos seres humanos baseada no conceito de indivíduos livres operando em mercados livres, tendo a competição como o seu componente principal. Nesse sentido, os indicadores de desenvolvimento estudados tinham natureza estritamente econômica, como o Produto Nacional Bruto (PNB), o PNB per capita, o Produto Interno Bruto (PIB), o PIB per capita e a renda per capita. O entendimento comum era de que o crescimento do PNB e do PIB e o aumento da renda per capita permitiria que o desenvolvimento fosse gradualmente refletido em ganhos para a população em geral, atingindo também a parcela mais pobre (REDDI, 2011).

No entanto, ao longo do tempo foram surgindo evidências de que esta premissa não era sempre válida, reconhecendo-se a ocorrência de casos de aumento da distância entre ricos e pobres mesmo em situações de PIB crescente e percebendo-se a inadequação da utilização de médias em nível nacional como referência. Essa nova percepção gerou uma maior preocupação com o atendimento das necessidades específicas dos segmentos mais pobres da sociedade fazendo com que surgissem formulações mais amplas procurando garantir a qualidade inicial de vida para tantas pessoas quanto possível ao longo de múltiplas dimensões além da puramente econômica (BURREL e TOYAMA, 2009).

Um novo paradigma de desenvolvimento olhando para o processo através de uma abordagem mais humana e centrada nas pessoas foi emergindo ao longo das décadas de 1970, 1980 e 1990, principalmente a partir da obra dos economistas e pensadores, Mahbub ul Haq e Amartya Sen (REDDI, 2011).

Haq (1995) trata o aumento da renda como um meio essencial, mas não como o fim do desenvolvimento, e certamente não como o resumo da vida humana, oferecendo uma nova visão do bem estar humano para o século XXI onde o bem estar real é equiparado ao bem estar das pessoas em suas casas, seus empregos, suas comunidades e seu meio ambiente.

Sen (1999) argumenta que na liberdade individual encontra-se a capacidade de participação política, desenvolvimento econômico e progresso social. O objetivo de todo o desenvolvimento é propiciar o exercício da liberdade de fazer uma escolha e, conseqüentemente, a capacitação de um indivíduo para que ele consiga fazer as escolhas que determinam a sua qualidade de vida. Para os pobres, o exercício da liberdade de escolha é limitado pela inadequação da renda, da educação, da saúde e da igualdade social. Às vezes essas limitações também podem ser um resultado de status socioeconômico, sexo, religião e etnia.

A abordagem humana para o desenvolvimento da qual Haq e Sen foram pioneiros foi introduzida pelo *United Nations Development Programme* (UNDP) em 1990 e culminou com o estabelecimento dos Objetivos de Desenvolvimento do Milênio (ODM) em 2000, ratificados a partir Declaração do Milênio e sendo adotados pelos estados membros das Nações Unidas como um compromisso para ajudar os cidadãos nos países mais pobres do mundo a alcançar uma vida melhor até o ano de 2015 (REDDI, 2011). Os oito tópicos abordados pelos ODM são:

- Erradicar a extrema pobreza e a fome;
- Atingir o ensino básico universal;
- Promover a igualdade entre os sexos e a autonomia das mulheres;
- Reduzir a mortalidade na infância;
- Melhorar a saúde materna;
- Combater o HIV/Aids, a malária e outras doenças;
- Garantir a sustentabilidade ambiental;
- Estabelecer uma Parceria Mundial para o Desenvolvimento.

O estabelecimento dos Objetivos de Desenvolvimento do Milênio não encerrou, no entanto, o debate em torno do conceito de desenvolvimento. Burrell e Toyama (2009) citam a preocupação de se ir além das necessidades básicas, tais como, por exemplo, a importância dada a valores como a expressão criativa e o entretenimento que, apesar de não contemplados pelos ODM, são muitas vezes almejados pelos indivíduos que estes objetivos buscam auxiliar.

Além disso, o surgimento constante de novas questões após o estabelecimento dos ODM torna necessária uma revisão contínua da abrangência da área a ser tratada. Por

exemplo, Heeks (2009) cita três grandes questões que viram sua importância crescer na agenda de desenvolvimento nos anos desde a formalização dos ODMs e que são agrupadas no que ele chama de noção de "desenvolvimento resiliente":

- A segurança, incluindo o terrorismo.
- O crescimento econômico, incluindo a sua variabilidade e fragilidade, como demonstrada pela recessão pós-crise do crédito e a recuperação lenta que se seguiu.
- A sustentabilidade ambiental, particularmente associada às mudanças climáticas.

2.4 Tecnologias de Informação e Comunicação

De acordo com a definição adotada pela UNDP (2001), as TIC são ferramentas de manuseio de informação englobando um conjunto variado de produtos, aplicações e serviços que são usados para produzir, armazenar, processar, distribuir e trocar informações. A abrangência do termo inclui desde tecnologias mais “antigas” como rádio, televisão e telefone até as mais "novas" como computadores, satélites, a tecnologia sem fio e a internet. Com a capacidade de trabalhar conjuntamente, estas diferentes ferramentas formam o "mundo conectado", que consiste de uma gigantesca infraestrutura dos serviços de telefonia interligados, hardware de computação padronizado, a internet, o rádio e a televisão atingindo todos os cantos do planeta.

Segundo Reddi (2011) os dispositivos e aplicações de TIC são usados essencialmente na produção, armazenamento e compartilhamento de informação e conhecimento, consistindo de hardware rodando aplicativos de software que permitem o desenvolvimento de conteúdo e que juntos possibilitam a prestação de serviços ao usuário através de meios de conectividade. Subjacente a todo o processo existe um conjunto de ambientes de suporte físico, de políticas de uso, legais, regulamentares, e financeiras. Com essa definição ela elabora um modelo de tipologia das TIC conforme apresentado na Figura 1 e cujos componentes são detalhados a seguir.

Figura 1 - Tipologia das TIC

FONTE: REDDI, 2011

Hardware

Os equipamentos de Hardware podem ser classificados em dispositivos para:

- Captura de informação, como câmeras, teclados, microfones e gravadores, monitores e scanners.
- Armazenamento de informação, como servidores, discos rígidos, filmes, fitas, CDs, DVDs e cartões de memória.
- Compartilhamento de informação como rádios, TVs, telefones fixos e móveis, leitores de livros eletrônicos e computadores, incluindo laptops e tablets.

Software

Para utilização dos dispositivos de hardware, são necessárias aplicações de software que podem ser proprietários ou *open source*.

O software proprietário é o software licenciado sob o direito legal exclusivo do detentor do copyright, sendo dado ao titular da licença o direito de utilizar o software sob certas condições, restringindo atividades como modificações, distribuição a terceiros ou engenharia reversa. Já o software *open source* está disponível em forma de código fonte fornecido sob uma licença de software que permite aos usuários estudar, mudar e aprimorar o software. O software *open source* pode ser pago ou gratuito.

O software pode ser um sistema operacional ou um aplicativo:

- O sistema operacional é um conjunto de programas que ajudam no controle e gerenciamento do hardware e dos recursos do software de um computador ou outro dispositivo. Em computadores podemos destacar o Windows da Microsoft, o Mac OS da Apple e o *open source* Linux. Já em smartphones e tablets destacam-se o Android da Google, o iOS da Apple e o Windows Phone da Microsoft.
- Os aplicativos, também conhecidos como "apps" são programas projetados para ajudar o usuário executar tarefas únicas ou múltiplas como, por exemplo, o Microsoft Office, softwares corporativos, softwares de contabilidade, softwares gráficos e tocadores de media.

Conteúdo

Citado por Reddi (2011) como talvez o elemento mais importante para as atividades de ICT4D, o conteúdo é a informação ou material criativo que se utiliza de um meio para apresentação. Podendo ser desenvolvido e utilizado no âmbito global, nacional, regional e até local, o conteúdo é a substância que dá sentido ao meio. Ele pode ser definido como texto, áudio, vídeo, filme, gráfico ou animação.

Serviços

O conteúdo é desenvolvido para atender um conjunto de requisitos ou serviços que podem ser classificados como serviços de informação, de educação e de entretenimento. No contexto de ICT4D o foco maior é dado para a informação e a educação, mas alguns autores já demonstram uma preocupação com a necessidade de inclusão do entretenimento na discussão (BURREL e TOYAMA, 2009). Os vários tipos de serviços que podem ser fornecidos através da utilização das TIC incluem:

- Governo para Governo (G2G): sistema conectando escritórios governamentais buscando integrar diversos ramos do governo.
- Governo para *Business* (G2B): interação com o setor privado como, por exemplo, no controle de obras públicas e serviços fornecidos por empresas privadas.
- Governo para Cidadão (G2C): interação do governo com os cidadãos para quem o governo oferece serviços em troca de impostos. O G2C inclui divulgação de informações ao público, serviços básicos ao cidadão, como renovações de licenças, emissões de certidões, entrega de imposto de renda, bem como atendimento ao cidadão para serviços básicos como educação, saúde, entre outros.
- *Business* para *Business* (B2B): troca de produtos, serviços ou informações entre empresas do setor privado.
- *Business* para Cidadão (B2C): fornecimento de produtos ou serviços diretos ao consumidor.
- Cidadão para Cidadão (C2C): uso das TIC por cidadãos e grupos de cidadãos para se comunicar e interagir para uma causa específica ou para compartilhar informações sobre temas de interesse público.

Meios de conectividade

O poder das TIC vem da conexão de diferentes partes em uma rede integrada e convergente através dos meios de conectividade. Os links de telecomunicações que incluem linhas terrestres de telefonia fixa, transmissores de rádio de linha visada, redes sem fio, satélites e cabos de fibra óptica são parte de uma extensa rede de conexão. Para o usuário final os meios de conectividade incluem cabos, modems e roteadores (internos ou externos) para um computador ou telefone celular.

Ambientes de suporte: Físico

Em paralelo com a instalação de torres de transmissão, cabos de fibra óptica e infraestruturas semelhantes de tecnologia sem fio, o cuidado deve ser tomado para assegurar que energia elétrica estável seja fornecida, para que os terminais terrestres e centrais telefônicas e pontos de acesso sejam construídos, e para que os computadores, modems e roteadores também sejam fornecidos. Junto com o estabelecimento de infraestrutura, a manutenção e atualização da tecnologia é um grande problema nos países em desenvolvimento, juntamente com a construção de capacidade técnica para criar, operar, manter e atualizar a tecnologia.

Ambientes de suporte: Social e Regulatório

Os governos têm políticas, leis e regulamentos que regem diferentes setores da sociedade. A política pode ser definida como um princípio orientador, um procedimento ou curso de ação com a intenção de influenciar e determinar decisões e ações para atender aos interesses da sociedade em geral.

Com a dificuldade de previsão e planejamento para o surgimento de novas tecnologias, os governos muitas vezes têm de desenvolver políticas e procedimentos muito tempo depois da difusão da tecnologia na sociedade. Mesmo assim, é tarefa dos governos estabelecer estas políticas e procedimentos através dos quais a tecnologia será introduzida e implementada, de forma que as TIC operem dentro de um conjunto de normas regulamentando o setor.

Sem uma política clara para ICT4D fica difícil priorizar e promover a utilização das TIC para o cumprimento de metas de desenvolvimento. Sem uma política e estrutura de governança, os governos não conseguem medir e aferir o seu progresso em relação ao de outros países.

2.5 O papel das Tecnologias de Informação e Comunicação no Desenvolvimento

Os gastos na implantação de TIC nos países em desenvolvimento alcançam valores bastante expressivos, que em sua grande parte acabam tendo como destino final os países desenvolvidos onde a maioria das sedes das multinacionais da área de TIC se encontra (HEEKS, 2010). Todo este investimento levanta a questão de qual é realmente o grau de contribuição para que os objetivos de alavancar o desenvolvimento sejam alcançados.

Assim como há divergências sobre o que deve ser entendido como "desenvolvimento" e quais as formas de alcançá-lo, também existem divergências sobre a real capacidade das TIC para auxiliar no desenvolvimento (ALAMPAY, 2006).

Para os mais otimistas existe a confiança total na capacidade da tecnologia em resolver problemas sociais profundos como a pobreza. Nicholas Negroponte, fundador do projeto "*One Laptop Per Child*" defende a ideia de que um laptop por criança vai resolver os problemas da educação no mundo em desenvolvimento, enquanto Michael Best, um dos pioneiros do ICT4D intervencionista, chega a afirmar que a internet deveria ser considerada um direito humano colocando-a no mesmo patamar de necessidades como a comida, a água e a segurança física (TOYAMA, 2011). Defende-se que as TIC podem ajudar na redução da pobreza, na expansão das oportunidades para o desenvolvimento econômico e até possibilitar que etapas de desenvolvimento sejam "puladas" acelerando a capacidade dos países em desenvolvimento chegarem ao patamar dos mais desenvolvidos (ALAMPAY, 2006). Este tipo de perspectiva otimista é conhecido nos círculos acadêmicos como determinismo tecnológico, a ideia de que a mera presença da tecnologia resulta em aplicações familiares e padrões desta tecnologia, que por sua vez provocam mudanças sociais (WARSCHAUER, 2003).

Heeks (2009) lista três formas de ações de desenvolvimento que podem contribuir para tratar das necessidades dos pobres e que são incorporadas nas iniciativas de ICT4D:

- Inclusiva: no sentido de aprimorar oportunidades e serviços para que atinjam todas as pessoas, incluindo os pobres como, por exemplo, através de iniciativas de governo eletrônico buscando aumentar a entrega de serviços públicos via internet.
- Capacitação: dando suporte às políticas ou ao contexto no sentido de melhorar a vida dos mais pobres como, por exemplo, buscar incluir os interesses de países mais pobres nos regimes globais que controlam a internet, o tráfego e as tarifas de telecomunicações.
- Foco: visando especificamente os direitos, interesses e necessidades através de iniciativas no sentido de levar diretamente as TIC para comunidades mais pobres em serviços tratando problemas relacionados com a pobreza, saúde, educação e igualdade entre os gêneros.

Os mais pessimistas, no entanto, consideram que existe uma série de problemas que usualmente não são bem resolvidos nas intervenções voltadas ao uso da tecnologia para o desenvolvimento e que resultam no fracasso de projetos como, por exemplo, não projetar uma tecnologia apropriada ao contexto, não estabelecer um bom relacionamento ou parceria com organizações ou governos locais, falhar na aderência a normas socioculturais, desconsiderar a existência de uma infraestrutura deficiente, não buscar a participação da comunidade, entre outros. Nesta linha de entendimento defende-se que a utilização da tecnologia por si só raramente resulta em algum ganho significativo (TOYAMA, 2011).

Sob o ponto vista do impacto, alguns autores consideram que, apesar de suas notáveis capacidades teóricas para estudar a inovação tecnológica em relação ao contexto socioeconômico, existe uma quantidade muito limitada de resultados de pesquisa que demonstram maiores impactos e, conseqüentemente, a contribuição para formação de argumentos convincentes para defesa das TIC como aceleradoras do desenvolvimento socioeconômico é deficiente (AVGEROU, 2010; SEY, 2008).

Heeks (2010) considera, no entanto, que as TIC vem realizando uma contribuição e já foi superada a situação de alguns anos atrás em que as publicações falavam principalmente sobre a capacidade potencial, passando hoje a tratar também de

elementos da cadeia de valor de ICT4D (acessibilidade, infraestrutura e utilização) com destaque de aplicações no:

- Desenvolvimento como crescimento econômico: as TIC contribuindo para o aumento da capacidade de pessoas de comunidades de baixa renda não só economizarem dinheiro, como também aumentarem sua renda.
- Desenvolvimento como meio ambiente sustentável: as TIC auxiliando o desenvolvimento de recursos adicionais e na implementação de novas estratégias ambientais.
- Desenvolvimento como liberdade: as TIC contribuindo para valorização da capacidade de escolha individual em países em desenvolvimento.

É fato que por si só as iniciativas de ICT4D não são capazes de fornecer comida, água potável, acesso à saúde, os direitos civis, ou a paz. No entanto, as tecnologias que facilitam a comunicação aumentam a capacidade das pessoas para aprender e interagir, permitindo que a informação se espalhe ao longo do tempo e do espaço em taxas cada vez mais rápidas (RAITI, 2007).

Em seu estudo realizado com especialistas possuindo anos de experiência na área de ICT4D, Patra et al. (2009) identificaram um sentimento de que o potencial de impacto para desenvolvimento é alto mas que os resultados alcançados até o momento da realização do estudo eram baixos. Isso ficou demonstrado pelo fato de que um grande número de entrevistados preferiu não responder a perguntas sobre os impactos percebidos de ICT4D, e também pelo fato de que poucos entrevistados acreditavam que o trabalho existente teve um impacto significativo até o momento. Quebrando o estudo em possíveis áreas de aplicação, as que foram identificadas pela maioria dos entrevistados como aquelas onde a aplicação das TIC possui potencial de impacto para desenvolvimento (em qualquer nível) foram as áreas de educação, saúde, comunicações, e uma melhor governança. A área onde a esmagadora maioria acredita que o impacto seja significativo foi a de tecnologias de comunicação. Outras áreas de desenvolvimento, tais como meio ambiente sustentável e erradicação da fome foram identificadas como fora do âmbito de contribuição possível das TIC. Uma tendência interessante também identificada foi que 20% dos entrevistados consideraram que a área de ICT4D teve um impacto negativo global na redução da pobreza, e uma pequena porcentagem sentiu que havia um impacto negativo também sobre a governança. Este

fato levanta a questão de um possível risco de impacto negativo da tecnologia podendo resultar em um aumento das desigualdades.

Indo de acordo com este risco de potencial negativo, Toyoma (2011) sugere que o impacto das tecnologias no desenvolvimento internacional funciona como um amplificador de condições previamente existentes e que a tecnologia não é transformadora por si só. As consequências de uma teoria de amplificação para ICT4D são que:

- A tecnologia não pode substituir a falta de capacidade institucional e intenção humana.
- A tecnologia tende a amplificar desigualdades existentes.
- Os projetos de tecnologia no desenvolvimento global têm mais sucesso quando eles amplificam esforços de desenvolvimento já bem sucedidos, ao invés de procurar corrigir, fornecer ou substituir elementos institucionais problemáticos ou inexistentes.

2.6 A pesquisa na área

O maior desafio da pesquisa na área de ICT4D talvez seja conseguir acompanhar a velocidade das mudanças tanto tecnológicas como na sociedade. O surgimento de novas questões na agenda de desenvolvimento nos anos após a formalização dos ODM como o caso do terrorismo, da crise de crédito e da sustentabilidade ambiental citados anteriormente (HEEKS, 2009) são exemplos neste sentido no campo social. No campo tecnológico, a velocidade das mudanças técnicas assim como da difusão da tecnologia implicam a impossibilidade de simplesmente extrapolar os resultados do passado para uma situação atual, com novas avaliações de impacto sendo sempre necessárias em função da evolução da contribuição das TIC para o desenvolvimento (HEEKS, 2010).

Outra questão importante que deve ser considerada na pesquisa de ICT4D é a necessidade de entendimento da contextualização específica de cada trabalho evitando uma extrapolação excessiva dos resultados para outras situações. Os resultados de uma amostragem estatística aleatória são generalizáveis apenas para a população maior a partir da qual a amostragem foi realizada. Por exemplo, não há qualquer garantia de que

um resultado que é válido para a população da Índia, por maior que seja a amostragem, será aplicável para a África Oriental. Da mesma forma, um estudo de caso único raramente é suficiente para fazer afirmações sobre o que está acontecendo em outras regiões, embora possa contribuir para modelos ou hipóteses que ofereçam novas formas de pensar sobre um problema (BURREL e TOYAMA, 2009).

A multidisciplinaridade característica da área também impõe o desafio de obter uma comunicação eficiente em um ambiente com pesquisadores com bases diferentes de conceitos, termos e metodologias (BURREL e TOYAMA, 2009). Raiti (2007) chega a identificar como problema o fato de diversos autores possuírem mais conhecimento da literatura sobre desenvolvimento do que sobre TIC, defendendo a necessidade de uma combinação maior de ambas na troca de ideias entre parceiros multidisciplinares.

Finalmente, o trabalho como pesquisador muitas vezes implica interagir diretamente com as pessoas de um estrato socioeconômico muito diferente de sua origem e em localidades distantes, o que exige uma especial atenção no tratamento de diferenças políticas, econômicas, culturais e ecológicas em relação a experiências pessoais anteriores (BURREL e TOYAMA, 2009).

As características de ICT4D apresentadas, onde dois conceitos abrangentes como as TIC e o desenvolvimento são combinados sob a visão de diferentes disciplinas fazem desta uma área fértil que pode gerar uma lista extensa de focos diferentes de pesquisa como, por exemplo, os citados por Patra et al. (2009): desigualdade digital, a tecnologia e o desenvolvimento setorial e mudança macroeconômica, a tecnologia e a mudança urbana, o potencial da tecnologia para o desenvolvimento, o impacto dos projetos de ICT4D e os estágios da pesquisa de ICT4D.

Já Donner e Toyama (2009) citam três diferentes temas de agrupamento para o foco das pesquisas em ICT4D:

- Definição de usuários: Muitas vezes no centro das avaliações e definições de políticas sobre o relativo sucesso ou fracasso de uma tecnologia, a contagem de usuários tem o desafio de ir além de considerar o número de assinaturas de tecnologias como celulares e internet uma vez que estes números podem tanto superestimar como subestimar a real quantidade de usuários que tem acesso ou fazem uso da tecnologia.

- Diminuição da divisão digital: Sendo talvez o conceito mais popular dentro da área de ICT4D, a divisão digital está relacionada a qualquer situação onde um conjunto de pessoas está usando ou acessando menos de uma dada tecnologia do que outro grupo de pessoas considera desejável. Frequentemente, o conceito está associado à noção de que a divisão leva ou reforça as desigualdades de vários tipos, embora muitos considerem a divisão digital como sintoma e não uma causa das desigualdades. Outra questão bastante discutida é que muitas vezes as análises tendem a considerar o quanto uma tecnologia é utilizada ao invés do uso que é dado para ela na prática.
- Avaliação de impacto: Aponta para o principal pressuposto implícito associado à área de ICT4D de que há um impacto positivo associado à implantação e utilização de tecnologias. Neste sentido ainda existe muito a ser demonstrado na prática de forma a chegar a um claro entendimento do que funciona e o que não funciona para contribuir para o desenvolvimento.

A grande extensão territorial associada ao elevado grau de desigualdade social existente no país tornam o Brasil um caso de particular complexidade e interesse no que se refere à linha de pesquisa que trata da compreensão da divisão digital, sendo este o foco que as questões de pesquisa deste trabalho se propõem a tratar.

3 DIVISÃO DIGITAL

A preocupação com a separação social entre aqueles que têm acesso a computadores e à internet e os que não têm ganhou destaque entre os responsáveis por políticas públicas e os cientistas sociais a partir do início dos anos 90. Inicialmente utilizada pelo departamento *U.S. National Telecommunications and Information Administration* do governo americano, a expressão “digital divide” ganhou popularidade passando a ser empregada internacionalmente para descrever as diferenças de estágios da aplicação da tecnologia da informação entre os países (WARSCHAUER, 2003).

No Brasil, tanto na literatura acadêmica como nos relatórios dos órgãos responsáveis por políticas públicas e nos meios de comunicação em geral, a expressão que mais frequentemente aparece para o tema “*digital divide*” é a “exclusão digital”, assim como a expressão “inclusão digital” é utilizada para as iniciativas buscando sua diminuição. Na opinião deste autor estas expressões não parecem ser as mais adequadas por sugerirem uma classificação binária entre ser “excluído” ou “incluído”, que certamente não atende à complexidade do fenômeno. Além disso, conotações associadas ao termo “exclusão” como “recusa”, “rejeição” e “expulsão” podem falsamente sugerir a necessidade obrigatória de um agente intencionalmente responsável pela sua ocorrência. Assim, traduções menos utilizadas como “divisão digital” ou “fosso digital” parecem mais adequadas, sendo que a primeira foi escolhida para utilização preferencial ao longo deste trabalho.

A divisão digital é um conceito complexo, dinâmico e multifacetado e que, apesar de já ter sido muito debatido, ainda é definido de formas distintas por diferentes correntes de pesquisa, sem um consenso universalmente aceito (BARZILAI-NAHON, 2006; BRUNO et al., 2010). Em uma análise da evolução das definições, White et al. (2011) identificam dois temas como recorrentes: as desigualdades no acesso às TIC e as diferenças na capacidade de obter resultados a partir da sua utilização.

A intenção original que tinha foco em descrever as diferenças existentes entre países foi ampliada para tratar também das disparidades em contextos menores, com a OECD (2001) definindo a divisão digital como diferenças entre indivíduos, domicílios, empresas ou regiões relacionadas ao acesso e uso das TIC. Dewan e Riggins (2005) consideram três níveis diferentes de análise da exclusão digital: individual,

organizacional, e global. Norris (2003) enxerga três dimensões de divisão: global, social e democrática. A divisão global é a diferença no acesso à internet entre os países industrializados e não industrializados, a divisão social é a diferença entre aqueles com e sem acesso à informação dentro de um país, e a divisão democrática representa a diferença entre os indivíduos que utilizam e os que não utilizam recursos digitais para engajamento e mobilização na vida pública.

Apesar de adotada por diversos autores, a abordagem contrastando aqueles “que têm” acesso em relação aos “que não têm” se demonstra muito simplista para a análise da complexidade da realidade da relação que diferentes pessoas, vivendo em contextos bastante distintos, têm com o uso da tecnologia (BRUNO et al., 2010). Conforme um exemplo apresentado por Warschauer (2003), um norte-americano que acessa a internet uma vez por mês em uma biblioteca local pode facilmente ser enquadrado como um “que não têm”, enquanto uma pessoa com o mesmo perfil vivendo em um país pobre pode ser considerado um “que têm”. Para ele, uma distinção simplesmente binária não é suficiente, sugerindo uma interessante analogia entre a exclusão digital e a alfabetização:

A alfabetização não existe em uma divisão bipolar entre aqueles que absolutamente conseguem e não conseguem ler. Existem níveis de alfabetização para objetivos funcionais, profissionais, cívicos, literários e acadêmicos. E as pessoas se tornam alfabetizadas não apenas através do acesso físico aos livros, mas através da educação, comunicação, relações de trabalho, apoio familiar e assistência de redes sociais (WARSCHAUER, 2003)

Desta forma a divisão digital não pode ser tratada simplesmente como um problema de fornecimento de acesso, devendo-se levar também em consideração vários fatores de influência tais como questões históricas, socioeconômicas, geográficas, educacionais, comportamentais, de geração, ou da incapacidade física dos indivíduos (VEHOVAR et al., 2006).

Para Alampay (2006), a abordagem de capacidades de Sen (1999) se aplica a divisão digital uma vez que, apesar do acesso constituir uma boa base e ser pré-requisito fundamental, diferenças individuais de capacidades e escolhas desempenham um papel importante sobre as formas de uso, as aplicações e o valor dado pelas pessoas à tecnologia.

A multiplicidade de divisões pode ser mais ou menos estruturada em algumas categorias centrais. A primeira e mais usual divisão digital é geralmente definida como simples diferenças na disponibilidade de acesso e frequência de uso da internet ou alguma outra tecnologia. Em um segundo estágio, os não usuários podem ser ainda diferenciados com relação aos obstáculos que os levam a não utilizar a internet, tais como a falta de interesse ou falta de conhecimento do que os computadores podem oferecer. Por outro lado, os usuários da internet também podem ser segmentados com base em habilidades e experiência. Aqueles que sabem como usar as opções mais avançadas tem uma vantagem sobre aqueles que só usam serviços mais simples. O primeiro nível de exclusão que se refere a diferenças no acesso e uso inevitavelmente desaparece quando a internet se torna universalmente acessível. No entanto, a divisão digital relacionada com a experiência e uso avançado continuará existindo. Os usuários da internet também podem ser diferenciados de acordo com a tecnologia utilizada. Aqueles com acesso mais rápido, como a banda larga, mais conveniente, como através do telefone celular ou com acesso a dispositivos mais avançados e caros tem maior capacidade de usufruir dos serviços e tecnologias de nova geração do que aqueles com dispositivos mais primitivos e conexões discadas (VEHOVAR et al., 2006).

Sob o ponto de vista teórico há uma clara falta de um framework conceitual estabelecido. Segundo Vehovar et al. (2006) algumas teorias com origem em outras áreas podem em princípio serem aplicadas, como por exemplo a teoria da difusão da inovação (ROGERS, 1995), teorias de audiência, particularmente a teoria dos usos e gratificações (BLUMLER; KATZ, 1974), ou teoria das diferenças de conhecimento (TICHENOR et al., 1970). Teorias que estudam os níveis individuais e estruturais de exclusão social e desigualdade social (por exemplo, de gênero, econômica, racial ou étnica) também podem ser usadas (ATKINSON; HILLS, 1998; MUFFELS et al., 2002; RODGERS et al., 1995; SILVER, 1994). No entanto, a relação entre a exclusão social e a digital ainda tem de ser explicada (HUSING et al., 2006) uma vez que ainda não está claro se a exclusão digital é um tipo particular de exclusão social ou se as TIC funcionam como um fator de diminuição da exclusão social. Infelizmente, muitas vezes parece que o crescimento explosivo da internet na verdade agrava as desigualdades existentes (MENOU, 2001; NORRIS, 2003). Além disso, Parayil (2005) argumenta que a exclusão digital é tanto um sintoma como uma causa da desigualdade social e econômica. De qualquer forma, é consenso geral na literatura acadêmica que a divisão

digital está intimamente relacionada com a desigualdade social (ATTEWELL, 2001; BONFADELLI, 2002; DIMAGGIO et al., 2001; MENOU, 2001)

Na literatura existente sobre divisão digital são encontrados dois principais tipos de estudo. O primeiro tem foco em quantificar a extensão e evolução da divisão através da elaboração de índices compostos e *rankings* de países. Já o segundo busca entender e/ou explicar o porquê da existência digital através da identificação e avaliação de fatores de influência determinantes (BILLON et al., 2009; VICENTE; LÓPEZ, 2011).

Cada uma das duas questões de pesquisa deste trabalho está relacionada diretamente com um destes focos, que são detalhados abaixo.

3.1 Indicadores

Entre os estudos que buscam quantificar a divisão digital, tem sido cada vez mais frequente a tentativa de capturar a multidimensionalidade do fenômeno através da elaboração de indicadores ou índices compostos que têm se demonstrado úteis para resumir a complexidade da divisão digital entre regiões, facilitando a interpretação das lacunas através da construção de *rankings* (VICENTE; LÓPEZ, 2011).

Os indicadores podem ser agrupados em categorias de acordo com seu significado, a partir de agregações realizadas com base em algum modelo matemático subjacente. Desta forma os índices compostos diferem entre si principalmente em função da base dos indicadores selecionados e da metodologia empregada na agregação. A escolha dos indicadores a serem considerados usualmente depende do framework conceitual adotado para descrever e representar o processo de penetração e difusão das TIC entre os países (BRUNO et al., 2010). No entanto, Barzilai-Nahon (2006) destaca que uma série de índices propostos não define e conceitualiza a exclusão digital previamente, iniciando o seu processo de design a partir das variáveis e níveis de indicadores, tentando chegar a fatores que são mensuráveis e ignorando o que é realmente significativo em um determinado contexto específico. Vehovar et al. (2006) também ressaltam a necessidade de uma justificação sólida das escolhas de indicadores e cálculos empregados uma vez que a flexibilidade pode resultar em valores diferentes e até conflitantes, além de facilitar a ocorrência de manipulação de resultados. Um exemplo por eles apresentado de que isto pode ocorrer é a comparação de dois países

européus (Eslovénia e Estónia) de acordo com três indicadores de TIC. Considerando-se o uso da internet, a diferença entre os dois países fica relativamente alta (de 37% na Eslovénia contra 52% na Estónia), sendo um pouco menor entre os usuários de computador (de 49% na Eslovénia contra 55% na Estónia). No entanto, considerando-se apenas a quantidade de usuários de internet a partir de suas casas, a Eslovénia passa a ficar em melhor posição (34% contra 27%). Desta forma um indicador simplista aplicando-se a média das três variáveis resultaria em uma diferença aparente muito menor do que a real: 40% (Eslovénia) versus 45% (Estônia).

Sob o ponto de vista da técnica utilizada, muitos autores têm incorporado as várias dimensões do desenvolvimento digital em um índice por meio de análise fatorial, como por exemplo, Corrocher e Ordanini (2002) que identificam seis fatores posteriormente agregados em um índice sintético aplicado para medir a divisão digital entre dez países desenvolvidos e Vicente e López (2011) que utilizam análise fatorial a fim de classificar os níveis de desenvolvimento digital entre os países da União Europeia. Alguns outros exemplos de técnicas multivariadas utilizadas para medir a exclusão digital incluem os trabalhos de Çilan et al. (2009), que utilizam a análise de variância multivariada e a análise discriminante para avaliar a divisão digital entre os estados membros e candidatos da União Europeia, e de White et al. (2011), que utilizam análise de clusterização de múltiplos estágios baseada em modelo para gerar um mapa mundial de exclusão digital.

Dentre os indicadores que tiveram maior destaque entre os utilizados por organizações internacionais na elaboração de políticas públicas, Vehovar et al. (2006) citam o *Digital Access Index* (DAI), o *Digital Opportunity Index* (DOI), o *Networked Readiness Index* (NRI), o *Technology Achievement Index* (TAI), o *Information Society Index* (ISI), o *Internet Connectedness Index* (ICI), o Infostate e o *Digital Divide Index* (DIDIX) desenvolvido pelo projeto SIBIS (*Statistical Indicators Benchmarking Information Society*).

Em um estudo mais recente, Bruno et al. (2010) traçam um histórico da evolução dos índices propostos por organizações como a *International Telecommunication Union* (ITU), conforme apresentado no Quadro 1:

Quadro 1 – Características dos principais índices para medida de exclusão digital

Índice	Nome	Ano	Instituição	Número de categorias	Número de indicadores	Metodologia
DAI	<i>Digital Access Index</i>	2003	ITU	5	8	Média das pontuações das categorias que são médias dos respectivos indicadores
Infostate	Infostate	2003	Orbicom	2	21	Média geométrica da pontuação das categorias que são médias geométricas dos respectivos indicadores
DOI	<i>Digital Opportunity Index</i>	2005	ITU	3	11	Média das pontuações das categorias que são médias dos respectivos indicadores
ICT-OI	<i>Information and Communication Technologies- Opportunity Index</i>	2005	ITU- Orbicom	2	10	Média geométrica da pontuação das categorias que são médias geométricas dos respectivos indicadores
IDI	<i>Information and Communication Technologies Development Index</i>	2009	ITU	3	11	Média das pontuações das categorias que são médias dos respectivos indicadores

FONTE: BRUNO et al. (2010)

Dos índices citados, o *ICT Development Index* (IDI) é o mais recente, tendo sido escolhido como base para a análise realizada neste trabalho. Ele representa a evolução e a síntese dos índices anteriores sugeridos pelo ITU (BRUNO et al. ,2010), sendo composto por uma combinação de 11 indicadores em uma medida de referência para monitorar e comparar a evolução das TIC em todos os países.

O IDI foi desenvolvido pelo ITU em 2008 e primeiramente apresentado na edição 2009 do relatório *Measuring the Information Society*. Ele foi criado como resposta do ITU ao pedido de seus países membros para desenvolver um "índice único" e publicá-lo regularmente. Os 11 indicadores são distribuídos entre três categorias voltadas a medir acesso, uso e habilidades. Os seus componentes assim como os respectivos pesos no cálculo do índice são apresentados no Quadro 2. Para maiores detalhes sobre as definições de cada um dos indicadores assim como a metodologia utilizada sugere-se consultar a edição 2011 do relatório *Measuring the Information Society* (ITU, 2011).

Quadro 2– Indicadores usados no cálculo do ICT Development Index

Categoria	Peso da categoria	Indicadores	Peso do Indicador
Acesso às TIC	0,4	Assinaturas de telefone fixo por 100 habitantes	0,2
		Assinaturas de telefone celular por 100 habitantes	0,2
		Banda internacional de internet por usuário de internet	0,2
		Percentual de domicílios com computador	0,2
		Percentual de domicílios com acesso a internet	0,2
Uso das TIC	0,4	Percentual de indivíduos usando a internet	0,33
		Assinaturas de banda-larga fixa por 100 habitantes	0,33
		Assinaturas de banda-larga móvel ativas por 100 habitantes	0,33
Habilidades para TIC	0,2	Taxa de alfabetização de adultos	0,33
		Taxa de matrícula no secundário	0,33
		Taxa de matrícula no terciário	0,33

FONTE: BRUNO et al. (2010)

3.2 Fatores de influência

Segundo Vicente e López (2011) a técnica mais comumente empregada em estudos buscando explicar as causas da divisão digital é a análise de regressão múltipla, onde as variáveis dependentes selecionadas como indicadores representativos da adoção de uma determinada tecnologia mais frequentemente utilizadas são a quantidade de usuários da internet, (ANDRÉS et al., 2010; BEILLOCK; DIMITROVA, 2003; CHINN; FAIRLIE,

2010; GUILLÉN; SUÁREZ, 2005; HAWKINS; HAWKINS, 2003), a quantidade de hosts (HARGITTAI, 1999, KIISKI; POHJOLA, 2002), telefones celulares (QUIBRIA et al., 2003; WONG, 2002), hosts de e-commerce seguros (WONG, 2002), as despesas com hardware de computador per capita (POHJOLA, 2003), as importações de equipamentos de informática (CASELLI; COLEMAN, 2001), e a penetração da banda larga (CAVA-FERRERUELA; ALABAU-MUÑOZ, 2006; DISTASO et al., 2006). Já Billon et al. (2009) utilizam a correlação canônica para analisar simultaneamente um grande número de variáveis dependentes (usuários de computadores, usuários de internet, assinantes de banda larga, telefones celulares, linhas telefônicas e servidores de internet seguros).

A partir da seleção dos indicadores representativos são realizadas análises buscando determinar o grau de influência de uma série de fatores como renda, ocupação, sexo, idade, educação, localização geográfica, etnia e raça, religião, linguagem, estrutura familiar, capacidade física, experiência, autonomia, acessibilidade, estrutura competitiva do mercado, a propriedade e a densidade de computadores e web sites, infraestrutura de comunicação, equipamentos, apoio social, a estrutura política (HANAFIZADEH, 2009). Alguns dos fatores de mais destaque na literatura são detalhados abaixo:

Renda

Estudos empíricos demonstram que o principal fator subjacente às disparidades na adoção das TIC são as diferenças de riqueza econômica tanto entre regiões como entre indivíduos (ANDONOVA, 2006; CRENSHAW; ROBISON, 2006; DIMAGGIO et al., 2004; FAIRLIE, 2004; HARGITTAI, 1999 ; KIISKI; POHJOLA, 2002; DEWAN et al., 2005; NORRIS, 2003; POHJOLA, 2003). Alguns estudos como os de Chinn e Fairlie (2010) e Beilock e Dimitrova (2003) identificam a renda per capita como o fator de maior importância para explicar as diferenças nas taxas de penetração das TIC.

Sexo

Estudos como os de Li e Kirkup (2007) e Hills e Argyle (2003) identificam diferenças no nível de utilização da internet relacionada ao sexo e fornecem evidência de um predomínio masculino e uma resistência à participação feminina. Já Punamaki et al.

(2007) identificaram em um estudo realizado com adolescentes na Finlândia que enquanto os meninos usavam o computador para jogos, escrever e enviar e-mails e navegar pela internet com mais frequência do que as meninas, estas eram usuárias mais frequentes de telefones celulares.

Em função dos contextos sociais e culturais que afetam o uso e desenvolvimento das TIC, estas nunca se demonstram neutras em relação ao sexo de seus usuários. Segundo Alampay (2006), tem se demonstrado que homens e mulheres não partem de uma base de igualdade no que diz respeito ao estabelecimento de direitos e cidadania social, principalmente pelo desequilíbrio em função das maiores responsabilidades que as mulheres tem em casa e desta forma é esperado que o acesso às facilidades de comunicação também seja afetado, com os homens tendo mais oportunidades.

Educação

De acordo com Vicente e López (2011), a maioria dos estudos aponta que a adoção de inovações é mais rápida entre pessoas com níveis educacionais mais altos. Tecnologias que requerem uma maior interatividade com o usuário (como é o caso da internet) exigem que o usuário possua competências adequadas para realizar a busca, análise e uso da informação (VICENTE e LÓPEZ, 2006). Desta forma, regiões com proporções maiores de pessoas de nível educacional mais elevado estarão mais propensas a apresentar uma penetração maior das TIC do que naquelas que contam com menos indivíduos qualificados (CRENSHAW; ROBISON, 2006; HARGITTAI, 1999; VICENTE e LÓPEZ, 2006). Billon et al. (2007) identificam uma correlação positiva entre a proporção regional da população com estudo superior e a adoção da internet e do e-commerce.

Idade

Segundo Alampay (2006) espera-se que os jovens tenham maior motivação para utilizar as TIC. A pesquisa empírica com base em micro dados tem demonstrado que a adoção de TIC está associada principalmente às gerações mais jovens (FAIRLIE, 2004). A diferença de idade no uso das TIC pode ser o resultado de uma série de fatores, incluindo habilidades e atitudes, assim como as necessidades percebidas (COHENDET,

2003; SCIADAS, 2002), além do aspecto de que o uso para fins educativos pode também gerar uma forte demanda por tecnologia entre as crianças (FAIRLIE et al., 2010). Neste contexto, espera-se que regiões com proporção maior de população jovem tenham maiores taxas de penetração da internet (CHINN; FAIRLIE, 2007).

Localização

Em relação à localização, tem sido demonstrado que a divisão digital não existe apenas entre países, mas também entre áreas urbanas e rurais. O tamanho da população, a sua densidade e o grau de urbanização apresentam correlação com a exclusão digital (VICENTE e LÓPEZ, 2011). A teoria da densidade urbana destaca a associação negativa entre o custo de adoção das TIC e o tamanho e densidade da população (FORMAN et al., 2005). O acesso a estas tecnologias é mais fácil e mais barato nas cidades do que em áreas rurais em função de uma melhor infraestrutura de telecomunicações e de custos menores para implantação de novas infraestruturas. Além disso, as cidades tendem a concentrar recursos mais altamente qualificados para o trabalho (BAPTISTA, 2001; KAUFMANN et al., 2003).

Sob a perspectiva do aprendizado social, as pessoas que vivem nas cidades podem ter custos mais baixos para começar a usar a internet, uma vez que podem aprender a utilizá-la com vizinhos e conhecidos (AGARWAL et al., 2009) influenciando positivamente a difusão (BILLON et al., 2007). Assim, um maior grau de difusão das TIC é esperado em regiões com maior proporção de população urbana (CRENSHAW; ROBISON, 2006; LIU; SAN, 2006).

Raça e Idioma

Diversos estudos têm investigado os determinantes da exclusão digital entre os grupos raciais (FAIRLIE, 2004). Tais lacunas parecem estar relacionadas principalmente, mas não exclusivamente, a desigualdades de renda e educação (FAIRLIE, 2004).

Em relação ao idioma, tem sido dada atenção especial ao papel do Inglês uma vez que esta é a língua da maioria dos conteúdos on-line, gerando uma expectativa de que competências linguísticas individuais pudessem influenciar o interesse na internet

(HARGITTAI, 1999). No entanto, a investigação não produziu quaisquer resultados conclusivos (BILLON et al., 2009; CHINN; FAIRLIE, 2007; HARGITTAI, 1999, KIISKI; POHJOLA, 2002; LIU; SAN, 2006; HAWKINS; HAWKINS, 2003).

Institucionais

O ambiente institucional legal de um país é outro fator chave que contribui para exclusão. Hargittai (1999) apresenta evidências de que países com um monopólio no setor de telecomunicações apresentam níveis significativamente mais baixos de adoção da internet enquanto Guillén e Suárez (2005) mostram que a concorrência no serviço de telefonia local aumenta a penetração da internet. Mais recentemente, Andres et al. (2010) identificam uma correlação positiva entre o número de provedores de serviços de internet e as taxas de adoção.

A qualidade das instituições e eficácia do governo também estão associados com a penetração das TIC com vários autores identificando o estado de direito como um fator significativo para explicar diferenças entre países na adoção da internet (BILLON et al., 2009; CHINN; FAIRLIE, 2007; CHINN; FAIRLIE, 2010). A proteção dos direitos de propriedade (CASELLI; COLEMAN, 2001), das liberdades civis (ANDONOVA, 2006; BEILOCK; DIMITROVA, 2003), a estabilidade política (LIU; SAN, 2006), os direitos políticos e o clima de investimento também afetam a difusão da internet (ANDONOVA, 2006).

Da mesma forma, o nível de abertura comercial da economia pode ser relevante para a adoção das TIC, esperando-se um impacto positivo e significativo. No entanto, a evidência empírica tem trazido poucas respostas em relação a esta questão (CASELLI; COLEMAN, 2001, CHINN; FAIRLIE, 2007; CHINN; FAIRLIE, 2010; CRENSHAW; ROBISON, 2006; HARGITTAI, 1999; DEWAN et al., 2005; VICENTE e LÓPEZ, 2006).

3.3 A divisão digital no Brasil

Em um dos poucos estudos encontrados que incluem uma análise dos fatores de influência na divisão digital no Brasil, Hilbert (2010a) observa que os determinantes amplamente reconhecidos da desigualdade socioeconômica são também os que definem a exclusão digital, incluindo renda, educação, habilidades, emprego, geografia, idade, gênero e etnia, entre outros. Dentre estes, ele afirma que a grande desigualdade da distribuição de renda na América Latina faz com que não seja surpresa o destaque dado às questões estruturais da renda.

Com base em uma pesquisa domiciliar de 2007 no Brasil, o estudo de Hilbert (2010a) seleciona 10 variáveis potenciais para explicar o acesso das famílias à internet: renda domiciliar per capita, nível pessoal de educação, tamanho domiciliar (famílias de uma ou duas pessoas em relação a famílias maiores), idade, matrícula corrente em escola ou instituição de ensino, categoria de trabalho, região geográfica (urbana/rural), instalação de TV em cores no domicílio, gênero (masculino/feminino), e etnia (como auto identificação de pertencer a um grupo indígena ou não). Os resultados do estudo demonstram que o nível de renda é de longe o mais forte determinante para o acesso à internet, seguido pelo nível de educação, o tamanho do agregado familiar e a idade. O mesmo exercício realizado para dados da pesquisa domiciliar para o Brasil em 2005 e 2002 mostrou que estes resultados foram estáveis ao longo do tempo.

3.4 Fontes de dados disponíveis

As principais fontes de dados para análise da divisão digital são os levantamentos realizados com base nas orientações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), do Instituto de Estatísticas da Comissão Europeia (Eurostat), e do Observatório para a Sociedade da Informação na América Latina e Caribe (Osilac), que pertence à Comissão Econômica para a América Latina e Caribe das Nações Unidas (Cepal).

Dados consolidados destes levantamentos podem ser obtidos a partir de bases de dados fornecidas por diferentes organizações, com destaque para o *World Telecommunication*

Development database do *International Telecommunication Union* (ITU) e as bases de dados disponibilizadas pelo Eurostat.

No Brasil, um dos principais responsáveis pela realização de diversos projetos de pesquisa visando a medição e o acompanhamento da expansão das TIC no país é o Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (Cetic.br) do Comitê Gestor da Internet no Brasil (CGI.br). Criado em 1995, o CGI.br coordena projetos voltados para o desenvolvimento da internet no país, fomentando a coleta, a organização e a disseminação de informações, análises, indicadores e estatísticas sobre as TIC no Brasil.

O presente trabalho utilizou as informações disponibilizadas pelo ITU para análise da divisão digital em relação a diferentes países (ITU, 2011) e os dados do levantamento TIC Domicílios e Empresas 2010 - Pesquisa sobre uso das tecnologias da informação e comunicação no Brasil (CGI.BR, 2010). Além disso, foram utilizados também dados obtidos a partir de relatórios disponibilizados na internet pela ANATEL e pelo IBGE. É importante destacar que o estudo buscou manter o máximo de compatibilidade possível entre as diferentes fontes no que se refere aos períodos de referência em que os dados foram obtidos, com praticamente todas as informações correspondendo ao ano de 2010.

4 MÉTODO

4.1 Contextos Metodológicos

Este é um trabalho de natureza quantitativa que emprega métodos baseados em técnica de mineração de dados e aprendizado de máquina na análise de dados de natureza secundária, uma vez que foram originalmente coletados por outras pessoas com distintas finalidades e propósitos.

Alguns estudos demonstram a possibilidade de aplicação das Redes Bayesianas com resultados equivalentes ou até superiores a técnicas mais comumente encontradas em estudos de natureza semelhante a este, como é o caso da regressão logística e da análise fatorial (KARCHER, 2009, CONRADY; JOUFFE, 2011c). A opção por esta técnica se deu principalmente em função da possibilidade de utilização dos modelos obtidos para simulações e inferência a partir de dados incompletos. Além disso, a obtenção de uma representação gráfica como resultado de sua aplicação permite análises mais simples e intuitivas.

A pesquisa pode ser classificada como exploratória, que é aquela realizada para esclarecer situações ambíguas e determinar a dimensão de um problema, sem a intenção de apresentar uma conclusão definitiva, mas fornecer uma base para orientar e aperfeiçoar esforços de investigação posteriores (ZIKMUND; BABIN, 2006 p.51).

4.2 O processo de *Knowledge Discovery* e *Data Mining* (KDDM)

Usados muitas vezes como sinônimos para o processo de extração de informações a partir de dados, os termos *Data Mining* (DM), *Knowledge Discovery* (KD), e *Knowledge Discovery in Databases* (KDD) na realidade se referem a partes diferentes de um processo mais amplo.

A expressão *Knowledge Discovery* se refere a um processo amplo que busca conhecimento sobre um domínio de aplicação. Ele consiste em diversas etapas (uma delas sendo o *Data Mining*), cada uma com o objetivo de completar uma tarefa

particular de descoberta e sendo alcançado através de um método de descoberta (KLOSGEN; ZYTKOW, 1996). *Knowledge Discovery in Databases* é a aplicação deste processo a bases de dados, (KLOSGEN; ZYTKOW, 1996) enquanto *Data Mining* se refere apenas à etapa do processo em que se utiliza a aplicação de algoritmos específicos para extrair padrões dos dados (FAYYAD et al., 1996).

Neste contexto, o termo *Knowledge Discovery and Data Mining* vem sendo proposto como uma alternativa mais abrangente e adequada para se referir ao processo geral da aplicação do *Knowledge Discovery* a qualquer fonte de dados (CIOS; KURGAN, 2005).

Fayyad et al. (1996) apresentaram um primeiro modelo para o processo de KDDM constituído por uma sequência de nove etapas, enfatizando a importância de que o produto final apresente um significado útil para o usuário e a natureza iterativa do processo podendo-se retornar a qualquer etapa anterior para retrabalho. A sequência proposta consiste em:

- 1) Desenvolver uma compreensão do domínio da aplicação e do conhecimento prévio relevante e identificar o objetivo do processo de KDDM.
- 2) Determinar um conjunto de dados alvo: selecionar um grupo de dados ou um subconjunto de variáveis ou amostras de dados onde a descoberta será executada.
- 3) Limpeza de dados e pré-processamento: operações básicas como a remoção de ruído e decisão sobre estratégias para tratamento de dados ausentes.
- 4) Redução de dados e projeções: buscar métodos de transformação ou redução de dimensionalidade para diminuir o número de variáveis de análise.
- 5) Adequar objetivos do processo KDDM definidos na etapa 1 ao método de *Data Mining* específicos: por exemplo, sumariação, classificação, regressão, clustering.
- 6) Escolher os algoritmos de mineração de dados: a seleção de Método(s) a ser usado para procurar por padrões de dados.
- 7) *Data Mining*: aplicação dos algoritmos de *Data Mining*.
- 8) Interpretar os padrões encontrados, possivelmente retornando a qualquer um dos passos 1 a 7 para mais iterações. Esta etapa também pode envolver a visualização dos padrões ou modelos extraídos.

- 9) Consolidação da descoberta de conhecimento: incorporando este conhecimento em outro sistema para novas ações, ou simplesmente documentando-o e informando às partes interessadas. Isso também inclui verificar e resolver conflitos potenciais com conhecimento anterior.

Kurgan e Musilek (2006) afirmam que a partir deste modelo inicial os esforços de investigação se concentraram em propor novos modelos, ao invés de melhorar a concepção de um modelo único. Eles realizam uma comparação entre os principais modelos existentes e apresentam um modelo genérico simplificado constituído por seis etapas:

- 1) Compreensão do domínio da aplicação;
- 2) Compreensão dos dados;
- 3) Preparação dos dados e identificação da tecnologia de *Data Mining*;
- 4) *Data Mining*;
- 5) Avaliação do conhecimento descoberto;
- 6) Consolidação e uso do conhecimento descoberto.

Os objetivos de descoberta de conhecimento são definidos de acordo com a utilização prevista do sistema, podendo-se distinguir dois tipos de objetivos: “Verificação”, onde o sistema limita-se a verificar a hipótese do usuário, e a “Descoberta”, onde o sistema encontra novos padrões de forma autônoma. O objetivo de “Descoberta” pode ainda ser subdividido em “Previsão”, onde o sistema encontra padrões com o objetivo de prever o comportamento futuro; e “Descrição”, onde o sistema encontra padrões de forma a apresentá-los ao usuário de uma forma compreensível para os seres humanos (FAYYAD et al.,1996).

Os métodos utilizados no processo de alcançar os objetivos de “Descoberta” incluem:

- Classificação: aprender uma função que classifica um item de dados em uma de várias classes pré-definidas.
- Regressão: aprender uma função que mapeia um item de dados para uma variável de previsão de valores reais e a descoberta de relações funcionais entre variáveis.
- Clustering: identificação de um conjunto finito de categorias ou clusters para descrever os dados.

- Sumariação: encontrar uma descrição compacta para um subconjunto de dados, como por exemplo, a derivação de resumo ou associação de regras e a utilização de técnicas de visualização multivariada.
- Modelagem de dependências: encontrar um modelo que descreve dependências significativas entre as variáveis.
- Detecção de mudança e desvios: descobrir as mudanças mais significativas em relação a dados prévios ou valores normativos.

Em função dos objetivos deste trabalho, a aplicação se deu no sentido da “Descoberta” utilizando modelos tanto para previsão como para descrição. Na análise foram utilizados métodos de classificação, clustering e modelagem de dependências.

Além dos métodos apresentados, a implementação de algoritmos de *Data Mining* se baseia em três componentes:

Representação do Modelo: a linguagem usada para descrever os padrões detectáveis. Se a representação é muito limitada, nenhuma quantidade de tempo de treinamento ou exemplos será capaz de produzir um modelo preciso para os dados. Por outro lado um poder excessivo na representação pode resultar em *overfitting* de dados. É importante que um analista de dados possa compreender totalmente as premissas de representação que podem ser inerentes a um método específico.

Critérios de avaliação do Modelo: funções de ajuste (“*fit functions*”) de quanto um determinado padrão (englobando um modelo e seus parâmetros) cumpre os objetivos do processo de KDDM. Por exemplo, modelos preditivos muitas vezes são julgados pela precisão de previsão empírica em algum conjunto de teste. Modelos descritivos podem ser avaliados juntamente com as dimensões de precisão preditiva, novidade, utilidade e compreensibilidade do modelo ajustado.

Método de busca: consiste em dois componentes: busca de parâmetros e busca de modelo. Uma vez que a representação do modelo (ou família de representações) e os critérios de avaliação do modelo estejam definidos, o problema de mineração de dados torna-se reduzido a apenas uma tarefa de otimização: localizar os parâmetros/modelos da família selecionada que otimizam os critérios de avaliação. Na busca de parâmetros o algoritmo deve procurar os parâmetros que otimizam os critérios de avaliação do modelo de acordo com os dados observados e uma representação do modelo fixo. A busca de modelo ocorre como um loop sobre o método de pesquisa de parâmetros: a

representação do modelo é alterada de forma que uma família de modelos é considerada (FAYYAD et al.,1996).

O item seguinte detalha a técnica de mineração de dados utilizada neste trabalho, as Redes Bayesianas.

4.3 Redes Bayesianas

Redes Bayesianas são modelos gráficos probabilísticos que representam um conjunto de variáveis aleatórias relacionadas a um determinado problema e as relações probabilísticas entre elas.

A sua aplicação resulta em uma representação gráfica do conhecimento, possibilitando uma análise mais intuitiva e de compreensão mais simples para o cérebro humano. As Redes Bayesianas possibilitam a obtenção de conhecimento sobre o domínio de um problema e a previsão das consequências de uma intervenção, verificando-se qual o comportamento das probabilidades de ocorrência das demais variáveis ao se atuar em um determinado elemento da rede.

Os modelos podem ser obtidos a partir do conhecimento de especialistas, a partir de aprendizagem utilizando bases de dados ou combinando conhecimento prévio com aquele o obtido a partir dos dados. No caso da aprendizagem a partir dos dados, os métodos empregados oferecem uma abordagem eficiente para que seja evitado um *overfitting* (HECKERMAN, 2008).

A estrutura de uma Rede Bayesiana é representada por um gráfico acíclico direcionado (DAG do inglês *directed acyclic graph*) conforme exemplificado na Figura 2, em que os nós representam variáveis e as ligações expressam as dependências existentes entre elas (PEARL, 1988). Um gráfico acíclico é uma rede de nós correspondentes a atributos que são conectados por bordas direcionadas de modo que não ocorra nenhum ciclo (WITTEN et al., 2011). Os nós podem representar qualquer tipo de variável aleatória, seja ela um parâmetro medido, uma variável latente ou uma hipótese. Variáveis aleatórias são variáveis que podem assumir um valor a partir de um conjunto definido, sendo que a ocorrência de cada valor está associada a uma determinada probabilidade

(DALY et al., 2011). Esta característica representa uma das limitações da aplicação das Redes Bayesianas uma vez que o conjunto definido de valores implica variáveis discretas. Apesar de ser possível a utilização de variáveis contínuas, nestes casos é necessária a realização de algum processo de discretização implicando limitações na utilização das Redes Bayesianas (FRIEDMAN; GOLDSZMIDT, 1996).

Figura 2 - Gráfico acíclico direcionado

FONTE: DALY et al., 2011

Modelos probabilísticos baseados em gráficos acíclicos direcionados possuem uma longa e rica tradição, a partir do trabalho realizado pelo geneticista Sewall Wright na década de 1920. O nome Rede Bayesiana homenageia Thomas Bayes (1702-1761), responsável pelo teorema que especifica uma regra de atualização das probabilidades à luz de novas evidências e é a base da abordagem utilizada. A capacidade para inferência bidirecional, combinada com uma base probabilística rigorosa, levou a rápida adoção de Redes Bayesianas como o método de escolha para o raciocínio sob incerteza no campo da inteligência artificial e dos sistemas especialistas, substituindo as técnicas baseadas em regras anteriormente utilizadas (CHENG et al., 2002).

Pearl e Russell (2002) sugerem que o aspecto mais importante das Redes Bayesianas está em elas serem representações diretas do mundo real e não apenas de processos de raciocínio. As setas existentes no diagrama representam conexões causais reais e não o fluxo de informações durante o processo de raciocínio, como acontece nos sistemas baseados em regras e nas redes neurais. Desta forma, os processos de raciocínio podem

ser realizados propagando as informações em qualquer direção das conexões de uma Rede Bayesiana.

Teorema de Bayes

A base fundamental da estatística Bayesiana se encontra no conceito da probabilidade condicional. Sendo A e B dois eventos tais que $P(B) \neq 0$ então a probabilidade condicional do evento A dado o evento B, denotada por $P(A|B)$, é dada pela expressão:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

A partir da expressão acima:

$$\text{Com } P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ e } P(B|A) = \frac{P(B \cap A)}{P(A)} \text{ temos}$$

$$P(A|B) * P(B) = P(B|A) * P(A)$$

Assim, dados dois eventos A e B tais que $P(A) \neq 0$ e $P(B) \neq 0$ temos o teorema de Bayes definido pela expressão:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$

O teorema de Bayes é empregado quando não é possível determinar a probabilidade condicional de interesse de forma direta. A utilização do teorema determina o que é chamado de inferência Bayesiana (NEAPOLITAN, 2004).

Suposições de independência

Segundo Charniak (1991) uma objeção ao uso das teorias de probabilidade está no fato de que uma especificação completa de uma distribuição de probabilidade requer uma quantidade numérica absurda de forma a poder tratar todo o universo de possibilidades. Por exemplo, num problema com n variáveis aleatórias binárias seria necessário especificar as probabilidades de todas as combinações destas variáveis para poder tratá-lo, correspondendo a $(2^n - 1)$ combinações de valores. As Redes Bayesianas possuem premissas de independência entre as variáveis que fazem com que a quantidade de

valores necessários para especificar uma distribuição das probabilidades caia dramaticamente.

Cada variável da Rede Bayesiana é condicionalmente independente do conjunto de todos os seus não descendentes dado o conjunto de todos os seus pais. Esta regra é chamada de condição de Markov (NEAPOLITAN, 2004).

Uma forma de checar as independências de forma gráfica em uma Rede Bayesiana é através de um conceito conhecido por D-separação.

Em uma Rede Bayesiana dizemos que dois vértices distintos X e Y estão d-separados se para todos os caminhos entre X e Y existe um vértice intermediário V (distinto de X e Y) tal que a conexão entre X e Y através de V:

- É linear ou divergente e V recebeu uma evidência ou;
- É convergente e nem V nem algum de seus descendentes receberam uma evidência.

Em outras palavras, a D-separação ocorre quando o conhecimento do estado de uma determinada variável “bloqueia” um caminho causal entre duas outras variáveis de forma que o conhecimento de evidências de um lado do bloqueio não traz nenhuma informação em relação ao estado do outro lado do bloqueio.

Figura 3 - Tipos de conexão

FONTE: CHARNIAK, 1991

A Figura 3 apresenta os tipos de conexões existentes em Redes Bayesianas. Para as conexões do tipo linear e divergente, as variáveis A e C são d-separadas a partir de evidências na variável B. Se não sabemos o estado de B, então uma mudança de A causa uma alteração em B que, por sua vez causará uma alteração em C. Uma vez conhecido o estado da variável B, o conhecimento do estado da variável A não nos dirá

nada mais sobre o estado de C, assim como o conhecimento de C de nada esclarecerá sobre A. Já para a conexão convergente, as variáveis A e C são d-separadas por falta de evidências sobre B. Se não sabemos o estado de B, então A e C são condicionalmente independentes. No entanto, uma vez conhecido o estado de B, as variáveis A e C tornaram-se dependentes condicionalmente.

4.3.1 A aprendizagem de Redes Bayesianas

O aprendizado de uma Rede Bayesiana a partir de dados envolve duas tarefas distintas: o aprendizado da estrutura da rede determinando as relações de dependências entre as variáveis e o aprendizado dos parâmetros, que refletem a força dessas dependências (CHENG et al., 2002).

Há dois cenários possíveis a serem considerados para a aprendizagem de Redes Bayesianas a partir de uma amostra de casos:

- A estrutura da rede é conhecida e parâmetros devem ser estimados;
- A estrutura da rede não é conhecida e deve ser encontrada e os parâmetros devem ser estimados.

No primeiro cenário, se os dados forem totalmente observáveis pode ser aplicada a estimativa máxima a posteriori (MAP) para se calcular os parâmetros da Rede Bayesiana. Para redes discretas, as estimativas MAP são calculadas através da contagem de frequências relativas de ocorrência dos valores de cada variável dada cada um das configurações dos valores dos pais (LAURÍA; DUCHESSI, 2007).

No segundo cenário, a estrutura subjacente da Rede Bayesiana é desconhecida e também deve ser determinada, o que implica a especificação das premissas de independência condicional entre as variáveis do modelo, assim como também dos parâmetros. Esta tarefa pode ser realizada através de uma série de diferentes algoritmos de aprendizagem.

Os algoritmos para aprendizado de Redes Bayesianas a partir de dados podem ser subdivididos em duas abordagens básicas: métodos baseados em testes de independência condicional e métodos baseados em uma função de pontuação (*score*

function) e um processo de busca. Além destas abordagens existem algoritmos híbridos que usam uma combinação de métodos baseados na independência condicional e em *score*. Os algoritmos baseados em testes de independência realizam um estudo qualitativo sobre a dependência e independência das relações entre as variáveis do domínio e tentam encontrar uma rede que representa essas relações da melhor forma possível. Assim, eles partem de uma lista de relacionamentos de independência condicional obtida a partir dos dados por meio de testes de independência condicional e geram uma rede que representa a maior parte dessas relações.

Os algoritmos baseados em uma função de pontuação (também chamada de métrica) tentam encontrar um modelo gráfico que maximiza a pontuação da função selecionada, sendo que esta função é geralmente definida como uma medida de ajuste entre o modelo gráfico e os dados. A função de pontuação é combinada com um método de busca para medir a qualidade de cada estrutura explorada dentro do espaço de soluções possíveis. Durante o processo de exploração, a função pontuação é aplicada de forma a avaliar a adequação de cada estrutura candidata aos dados. Cada algoritmo caracteriza-se por uma função específica de pontuação e o processo de pesquisa usado. As funções de pontuação são baseadas em princípios diferentes, tais como a entropia e abordagens Bayesianas, sendo que algumas das métricas disponíveis para comparação dos modelos incluem o Kullback–Leibler (KL), K2, o BDeu, a métrica BIC e a *Minimum Description Length* (MDL) (ACID et al., 2004 e LAURÍA; DUCHESSI, 2007).

No caso do software BayesiaLab utilizado neste trabalho, todos os algoritmos de aprendizagem estrutural baseiam-se em princípios de MDL utilizando conceitos da Teoria da Informação como entropia e informação mútua.

De acordo com Conrady e Jouffe (2011c) a entropia mede a incerteza inerente à distribuição de uma variável aleatória. A entropia de $H(x)$ de uma variável aleatória x é definida como:

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x),$$

A letra minúscula x representa os estados que a variável X pode assumir. Observe que o log está na base 2 e o valor da entropia é expressa em bits (0/1).

Como exemplo, Conrady e Jouffe (2011c) apresentam o caso de lançamento de uma moeda onde X pode ter um dos dois estados, cara e coroa, com o conjunto de resultados possíveis sendo $X = \{\text{cara}, \text{coroa}\}$ e a probabilidade de cara e coroa sendo 0,5, ou seja, $p(\text{cara}) = 0,5$ e $p(\text{coroa}) = 0,5$. A entropia calculada neste caso será:

$$\begin{aligned} H(X) &= -p(\text{cara}) \log_2 p(\text{cara}) - p(\text{coroa}) \log_2 p(\text{coroa}) = \\ &= -0,5 \log_2 0,5 - 0,5 \log_2 0,5 = 0,5 + 0,5 = 1 \text{ bit}. \end{aligned}$$

Isso significa que nossa incerteza antes um sorteio justo equivale a um valor de entropia de 1 bit, que é a entropia máxima dada a distribuição uniforme da variável com dois estados. No caso de uma moeda viciada com $p(\text{cara}) = 0,7$ e $p(\text{coroa}) = 0,3$, é intuitivo perceber que a incerteza seria menor já que um dos estados do sorteio será mais provável e a entropia $H(X_{\text{viciada}})$ resulta em um valor mais baixo:

$$H(X_{\text{viciada}}) = -0,7 \log_2 0,7 - 0,3 \log_2 0,3 = 0,881.$$

Desenhando-se o gráfico 1 para uma variação entre diferentes probabilidades de uma moeda viciada, vemos que qualquer probabilidade diferente do valor de 0,5 para cada estado da moeda perfeita reduz a entropia.

Gráfico 1 - Entropia do lançamento da moeda

FONTE: CONRADY; JOUFFE, 2011c

Do conceito de entropia derivam algumas das métricas utilizadas pelo software na aprendizagem e análise das Redes Bayesianas:

Informação mútua

A informação mútua $I(X, Y)$ mede o quanto em média a observação da variável aleatória Y nos diz sobre a incerteza de X , ou seja, o quanto a entropia de X é reduzida se nós temos informações sobre o Y .

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Cabe notar que a informação mútua é uma métrica simétrica que reflete a redução de incerteza de X pelo conhecimento de Y assim como a de Y pelo conhecimento de X .

Entropia relativa (divergência de Kullback-Leibler)

A entropia relativa ou divergência de Kullback-Leibler é uma medida da diferença entre duas distribuições de probabilidade p e q . Para distribuições de probabilidade p e q de uma variável aleatória discreta X , a divergência K-L é definida como:

$$D_{KL} = (p(X) || q(X)) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)}$$

A entropia relativa é o valor esperado da diferença logarítmica entre as distribuições de probabilidade conjunta $p(X)$ e $q(X)$, e ao contrário da informação mútua, a entropia relativa não é simétrica. A força do arco utilizada na análise de Redes Bayesianas através do software BayesiaLab é diretamente proporcional à entropia relativa e descreve a força de ligação direcional entre duas variáveis. Para maiores detalhes e exemplos de cálculos da entropia relativa sugere-se consultar Conrady e Jouffe (2011c).

4.3.2 Algoritmos

O software BayesiaLab utilizado neste trabalho apresenta uma série de diferentes algoritmos agrupados em três diferentes tipos como opções para utilização:

Aprendizagem não supervisionada

Estes algoritmos trabalham buscando descobrir associações entre variáveis a partir de relações probabilísticas encontradas nos dados utilizados na aprendizagem. Uma vez que a quantidade de redes possíveis de serem implementadas com um dado número de variáveis facilmente chega a valores impossíveis de serem avaliados por uma busca exaustiva pela de melhor desempenho, os algoritmos de aprendizagem se baseiam em conjuntos diferentes de heurísticas que permitem a redução do espaço de busca.

O software permite a escolha entre alguns algoritmos de aprendizagem conceitualmente diferentes e com tempos de execução bastante variados. Com heurísticas diferentes os

resultados de cada método podem variar, mas como todos os métodos utilizam a mesma métrica (a pontuação MDL), as redes resultantes podem ser facilmente comparadas.

Os algoritmos pertencentes a este grupo são:

Maximum spanning tree: É o algoritmo mais rápido para aprendizado não supervisionado, utilizando apenas duas passagens. Mesmo com uma rede resultante não ótima, este algoritmo pode ser útil para uma primeira imputação dos valores em falta e gerar uma rede inicial, para análise de clusterização de variáveis, ou antes de usar os métodos Tabu ou EQ para refinar o modelo.

Tabu: Este método é particularmente útil para refinar uma rede construída por especialistas humanos ou para a atualização de uma rede aprendida a partir de um conjunto de dados diferente.

EQ: Este método é muito eficiente, conseguindo reduzir fortemente o tamanho do espaço de busca. Mais detalhes sobre este algoritmo pode ser encontrado em Munteanu e Bendou (2001).

SopLEQ : Um método de pesquisa de estrutura baseado numa caracterização global dos dados e na exploração das propriedades de equivalência de Redes Bayesianas.

Tabu Order : Método de aprendizagem utilizando Tabu, busca no espaço da ordem dos nós da Rede Bayesiana para encontrar a melhor rede para uma ordem de nós fixos. Este é o método de busca mais completo, mas também o mais demorado.

Aprendizagem supervisionada

São algoritmos que permitem a aprendizagem de Redes Bayesianas buscando uma estrutura inteiramente dedicada à caracterização de uma variável alvo, dentre os quais se destacam aqueles baseados nas arquiteturas de rede:

Naive Bayes: Rede Bayesiana com uma arquitetura pré-definida na qual o nó de destino é o pai de todos os outros nós. Esta estrutura assume que o nó alvo é a causa de todos os outros nós e que o conhecimento do seu valor faz com que os demais nós se tornem independentes entre si. Apesar de fortes suposições, falsas na maioria dos casos, o baixo

número de probabilidades a serem estimadas torna esta estrutura muito robusta, com um tempo de aprendizagem curto.

Augmented Naive Bayes: possuindo a estrutura parcialmente pré-definida, esta arquitetura é composta por uma rede *Naive Bayes*, enriquecida pelas relações entre os nós filhos, sabendo o valor do nó de destino (o comum pai). A precisão da previsão deste algoritmo é melhor do que os obtidos pela arquitetura *Naive*, mas a busca das relações entre os filhos pode levar mais tempo.

Clusterização

São algoritmos que permitem o agrupamento de dados de maneira não supervisionada, a fim de encontrar partições contendo elementos homogêneos. Esses algoritmos são baseados na arquitetura *Naive Bayes* descrita no item anterior em que um nó CLUSTERS definindo as partições é obtido como pai de todos os demais nós. Ao contrário da aprendizagem supervisionada, os valores do nó CLUSTERS nunca são observados em uma base de dados, sendo estimados a partir de métodos de *Expectation-Maximization*.

4.3.3 Comparação com outras técnicas de modelagem

Segundo Lauría e Duchessi (2007), são três as principais técnicas de modelagem utilizadas em pesquisas: a análise de regressão multivariada e as duas vertentes de modelagem de equações estruturais (SEM): LISREL e *partial least squares* (PLS).

O objetivo geral da regressão múltipla é conhecer melhor a relação entre diversas variáveis independentes e uma variável dependente. A análise de regressão múltipla requer que as variáveis apresentem uma distribuição normal e através dela pode-se apenas verificar relações, mas nunca ter certeza sobre um mecanismo causal subjacente.

A modelagem de equações estruturais é uma abordagem de modelagem causal que combina informações de causa-efeito com dados estatísticos para fornecer uma avaliação quantitativa das relações entre as variáveis estudadas. Ela se baseia em relações causais hipotéticas entre as variáveis, representando graficamente a natureza e

a força dessas relações. Se as relações são significativas, a construção teórica é considerada válida e pode ser usada para fornecer orientações para a aplicação do modelo na prática. O LISREL tem foco em maximizar a covariação explicada entre os vários construtos, enquanto o PLS, assim como a regressão múltipla, maximiza a variação explicada entre eles. Apesar das técnicas SEM serem úteis para validação empírica de relações causais baseadas em teoria e de certo modo também para previsão, elas não se demonstram adequadas para um diagnóstico de situação, tendo limitações no apoio a decisões gerenciais. Além disso, as SEM tratam principalmente de modelos de relações lineares. Se as relações forem não lineares, o efeito potencial das variáveis independentes em explicar a variação nas variáveis dependentes não pode ser conhecido com precisão, resultando em previsões e diagnósticos pobres (GUPTA; KIM, 2008).

Ao invés de formular uma rede ou modelo com base em uma teoria para então testá-lo como ocorre nas técnicas de SEM, as técnicas de Redes Bayesianas identificam a rede que melhor se encaixa aos dados, permitindo também que sejam incorporadas relações hipotéticas entre as variáveis para restringir o universo de pesquisa e garantir a validade do modelo.

Em contraposição aos outros métodos, as Redes Bayesianas são especialmente adequadas para previsão e diagnóstico e podem ser treinadas na mesma estrutura com novos dados. Além disso, elas também podem ser utilizadas para a modelagem de relações não lineares e são bastante úteis para avaliar o impacto de mudanças na situação modelada.

No entanto, Gupta e Kim (2008) afirmam que as Redes Bayesianas apresentam certas limitações na modelagem causal sob o ponto de vista de pesquisa de ciências sociais uma vez que para o estabelecimento de um nexo de causalidade, três critérios devem ser atendidos: a ordem temporal, a associação e a eliminação de alternativas plausíveis. Na modelagem Bayesiana, os relacionamentos são baseados na associação (independência condicional) e, até um determinado grau, na ordem temporal. No entanto, o terceiro critério de eliminação de alternativas plausíveis não é atendido. O resultado é que as Redes Bayesianas não diferenciam entre uma relação causal e uma relação espúria porque relações causais não podem ser determinadas apenas de dados estatísticos (PEARL, 1998). Já segundo Lauría e Duchessi (2007), as Redes Bayesianas também podem ser interpretadas de forma causal da mesma forma como ocorre nos modelos

SEM, mas a noção de causalidade é aberta a interpretação e tem sido objeto de controvérsia e debate filosófico.

Anderson e Vastag (2004) afirmam que se o objetivo for apenas uma descrição de construtos teóricos, mas sem nenhum interesse em inferência presente ou futura em relação às variáveis observáveis, então SEM se demonstra o melhor candidato a método de escolha. Já se os objetivos incluem previsão e diagnóstico das variáveis observadas, como é o caso deste trabalho, então a abordagem de Redes Bayesianas se demonstra um melhor candidato.

O Quadro 3 apresenta um resumo comparativo entre as modelagens através de regressão, LISREL, PLS e Redes Bayesianas.

Quadro 3– Comparação entre regressão, LISREL, *partial least squares* e Redes Bayesianas

Categorias	Regressão	LISREL	<i>Partial least squares</i>	Redes Bayesianas
Objetivos de pesquisa	Explicação de variância	Confirmação de teoria	Explicação causal e previsão	Explicação causal e inferência estatística
Forma funcional	Linear	Linear	Linear	Não linear
Premissas de distribuição	Normal	Normal	Normal	Sem restrições, mas tipicamente multinomial
Tamanho da amostra	pequena-grande	média-grande	pequena-grande	pequena-grande
Dependência de teoria	Modelos baseados em teoria	Modelos baseados em teoria	Modelos baseados em teoria	Sem restrições: teoria ou através de dados
Adequação do modelo	Significância estatística do modelo	Significância estatística do modelo	Significância estatística do modelo	Funções de <i>score</i> , testes de independência condicional e testes de cross-validação
Abordagem subjacente	Minimizar variação não explicada	Minimizar covariação	Minimizar variação não explicada	Modelos gráficos e aprendizagem de máquina estatística em inteligência artificial
Curva de aprendizado	acentuada	moderada	moderada	plana
Disponibilidade de ferramentas	amplamente disponíveis	disponíveis	disponíveis	disponíveis

FONTE: Lauría e Duchessi, 2007

4.4 Aplicações de Redes Bayesianas

Bastante difundida na área médica e apresentando também aplicações em uma diversidade de áreas como biologia, previsão de tempo, análise de texto, entre outras (DALY et al., 2011), as Redes Bayesianas ainda não encontram muito destaque nas ciências sociais aplicadas, apesar de possuírem uma série de características adequadas para tratar elementos encontrados em situações estudadas nesta área do conhecimento (LYNCH, 2010, p.71), conforme apresentado nos tópicos anteriores.

Na pesquisa bibliográfica foram encontrados dois artigos que utilizam a aprendizagem de Redes Bayesianas na análise de dados em contextos relacionados ao desenvolvimento ou a adoção de tecnologia.

Sawhney (2001) explora o uso de técnicas de estimação bayesiana para modelagem de indicadores relacionados ao Desenvolvimento Humano, identificando a importância de indicadores de emprego, educação e saúde das mulheres com relação a fecundidade e mortalidade infantil, em comparação com efeitos mais fracos das variáveis associadas ao progresso econômico geral (como densidade telefônica e PIB).

Nedevschi et al. (2006) examinam o uso de Redes Bayesianas como ferramentas para revelar a estrutura das relações entre os fatores demográficos, sociais e econômicos, e a penetração de várias tecnologias, identificando pelo menos três exemplos de casos em que elas podem ser úteis para o cientista social:

- Confirmação de hipóteses, confirmando que diversos fatores sociais afetam a utilização do computador pela população em geral, como idade, sexo e renda.
- Detecção de ligações não intencionais nos dados, identificando um viés indesejável no conjunto de dados associando a idade e o sexo.
- Geração de novas hipóteses a partir de dados complexos, sugerindo a possível existência de uma hierarquia de adoção de tecnologias.

4.5 Passos para desenvolvimento do trabalho

Este projeto utilizou como base a metodologia genérica de seis etapas para o processo de *Knowledge Discovery and Data Mining* conforme descrita por Kurgan e Musilek (2006) associada à metodologia específica para construção de uma Rede Bayesiana a partir de dados sugerida por Lauría e Duchessi (2007), que coincidentemente também possui seis etapas: reduzir a dimensionalidade; redimensionar e discretizar as variáveis; definir restrições semânticas; busca de modelos candidatos; comparação dos modelos candidatos; e aplicar o modelo selecionado.

No Quadro 4 é apresentado um paralelo da correspondência das etapas de ambas as metodologias.

Quadro 4 – Paralelo de correspondência entre as etapas das metodologias de KDDM de Kurgan e Musilek (2006) e Redes Bayesianas de Lauría e Duchessi (2007)

Knowledge Discovery (Kurgan e Musilek,2006)	Redes Bayesianas (Lauría e Duchessi,2007)
1. Compreensão do domínio da aplicação	
2. Compreensão dos dados	
3. Preparação dos dados e identificação da tecnologia de <i>Data Mining</i>	1. Redução de dimensionalidade 2. Redimensionamento e discretização das variáveis
4. <i>Data Mining</i>	3. Definição de restrições semânticas (requer compreensão do domínio) 4. Busca de modelos candidatos 5. Comparação dos modelos candidatos
5. Avaliação do conhecimento descoberto	
6. Consolidação e uso do conhecimento descoberto	6. Aplicação do modelo selecionado

4.5.1 Compreensão do domínio da aplicação

Etapa base do processo de KDDM, a compreensão do domínio da aplicação corresponde a um entendimento dos objetivos e requisitos de negócio de forma a convertê-los em uma definição de problema para a aplicação do *Data Mining*.

Servindo como pano de fundo direcionador de todas as etapas seguintes, a compreensão do domínio da aplicação tem um destaque ainda maior devido à natureza das Redes

Bayesianas facilitar a combinação de conhecimento prévio com conhecimento adquirido a partir dos dados.

4.5.2 Compreensão dos dados

Esta etapa consiste na identificação de problemas de qualidade de dados, exploração de dados e seleção de subconjuntos interessantes para o processo de análise subsequente. No contexto deste projeto de pesquisa a qualidade dos dados está principalmente relacionada com a metodologia empregada no levantamento TIC Domicílios e Empresas 2010 - Pesquisa sobre uso das tecnologias da informação e comunicação no Brasil (CGI.BR, 2010). A análise das questões empregadas no questionário sob a ótica da revisão da literatura de divisão digital, assim como uma exploração dos dados correspondentes as respostas formaram a base para seleção de subconjuntos de dados e variáveis.

4.5.3 Preparação dos dados e identificação da tecnologia de Data Mining

Consiste na preparação do conjunto de dados final, que será alimentado na ferramenta de *Data Mining* e inclui a seleção de dados e atributos, limpeza, construção de novos atributos e transformações dos dados a partir do que foi identificado na etapa anterior.

Dentro da tecnologia de Redes Bayesiana para a mineração dos dados foi realizada uma análise de softwares disponíveis, optando-se pela utilização do software pago BayesiaLab em função de um maior conjunto de funcionalidades e melhor performance quando comparado com as diversas opções de softwares livres disponíveis para download na internet. Nesta etapa também foi importante a determinação dos algoritmos do software mais adequados para cada caso analisado.

Outras considerações relativas à aplicação de Redes Bayesianas realizadas nesta etapa incluem a redução de dimensionalidade e o redimensionamento e discretização de variáveis.

Redução de dimensionalidade

A recomendação do uso de uma técnica de redução de dimensionalidade para diminuir o número de variáveis e parâmetros a serem estimados se dá em função do número de casos necessários para estimar essas variáveis e parâmetros crescer exponencialmente conforme o número de variáveis e parâmetros de cresce. Assim, para um tamanho de amostra determinado, há um número máximo de variáveis e parâmetros que podem ser estimados com precisão. Existem diversas técnicas para redução da dimensionalidade, incluindo análise fatorial, análise dos componentes principais (PCA), escalonamento multidimensional (MDS), entre outras (LAURÍA; DUCHESSI, 2007).

Neste estudo a quantidade de variáveis não se demonstrou grande o suficiente para que fosse necessária a realização de reduções de dimensionalidade em nenhum dos casos tratados.

Redimensionamento e discretização de variáveis

Para atender aos requisitos da maior parte dos algoritmos para construção de Redes Bayesianas que exigem variáveis discretas é necessário redimensionar distribuições contínuas e/ou discretizar variáveis contínuas de acordo com algum critério. Neste estudo esta etapa se demonstrou importante principalmente na análise dos indicadores representativos da divisão digital entre diferentes países.

4.5.4 *Data Mining*

Nesta etapa o método selecionado de *Data Mining* é aplicado nos dados preparados. Alguns aspectos relacionados especificamente às Redes Bayesianas são discutidos a seguir.

Definição de restrições semânticas

A definição de restrições semânticas reduz o espaço de busca pela melhor rede permitindo que o algoritmo considere apenas os modelos que atendem a condições de precedência temporal ou outros requisitos de dependência entre as variáveis de forma a tornar a busca mais rápida e eficiente.

Busca de modelos candidatos

Nesta etapa são aplicados os algoritmos de busca e *score* para pesquisar o espaço das redes, identificando vários candidatos potenciais e estimando os parâmetros associados. As funções de pontuação e algoritmos de busca transformam a aprendizagem da estrutura de rede em um problema de otimização, onde o objetivo é encontrar no espaço de estruturas de rede aquela que maximiza a pontuação e representa a melhor adequação ao conjunto de dados de treinamento.

Comparação dos modelos candidatos

A partir dos modelos obtidos no item anterior é necessário determinar uma forma de comparação comum que permita identificar qual deles melhor representa o processo sendo modelado.

No caso do software Bayesialab todos os algoritmos utilizam a minimização do MDL como base para escolha do melhor modelo, conforme já detalhado anteriormente no item 4.3.2 sobre algoritmos.

4.5.5 Avaliação do conhecimento descoberto

Esta etapa consiste na interpretação do resultado do processo de *Data Mining* sob a ótica da compreensão do domínio da aplicação e do problema estipulado na primeira etapa.

A natureza visual das Redes Bayesianas que permite uma interpretação intuitiva do resultado constitui um grande facilitador para esta etapa.

Novamente a natureza cíclica e iterativa do processo de KDDM permite que, a partir dos resultados aqui analisados, se decida realizar ajustes e retornar a alguma etapa anterior.

4.5.6 Uso do conhecimento descoberto

É a etapa final onde se incorpora o conhecimento descoberto em um sistema final, checando e resolvendo conflitos potenciais com algum conhecimento aceito anteriormente.

A Figura 4 apresenta o diagrama combinado das metodologias conforme utilizadas neste trabalho.

Figura 4 - Diagrama metodológico adaptado de Kurgan e Musilek (2006) e Lauría e Duchessi (2007)

4.6 Dados utilizados

Na obtenção do modelo para análise do posicionamento do Brasil e suas subdivisões regionais no contexto internacional da divisão digital foram utilizados os valores dos indicadores de cálculo do índice IDI disponibilizados pelo ITU para 152 diferentes países (ITU, 2011). Para o processo da análise a partir do modelo obtido foram utilizados dados levantados através de relatórios disponibilizados na internet pela ANATEL e pelo IBGE, além de dados do levantamento TIC Domicílios e Empresas 2010 - Pesquisa sobre uso das tecnologias da informação e comunicação no Brasil (CGI.BR, 2010). As tabelas completas com estes valores são apresentadas nos Apêndices 1 e 2.

A base utilizada para obtenção dos modelos de análise dos fatores que influenciam a divisão digital no Brasil foi gerada a partir dos dados do levantamento realizado para elaboração do relatório TIC Domicílios e Empresas 2010 - Pesquisa sobre uso das tecnologias da informação e comunicação no Brasil (CGI.BR, 2010). O levantamento contou com uma amostra principal de 23.107 domicílios distribuídos por todo o território nacional, com a coleta de dados ocorrendo no período entre agosto e outubro de 2010. Além disso, foi realizada uma oversample de 1500 usuários de internet a fim de reduzir a variabilidade dos indicadores de uso da rede, totalizando 9932 usuários de internet. As entrevistas relativas à amostra principal de domicílios foram realizadas presencialmente nos domicílios em 307 municípios, com indivíduos de 10 anos de idade ou mais. Para maiores detalhes sobre a amostra e a metodologia utilizada sugere-se consultar o relatório completo disponível na internet (CGI.BR, 2010).

Para este trabalho foi desconsiderada uma pequena quantidade dos registros em que o respondente não declarou a raça (279 domicílios e 110 usuários de internet), não soube responder sobre o uso do celular (158 indivíduos) ou não soube responder se efetuou compras na internet (três usuários de internet). Desta forma a amostra final resultante utilizada nas análises totalizou 22.770 domicílios/indivíduos e 9.932 usuários de internet, sem ocorrência de *missing values* em nenhuma das variáveis consideradas.

A Tabela 1 apresenta a distribuição da amostra final da base de dados em função das variáveis utilizadas na análise.

Tabela 1 – Distribuição da base de dados para análise dos fatores que influenciam a divisão digital no Brasil

Variáveis		Domicílios/Indivíduos	Usuários de internet
Área	Rural	3522	894
	Urbana	19248	8925
Região	Centro-Oeste	2268	1287
	Nordeste	6841	2777
	Norte	3817	1732
	Sudeste	5269	2409
	Sul	4575	1614
Estado	Acre	422	157
	Alagoas	428	132
	Amapá	438	153
	Amazonas	473	194
	Bahia	1714	661
	Ceará	1199	628
	Distrito Federal	786	471
	Espírito Santo	610	171
	Goiás	514	286
	Maranhão	525	141
	Mato Grosso	486	266
	Mato Grosso do Sul	482	264
	Minas Gerais	1241	648
	Paraná	1371	758
	Paraíba	446	115
	Pará	1250	674
	Pernambuco	1215	612
	Piauí	513	174
	Rio Grande do Norte	428	174
	Rio Grande do Sul	1877	575
	Rio de Janeiro	1263	695
	Rondônia	410	199
	Roraima	437	197
	Santa Catarina	1327	281
	Sergipe	373	140
	São Paulo	2155	895
	Tocantins	387	158
Sexo	Feminino	11716	4924
	Masculino	11054	4895
Grau de instrução	Analfabeto Até Pré-primário	1976	110
	Fundamental I	6070	1023
	Fundamental II	5713	2468
	Médio	6472	3854
	Superior	2539	2364
Faixa etária	De 10 a 15 anos	4345	2711
	De 16 a 24 anos	3586	2410
	De 25 a 34 anos	4089	2194
	De 35 a 44 anos	3842	1385
	De 45 a 59 anos	3795	872
	De 60 anos ou mais	3113	247
Classe Social	Classe A	232	229
	Classe B	4025	3147
	Classe C	12378	5537
	Classe D e E	6135	906
Situação de emprego	Aposentado(a)	1922	183

Variáveis		Domicílios/Indivíduos	Usuários de internet
	Desempregado(a)	450	231
	Dona de casa	1745	381
	Estudante	4221	2806
	Trabalhando	14432	6218
Raça declarada	Amarela	337	157
	Branca	8860	4008
	Indígena	241	89
	Parda	10309	4426
	Preta/ Negra	3023	1139
Total		22770	9819

FONTE: CGLBR (2010)

5 MEDINDO A DIVISÃO DIGITAL BRASILEIRA

A partir dos passos apresentados no item anterior e tendo como pano de fundo o objetivo de avaliar o grau da divisão digital brasileira dentro de um contexto internacional, a etapa de compreensão do domínio da aplicação consistiu na revisão da literatura apresentada anteriormente, onde a construção de *rankings* a partir de índices compostos foi identificada como uma forma adequada e bastante utilizada na análise da complexidade da divisão digital entre regiões (VICENTE; LÓPEZ, 2011).

Utilizado pelo *International Telecommunication Union* (ITU) no ranqueamento da divisão digital entre 152 países (ITU, 2011), identificado como uma evolução e síntese de índices anteriores (BRUNO et al., 2010) e com disponibilidade de acesso aos dados utilizados em seu cálculo, o *ICT Development Index* (IDI) foi selecionado como base de referência internacional para a proposta de avaliação da divisão digital entre diversas subdivisões regionais brasileiras. A tabela com os valores normalizados dos indicadores utilizados no cálculo do índice, assim como o valor do IDI e a posição no *ranking* de cada país está apresentada no Apêndice 1.

Com a importação dos dados da tabela no software Bayesialab foi possível realizar uma primeira análise geral das relações entre as variáveis componentes do IDI a partir da obtenção de um modelo de Rede Bayesiana utilizando o algoritmo EQ de aprendizado não supervisionado, conforme sugerido por Conrady e Jouffe (2011a)

A natureza classificatória do problema sugeriu a busca de um modelo representado por uma Rede Bayesiana classificadora, optando-se pela utilização do algoritmo de clusterização de dados do software Bayesialab que resulta em uma Rede Bayesiana do tipo *Naive*.

Como o IDI utiliza uma quantidade de apenas 11 variáveis no seu cálculo, não foi necessária a realização de nenhum método de redução de dimensionalidade. No entanto, o processo de importação dos dados para o Bayesialab exigiu a aplicação de um método de discretização para atender o requisito base do aprendizado de Redes Bayesianas. A realização de diversos testes demonstrou que a discretização de cada uma das 11 variáveis em quatro faixas iguais foi a que apresentou resultados mais interessantes na distribuição dos 152 países em clusters. Com um número pequeno de elementos a serem

clusterizados, quantidades menores ou maiores de faixas de discretização resultavam em classificações menos esclarecedoras em função da concentração ou dispersão excessiva dos elementos nos clusters.

A simplicidade do modelo de Rede Bayesiana *Naive* não exigiu que nenhuma restrição semântica fosse empregada, e a escolha do melhor modelo foi realizada automaticamente pelo algoritmo de clusterização de dados do Bayesialab. A parametrização utilizada no processo instruiu o programa a selecionar automaticamente um número de clusters entre 5 e 25. Também foram atribuídos pesos diferentes para cada indicador, adaptando os pesos utilizados no cálculo do IDI à forma de trabalho do algoritmo de clusterização utilizado no Bayesialab. Assim o peso resultante dos indicadores de acesso às TIC ($0,4 \times 0,2 = 0,08$), uso das TIC ($0,4 \times 0,33 = 0,133$) e habilidades para TIC ($0,2 \times 0,33 = 0,066$) foram normalizados respectivamente para 0,6 (acesso), 1 (uso) e 0,5 (habilidades), conforme mostrado no Quadro 5:

Quadro 5 – Peso atribuído aos indicadores de cálculo do IDI para clusterização no Bayesialab

Categoria	Indicadores	Abreviação	Pesos		
			Categoria	Indicadores	Bayesialab
Acesso às TIC	Assinaturas de telefone fixo por 100 habitantes	FIX	0,4	0,2	0,6
	Assinaturas de telefone celular por 100 habitantes	CEL		0,2	0,6
	Banda internacional de internet por usuário de internet	BAND		0,2	0,6
	Percentual de domicílios com computador	COMP		0,2	0,6
	Percentual de domicílios com acesso a internet	INT		0,2	0,6
Uso das TIC	Percentual de indivíduos usando a internet	USR	0,4	0,33	1
	Assinaturas de banda-larga fixa por 100 habitantes	INT FIX		0,33	1
	Assinaturas de banda-larga móvel ativas por 100 habitantes	INT MOV		0,33	1
Habilidades para TIC	Taxa de alfabetização de adultos	ALF	0,2	0,33	0,5
	Taxa de matrícula no secundário	SEC		0,33	0,5
	Taxa de matrícula no terciário	TER		0,33	0,5

Com o objetivo de avaliar o modelo obtido e a consequente distribuição dos 152 países em clusters foi realizada uma comparação com o próprio ranqueamento utilizando o IDI (apresentado no Apêndice 1). Além disso, uma vez que as diferenças de riqueza econômica são apontadas como o principal fator associado ao grau de divisão digital (ANDONOVA, 2006; CRENSHAW; ROBISON, 2006; DIMAGGIO et al., 2004; FAIRLIE, 2004; HARGITTAI, 1999 ; KIISKI; POHJOLA, 2002; DEWAN et al., 2005; NORRIS, 2003; POHJOLA, 2003), também foi realizada uma comparação em relação ao ranqueamento dos países através do PIB per capta (apresentado no Apêndice 3).

A etapa seguinte consistiu na aplicação da Rede Bayesiana obtida para inferir a que cluster mais provavelmente uma subdivisão regional do Brasil pertence, identificando assim os grupos de países correspondentes a cada uma delas. Para isso foram levantados dados disponíveis correspondentes aos indicadores componentes do IDI para diversas subdivisões regionais brasileiras (regiões, estados, regiões metropolitanas/interior e rural/urbana) em diversas fontes como a ANATEL, o IBGE e o CGI.br.

Nota-se aqui uma das grandes motivações da utilização de Redes Bayesianas nesta análise, uma vez que a facilidade de obtenção de dados é diferente para cada sub-região, resultando na disponibilidade de diferentes conjuntos incompletos de indicadores para a realização da inferência. Os dados que puderam ser obtidos são apresentados em uma tabela no Apêndice 2, onde se observa também a quantidade significativa de células em branco representando os dados faltantes.

5.1 Visão Geral dos dados

Para uma visão geral da base de dados utilizados pelo International Telecommunication Union (ITU) no ranqueamento da divisão digital entre 152 países (ITU, 2011) obteve-se a Rede Bayesiana apresentada na Figura 5 a partir do algoritmo EQ, possibilitando uma primeira análise das relações probabilísticas entre os indicadores de cálculo do IDI. Na figura são apresentados os valores correspondentes à informação mútua (em preto) e à entropia relativa de cada arco (valores em azul correspondem ao valor no sentido da seta e os vermelhos ao do sentido oposto).

Figura 5 - Rede Bayesiana dos indicadores de cálculo do IDI (informação mútua e entropia relativa)

Podemos observar valores de entropia relativa elevados para diversos arcos, indicando que o conhecimento do estado de determinados nós reduz drasticamente a incerteza em relação ao valor dos nós a eles conectados pelo arco de ligação. Chegando a mais de 50% em alguns casos, a entropia relativa indica uma forte relação probabilística entre os dois valores.

A mesma rede é apresentada na Figura 6, desta vez exibindo os valores correspondentes às correlações de Pearson entre cada par de nós conectados.

Figura 6 - Rede Bayesiana dos indicadores de cálculo do IDI (correlações de Pearson)

Embora a correlação não seja uma métrica central para análise de Redes Bayesianas uma vez que trata apenas de relações lineares, este parece ser o caso da relação entre variáveis como, por exemplo, o percentual de domicílios com computador (COMP), o percentual de domicílios com acesso à internet em casa (INT) e o percentual de indivíduos que utilizam a internet (USR). Os elevados valores de correlação de Pearson, informação mútua e entropia relativa parecem sugerir que estes indicadores estão de tal maneira associados que torna questionável a necessidade de utilização de mais de um deles na elaboração de um índice composto.

Aplicando-se o algoritmo de clusterização de variáveis disponível no software e que permite o agrupamento de nós em função das relações probabilísticas entre eles, foram identificados três clusters destacados por cores diferentes Figura 7.

A obtenção destes clusters sugere a existência de fatores subjacentes a cada um dos grupos de nós que poderiam substituir as variáveis na composição de um novo índice

menos sujeito a um eventual desbalanceamento causado pela utilização de variáveis redundantes no cálculo.

Figura 7 - Identificação por cores dos clusters de variáveis da Rede Bayesiana dos indicadores de cálculo do IDI

Baseando-se na experiência prévia deste autor, algumas conexões obtidas parecem fazer bastante sentido, como as que formam os ramos COMP/INT/INT MOV (tecnologias associadas), COMP/USR/BAND (infraestrutura associada ao volume de usuários) e INT FIX/FIX (tecnologias associadas). Já as conexões formando o ramo SEC/CEL/ALF (indicando uma relação inesperada entre índices associados à alfabetização e à educação e a quantidade de celulares) e o ramo INT FIX/TER parecem mais surpreendentes e, apesar de apresentarem valores de informação mútua mais baixos, talvez mereçam uma avaliação mais detalhada em trabalhos futuros.

Uma vez que o questionamento dos componentes de cálculo do IDI ou a proposição de um novo índice não fazem parte do escopo deste trabalho, foi dado prosseguimento na análise aplicando-se o algoritmo de clusterização de dados para obtenção de uma Rede Bayesiana classificadora.

5.2 Clusterização e inferência

A clusterização dos indicadores componentes do IDI resultou na Rede Bayesiana *Naive* apresentada na Figura 8, onde oito clusters foram identificados pelo algoritmo e associados ao nó central resultante:

Figura 8 - Rede Bayesiana gerada a partir da clusterização dos indicadores de cálculo do IDI

A Figura 9 apresenta as probabilidades a priori para cada um dos estados possíveis de cada nó, descritos aqui pelo nome da coluna no banco de dados (fixed=FIX, fixed_broadband=INT FIX, mobile=CEL, mobilebroadband=INT MOV, bandwidth=BAND, secondary=SEC, computer=COMP, Tertiary=TER, casas_internet=INT, literacy=ALF e pessoas_internet=USR). A partir do conhecimento do valor de cada um dos nós correspondentes aos indicadores de cálculo do IDI para uma determinada sub-região, pode-se verificar o efeito sobre as probabilidades de ela pertencer a cada um dos clusters correspondentes ao nó central.

Figura 9 - Probabilidades associadas aos estados possíveis dos nós da Rede Bayesiana obtida pela clusterização dos indicadores de cálculo do IDI

Os clusters com a distribuição dos países são apresentados no Quadro 6, juntamente com a distribuição das sub-regiões brasileiras obtidas por inferência alimentando-se a rede com os dados disponíveis para cada uma delas. A partir de uma análise dos países pertencentes a cada cluster foi possível realizar um ranqueamento dos clusters indo daquele composto por países mais avançados na utilização das TIC (número 1) até os menos avançados (número 8).

Quadro 6 – Distribuição dos países e sub-regiões brasileiras entre os clusters determinados

	1	2	3	4	5	6	7	8			
Países	Alemanha	Austrália	Espanha	Bahrain	Antígua e Barbuda	Albânia	México	África do Sul	Síria	Angola	Lao
	Bélgica	Áustria	Estônia	Barbados	Argentina	Arábia Saudita	Moldávia	Argélia	Sri Lanka	Bangladesh	Madagáscar
	Canadá	Cingapura	Grécia	Brunei Darussalam	Bielorrússia	Armênia	Oman	Bolívia	Suazilândia	Benin	Mali
	Coreia do Sul	Estados Unidos	Irlanda	Catar	Bósnia e Herzegovina	Azerbaijão	Panamá	Botswana	Tailândia	Burkina Faso	Mauritânia
	Dinamarca	Japão	Israel	Chipre	Bulgária	Brasil	Peru	Cuba	Turcomenistão	Butão	Moçambique
	Finlândia	Nova Zelândia	Itália	Croácia	Chile	Cabo Verde	Quênia	El Salvador	Ucrânia	Camarões	Nepal
	França		Portugal	Emirados Árabes	Romênia	Cazaquistão	Rep. Dominicana	Equador	Uzbequistão	Camboja	Níger
	Holanda			Eslovênia	Rússia	China	Sérvia	Fiji		Chade	Nigéria
	Hong Kong			Hungria	Trinidad & Tobago	Colômbia	Seychelles	Gabão		Comores	Papua Nova Guiné
	Islândia			Letônia	Turquia	Costa Rica	Suriname	Guatemala		Congo	Paquistão
	Luxemburgo			Lituânia	Uruguai	Egito	Tunísia	Honduras		Costa do Marfim	Ruanda
	Noruega			Macau, China		Filipinas	Venezuela	Indonésia		Djibouti	Senegal
	Reino Unido			Macedônia		Geórgia	Vietnã	Irã		Eritreia	Tanzânia
	Suécia			Malásia		Guiana		Marrocos		Etiópia	Togo
	Suíça			Malta		Jamaica		Mongólia		Gâmbia	Uganda
			Montenegro		Jordânia		Namíbia		Gana	Zâmbia	
			Polónia		Líbano		Nicarágua		Guiné	Zimbábue	
			República Checa		Maldivas		Paraguai		Iémen		
			República Eslovaca		Maurício		Quirguistão		Índia		
Zona						Urbana				Rural	
Regiões						Centro-Oeste Nordeste Norte	Sudeste Sul				
Estados				Distrito Federal	Paraná São Paulo	Acre Amapá Amazonas Bahia Ceará Goiás Mato Grosso Mato Grosso do Sul Minas Gerais Pará	Pernambuco Piauí Rio de Janeiro Rio Grande do Norte Rio Grande do Sul Rondônia Roraima Santa Catarina Sergipe Tocantins	Espírito Santo	Alagoas Maranhão Paraíba		
Regiões metropolitanas					RM Belo Horizonte RM Curitiba RM Rio de Janeiro RM Salvador RM São Paulo São Paulo - Interior	Bahia - Interior MG - Interior Paraná - Interior RJ - Interior Rio Grande do Sul - Interior	RM Belém RM Fortaleza RM Porto Alegre RM Recife		Ceará - Interior Pará - Interior Pernambuco - Interior		

5.3 Análise dos resultados

Avaliando o Brasil em relação aos demais países do mundo vemos que a classificação no sexto cluster demonstra um aproveitamento das TIC ainda bastante aquém do desejado para um país integrante das dez maiores economias do mundo, ficando bem atrás de diversos países da Europa, Ásia e América do Norte. A comparação é melhor quando considerados os demais países membros do grupo BRICS que designa um grupo de países emergentes de destaque na economia mundial, com o Brasil se posicionando atrás apenas da Rússia, e ficando no mesmo cluster da China e à frente dos membros África do Sul e Índia. O mesmo não ocorre em relação aos vizinhos sul-americanos Argentina, Chile e Uruguai, que se classificam à frente do Brasil.

No caso da divisão digital interna, observa-se que a maioria das sub-regiões compartilha a sexta posição do Brasil, com alguns poucos destaques como o Distrito Federal classificado isoladamente como o mais avançado no cluster 4, assim como os estados de São Paulo e Paraná e algumas grandes regiões metropolitanas como Rio de Janeiro, Belo Horizonte e Salvador um pouco à frente do resto do país, classificados no cluster 5. Como destaques negativos aparecem as regiões rurais, os estados de Alagoas, Maranhão e Paraíba além do interior dos estados do Pará, Ceará e Pernambuco figurando no último cluster, dois níveis atrás da média nacional.

Comparando-se a distribuição dos países em clusters com o valor e a posição no *ranking* IDI, verificamos correlações fortes e estatisticamente significantes entre o número do cluster de cada país e o valor do seu IDI (coeficiente de correlação de Spearman = -0,970) e o da sua posição no *ranking* de IDI (coeficiente de correlação de Pearson = 0,941) sugerindo que, apesar de apresentar uma classificação limitada a oito níveis, a distribuição dos países em clusters obtida é adequada como uma forma alternativa ao próprio IDI para medição da divisão digital entre eles.

Como avaliação complementar verificou-se a correlação entre o número do cluster e o PIB per capita, indicador utilizado como representativo de riqueza econômica que, conforme mencionado anteriormente, é um fator citado em diversos trabalhos na literatura de ICT4D como de forte influência na adoção e disseminação das TIC. A verificação do coeficiente de correlação de Spearman = -0,823 entre o PIB per capita de

cada país (apresentado no Apêndice 3) e o *ranking* de seu cluster demonstra uma forte correlação. Já para as sub-regiões brasileiras, o coeficiente de correlação de Spearman entre os clusters e o PIB per capita de cada sub-região (apresentados no Apêndice 4) foi de -0,513, indicando uma correlação moderada. Em ambos os casos a correlação se demonstrou estatisticamente significativa, confirmando a relação entre o avanço no grau de utilização das TIC e o PIB per capita de uma determinada região.

6 AVALIANDO OS FATORES DE INFLUÊNCIA NA DIVISÃO DIGITAL NO BRASIL

Seguindo novamente os passos da associação das metodologias de Kurgan e Musilek (2006) e Lauría e Duchessi (2007), desta vez com o objetivo de avaliar os fatores de influência na determinação da divisão digital nas diferentes subdivisões regionais brasileiras, a revisão da literatura para o entendimento do domínio da aplicação identificou a renda, o sexo, o grau de instrução, a idade, a localização geográfica, a raça, o idioma (particularmente o conhecimento do inglês) além de questões institucionais (VICENTE; LÓPEZ, 2011) como fatores bastante citados como tendo impacto na difusão das TIC. Já dentre as variáveis dependentes utilizadas como indicadores representativos da adoção tecnológica foram selecionadas algumas das mais recorrentes: uso do computador, uso da internet e uso do celular (BILLON et al., 2009; VICENTE; LÓPEZ, 2011).

Buscando avaliar também algumas das aplicações dadas à internet, foram incluídas na análise a adoção do comércio eletrônico e do internet banking pelos usuários de internet.

A maior parte das variáveis identificadas na literatura como possíveis fatores de influência faz parte do levantamento sobre uso das TIC no Brasil 2010 realizado pelo Comitê Gestor da Internet no Brasil (CGI.BR, 2010). As exceções são o conhecimento do idioma inglês e os fatores institucionais, que acabaram sendo eliminados da análise. Apesar de não mencionada nos trabalhos que fizeram parte da revisão de literatura, decidiu-se pela inclusão da variável correspondente à situação de emprego do entrevistado (Aposentado(a)/Desempregado(a)/Dona de casa/Estudante/Trabalhando) por parecer ser uma variável de possível impacto e em função da sua disponibilidade no levantamento do CGI.br .

Desta forma a lista de variáveis independentes escolhida para obtenção do modelo incluiu sexo, faixa etária, grau de instrução, situação de emprego, raça, classe social, zona rural/urbana, região e estado. A opção pela utilização da classe social como representativa da renda se deve ao elevado índice de ausência de resposta na questão correspondente à faixa de renda.

Para as variáveis dependentes correspondentes ao uso do computador, da internet e do celular foram consideradas as respostas confirmando o uso nos últimos três meses como positivas. Já o uso do comércio eletrônico foi considerado positivo para respondentes que afirmaram ter comprado ou encomendado produtos pela internet nos últimos 12 meses, enquanto que o internet banking considerou a realização de uma ou mais transações pela internet nos últimos três meses.

Conforme sugerido por Conrady e Jouffe (2011a) foi realizada uma primeira análise das relações entre as variáveis do estudo utilizando-se o algoritmo EQ de aprendizado não supervisionado. Foi obtida uma Rede Bayesiana representativa das variáveis associadas à adoção dos computadores, internet e celulares incluindo os respondentes da pesquisa do CGI.br de forma geral e outra rede para a adoção do comércio eletrônico e internet banking incluindo apenas os participantes que afirmaram utilizarem a internet.

Mesmo sendo possível utilizar a rede obtida a partir do aprendizado não supervisionado para a previsão de um nó designado como alvo, usualmente o aprendizado supervisionado se demonstra um método mais adequado quando existe uma variável alvo específica a ser analisada (CONRADY; JOUFFE, 2011b). Desta forma optou-se pela utilização do algoritmo de obtenção de rede *Augmented Naive Bayes* em função de seu melhor desempenho na avaliação de variáveis alvo quando comparado aos demais algoritmos supervisionados disponíveis no software Bayesialab, aplicando-o na obtenção de um modelo distinto para cada uma das variáveis dependentes consideradas.

Durante o processo de importação dos dados para a geração das redes utilizou-se uma quebra em duas amostras, a primeira incluindo 80% dos registros que foi utilizada para o aprendizado das redes e a segunda de 20%, utilizada para testes. Esta distribuição seguiu a sugestão padrão constante na documentação do software Bayesialab. Desta forma foi possível analisar o poder preditivo dos modelos gerados, utilizando-se a matriz de confusão na comparação entre as classificações previstas pelos modelos com as classificações corretas.

Não foi necessária a realização de nenhum método de redução de dimensionalidade e utilizou-se uma discretização das variáveis de acordo com os estados apresentados na distribuição da Tabela 1 apresentada no item 4.6. Nenhuma restrição semântica foi empregada, e a escolha do melhor modelo foi realizada automaticamente pelos algoritmos do software utilizado.

A análise dos modelos gerados consistiu em uma verificação da sua capacidade de prever a adoção ou não de cada uma das tecnologias a partir do conhecimento das variáveis independentes, seguida de uma avaliação do efeito que o conhecimento isolado das diferentes variáveis tem sobre a probabilidade da tecnologia ter sido adotada ou não.

6.1 Adoção de tecnologias

A Rede Bayesiana obtida a partir do algoritmo EQ para análise das relações probabilísticas entre a adoção das tecnologias e as variáveis independentes é apresentada na Figura 10, exibindo os valores correspondentes à informação mútua (em preto) e à entropia relativa de cada arco (valores em azul correspondem ao valor no sentido da seta e os vermelhos ao do sentido oposto).

Figura 10 - Rede Bayesiana com a relação entre a adoção de tecnologias e variáveis independentes

Em uma primeira análise baseada no conhecimento prévio deste autor não foi identificada nenhuma conexão conflitante de forma a sugerir a necessidade de criação de uma restrição semântica ou a existência de um possível viés do banco de dados (como seria o caso, por exemplo, de conexões diretas entre nós como sexo, idade e raça).

O valor elevado da informação mútua entre situação de emprego e idade se dá em função da lista de estados do nó emprego incluir os elementos “estudante” e “aposentado”, bastante influenciados por faixa etária. O mesmo efeito ocorre em menor grau na relação entre sexo e situação de emprego devido ao estado “dona-de-casa”. O caso do valor de 100% da entropia relativa no sentido estado brasileiro-região era esperado, uma vez que o conhecimento do estado brasileiro implica a redução da incerteza em relação à região do Brasil a zero.

A forte entropia relativa em ambos os sentidos da conexão entre o nó referente a computador e internet sugere que o seu uso no Brasil está associado de tal forma que o perfil de ambos se confunde, pelo menos no que diz respeito às demais variáveis incluídas neste trabalho. Este fato ficou ainda mais evidente no comportamento obtido em testes iniciais de modelos gerados isoladamente para ambos. Assim optou-se por incluir apenas o uso de internet neste trabalho de forma a evitar a apresentação de análises detalhadas redundantes.

Observando-se apenas as conexões diretas entre as variáveis consideradas como independentes e dependentes neste estudo é interessante notar que enquanto o conjunto formado pelos nós de forte relação internet e computador aparece como sendo influenciado diretamente por cinco variáveis independentes (idade, situação de emprego, classe social, instrução e zona), o nó correspondente ao uso do celular conecta-se diretamente apenas com idade e instrução. Além disso, as variáveis sexo, região do Brasil, estado brasileiro e raça não se conectam diretamente com nenhuma variável dependente.

Seguiu-se então com a análise detalhada de cada uma das variáveis dependentes, obtendo-se redes específicas para o uso de internet e uso do celular.

6.1.1 Internet

A rede *Augmented Naive Bayes* incluindo o uso de internet como nó alvo obtida a partir dos dados é apresentada na Figura 11, juntamente com a informação mútua e entropia relativa dos arcos de ligação:

Figura 11 - Rede *Augmented Naive Bayes* do uso de internet

A matriz de confusão obtida a partir a aplicação da rede na discriminação dos 20% dos registros utilizados na amostra de testes demonstra uma precisão total de 80,99% na identificação do uso ou não da internet, com precisão de 81,44% para identificação de “não usuários” e 80,32% para “usuários” (Tabela 2). O alto poder de previsão obtido indica que as variáveis independentes utilizadas no modelo constituem um conjunto

adequado na determinação do uso ou não de internet pelos indivíduos que participaram do levantamento.

Nas linhas da matriz de confusão são apresentadas as quantidades previstas como “Sim” ou “Não”, enquanto que as colunas mostram as quantidades reais existentes na base de dados. Os valores não inteiros se devem a existência de pesos diferentes associados a cada registro, determinados pela metodologia do levantamento do CGI.BR (2010). As células apresentam as ocorrências das previsões cruzadas com o valor real, a confiabilidade (ocorrências da célula sobre o total da linha) e a precisão (ocorrências da célula sobre o total da coluna).

Tabela 2 – Matriz de confusão para Rede *Augmented Naive Bayes* da adoção de internet

Precisão total: 80,99%		
Ocorrências		
Valor	Não (2719,47)	Sim (1828,86)
Não (2574,82)	2214,82	360
Sim (1973,51)	504,65	1468,86
Confiabilidade		
Valor	Não (2719,47)	Sim (1828,86)
Não (2574,82)	86,02%	13,98%
Sim (1973,51)	25,57%	74,43%
Precisão		
Valor	Não (2719,47)	Sim (1828,86)
Não (2574,82)	81,44%	19,68%
Sim (1973,51)	18,56%	80,32%

O Gráfico 2 apresenta a análise de sensibilidade para a utilização da internet por um indivíduo (INTERNET=Sim), fornecendo uma visão da amplitude do efeito que cada uma das variáveis tem nesta probabilidade. Com o gráfico apresentando uma probabilidade a priori de 40,84% e os valores máximo e mínimo de probabilidade que podem ser alcançados através do conhecimento de diferentes estados de cada variável é possível ter uma noção visual da amplitude da divisão digital existente dentro de cada uma das variáveis.

Gráfico 2 - Análise de sensibilidade para o uso da Internet=Sim

A análise do gráfico de sensibilidade demonstra que a variável que possui um efeito mais amplo nesta divisão é o nível de instrução, com a probabilidade de uso variando de 5,36% até 80,43%. No entanto, a classe social se apresenta como o elemento com maior capacidade isoladamente para aumentar a probabilidade de um indivíduo utilizar a internet (até uma probabilidade máxima de 85,65%), enquanto a idade apresenta a maior capacidade isolada para reduzir esta probabilidade (até uma probabilidade mínima de 4,3%).

Os diagramas de barra com a probabilidade associada a cada um dos possíveis estados das quatro variáveis de maior amplitude na divisão digital relacionada ao uso da internet são apresentados no Gráfico 3. Nestes diagramas vemos que a probabilidade diminui com a idade e aumenta em função do nível educacional e social. Para situação de emprego é interessante observar que saber se um indivíduo esta ou não empregado não traz nenhum ganho na identificação de uso da internet, enquanto a probabilidade aumenta bastante para estudantes e diminui para aposentados e donas-de-casa. Para os diagramas de todas as variáveis consulte o Apêndice 5.

P(INTERNET = Sim | ...)

Gráfico 3 - Probabilidades de uso da internet para os estados das variáveis grau de instrução, faixa etária, classe social e situação de emprego

A Tabela 3 apresentada na página seguinte exibe o resultado da avaliação da otimização realizada pelo software para os estados “Não” e “Sim” do uso de internet, exibindo os nós e estados em ordem decrescente a partir de um *score* calculado que representa o quanto cada estado contribui isoladamente para alteração da probabilidade de uso da internet.

Podemos observar que idades acima de 60 anos e o analfabetismo são os estados de maior destaque na determinação de um indivíduo que não usa a internet, enquanto que a classe A e as pessoas com nível superior são as com mais probabilidade de serem usuários de internet.

Tabela 3 – *Ranking* de contribuições para alteração da probabilidade de uso da internet

INTERNET=Não		
Nó	Estado	Score
idade	De 60 anos ou mais	36,53564
instrução	Analfabeto Até Pré-primário	35,47891
emprego	Aposentado(a)	34,83725
classe	Classe D e E	27,56372
instrução	Fundamental I	25,20507
zona	RURAL	24,25581
emprego	Dona de casa	23,75126
estado	Alagoas	23,27974
estado	Maranhão	22,82591
idade	De 45 a 59 anos	20,85603
estado	Paraíba	19,28940
estado	Espírito Santo	17,22541
estado	Ceará	15,78788
região	Nordeste	13,04979
estado	Pernambuco	10,40719
estado	Piauí	9,93837
estado	Pará	9,92057
estado	Acre	9,45048
idade	De 35 a 44 anos	7,76959
estado	Rio Grande do Norte	7,50808
estado	Sergipe	7,50271
estado	Amazonas	7,50182
estado	Tocantins	6,83056
região	Norte	6,55102
estado	Bahia	6,01030
estado	Amapá	4,74970
estado	Rio Grande do Sul	4,51361
raça	Preta/ Negra	4,19838
raça	Parda	3,66830
estado	Minas Gerais	2,49997
raça	Indígena	1,31065
sexo	Feminino	0,46838
emprego	Desempregado(a)	0,37021
estado	Santa Catarina	0,24429

INTERNET=Sim		
Nó	Estado	Score
classe	Classe A	44,80905
instrução	Superior	39,58969
classe	Classe B	31,92154
emprego	Estudante	28,30384
idade	De 10 a 15 anos	25,44952
idade	De 16 a 24 anos	22,65623
estado	Distrito Federal	18,34115
instrução	Médio	17,51647
estado	São Paulo	12,07604
idade	De 25 a 34 anos	11,36193
região	Centro-Oeste	10,26181
estado	Goiás	9,69321
estado	Rondônia	8,30387
estado	Mato Grosso	7,73754
estado	Mato Grosso do Sul	7,28398
estado	Rio de Janeiro	6,90883
região	Sudeste	6,40156
estado	Paraná	6,20418
raça	Branca	4,67561
zona	URBANA	4,62734
instrução	Fundamental II	3,01492
estado	Roraima	2,18009
raça	Amarela	1,59528
classe	Classe C	1,31102
sexo	Masculino	0,50439
região	Sul	0,41625
emprego	Trabalhando	0,00850

6.1.2 Celular

A rede *Augmented Naive Bayes* incluindo o uso de celular como nó alvo obtida a partir dos dados é apresentada na Figura 12, juntamente com a informação mútua e entropia relativa dos arcos de ligação:

Figura 12 - Rede *Augmented Naive Bayes* do uso de celular

A matriz de confusão obtida a partir a aplicação da rede na discriminação dos 20% dos registros utilizados na amostra de testes demonstra uma precisão total de 79,94% na identificação do uso ou não do celular, com precisão de 93,99% para identificação de “usuários”, mas apenas 33,21% para “não usuários” (Tabela 4). A baixa precisão na previsão de não usuários indica que o conjunto de variáveis utilizadas não é suficiente caracterizar fortemente a não adoção do uso do celular.

Tabela 4 – Matriz de confusão para Rede *Augmented Naive Bayes* da adoção de celular

Precisão total: 79,94%		
Ocorrências		
Valor	Não (1051,69)	Sim (3496,64)
Não (559,32)	349,31	210,02
Sim (3989)	702,39	3286,62
Confiabilidade		
Valor	Não (1051,69)	Sim (3496,64)
Não (559,32)	62,45%	37,55%
Sim (3989)	17,61%	82,39%
Precisão		
Valor	Não (1051,69)	Sim (3496,64)
Não (559,32)	33,21%	6,01%
Sim (3989)	66,79%	93,99%

Para verificar se um eventual problema de adequação da técnica poderia ser a fonte da baixa qualidade de previsão do modelo, foi realizada uma regressão logística utilizando o método Enter para comparação. Utilizando-se as mesmas amostras que foram utilizadas para aprendizado e teste do modelo para a rede ANB, a precisão geral obtida através da regressão foi de 78,4%, sendo 24,9% para “não” e 94,6% para “sim”. Estes resultados reforçam a conclusão de que realmente as variáveis socioeconômicas utilizadas não são suficientes para uma boa previsão do uso de celular.

A previsão do modelo se baseia em combinações de estados de variáveis que façam com que a probabilidade de adoção do celular se torne maior que 50%. Aparentemente a alta probabilidade de utilização do celular, partindo de uma probabilidade a priori de 78,94% para o uso de celular e com poucas combinações de estados de variáveis obtendo probabilidades abaixo de 50%, torna mais difícil a previsão de não adoção a partir das variáveis do modelo.

No entanto, ainda que as variáveis constantes no modelo obtido não resultem em combinações preditivas precisas, este fato não invalida a possibilidade de análise a partir das relações probabilísticas obtidas entre cada uma delas e o nó alvo.

Gráfico 4 - Análise de sensibilidade para o uso da Celular=Sim

Como pode ser notado no Gráfico 4 da análise de sensibilidade para o uso do celular em função de cada uma das variáveis utilizadas no modelo, temos neste caso uma probabilidade a priori bem maior do que a obtida para o uso da internet. Apenas a variável de situação de emprego consegue isoladamente diminuir a probabilidade a um valor menor do que 50%, e ainda assim marginalmente (49,34% para aposentados).

A variável de maior amplitude novamente é a instrução, mas abrangendo uma faixa bem menor com probabilidades de uso variando de 53,27% até 95,24%. A segunda posição em amplitude ficou com a faixa etária, indo de acordo com o modelo obtido pelo algoritmo EQ onde apenas o grau de instrução e a faixa etária se conectavam diretamente com o nó de uso de celular. A classe social, apesar de cair para quarta posição em amplitude, também é neste caso a variável com maior poder isolado de aumentar a probabilidade de uso do celular (até uma probabilidade máxima de 96,00%).

As menores amplitudes indicam que, no que diz respeito às variáveis incluídas no modelo, a divisão digital em relação à adoção do celular se dá em um nível menor que o da internet.

Os diagramas de barra com a probabilidade associada a cada um dos possíveis estados das quatro variáveis de maior amplitude na divisão digital relacionada ao uso da internet são apresentados no Gráfico 5. Nestes diagramas vemos que a probabilidade aumenta

em função do nível educacional e social, mas ao contrário da internet onde a probabilidade do uso cai em função da idade, no caso do celular ocorre um aumento da faixa de 10 a 15 anos para a de 16 a 24 anos, passando a decrescer para faixas etárias maiores. Outro aspecto contrastante é que o fato de ser estudante praticamente não apresenta alterações em relação à probabilidade a priori. Para os diagramas de barras de todas as variáveis do modelo consulte o Apêndice 6.

P(USOU CELULAR = Sim | ...)

Gráfico 5 - Probabilidades de uso do celular para os estados das variáveis grau de instrução, faixa etária, classe social e situação de emprego

A Tabela 5 apresenta o resultado da avaliação da otimização da variável correspondente ao uso do celular tanto para “Não” como para “Sim”, listando os estados das variáveis em função do seu poder isolado de afetar a probabilidade de adoção do celular. Podemos observar o destaque do efeito negativo no uso do celular a partir dos *scores* obtidos para aposentados, maiores de 60 anos e pessoas com instrução até o pré-primário.

Tabela 5 – Ranking de contribuições para alteração da probabilidade de adoção de celular

CELULAR=Não		
Nó	Estado	Score
emprego	Aposentado(a)	29,59544
idade	De 60 anos ou mais	28,57499
instrução	Analfabeto Até Pré-primário	25,66746
classe	Classe D e E	17,00694
estado	Amazonas	14,25120
instrução	Fundamental I	12,78267
zona	RURAL	11,93658
estado	Minas Gerais	8,50192
estado	Piauí	7,30432
emprego	Dona de casa	7,08745
estado	Maranhão	6,92517
estado	Espírito Santo	5,70391
estado	Bahia	4,99381
raça	Preta/ Negra	4,40826
estado	Sergipe	2,84862
estado	Ceará	2,46792
idade	De 45 a 59 anos	2,30189
região	Nordeste	2,05903
estado	Pernambuco	1,91696
estado	Santa Catarina	0,97290
idade	De 10 a 15 anos	0,90575
sexo	Masculino	0,83359
raça	Parda	0,75568
estado	Paraná	0,39824
região	Sudeste	0,07185

CELULAR=Sim		
Nó	Estado	Score
classe	Classe A	17,05953
instrução	Superior	16,30608
instrução	Médio	12,79509
idade	De 16 a 24 anos	12,78773
classe	Classe B	12,65183
estado	Acre	12,60090
idade	De 25 a 34 anos	11,87609
estado	Mato Grosso	9,06757
estado	Distrito Federal	9,03227
estado	Tocantins	8,55297
estado	Rondônia	7,62802
região	Centro-Oeste	6,68329
estado	Amapá	6,61053
idade	De 35 a 44 anos	6,28980
estado	Paraíba	6,14814
estado	Goiás	5,97699
estado	Alagoas	5,46045
raça	Indígena	5,45955
emprego	Trabalhando	5,21981
estado	Roraima	4,35078
raça	Amarela	4,24303
classe	Classe C	3,63750
estado	Rio de Janeiro	3,57010
estado	Mato Grosso do Sul	3,25988
estado	Rio Grande do Norte	3,06848
estado	São Paulo	2,69464
zona	URBANA	2,27740
estado	Rio Grande do Sul	2,11009
raça	Branca	1,82721
instrução	Fundamental II	1,80366
emprego	Desempregado(a)	1,78556
estado	Pará	1,75566
sexo	Feminino	0,76895
região	Sul	0,50138
região	Norte	0,34802
emprego	Estudante	0,28445

6.2 Aplicações da internet

A Rede Bayesiana obtida a partir do algoritmo EQ para análise das relações probabilísticas entre o uso do comércio eletrônico (eCOMMERCE) e do internet banking (BANKING) e as variáveis independentes é apresentada na Figura 13, exibindo os valores correspondentes à informação mútua (em preto) e à entropia relativa de cada arco (valores em azul correspondem ao valor no sentido da seta e os vermelhos ao do sentido oposto).

Figura 13 - Rede Bayesiana com a relação entre uso do comércio eletrônico e do internet banking e variáveis independentes

Novamente uma primeira análise baseada no conhecimento prévio deste autor não identificou nenhuma conexão conflitante ou a existência de um possível viés do banco de dados.

Assim como aconteceu para o modelo de adoção de tecnologias, foram obtidas as relações esperadas entre situação de emprego e idade, sexo e situação de emprego e estado brasileiro e região.

Observando-se apenas as conexões diretas entre as variáveis consideradas como independentes e dependentes neste estudo vemos que enquanto a variável do uso do comércio eletrônico aparece sendo influenciada diretamente por três variáveis independentes (classe social, instrução e sexo), o nó correspondente ao uso do internet banking conecta-se diretamente apenas com a idade. Além disso, as variáveis região do Brasil, estado brasileiro, zona e raça não se conectam diretamente com nenhuma variável dependente.

Seguiu-se com a análise detalhada de cada uma das variáveis dependentes, obtendo-se redes específicas para a o uso do comércio eletrônico e do internet banking.

6.2.1 Comércio eletrônico

A rede Augmented Naive Bayes incluindo o uso do comércio eletrônico como nó alvo obtida a partir dos dados é apresentada na Figura 15, juntamente com a informação mútua e entropia relativa dos arcos de ligação:

Figura 14 - Rede Augmented Naive Bayes do uso de comércio eletrônico

A matriz de confusão obtida a partir a aplicação da rede na discriminação dos 20% dos registros utilizados na amostra de testes demonstra uma precisão total de 75,85% na identificação de compra ou não através da internet, com precisão de 97,20% para identificação de “não compradores”, mas com apenas o valor bastante baixo de 12,65% para “compradores” (Tabela 6). A baixa precisão na previsão de “compradores” indica que o conjunto de variáveis utilizadas não é suficiente caracterizar fortemente a utilização do comércio eletrônico.

Novamente foi realizada uma regressão logística utilizando o método Enter para comparação, resultando em uma precisão geral de 75,9%, com 97,6% para “não” e 11,7% para sim, reforçando a inadequação do conjunto de variáveis para previsão.

Ao contrário do uso do celular onde a probabilidade de adoção parte de um valor a priori bastante elevado, para o uso do comércio eletrônico o valor a priori é de apenas 22,21%, com poucas combinações de estados de variáveis resultando em probabilidades maiores do que 50%.

Tabela 6 - Matriz de confusão para Rede *Augmented Naive Bayes* de uso do comércio eletrônico

Precisão total: 75,85%		
Ocorrências		
Valor	Não (1379,43)	Sim (466,11)
Não (1748,01)	1340,85	407,15
Sim (97,54)	38,58	58,96
Confiabilidade		
Valor	Não (1379,43)	Sim (466,11)
Não (1748,01)	76,71%	23,29%
Sim (97,54)	39,55%	60,45%
Precisão		
Valor	Não (1379,43)	Sim (466,11)
Não (1748,01)	97,20%	87,35%
Sim (97,54)	2,80%	12,65%

A análise do gráfico de sensibilidade (Gráfico 6) demonstra que a variável que possui um efeito mais amplo na divisão do comércio eletrônico é a classe social, se destacando por uma amplitude bem maior que as demais variáveis, com a probabilidade variando de 9,18% até 64,62%. Este valor também é o máximo alcançado por uma variável isoladamente. A variável correspondente aos estados brasileiros ganhou mais destaque neste caso do que na adoção de celular e internet, apresentando o valor mínimo possível de ser obtido isoladamente por uma variável (1,6% para o estado do Alagoas) e a terceira maior amplitude, substituindo a situação de emprego entre as quatro principais variáveis neste quesito.

Gráfico 6 - Análise de sensibilidade para o uso de comércio eletrônico=Sim

Os diagramas de barra com a probabilidade associada a cada um dos possíveis estados das quatro variáveis de maior amplitude na divisão digital relacionada ao uso do comércio eletrônico são apresentados nos Gráfico 7 e Gráfico 8. Observa-se que da mesma forma que ocorre para o uso de internet, a probabilidade de uso do comércio eletrônico aumenta em função do nível educacional e social, mas ao contrário da internet onde a probabilidade do uso cai em função da idade, para o comércio eletrônico ela aumenta até a faixa de 45 a 59 anos, sofrendo uma queda brusca a partir dos 60 anos. As grandes diferenças encontradas entre alguns estados não apresentam uma justificativa aparente de fácil identificação, como seria o caso de uma influência de PIB ou da localização geográfica, possivelmente estando ligada a fatores institucionais e/ou culturais e merecendo uma investigação maior em trabalhos futuros.

Para os diagramas de todas as variáveis consulte o Apêndice 7.

P(eCOMMERCE = Sim | ...)

Gráfico 7 - Probabilidades de uso do comércio eletrônico para os estados das variáveis grau de instrução, faixa etária e classe social

Gráfico 8 - Probabilidades de uso do comércio eletrônico para os estados da variável estado brasileiro

A Tabela 7 apresentada na página seguinte exhibe o resultado da avaliação da otimização da variável correspondente ao uso do comércio eletrônico tanto para “Não” como para “Sim” onde vemos o destaque do efeito positivo de pertencer a classe A no uso do comércio eletrônico.

Tabela 7 – Ranking de contribuições para alteração da probabilidade de uso de comércio eletrônico

eCommerce=Não		
Nó	Estado	Score
estado	Alagoas	20,60911
instrução	Analfabeto Até Pré-primário	16,46111
estado	Rio Grande do Norte	16,13380
idade	De 10 a 15 anos	15,14494
estado	Paraíba	14,74216
classe	Classe D e E	13,02618
instrução	Fundamental I	11,79476
estado	Goiás	11,78418
instrução	Fundamental II	11,31891
emprego	Estudante	11,09872
estado	Rio Grande do Sul	10,84496
estado	Sergipe	9,92336
estado	Bahia	8,56636
zona	RURAL	8,00562
estado	Pernambuco	7,83579
região	Nordeste	7,36541
classe	Classe C	6,74052
estado	Santa Catarina	5,89051
região	Sul	5,58084
emprego	Desempregado(a)	5,27125
emprego	Aposentado(a)	5,17196
estado	Espírito Santo	4,71275
raça	Preta/ Negra	4,25576
estado	Amazonas	3,71663
raça	Parda	3,10268
estado	Ceará	2,92184
região	Centro-Oeste	2,21800
sexo	Feminino	1,78303
idade	De 16 a 24 anos	1,42285
raça	Indígena	1,26956
raça	Amarela	1,02371
estado	Pará	0,91344
idade	De 60 anos ou mais	0,71086
estado	Piauí	0,63145
emprego	Dona de casa	0,49546
estado	Paraná	0,11263

eCommerce=Sim		
Nó	Estado	Score
classe	Classe A	42,41451
instrução	Superior	14,71302
classe	Classe B	10,60390
idade	De 45 a 59 anos	8,71533
estado	Rondônia	8,24422
idade	De 35 a 44 anos	8,12521
estado	Acre	7,86763
estado	Minas Gerais	7,64878
estado	Tocantins	7,61321
estado	Amapá	6,56461
estado	Rio de Janeiro	6,53626
estado	Distrito Federal	5,63278
região	Sudeste	4,85985
estado	Mato Grosso	4,78331
idade	De 25 a 34 anos	4,62904
estado	Maranhão	4,43785
emprego	Trabalhando	4,37068
estado	São Paulo	3,94157
raça	Branca	3,41456
sexo	Masculino	1,87468
estado	Mato Grosso do Sul	1,16188
estado	Roraima	0,95095
região	Norte	0,92841
zona	URBANA	0,44152
instrução	Médio	0,16639

6.2.2 Internet banking

A rede *Augmented Naive Bayes* incluindo o uso do internet banking como nó alvo obtida a partir dos dados é apresentada na Figura 15, juntamente com a informação mútua e entropia relativa dos arcos de ligação:

Figura 15 - Rede *Augmented Naive Bayes* do uso de internet banking

A matriz de confusão obtida a partir a aplicação da rede na discriminação dos 20% dos registros utilizados na amostra de testes demonstra uma precisão total de 80,55% na identificação da realização ou não de transações bancárias através da internet, com precisão de 96,57% para identificação de “não usuários”, mas com apenas o valor bastante baixo de 7,54% para “usuários” (Tabela 8). Da mesma forma como ocorre para o comércio eletrônico, a baixa precisão na previsão de “usuários” indica que o conjunto de variáveis utilizadas não é suficiente para caracterizar fortemente a utilização do internet banking. Este fato foi confirmado pela realização de regressão logística comparativa onde se obteve precisão geral de 81,9%, sendo 98,9% para “não” e 4,2% para “sim”.

Tabela 8 - Matriz de confusão para Rede *Augmented Naive Bayes* de uso do internet banking

Precisão total: 80,55%		
Ocorrências		
Valor	Não (1513,34)	Sim (332,2)
Não (1768,65)	1461,49	307,16
Sim (76,89)	51,85	25,04
Confiabilidade		
Valor	Não (1513,34)	Sim (332,2)
Não (1768,65)	82,63%	17,37%
Sim (76,89)	67,44%	32,56%
Precisão		
Valor	Não (1513,34)	Sim (332,2)
Não (1768,65)	96,57%	92,46%
Sim (76,89)	3,43%	7,54%

De todas as variáveis dependentes analisadas, o uso do internet banking foi a que apresentou menores valores de probabilidade de adoção, com uma probabilidade de utilização a priori de 16,87% e sem que nenhuma variável consiga isoladamente elevar a probabilidade para mais do que 40%.

O gráfico de sensibilidade (Gráfico 9) demonstra novamente a classe social com o efeito isoladamente mais amplo na divisão do internet banking, com a probabilidade variando de 5,37% até 39,34%, sendo este valor também o máximo alcançado por uma variável isoladamente. Novamente como no comércio eletrônico, a variável correspondente aos

estados brasileiros ganha destaque, apresentando o valor mínimo possível de ser obtido isoladamente por uma variável (0% para o estado do Alagoas) e a segunda maior amplitude.

Gráfico 9 - Análise de sensibilidade para o uso de internet banking=Sim

Os diagramas de barra com a probabilidade associada a cada um dos possíveis estados das quatro variáveis de maior amplitude na divisão digital relacionada ao uso do comércio eletrônico são apresentados nos Gráfico 10 e Gráfico 11. Novamente vemos a probabilidade de uso crescendo em função do nível educacional e social. Semelhante ao encontrado no caso do comércio eletrônico, para o internet banking também ocorre uma variação grande das probabilidades de utilização em diferentes estados brasileiros sem uma justificativa aparente de fácil identificação.

Para os diagramas de todas as variáveis consulte o Apêndice 8.

Um aspecto interessante neste caso foi o comportamento em função da idade, que a partir dos 25 anos apresenta as probabilidades oscilando pouco em torno de 24%, sendo este o único caso dentre os analisados neste trabalho onde não ocorre uma redução acentuada para faixa etária acima dos 60 anos.

P(BANKING = Sim | ...)

Gráfico 10 - Probabilidades de uso do internet banking para os estados das variáveis grau de instrução, faixa etária e classe social

Gráfico 11 - Probabilidades de uso do internet banking para os estados da variável estado brasileiro

A Tabela 9 apresentada na página seguinte exhibe o resultado da avaliação da otimização da variável correspondente ao uso do internet banking tanto para “Não” como para “Sim”.

Tabela 9 – Ranking de contribuições para alteração da probabilidade de uso de internet banking

eBanking=Não		
Nó	Estado	Score
estado	Alagoas	16,86690
instrução	Analfabeto Até Pré-primário	16,36387
estado	Paraíba	15,46964
idade	De 10 a 15 anos	14,80453
emprego	Estudante	13,39868
estado	Rio Grande do Norte	12,21273
classe	Classe D e E	11,49279
estado	Acre	11,41732
instrução	Fundamental I	10,36146
instrução	Fundamental II	9,60824
estado	Pernambuco	8,39390
estado	Goiás	7,81955
região	Nordeste	7,05435
estado	Pará	6,74770
estado	Amazonas	6,42174
estado	Maranhão	6,33410
região	Norte	5,33812
classe	Classe C	5,12820
estado	Ceará	4,64722
estado	Piauí	4,62004
emprego	Dona de casa	4,43000
estado	Bahia	4,33162
estado	Tocantins	4,19878
idade	De 16 a 24 anos	4,09559
estado	Sergipe	3,59136
estado	Rio Grande do Sul	3,50255
zona	RURAL	3,33271
emprego	Desempregado(a)	3,17948
estado	Santa Catarina	3,16638
raça	Indígena	3,03634
região	Sul	2,77230
sexo	Feminino	2,39094
raça	Parda	1,91264
estado	Paraná	1,79492
estado	Roraima	0,70007
região	Centro-Oeste	0,61751
instrução	Médio	0,53629
estado	Rondônia	0,29377

eBanking=Sim		
Nó	Estado	Score
classe	Classe A	22,47854
estado	Espírito Santo	16,79364
instrução	Superior	13,85023
classe	Classe B	9,20216
estado	Rio de Janeiro	8,87808
idade	De 25 a 34 anos	8,41295
idade	De 45 a 59 anos	7,95160
estado	Distrito Federal	7,81704
emprego	Aposentado(a)	7,64381
idade	De 35 a 44 anos	6,02121
idade	De 60 anos ou mais	5,74279
emprego	Trabalhando	4,80828
região	Sudeste	4,39587
estado	São Paulo	2,85491
sexo	Masculino	2,51384
estado	Mato Grosso	2,44257
estado	Minas Gerais	1,86832
raça	Branca	1,44077
estado	Amapá	1,25178
raça	Amarela	0,55707
estado	Mato Grosso do Sul	0,53217
zona	URBANA	0,18380
raça	Preta/ Negra	0,14219

6.3 Análise dos resultados

As variáveis selecionadas para construção dos modelos de previsão de adoção só demonstraram ser um conjunto adequado para a previsão do uso de internet. Uma hipótese para isto acontecer pode ser a taxa de penetração da tecnologia na população em geral, levando tecnologias com baixas taxas de adoção (caso do comércio eletrônico e internet banking) e altas taxas de adoção (celular) a apresentarem maior dificuldade de discriminação a partir das variáveis socioeconômicas.

No que se refere à influência isolada de cada um dos fatores, segue uma análise para cada uma das variáveis independentes consideradas no estudo:

Renda

Representada neste estudo pela classe social, a renda demonstrou grande importância na determinação das quatro variáveis dependentes analisadas, com um aumento da probabilidade de adoção para níveis sociais mais elevados.

Indo de acordo com estudos como os de Chinn e Fairlie (2010) e Beilock e Dimitrova (2003), ela se demonstrou o fator isoladamente de maior importância no aumento da probabilidade de uso de uma determinada tecnologia. No entanto, em nenhum caso ela demonstrou a maior capacidade isolada para diminuição desta probabilidade, sugerindo um papel de maior “incentivadora”, mas não de maior “obstáculo”.

Considerando-se a amplitude das probabilidades de adoção determinadas pela classe social, ela fica em primeiro lugar para determinação do uso do comércio eletrônico e internet banking, caindo para segundo lugar na adoção da internet e para quarto no caso do celular. Comparando a probabilidade a priori de cada uma dessas variáveis vemos que a classe social parece diminuir de importância em função do aumento da probabilidade de adoção considerando-se a população total da amostra. Uma hipótese que pode ser levantada em função disto é a renda ser a primeira barreira de entrada na adoção de uma nova tecnologia, perdendo importância conforme aumenta a penetração da tecnologia na população em geral.

Grau de instrução

Da mesma forma que a classe social, o grau de instrução também demonstrou ser de grande importância na determinação das quatro variáveis dependentes analisadas, com um aumento da probabilidade de adoção para graus de instrução mais elevados, indo de acordo com o esperado por diversos estudos anteriores (CRENSHAW; ROBISON, 2006; HARGITTAL, 1999; VICENTE e LÓPEZ, 2006; Billon et al., 2007).

O grau de instrução assumindo a primeira posição em importância considerando-se a amplitude das probabilidades para a adoção da internet e do celular sugere um possível papel desta variável como principal barreira de entrada para o uso de tecnologias com uma taxa mais alta de penetração.

Faixa etária

Variável com bastante influência na determinação das probabilidades de adoção das quatro variáveis dependentes, a idade se destacou por não apresentar um comportamento uniforme.

Para uso da internet e do celular observou-se uma diminuição das probabilidades em função da idade, com exceção da faixa até 15 anos no caso do celular, provavelmente limitada pela dependência da decisão ou capacidade financeira dos pais em fornecer o celular para os filhos. Este resultado está de acordo com estudos prévios de Alampay (2006) e Chinn e Fairlie (2007).

Para o comércio eletrônico ocorreu um aumento das probabilidades em função da idade, com uma queda brusca a partir dos 60 anos, enquanto que no caso do internet banking a probabilidade de uso cresce até os 25 anos e a partir daí oscila próximo a 24%, sem queda mesmo após os 60 anos. Uma hipótese possível para este fenômeno seria o fato de que enquanto a compra de produtos é uma opção, transações bancárias são necessárias e questões como problemas de mobilidade associadas a idades mais avançadas podem servir como incentivo para romper as barreiras existentes ao uso do comércio eletrônico para o caso do internet banking.

De qualquer forma, as quedas observadas em idades mais avançadas não estão associadas a questões de educação ou de renda, mas ainda assim deixam de fora uma

parcela da população com grau de instrução e poder de consumo que seriam suficientes para a utilização das tecnologias. Neste sentido, as iniciativas de inclusão parecem requerer uma abordagem de aspectos de natureza cultural, motivação e interesse pessoal.

Situação de emprego

Com algumas exceções, a análise do comportamento em função da situação de emprego pode estar associada a outras variáveis, como é o caso da relação entre a idade acima de 60 anos e ser aposentado.

Um destaque se dá para a condição de estudante que, apesar de não afetar significativamente as demais variáveis dependentes, aumenta drasticamente a probabilidade de uso da internet, indo de acordo com a afirmação de que fins educativos podem também gerar uma forte demanda por tecnologia (FAIRLIE et al., 2010).

Outra observação interessante é a diminuição da probabilidade de uso da internet e do celular no caso das donas-de-casa, confirmando a afirmação de Alampay (2006) de que em função das maiores responsabilidades que as mulheres tem em casa é esperado que o acesso as facilidades de comunicação também seja afetado.

Localização (zona rural/urbana, região e estado)

Ainda que apresentando menor amplitude de variação do que as variáveis citadas anteriormente, de uma forma geral a localização em zona rural contribui para uma queda da probabilidade de adoção de todas as variáveis dependentes, indo de acordo com estudos anteriores (CRENSHAW; ROBISON, 2006 ; LIU ;SAN, 2006).

No caso específico do uso do comércio eletrônico e internet banking, a influência dos estados brasileiros se demonstrou forte, sem que fosse possível identificar uma explicação evidente, uma vez que estados com perfis de PIB, localização geográfica e proporção rural/urbana semelhantes (como por exemplo, Alagoas e Maranhão no caso do comércio eletrônico) apresentaram resultados opostos. Este aspecto merece investigações mais detalhadas em trabalhos futuros, buscando identificar fatores que expliquem o comportamento observado.

Sexo

O sexo do respondente demonstrou uma influência de baixa amplitude na determinação das probabilidades de adoção das tecnologias, sendo que apenas para o uso de celular ocorre um aumento de probabilidade de uso pelas mulheres, enquanto que na adoção das demais se observou um aumento para os homens. Isto vai de acordo com o estudo de Punamaki et al. (2007) que identificaram os meninos como usuários do computador para navegar pela internet com mais frequência do que as meninas, enquanto estas são usuárias mais frequentes de telefones celulares.

Raça

De uma forma geral a amplitude de variação das probabilidades em função da raça não foi muito elevada, mas de forma consistente indicou a raça branca como a de maior probabilidade de utilização de uma tecnologia, enquanto que a raça negra quase sempre ficou com a menor probabilidade. A exceção foi o caso de internet banking onde a menor probabilidade de uso se deu para os indígenas.

7 CONCLUSÃO

Seguindo os objetivos deste trabalho apresentados no capítulo 1, foi obtido um modelo de classificação em clusters que possibilitou a identificação do grau de desigualdade da divisão digital entre as diferentes sub-regiões brasileiras, permitindo a comparação em um contexto internacional através da associação com os países pertencentes ao mesmo cluster da sub-região.

A forte concentração das diferentes sub-regiões em um mesmo cluster demonstra certa uniformidade na divisão digital regional do Brasil, situando-o em um nível abaixo do desejado para um país de destaque na economia mundial. Esta forte concentração também demonstra uma das limitações do modelo obtido no sentido de não permitir uma identificação mais precisa das diferenças entre estas regiões.

Um aspecto interessante da utilização de clusters é que, apesar de apresentar uma classificação menos precisa de cada país, representa de melhor forma a proximidade entre eles. Assim, enquanto a posição 64 no *ranking* do IDI (conforme apresentado no Apêndice 1) sugere o Brasil em uma posição à frente de mais da metade dos 152 países incluídos, a posição no sexto de oito clusters indica que apenas dois degraus foram galgados, faltando ainda 5 para alcançar o grupo de países com melhor aproveitamento das TIC.

No que diz respeito à técnica de Redes Bayesianas utilizada, a clusterização em um modelo que possibilita a inferência a partir de um conjunto limitado de informações demonstrou uma possibilidade de sua aplicação como forma complementar de análise em situações de incerteza com relação aos dados. Assim foi possível realizar análises mesmo em um cenário de falta de informações completas sobre diferentes sub-regiões brasileiras.

Desta forma, os resultados permitiram responder satisfatoriamente a primeira questão de pesquisa proposta, posicionando o Brasil no contexto internacional da divisão digital.

Quanto ao objetivo de identificar o grau de influência de diferentes fatores na determinação da divisão digital no Brasil, a escolha de quatro representantes das tecnologias assim como de diversas variáveis independentes baseando-se na revisão da literatura permitiu uma identificação de fatores com bastante influência na adoção das

TIC, com um destaque maior para a classe social e o grau de instrução, conforme o esperado de acordo com a literatura.

No entanto, a segunda questão de pesquisa sobre os fatores que influenciam a divisão digital no Brasil não pode ser totalmente respondida, uma vez que o conjunto de fatores selecionado demonstrou ser bastante preciso apenas na identificação da divisão digital em relação ao uso de internet. A utilização do celular, do comércio eletrônico e do internet banking não puderam ser totalmente explicadas, sugerindo a necessidade de mais estudos no sentido de identificar variáveis complementares que influenciem a adoção.

Atendendo à natureza exploratória deste trabalho foram identificadas algumas hipóteses que podem ser estudadas em trabalhos futuros, como a necessidade de mais variáveis para caracterizar a adoção de tecnologias com taxas de penetração muito baixas ou muito altas, a posição da renda como uma primeira barreira de entrada para adoção de uma nova tecnologia que a partir de uma maior disseminação é substituída pela educação e a obrigatoriedade da utilização de serviços bancários funcionar como incentivo para quebra da barreira de adoção do internet banking em idades mais avançadas.

Dentre as contribuições deste trabalho podemos destacar a utilização de uma nova abordagem de análise dos dados associados à divisão digital. Podendo ser replicada em trabalhos futuros que busquem avaliar a evolução da divisão digital ao longo do tempo, a metodologia demonstra grande potencial na análise de índices compostos de qualquer natureza e também para utilização em estudos de áreas específicas de aplicação das TIC, como a saúde e a educação.

Além disso, os resultados apresentados por este estudo oferecem contribuições práticas para auxiliar gestores responsáveis por políticas públicas e organismos internacionais a obterem uma melhor compreensão da divisão digital no Brasil.

REFERÊNCIAS

- ACID, S.; DE CAMPOS, L. M.; FERNÁNDEZ-LUNA, J. M. et al. *A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service*. **Artificial Intelligence in Medicine**, v. 30, n. 3, p. 215-232, 2004. Acesso em: 12/11/2011.
- AGARWAL, R.; ANIMESH, A.; PRASAD, K. *Social Interactions and the “Digital Divide”: Explaining Variations in Internet Use*. **Information Systems Research**, v. 20, n. 2, p. 277-294, 2009. Acesso em: 1/5/2012.
- ALAMPAY, E. *Beyond access to ICTs: Measuring capabilities in the information society*. **International Journal of Education and Development using ICT**, v. 2, n. 3, 2006.
- ANDERSON, R. D.; VASTAG, G. *Causal modeling alternatives in operations research: Overview and application*. **European Journal of Operational Research**, v. 156, n. 1, p. 92-109, 2004. Acesso em: 6/11/2011.
- ANDONOVA, V. *Mobile phones, the Internet and the institutional environment*. **Telecommunications Policy**, v. 30, n. 1, p. 29-45, 2006. Acesso em: 1/5/2012.
- ANDRÉS, L.; CUBERES, D.; DIOUF, M.; SEREBRISKY, T. *The diffusion of the Internet: A cross-country analysis*. **Telecommunications Policy**, v. 34, n. 5-6, p. 323-340, 2010.
- ATKINSON, A. B.; HILLS, J. *Exclusion, Employment and Opportunity*. CASE Paper, Centre for Analysis of Social Exclusion, LSE, 1998.
- ATTEWELL, P. A. *The first and second digital divides*. **Sociology of Education**, v. 74, n. 3, p. 252-259, 2001.
- AVGEROU, C. *Discourses on ICT and development*. **Information technologies and international development**, v. 6, n. 3, p. 1-18, 2010. Acesso em: 6/11/2011.
- BAPTISTA, R. *Geographical Clusters and Innovation Diffusion*. **Technological Forecasting and Social Change**, v. 66, n. 1, p. 31-46, 2001. Acesso em: 1/5/2012.
- BARZILAI-NAHON, K. *Gaps and bits: Conceptualizing measurements for digital divides*. **Information Society**, v. 22, n. 5, p. 269-278, 2006.
- BEILLOCK, R.; DIMITROVA, D. V. *An exploratory model of inter-country Internet diffusion*. **Telecommunications Policy**, v. 27, n. 3-4, p. 237-252, 2003.
- BELL, D. *The coming of post-industrial society: a venture in social forecasting*. Basic Books, 1976.
- BILLÓN, M.; EZCURRA, R.; LERA-LÓPEZ, F. *The Spatial Distribution of the Internet in the European Union: Does Geographical Proximity Matter?* **European Planning Studies**, v. 16, n. 1, p. 119-142, 2007.

- BILLON, M.; MARCO, R.; LERA-LOPEZ, F. *Disparities in ICT adoption: A multidimensional approach to study the cross-country digital divide*. **Telecommunications Policy**, v. 33, n. 10–11, p. 596-610, 2009. Acesso em: 21/1/2012.
- BLUMLER, J. G.; KATZ, E. *The uses of mass communications: Current perspectives on gratifications research*. Sage, 1974.
- BONFADELLI, H. *The Internet and Knowledge Gaps A Theoretical and Empirical Investigation*. **European Journal of Communication**, v. 17, n. 1, p. 65-84, 2002. Acesso em: 5/5/2012.
- BRUNO, G.; ESPOSITO, E.; GENOVESE, A.; GWEBU, K. L. *A Critical Analysis of Current Indexes for Digital Divide Measurement*. **The Information Society**, v. 27, n. 1, p. 16-28, 2010.
- BURRELL, J.; TOYAMA, K. *What constitutes good ICTD research*. **Information Technologies and International Development**, v. 5, n. 3, p. 82–94, 2009.
- CASELLI, F.; COLEMAN II, W. J. *Cross-country technology diffusion: The case of computers*. **American Economic Review**, v. 91, n. 2, p. 328-335, 2001.
- CASTELLS, M. *The Network Society: The Information Age: Economy, Society and Culture. Vol. I: The Rise of the Network Society*. Cambridge, MA, 1996.
- CAVA-FERRERUELA, I.; ALABAU-MUÑOZ, A. *Broadband policy assessment: A cross-national empirical analysis*. **Telecommunications Policy**, v. 30, n. 8–9, p. 445-463, 2006.
- CGI.BR. **TIC Domicílios e Empresas 2010 - Pesquisa sobre uso das tecnologias da informação e comunicação no Brasil**. ,2010. Disponível em: <<http://www.cetic.br/tic/2010/index.htm>>. Acesso em: 3/11/2011.
- CHARNIAK, E. *Bayesian Networks without Tears*. **AI magazine**, v. 12, n. 4, p. 50, 1991.
- CHENG, J.; GREINER, R.; KELLY, J.; BELL, D.; LIU, W. *Learning Bayesian networks from data: An information-theory based approach*. **Artificial Intelligence**, v. 137, n. 1-2, p. 43-90, 2002.
- CHINN, M. D.; FAIRLIE, Robert W. *ICT Use in the Developing World: An Analysis of Differences in Computer and Internet Penetration*. **Review of International Economics**, v. 18, n. 1, p. 153-167, 2010. Acesso em: 21/4/2012.
- CHINN, M. D.; FAIRLIE, Robert W. *The Determinants of the Global Digital Divide: A Cross-Country Analysis of Computer and Internet Penetration*. **Oxford Economic Papers**, v. 59, n. 1, p. 16-44, 2007. Acesso em: 1/5/2012.
- ÇILAN, Ç. A.; BOLAT, B. A.; COŞKUN, E. *Analyzing digital divide within and between member and candidate countries of European Union*. **Government Information Quarterly**, v. 26, n. 1, p. 98-105, 2009. Acesso em: 30/4/2012.

CIMOLI, M.; HOFMAN, A. A.; MULDER, N. (ORGS.). *Innovation and Economic Development: The Impact of Information and Communication Technologies in Latin America*. Edward Elgar Pub, 2010.

CIO, K. J.; KURGAN, L.A. *Trends in data mining and knowledge discovery. Advanced Techniques in Knowledge Discovery and Data Mining*, p. 1–26, 2005.

COHENDET, P. *The digital divide in the European enlarged economic scenario: an assessment of the socio-economic effects*. **Strasbourg: University Louis Pasteur**, 2003.

CONRADY, S.; JOUFFE, L. *Breast Cancer Diagnostics with Bayesian Networks*. ,2011a. Disponível em: <http://www.conradyscience.com/white_papers/WBCD_V7.pdf>. Acesso em: 1/7/2012.

CONRADY, S.; JOUFFE, L. *Knowledge Discovery in the Stock Market*. ,2011b. Disponível em: <<http://www.conradyscience.com/index.php/knowledgediscovery>>. Acesso em: 27/6/2012.

CONRADY, S.; JOUFFE, L. *Probabilistic Latent Factor Induction and Statistical Factor Analysis*. ,2011c. Disponível em: <http://www.conradyscience.com/white_papers/Factor_Analysis_V13.pdf>. Acesso em: 13/7/2012.

CORROCHER, N.; ORDANINI, A. *Measuring the digital divide: A framework for the analysis of cross-country differences*. **Journal of Information Technology**, v. 17, n. 1, p. 9-19, 2002.

CRENSHAW, E. M.; ROBISON, K. K. *Globalization and the Digital Divide: The Roles of Structural Conduciveness and Global Connection in Internet Diffusion*. **Social Science Quarterly**, v. 87, n. 1, p. 190-207, 2006. Acesso em: 1/5/2012.

DALY, R.; SHEN, Q.; AITKEN, S. *Learning Bayesian networks: approaches and issues*. **The Knowledge Engineering Review**, v. 26, n. 2, p. 99, 2011.

DEWAN, S.; GANLEY, D.; KRAEMER, K. L. *Across the digital divide: a cross-country multi-technology analysis of the determinants of IT penetration*. **Journal of the Association for Information Systems**, v. 6, n. 12, p. 409–432, 2005.

DEWAN, S.; RIGGINS, F. J. *The digital divide: Current and future research directions*. **Journal of the Association for information systems**, v. 6, n. 12, p. 298–337, 2005.

DIMAGGIO, P.; HARGITTAI, E.; CELESTE, C.; SHAFER, S. *Digital inequality: From unequal access to differentiated use*. **Social inequality**, p. 355–400, 2004.

DIMAGGIO, P.; HARGITTAI, E.; OTHERS. *From the “digital divide” to “digital inequality”: Studying Internet use as penetration increases*. **Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University**, 2001.

DISTASO, W.; LUPI, P.; MANENTI, F. M. *Platform competition and broadband uptake: Theory and empirical evidence from the European union*. **Information Economics and Policy**, v. 18, n. 1, p. 87-106, 2006. Acesso em: 1/5/2012.

DONNER, J.; TOYAMA, K. *Persistent themes in ICT4D Research: priorities for inter-methodological exchange*. **Proceedings of the 57th session of the International Statistics Institute**, p. 1-10, 2009.

DOTFORCE. *Digital opportunities for all: meeting the challenge*. **Report of the Digital Opportunity Task Force**, 2001.

FAIRLIE, Robert W. *Race and the digital divide*. **Contributions to Economic Analysis & Policy**, v. 3, n. 1, p. 1-38, 2004.

FAIRLIE, Robert W.; BELTRAN, D. O.; DAS, K. K. *Home computers and educational outcomes: Evidence from the NLSY97 and cps*. **Economic Inquiry**, v. 48, n. 3, p. 771-792, 2010.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P.; OTHERS. *Knowledge discovery and data mining: Towards a unifying framework*. ,1996.

FORMAN, C.; GOLDFARB, A.; GREENSTEIN, S. *How did location affect adoption of the commercial Internet? Global village vs. urban leadership*. **Journal of Urban Economics**, v. 58, n. 3, p. 389-420, 2005. Acesso em: 1/5/2012.

FREEMAN, C.; LOUÇÃ, F. *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press, USA, 2001.

FRIEDMAN, N.; GOLDSZMIDT, M. *Discretizing continuous attributes while learning Bayesian networks*. Machine Learning-International Workshop then Conference-. **Anais...** p.157-165, 1996.

GOMEZ, R.; BARON, L. F.; FIORE-SILFVAST, B. *The Changing Field of ICTD: Content analysis of research published in selected journals and conferences, 2000-2010*. ,2012.

GUILLÉN, M. F.; SUÁREZ, S. L. *Explaining the global digital divide: Economic, political and sociological drivers of cross-national internet use*. **Social Forces**, v. 84, n. 2, p. 681-708, 2005.

GUPTA, S.; KIM, H. W. *Linking structural equation modeling to Bayesian networks: Decision support for customer retention in virtual communities*. **European Journal of Operational Research**, v. 190, n. 3, p. 818-833, 2008. Acesso em: 5/11/2011.

HANAFIZADEH, M. R.; SAGHAEI, A.; HANAFIZADEH, P. *An index for cross-country analysis of ICT infrastructure and access*. **Telecommunications Policy**, v. 33, n. 7, p. 385-405, 2009. Acesso em: 21/1/2012.

HAQ, M. U. *Reflections on human development*. Oxford University Press, USA, 1995.

HARGITTAI, E. *Weaving the Western Web: explaining differences in Internet connectivity among OECD countries*. **Telecommunications Policy**, v. 23, n. 10–11, p. 701-718, 1999. Acesso em: 30/4/2012.

HAWKINS, E. T.; HAWKINS, K. A. *Bridging Latin America's Digital Divide: Government Policies and Internet Access*. **Journalism & Mass Communication Quarterly**, v. 80, n. 3, p. 646-665, 2003. Acesso em: 1/5/2012.

HECKERMAN, D. *A tutorial on learning with Bayesian networks*. **Innovations in Bayesian Networks**, p. 33–82, 2008.

HEEKS, Richard. *ICT4D 2.0: The Next Phase of Applying ICT for International Development*. **Computer**, v. 41, n. 6, p. 26-33, 2008.

HEEKS, Richard. *The ICT4D 2.0 Manifesto: Where Next for ICTs and International Development?* University of Manchester, Institute for Development Policy and Management, 2009.

HEEKS, Richard. *Do information and communication technologies (ICTs) contribute to development?* **Journal of International Development**, v. 22, n. 5, p. 625-640, 2010. Acesso em: 28/2/2012.

HILBERT, Martin. *When is cheap, cheap enough to bridge the digital divide? Modeling income related structural challenges of technology diffusion in Latin America*. **World Development**, v. 38, n. 5, p. 756–770, 2010a.

HILBERT, Martin. *Towards a conceptual framework for ICT for Development: lessons learned from Latin America*. ,2010b.

HILBERT, Martin; LÓPEZ, P.; VÁSQUEZ, C. *Information Societies or “ICT Equipment Societies?”* Measuring the Digital Information-Processing Capacity of a Society in Bits and Bytes. **The Information Society**, v. 26, n. 3, p. 157-178, 2010.

HILLS, P.; ARGYLE, M. *Uses of the Internet and their relationships with individual differences in personality*. **Computers in Human Behavior**, v. 19, n. 1, p. 59-70, 2003. Acesso em: 30/4/2012.

HÜSING, T.; GAREIS, K.; KORTE, W. B. *The Impact of ICT on Social Cohesion: Looking Beyond the Digital Divide*. In: S. Dutta; A. Meyer; A. Jain; G. Richter (Orgs.); **The Information Society in an Enlarged Europe**. p.75-123, 2006. Springer Berlin Heidelberg. Acesso em: 5/5/2012.

ITU. *Measuring the information society: The ICT development index*. International Telecommunication Union, 2009.

ITU. *Measuring the Information Society 2011*. ,2011. Disponível em: <http://www.itu.int/ITU-D/ict/publications/idi/material/2011/MIS_2011_without_annex_5.pdf>. Acesso em: 24/4/2012.

KARCHER, C. **Redes Bayesianas aplicadas à análise do risco de crédito.**, 26. fev. 2009. Biblioteca Digitais de Teses e Dissertações da USP. Disponível em:

<<http://www.teses.usp.br/teses/disponiveis/3/3142/tde-25052009-162507/>>. Acesso em: 17/3/2011.

KAUFMANN, A.; LEHNER, P.; TÖDTLING, F. *Effects of the Internet on the spatial structure of innovation networks*. **Information Economics and Policy**, v. 15, n. 3, p. 402-424, 2003. Acesso em: 1/5/2012.

KIISKI, S.; POHJOLA, Matti. *Cross-country diffusion of the Internet*. **Information Economics and Policy**, v. 14, n. 2, p. 297-310, 2002. Acesso em: 1/5/2012.

KLÖSGEN, W.; ZYTKOW, J. M. *Advances in knowledge discovery and data mining*. In: U. M. Fayyad; Gregory Piatetsky-Shapiro; Padhraic Smyth; R. Uthurusamy (Orgs.); p.573-592, 1996. Menlo Park, CA, USA: American Association for Artificial Intelligence. Disponível em: <<http://dl.acm.org/citation.cfm?id=257938.258017>>. Acesso em: 6/11/2011.

KURGAN, Lukasz A.; MUSILEK, P. *A Survey of Knowledge Discovery and Data Mining Process Models*. **The Knowledge Engineering Review**, v. 21, n. 01, p. 1-24, 2006.

LAURÍA, E. J. M.; DUCHESSI, P. J. *A methodology for developing Bayesian networks: An application to information technology (IT) implementation*. **European Journal of Operational Research**, v. 179, n. 1, p. 234-252, 2007. Acesso em: 5/11/2011.

LI, N.; KIRKUP, G. *Gender and cultural differences in Internet use: A study of China and the UK*. **Computers & Education**, v. 48, n. 2, p. 301-317, 2007. Acesso em: 30/4/2012.

LIU, M.; SAN, G. *Social Learning and Digital Divides: A Case Study of Internet Technology Diffusion*. **Kyklos**, v. 59, n. 2, p. 307-321, 2006. Acesso em: 1/5/2012.

LYNCH, S. M. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. 1º ed. Springer, 2007.

MENOU, M. J. *Digital and Social Equity? Opportunities and threats on the road to empowerment*. LIDA-Libraries in the Digital Age: ANNUAL COURSE AND CONFERENCE. **Anais...**, 2001.

MOWLANA, H.; WILSON, L. J. *Communication technology and development*. **Reports and Papers on Mass Communication**, v. no. 101, 1988. Disponível em: <<http://agris.fao.org/agris-search/search/display.do?f=1989/XF/XF89026.xml;XF8878971>>. Acesso em: 28/3/2012.

MUFFELS, R.; TSAKLOGLOU, P.; MAYES, D. G. (ORGS.). *Social Exclusion in European Welfare States*. Edward Elgar Publishing Ltd, 2002.

MUNTEANU, P.; BENDOU, M. *The EQ Framework for Learning Equivalence Classes of Bayesian Networks*. Proceedings of the 2001 IEEE International Conference on Data Mining. **Anais...**, ICDM '01.. p.417-424, 2001. Washington, DC, USA: IEEE

Computer Society. Disponível em: <<http://dl.acm.org/citation.cfm?id=645496.657736>>. Acesso em: 13/7/2012.

NEAPOLITAN, R. E. *Learning bayesian networks*. Pearson Prentice Hall Upper Saddle River, NJ, 2004.

NEDEVSCI, S.; SANDHU, J. S.; PAL, J.; FONSECA, R.; TOYAMA, K. *Bayesian networks: an exploratory tool for understanding ict adoption*. Proceedings of International Conference: ICTD2006, at Berkeley. **Anais...**, 2006.

NEGROPONTE, N. *Being digital*. Vintage Books, 1995.

NORRIS, P. *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Taylor & Francis, 2003.

OECD. *Understanding the Digital Divide*. ,2001. OECD Publications. Disponível em: <<http://www.oecd.org/internet/interneteconomy/1888451.pdf>>. Acesso em: 11/9/2012.

PARAYIL, G. *The Digital Divide and Increasing Returns: Contradictions of Informational Capitalism*. **The Information Society**, v. 21, n. 1, p. 41-51, 2005.

PATRA, R.; PAL, J.; NEDEVSCI, S. *ICTD state of the union: Where have we reached and where are we headed*. 2009 International Conference on Information and Communication Technologies and Development (ICTD). **Anais...** p.357-366, 2009. IEEE.

PEARL, Judea; RUSSELL, S. Bayesian networks. *The Handbook of Brain Theory and Neural Networks: Second Edition*, ed. M. Arbib. MIT Press, 2002.

PEARL, Judea. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1º ed. Morgan Kaufmann, 1988.

PEARL, Judea. *Graphs, Causality, and Structural Equation Models*. **Sociological Methods & Research**, v. 27, n. 2, p. 226 -284, 1998. Acesso em: 7/11/2011.

PERES, W.; HILBERT, Martin. *Information Societies in Latin America and the Caribbean*. ,2010.

PEREZ, C. *Structural change and assimilation of new technologies in the economic and social systems*. **Futures**, v. 15, n. 5, p. 357-375, 1983. Acesso em: 14/3/2012.

PEREZ, C. *Technological revolutions, paradigm shifts and socio-institutional change. Globalization, Economic Development and Inequality: An Alternative Perspective*, Cheltenham, UK: Edward Elgar, p. 217–242, 2004.

PICK, J. B.; AZARI, R. *Global digital divide: Influence of socioeconomic, governmental, and accessibility factors on information technology*. **Information Technology for Development**, v. 14, n. 2, p. 91-115, 2008. Acesso em: 6/11/2011.

POHJOLA, M. *The adoption and diffusion of ICT across countries: Patterns and determinants*. **The new economy handbook**, p. 77–100, 2003.

PUNAMÄKI, R.-L.; WALLENIUS, M.; NYGÅRD, C.-H.; SAARNI, L.; RIMPELÄ, A. *Use of information and communication technology (ICT) and perceived health in adolescence: The role of sleeping habits and waking-time tiredness.* **Journal of Adolescence**, v. 30, n. 4, p. 569-585, 2007. Acesso em: 30/4/2012.

QUIBRIA, M. .; AHMED, S. N.; TSCHANG, T.; REYES-MACASAQUIT, M.-L. *Digital divide: determinants and policies with special reference to Asia.* **Journal of Asian Economics**, v. 13, n. 6, p. 811-825, 2003. Acesso em: 1/5/2012.

QVORTRUP, L. *The Nordic telecottages: Community teleservice centres for rural regions.* **Telecommunications Policy**, v. 13, n. 1, p. 59-68, 1989. Acesso em: 14/3/2012.

RAITI, G. C. *The Lost Sheep of ICT4D Literature.* **Information Technologies & International Development**, v. 3, n. 4, p. 1-8, 2007. Acesso em: 24/12/2011.

REDDI, U. R. V. *Primer 1: An Introduction to ICT for Development.* ,2011. UN-APCICT/ESCAP. Disponível em: <http://www.unapcict.org/ecohub/primer-1-an-introduction-to-ict-for-development/at_download/attachment1>. Acesso em: 28/3/2012.

RODGERS, G.; GORE, C. G.; FIGUEIREDO, J. B. (ORGS.). *Social Exclusion: Rhetoric, Reality, Responses.* International Labour Org, 1995.

ROGERS, E. M. *Diffusion of Innovations.* Simon and Schuster, 1995.

SAWHNEY, N. *Bayesian Model Selection for Human Development Indicators. A Hypertext History of Multiuser Dimensions,* 2001.

SAXENIAN, A. L. *Silicon Valley's new immigrant entrepreneurs.* Public Policy Inst of California, 1999.

SCHUMPETER, J. A. *Business cycles: a theoretical, historical, and statistical analysis of the capitalist process.* New York, McGraw-Hill, 1939.

SCIADAS, G. *Unveiling the Digital Divide.* Science, Innovation and Electronic Information Division, Statistics Canada, 2002.

SEN, A. K. *Development as freedom.* Oxford University Press, 1999.

SEY, A. *Public access to ICTs: A review of the literature.* **Center for Information & Society, University of Washington**, 2008.

SILVER, H. *Social exclusion and social solidarity: three paradigms.* **Int'l Lab. Rev.**, v. 133, p. 531, 1994.

TICHENOR, P. J.; DONOHUE, G. A.; OLIEN, C. N. *Mass Media Flow and Differential Growth in Knowledge.* **The Public Opinion Quarterly**, v. 34, n. 2, p. 159-170, 1970. Acesso em: 5/5/2012.

TOYAMA, Kentaro. *Technology as amplifier in international development.* Proceedings of the 2011 iConference. **Anais...** , iConference '11.. p.75-82, 2011. New

York, NY, USA: ACM. Disponível em: <<http://doi.acm.org/10.1145/1940761.1940772>>. Acesso em: 23/12/2011.

TOYAMA, Kentaro; DIAS, M. B. *Information and Communication Technologies for Development*. **Computer**, v. 41, n. 6, p. 22-25, 2008. Acesso em: 9/3/2012.

UNDP. *Information Communications Technology for Development*. **United Nations Development Programme Evaluation Office Essentials**, v. 5, n. 5, 2001.

VEHOVAR, V.; SICHERL, P.; HÜSING, T.; DOLNICAR, V. *Methodological Challenges of Digital Divide Measurements*. **The Information Society**, v. 22, n. 5, p. 279-290, 2006.

VERTOVEC, S. *Transnational networks and skilled labour migration*. **Ladenburg Diskurs “Migration”**. Gottlieb Daimler-und Karl Benz-Stiftung, Ladenburg, p. 14–15, 2002.

VICENTE, M. R.; LÓPEZ, A. J. *Patterns of ICT diffusion across the European Union*. **Economics Letters**, v. 93, n. 1, p. 45-51, 2006. Acesso em: 1/5/2012.

VICENTE, M. R.; LÓPEZ, A. J. *Assessing the regional digital divide across the European Union-27*. **Telecommunications Policy**, v. 35, n. 3, p. 220-237, 2011. Acesso em: 21/1/2012.

WARSCHAUER, M. *Digital Divide*. **Scientific American**, 2003.

WEBSTER, F. *Theories of the information society*. Psychology Press, 2002.

WHITE, D. S.; GUNASEKARAN, A.; SHEA, T. P.; ARIGUZO, G. C. *Mapping the global digital divide*. **International Journal of Business Information Systems**, v. 7, n. 2, p. 207-219, 2011. Acesso em: 18/4/2012.

WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd ed. Morgan Kaufmann, 2011.

WONG, P.-K. *ICT production and diffusion in Asia Digital dividends or digital divide?* **Information Economics and Policy**, v. 14, n. 2, p. 167-187, 2002. Acesso em: 1/5/2012.

WSIS. *Building the Information Society: a global challenge in the new Millennium*. ,2003. Disponível em: <http://www.itu.int/dms_pub/itu-s/md/03/wsis/doc/S03-WSIS-DOC-0004!!PDF-E.pdf>. Acesso em: 11/9/2012.

WSIS. *Report of the Tunis phase of the World Summit on the Information Society*. ,2005. Disponível em: <<http://www.itu.int/wsis/docs2/tunis/off/9rev1.pdf>>. Acesso em: 11/9/2012.

ZIKMUND, W. G.; BABIN, B. J. *Exploring Marketing Research*. Cengage Learning, 2006.

Apêndices

Apêndice 1 – *Ranking* de países em função do IDI

Rank	Economia	IDI	FIX	CEL	BAND	COMP	INT	USR	INT FIX	INT MOV	SEC	TER	ALF
1	Coreia do Sul	8,40	0,987	0,620	0,748	0,818	0,968	0,837	0,610	0,910	0,979	1,020	0,990
2	Suécia	8,23	0,892	0,668	0,986	0,895	0,883	0,900	0,527	0,840	1,000	0,680	0,990
3	Islândia	8,06	1,062	0,639	1,003	0,930	0,920	0,950	0,577	0,449	1,000	0,802	0,990
4	Dinamarca	7,97	0,788	0,732	0,945	0,880	0,861	0,887	0,623	0,547	1,000	0,853	0,990
5	Finlândia	7,87	0,388	0,920	0,923	0,820	0,805	0,869	0,485	0,781	1,000	0,980	0,990
6	Hong Kong	7,79	1,027	1,119	1,071	0,773	0,757	0,694	0,503	0,745	0,821	0,566	0,990
7	Luxemburgo	7,78	0,895	0,843	0,906	0,902	0,903	0,906	0,547	0,721	0,963	0,086	0,990
8	Suíça	7,67	0,977	0,727	0,953	0,869	0,864	0,839	0,637	0,438	0,968	0,511	0,990
9	Holanda	7,61	0,720	0,684	0,952	0,920	0,909	0,907	0,633	0,377	1,000	0,626	0,990
10	Reino Unido	7,60	0,895	0,766	0,940	0,826	0,796	0,850	0,523	0,560	0,976	0,557	0,990
11	Noruega	7,60	0,580	0,665	0,925	0,909	0,898	0,934	0,577	0,471	1,108	0,716	0,990
12	Nova Zelândia	7,43	0,713	0,676	0,786	0,839	0,790	0,830	0,415	0,662	1,192	0,832	0,990
13	Japão	7,42	0,532	0,561	0,769	0,885	0,854	0,800	0,448	0,878	1,000	0,609	0,990
14	Austrália	7,36	0,648	0,594	0,847	0,811	0,741	0,760	0,387	0,827	1,493	0,775	0,990
15	Alemanha	7,27	0,923	0,747	0,894	0,857	0,825	0,819	0,527	0,364	1,000	0,462	0,990
16	Áustria	7,17	0,645	0,858	0,893	0,762	0,729	0,727	0,398	0,674	1,004	0,564	0,990
17	Estados Unidos	7,09	0,812	0,529	0,838	0,755	0,716	0,790	0,438	0,540	0,942	0,833	0,990
18	França	7,09	0,935	0,586	0,889	0,764	0,736	0,801	0,565	0,358	1,000	0,543	0,990
19	Cingapura	7,08	0,650	0,845	0,962	0,840	0,820	0,700	0,412	0,697	0,741	0,438	0,947
20	Israel	6,87	0,737	0,783	0,716	0,770	0,690	0,672	0,418	0,622	0,893	0,615	0,990
21	Macau, China	6,84	0,513	1,214	0,824	0,801	0,800	0,568	0,402	0,561	0,919	0,629	0,935
22	Bélgica	6,83	0,722	0,668	0,922	0,767	0,727	0,793	0,525	0,231	1,000	0,636	0,990
23	Irlanda	6,78	0,775	0,619	0,882	0,765	0,717	0,699	0,380	0,473	1,000	0,612	0,990
24	Eslovénia	6,75	0,750	0,615	0,889	0,705	0,681	0,700	0,407	0,328	0,970	0,964	0,997
25	Espanha	6,73	0,720	0,658	0,870	0,687	0,591	0,665	0,383	0,557	1,213	0,730	0,977
26	Canadá	6,69	0,833	0,416	0,868	0,839	0,782	0,816	0,497	0,148	1,001	0,666	0,990
27	Portugal	6,64	0,700	0,837	0,948	0,595	0,537	0,511	0,323	0,725	1,061	0,610	0,949
28	Itália	6,57	0,595	0,796	0,879	0,648	0,590	0,537	0,368	0,594	1,000	0,720	0,989
29	Malta	6,43	0,990	0,643	0,786	0,731	0,704	0,630	0,458	0,310	1,000	0,339	0,924
30	Grécia	6,28	0,763	0,636	0,824	0,534	0,464	0,444	0,330	0,583	1,032	0,908	0,972
31	Croácia	6,21	0,707	0,850	0,850	0,600	0,565	0,603	0,305	0,393	0,955	0,540	0,988
32	Emirados Árabes Unidos	6,19	0,328	0,856	0,832	0,760	0,650	0,780	0,175	0,584	0,952	0,304	0,900
33	Estônia	6,16	0,600	0,725	0,801	0,692	0,678	0,741	0,405	0,081	0,997	0,641	0,998
34	Hungria	6,04	0,497	0,708	0,734	0,664	0,605	0,653	0,327	0,299	0,977	0,741	0,994
35	Lituânia	6,04	0,368	0,866	0,855	0,592	0,606	0,621	0,343	0,229	0,986	0,815	0,997
36	Chipre	5,98	0,627	0,551	0,777	0,605	0,537	0,530	0,293	0,613	0,985	0,453	0,979
37	República Checa	5,97	0,348	0,804	0,888	0,641	0,605	0,688	0,245	0,278	0,949	0,705	0,990
38	Polônia	5,95	0,412	0,707	0,840	0,690	0,634	0,623	0,220	0,310	1,003	0,732	0,995
39	República Eslovaca	5,94	0,335	0,638	0,745	0,722	0,675	0,794	0,268	0,270	0,919	0,652	0,990
40	Letônia	5,90	0,393	0,602	0,824	0,628	0,598	0,684	0,322	0,274	0,984	0,692	0,998
41	Barbados	5,83	0,838	0,754	0,792	0,614	0,510	0,702	0,343	0,000	0,948	0,570	0,990
42	Antígua e Barbuda	5,63	0,785	1,086	0,776	0,509	0,380	0,800	0,287	0,000	1,105	0,147	0,990
43	Brunei Darussalam	5,61	0,333	0,642	0,807	0,796	0,713	0,500	0,090	0,614	0,982	0,171	0,953
44	Catar	5,60	0,282	0,779	0,790	0,896	0,840	0,690	0,153	0,284	0,852	0,102	0,947
45	Bahrain	5,57	0,302	0,731	0,763	0,870	0,740	0,550	0,203	0,213	0,964	0,512	0,914
46	Arábia Saudita	5,42	0,253	1,105	0,817	0,573	0,544	0,410	0,092	0,578	0,968	0,328	0,861

Rank	Economia	IDI	FIX	CEL	BAND	COMP	INT	USR	INT FIX	INT MOV	SEC	TER	ALF
47	Rússia	5,38	0,523	0,978	0,824	0,500	0,421	0,430	0,183	0,174	0,843	0,825	0,996
48	Romênia	5,20	0,348	0,675	0,864	0,479	0,422	0,399	0,233	0,328	0,932	0,669	0,977
49	Bulgária	5,19	0,490	0,831	0,929	0,351	0,331	0,462	0,245	0,245	0,885	0,562	0,983
50	Sérvia	5,11	0,675	0,760	0,857	0,509	0,402	0,409	0,142	0,190	0,915	0,498	0,978
51	Montenegro	5,03	0,447	1,090	0,794	0,320	0,212	0,520	0,138	0,380	0,881	0,298	0,970
52	Bielorrússia	5,01	0,718	0,634	0,797	0,408	0,312	0,317	0,290	0,116	0,901	0,770	0,997
53	Macedônia	4,98	0,333	0,615	0,775	0,603	0,492	0,519	0,208	0,208	0,836	0,459	0,971
54	Uruguai	4,93	0,477	0,775	0,808	0,528	0,333	0,434	0,190	0,057	0,837	0,768	0,983
55	Chile	4,65	0,337	0,682	0,786	0,468	0,350	0,450	0,175	0,071	0,907	0,601	0,986
56	Argentina	4,64	0,412	0,834	0,815	0,400	0,213	0,360	0,160	0,128	0,845	0,694	0,977
57	Moldávia	4,47	0,542	0,521	0,834	0,369	0,347	0,400	0,125	0,153	0,881	0,383	0,985
58	Malásia	4,45	0,268	0,714	0,746	0,410	0,251	0,553	0,122	0,272	0,678	0,370	0,925
59	Turquia	4,42	0,372	0,499	0,785	0,442	0,416	0,398	0,163	0,178	0,818	0,440	0,908
60	Oman	4,38	0,170	0,974	0,677	0,456	0,277	0,626	0,032	0,107	0,913	0,264	0,866
61	Trinidad & Tobago	4,36	0,365	0,831	0,780	0,531	0,198	0,485	0,180	0,000	0,912	0,126	0,987
62	Ucrânia	4,34	0,475	0,698	0,745	0,307	0,207	0,230	0,135	0,043	0,945	0,891	0,997
63	Bósnia e Herzegovina	4,31	0,443	0,471	0,770	0,337	0,181	0,520	0,173	0,107	0,912	0,370	0,978
64	Brasil	4,22	0,360	0,612	0,752	0,349	0,271	0,407	0,120	0,106	0,996	0,400	0,900
65	Venezuela	4,11	0,407	0,566	0,703	0,173	0,097	0,356	0,090	0,208	0,821	0,786	0,952
66	Panamá	4,09	0,262	1,086	0,794	0,198	0,118	0,428	0,130	0,034	0,727	0,444	0,936
67	Maldivas	4,05	0,253	0,921	0,827	0,544	0,229	0,283	0,082	0,087	0,918	0,000	0,984
68	Cazaquistão	4,02	0,417	0,725	0,722	0,251	0,232	0,340	0,088	0,003	1,030	0,401	0,997
69	Maurício	4,00	0,497	0,539	0,738	0,334	0,246	0,249	0,105	0,221	0,872	0,305	0,879
70	Costa Rica	3,99	0,530	0,383	0,753	0,413	0,241	0,365	0,103	0,055	0,961	0,269	0,961
71	Seychelles	3,94	0,425	0,799	0,682	0,321	0,171	0,410	0,122	0,047	1,050	0,018	0,918
72	Armênia	3,87	0,318	0,735	0,727	0,178	0,116	0,370	0,045	0,052	0,931	0,501	0,995
73	Jordânia	3,83	0,130	0,629	0,698	0,514	0,221	0,380	0,053	0,024	0,891	0,432	0,922
74	Azerbaijão	3,78	0,272	0,582	0,751	0,215	0,353	0,360	0,090	0,009	0,994	0,191	0,995
75	México	3,75	0,292	0,474	0,709	0,298	0,222	0,310	0,167	0,083	0,918	0,292	0,934
76	Colômbia	3,75	0,245	0,552	0,736	0,261	0,193	0,365	0,095	0,056	0,946	0,370	0,932
77	Geórgia	3,65	0,228	0,432	0,794	0,182	0,166	0,270	0,085	0,188	1,085	0,255	0,997
78	Albânia	3,61	0,172	0,835	0,732	0,156	0,117	0,450	0,057	0,000	0,724	0,367	0,959
79	Líbano	3,57	0,350	0,400	0,602	0,317	0,297	0,310	0,078	0,000	0,821	0,525	0,896
80	China	3,55	0,365	0,376	0,620	0,354	0,237	0,343	0,157	0,020	0,782	0,245	0,940
81	Vietnã	3,53	0,312	1,031	0,687	0,142	0,081	0,276	0,068	0,128	0,669	0,122	0,928
82	Suriname	3,52	0,270	0,998	0,641	0,299	0,119	0,316	0,050	0,000	0,757	0,123	0,946
83	Peru	3,52	0,182	0,589	0,721	0,227	0,123	0,343	0,052	0,072	0,903	0,381	0,896
84	Tunísia	3,43	0,205	0,624	0,757	0,170	0,077	0,368	0,077	0,011	0,943	0,367	0,776
85	Jamaica	3,41	0,160	0,666	0,793	0,200	0,118	0,261	0,072	0,079	0,917	0,242	0,864
86	Mongólia	3,41	0,117	0,536	0,878	0,223	0,077	0,102	0,038	0,118	0,942	0,527	0,975
87	Irã	3,39	0,605	0,536	0,641	0,337	0,208	0,130	0,012	0,000	0,831	0,365	0,850
88	Equador	3,37	0,240	0,601	0,704	0,270	0,118	0,240	0,023	0,087	0,811	0,424	0,842
89	Tailândia	3,30	0,168	0,593	0,740	0,228	0,114	0,212	0,065	0,038	0,770	0,450	0,935
90	Marrocos	3,29	0,195	0,589	0,675	0,342	0,255	0,490	0,027	0,100	0,583	0,129	0,561
91	Egito	3,28	0,198	0,512	0,701	0,340	0,312	0,267	0,030	0,064	0,805	0,296	0,664
92	Filipinas	3,22	0,122	0,504	0,740	0,131	0,101	0,250	0,030	0,166	0,821	0,277	0,954
93	Rep. Dominicana	3,21	0,170	0,527	0,652	0,164	0,076	0,395	0,060	0,022	0,768	0,342	0,882
94	Fiji	3,16	0,265	0,684	0,677	0,268	0,188	0,148	0,032	0,014	0,796	0,153	0,951
95	Guiana	3,08	0,332	0,433	0,698	0,065	0,051	0,299	0,027	0,000	1,034	0,112	0,990
96	Síria	3,05	0,332	0,337	0,575	0,404	0,352	0,207	0,005	0,013	0,747	0,157	0,842
97	África do Sul	3,00	0,140	0,591	0,593	0,183	0,101	0,123	0,025	0,166	0,939	0,158	0,887

Rank	Economia	IDI	FIX	CEL	BAND	COMP	INT	USR	INT FIX	INT MOV	SEC	TER	ALF
98	El Salvador	2,89	0,270	0,731	0,589	0,142	0,075	0,150	0,047	0,050	0,628	0,256	0,841
99	Paraguai	2,87	0,105	0,539	0,645	0,163	0,090	0,236	0,010	0,041	0,665	0,307	0,946
100	Quirguistão	2,84	0,157	0,541	0,449	0,040	0,032	0,200	0,005	0,041	0,841	0,508	0,992
101	Indonésia	2,83	0,263	0,539	0,643	0,108	0,039	0,091	0,013	0,103	0,795	0,235	0,922
102	Bolívia	2,83	0,142	0,425	0,666	0,170	0,039	0,200	0,017	0,014	0,797	0,372	0,907
103	Argélia	2,82	0,137	0,544	0,717	0,200	0,100	0,125	0,042	0,000	0,866	0,306	0,727
104	Cabo Verde	2,81	0,242	0,441	0,609	0,113	0,035	0,300	0,050	0,000	0,815	0,149	0,848
105	Sri Lanka	2,79	0,287	0,489	0,646	0,123	0,059	0,120	0,017	0,096	0,878	0,053	0,906
106	Honduras	2,72	0,147	0,736	0,692	0,129	0,060	0,111	0,017	0,042	0,652	0,187	0,836
107	Cuba	2,69	0,172	0,052	0,434	0,034	0,006	0,151	0,000	0,000	0,896	1,178	0,998
108	Guatemala	2,65	0,173	0,739	0,660	0,158	0,030	0,105	0,030	0,037	0,588	0,241	0,745
109	Botswana	2,59	0,113	0,693	0,699	0,065	0,028	0,060	0,010	0,063	0,824	0,086	0,841
110	Uzbequistão	2,55	0,113	0,449	0,447	0,045	0,013	0,200	0,005	0,039	1,035	0,098	0,993
111	Turcomenistão	2,50	0,172	0,373	0,651	0,120	0,070	0,022	0,000	0,000	0,850	0,217	0,996
112	Gabão	2,42	0,033	0,629	0,865	0,076	0,060	0,072	0,003	0,000	0,584	0,066	0,877
113	Namíbia	2,36	0,112	0,395	0,669	0,154	0,039	0,065	0,007	0,075	0,664	0,089	0,885
114	Nicarágua	2,31	0,075	0,383	0,722	0,082	0,026	0,100	0,013	0,030	0,690	0,195	0,780
115	Quênia	2,29	0,018	0,362	0,603	0,080	0,040	0,259	0,000	0,056	0,595	0,041	0,870
116	Índia	2,01	0,048	0,361	0,691	0,061	0,042	0,075	0,015	0,009	0,622	0,150	0,628
117	Camboja	1,99	0,042	0,339	0,816	0,043	0,004	0,013	0,005	0,089	0,437	0,092	0,776
118	Suazilândia	1,93	0,062	0,364	0,507	0,107	0,036	0,080	0,002	0,000	0,558	0,043	0,869
119	Butão	1,93	0,060	0,319	0,647	0,064	0,049	0,136	0,020	0,003	0,617	0,080	0,528
120	Gana	1,90	0,018	0,421	0,602	0,091	0,004	0,086	0,003	0,006	0,572	0,086	0,666
121	Lao P.D.R.	1,90	0,028	0,380	0,617	0,069	0,034	0,070	0,003	0,004	0,437	0,204	0,727
122	Nigéria	1,85	0,012	0,324	0,375	0,154	0,090	0,284	0,002	0,029	0,294	0,116	0,608
123	Paquistão	1,83	0,033	0,348	0,626	0,130	0,080	0,168	0,005	0,006	0,331	0,064	0,555
124	Zimbábue	1,81	0,050	0,351	0,455	0,053	0,040	0,115	0,005	0,007	0,410	0,023	0,919
125	Senegal	1,78	0,045	0,395	0,621	0,057	0,045	0,160	0,010	0,026	0,330	0,080	0,497
126	Gâmbia	1,74	0,047	0,503	0,555	0,057	0,032	0,092	0,000	0,005	0,503	0,046	0,465
127	Iémen	1,72	0,072	0,271	0,567	0,040	0,029	0,109	0,005	0,000	0,455	0,112	0,624
128	Comores	1,67	0,048	0,132	0,669	0,057	0,024	0,051	0,000	0,000	0,501	0,052	0,742
129	Djibouti	1,66	0,035	0,109	0,757	0,130	0,022	0,065	0,015	0,000	0,305	0,035	0,730
130	Costa do Marfim	1,61	0,018	0,444	0,714	0,018	0,011	0,026	0,000	0,000	0,299	0,089	0,553
131	Mauritânia	1,58	0,035	0,466	0,630	0,030	0,014	0,030	0,003	0,031	0,249	0,038	0,575
132	Angola	1,58	0,027	0,275	0,514	0,071	0,057	0,100	0,002	0,056	0,145	0,094	0,700
133	Togo	1,57	0,058	0,239	0,666	0,037	0,013	0,054	0,002	0,000	0,415	0,060	0,569
134	Nepal	1,56	0,047	0,181	0,604	0,042	0,012	0,068	0,007	0,000	0,442	0,101	0,591
135	Benin	1,54	0,025	0,470	0,615	0,025	0,001	0,031	0,005	0,000	0,421	0,063	0,417
136	Camarões	1,53	0,042	0,245	0,480	0,054	0,015	0,040	0,000	0,009	0,415	0,090	0,707
137	Bangladesh	1,52	0,010	0,272	0,632	0,031	0,026	0,037	0,000	0,002	0,419	0,079	0,559
138	Tanzânia	1,51	0,007	0,275	0,522	0,026	0,007	0,110	0,000	0,021	0,274	0,017	0,729
139	Zâmbia	1,50	0,012	0,222	0,505	0,024	0,020	0,067	0,002	0,000	0,455	0,024	0,709
140	Uganda	1,49	0,017	0,226	0,535	0,021	0,004	0,125	0,002	0,006	0,274	0,041	0,714
141	Madagáscar	1,45	0,013	0,234	0,686	0,014	0,011	0,017	0,000	0,008	0,315	0,036	0,645
142	Ruanda	1,44	0,007	0,196	0,606	0,005	0,001	0,077	0,000	0,013	0,267	0,048	0,707
143	Papua Nova Guiné	1,38	0,030	0,164	0,642	0,035	0,027	0,013	0,002	0,000	0,227	0,154	0,601
144	Guiné	1,31	0,003	0,236	0,586	0,015	0,010	0,010	0,000	0,000	0,370	0,177	0,395
145	Moçambique	1,30	0,007	0,182	0,572	0,075	0,011	0,042	0,002	0,015	0,255	0,027	0,551
146	Mali	1,26	0,012	0,281	0,600	0,030	0,012	0,027	0,000	0,011	0,383	0,060	0,262
147	Congo	1,17	0,002	0,101	0,434	0,003	0,003	0,007	0,000	0,000	0,367	0,060	0,668
148	Eritreia	1,09	0,017	0,021	0,392	0,003	0,003	0,054	0,000	0,000	0,318	0,020	0,666

Rank	Economia	IDI	FIX	CEL	BAND	COMP	INT	USR	INT FIX	INT MOV	SEC	TER	ALF
149	Burkina Faso	1,08	0,015	0,204	0,650	0,021	0,002	0,014	0,002	0,000	0,214	0,034	0,287
150	Etiópia	1,08	0,018	0,046	0,684	0,014	0,003	0,008	0,000	0,003	0,344	0,044	0,298
151	Níger	0,92	0,008	0,144	0,618	0,012	0,002	0,008	0,000	0,000	0,133	0,014	0,287
152	Chade	0,83	0,008	0,137	0,368	0,006	0,002	0,017	0,000	0,000	0,241	0,020	0,336

FONTE: ITU (2011)

Apêndice 2 – Indicadores componentes do IDI para sub-regiões brasileiras (normalizados)

	FIX	CEL	BAND	COMP	INT	USR	INT FIX	INT MOV	SEC	TER	ALF
	ANATEL (2010)			CGI.br (2010)			ANATEL (2010)				IBGE (2010)
Área Rural/Urbana											
Rural				0,123	0,056	0,160					
Urbana				0,392	0,312	0,454					
Regiões											
Centro-Oeste		0,732		0,404	0,326	0,502	0,030	0,083			0,928
Nordeste		0,515		0,138	0,105	0,282	0,035	0,080			0,809
Norte		0,524		0,228	0,140	0,336	0,173	0,122			0,888
Sudeste		0,674		0,450	0,362	0,466	0,149	0,121			0,945
Sul		0,637		0,417	0,297	0,421	0,126	0,140			0,949
Estados											
Acre	0,157	0,540		0,233	0,182	0,311	0,047	0,091			0,835
Alagoas	0,132	0,508		0,019	0,016	0,167	0,031	0,086			0,757
Amapá	0,163	0,589		0,243	0,138	0,357	0,005	0,078			0,916
Amazonas	0,178	0,537		0,255	0,162	0,325	0,028	0,105			0,901
Bahia	0,197	0,498		0,207	0,171	0,350	0,041	0,083			0,834
Ceara	0,150	0,524		0,132	0,088	0,258	0,042	0,082			0,812
Distrito Federal	0,562	1,043		0,599	0,516	0,606	0,241	0,215			0,965
Espírito Santo	0,347	0,627		0,283	0,209	0,240	0,099	0,097			0,919
Goiás	0,313	0,655		0,371	0,299	0,480	0,105	0,116			0,920
Maranhão	0,103	0,358		0,082	0,052	0,186	0,016	0,039			0,791
Mato Grosso	0,237	0,646		0,366	0,284	0,475	0,084	0,115			0,915
Mato Grosso Do Sul	0,292	0,699		0,337	0,257	0,482	0,106	0,152			0,923
Minas Gerais	0,323	0,592		0,317	0,244	0,367	0,102	0,097			0,917
Para	0,125	0,476		0,207	0,113	0,304	0,023	0,069			0,883
Paraíba	0,138	0,520		0,091	0,074	0,232	0,037	0,064			0,781
Paraná	0,418	0,617		0,485	0,352	0,486	0,165	0,107			0,937
Pernambuco	0,205	0,619		0,130	0,099	0,300	0,039	0,097			0,820
Piauí	0,125	0,472		0,198	0,138	0,314	0,025	0,083			0,771
Rio De Janeiro	0,555	0,676		0,454	0,379	0,480	0,147	0,124			0,957
Rio Grande Do Norte	0,163	0,596		0,127	0,103	0,343	0,034	0,092			0,815
Rio Grande Do Sul	0,380	0,659		0,343	0,251	0,364	0,131	0,144			0,955
Rondônia	0,187	0,645		0,294	0,192	0,473	0,060	0,103			0,913
Roraima	0,183	0,514		0,319	0,189	0,450	0,014	0,084			0,897
Santa Catarina	0,423	0,631		0,432	0,286	0,409	0,153	0,105			0,959
São Paulo	0,630	0,718		0,522	0,422	0,524	0,223	0,135			0,957

	FIX	CEL	BAND	COMP	INT	USR	INT FIX	INT MOV	SEC	TER	ALF
Sergipe	0,170	0,585		0,179	0,131	0,329	0,040	0,101			0,816
Tocantins	0,175	0,589		0,169	0,127	0,343	0,052	0,086			0,869
Regiões Metropolitanas/Interior											
Bahia - Interior				0,149	0,121	0,290					
Ceará - Interior				0,068	0,035	0,152					
Minas Gerais - Interior				0,253	0,184	0,315					
Pará - Interior				0,126	0,040	0,210					
Paraná - Interior				0,461	0,319	0,446					
Pernambuco - Interior				0,088	0,046	0,197					
Rio de Janeiro - Interior				0,365	0,293	0,397					
Rio Grande do Sul - Interior				0,332	0,243	0,316					
RM Belém				0,357	0,248	0,478					
RM BH				0,507	0,422	0,522					
RM Curitiba				0,541	0,428	0,582					
RM Fortaleza				0,246	0,183	0,450					
RM Porto Alegre				0,360	0,265	0,438					
RM Recife				0,180	0,164	0,424					
RM RJ				0,479	0,402	0,502					
RM Salvador				0,360	0,306	0,507					
RM São Paulo				0,534	0,443	0,560					
São Paulo - Interior				0,511	0,403	0,491					

FONTE: ANATEL(2010), CGLbr (2010), IBGE (2010)

Apêndice 3 – PIB per capita dos países incluídos no *ranking* em função do IDI

Economia	PIB per capita (Dólar internacional)	Fonte	Ano
África do Sul	\$ 10.973,00	FMI	2011
Albânia	\$ 7.741,00	FMI	2011
Alemanha	\$ 37.897,00	FMI	2011
Angola	\$ 5.895,00	FMI	2011
Antígua e Barbuda	\$ 17.981,00	FMI	2011
Arábia Saudita	\$ 24.237,00	FMI	2011
Argélia	\$ 7.333,00	FMI	2011
Argentina	\$ 17.516,00	FMI	2011
Armênia	\$ 5.384,00	FMI	2011
Austrália	\$ 40.234,00	FMI	2011
Áustria	\$ 41.822,00	FMI	2011
Azerbaijão	\$ 10.202,00	FMI	2011
Bahrain	\$ 27.556,00	FMI	2011
Bangladesh	\$ 1.693,00	FMI	2011
Barbados	\$ 23.417,00	FMI	2011
Bélgica	\$ 37.737,00	FMI	2011
Benin	\$ 1.481,00	FMI	2011
Bielorrússia	\$ 15.028,00	FMI	2011
Bolívia	\$ 4.789,00	FMI	2011
Bósnia e Herzegovina	\$ 8.133,00	FMI	2011
Botswana	\$ 1.603,00	FMI	2011
Brasil	\$ 11.769,00	FMI	2011
Brunei Darussalam	\$ 49.384,00	FMI	2011

Economia	PIB per capita (Dólar internacional)	Fonte	Ano
Bulgária	\$ 13.597,00	FMI	2011
Burkina Faso	\$ 1.466,00	FMI	2011
Butão	\$ 6.112,00	FMI	2011
Cabo Verde	\$ 3.947,00	FMI	2011
Camarões	\$ 2.257,00	FMI	2011
Camboja	\$ 2.216,00	FMI	2011
Canadá	\$ 40.541,00	FMI	2011
Catar	\$ 102.943,00	FMI	2011
Cazaquistão	\$ 13.001,00	FMI	2011
Chade	\$ 1.865,00	FMI	2011
Chile	\$ 17.222,00	FMI	2011
China	\$ 8.382,00	FMI	2011
Chipre	\$ 29.074,00	FMI	2011
Cingapura	\$ 59.711,00	FMI	2011
Colômbia	\$ 10.249,00	FMI	2011
Comores	\$ 1.232,00	FMI	2011
Congo	\$ 348,00	FMI	201
Coreia do Sul	\$ 31.714,00	FMI	2011
Costa do Marfim	\$ 159,00	FMI	2011
Costa Rica	\$ 11.927,00	FMI	2011
Croácia	\$ 18.192,00	FMI	2011
Cuba	\$ 99,00	CIA	2010 est.
Dinamarca	\$ 37.152,00	FMI	2011
Djibouti	\$ 2.642,00	FMI	2011
Egito	\$ 654,00	FMI	2011
El Salvador	\$ 755,00	FMI	2011
Emirados Árabes Unidos	\$ 48.158,00	FMI	2011
Equador	\$ 8.492,00	FMI	2011
Eritreia	\$ 735,00	FMI	2011
Eslovênia	\$ 28.642,00	FMI	2011
Espanha	\$ 30.626,00	FMI	2011
Estados Unidos	\$ 48.387,00	FMI	2011
Estônia	\$ 2.038,00	FMI	2011
Etiópia	\$ 1.093,00	FMI	2011
Fiji	\$ 462,00	FMI	2011
Filipinas	\$ 4.073,00	FMI	2011
Finlândia	\$ 36.236,00	FMI	2011
França	\$ 35.156,00	FMI	2011
Gabão	\$ 16.183,00	FMI	2011
Gâmbia	\$ 1.943,00	FMI	2011
Gana	\$ 3.083,00	FMI	2011
Geórgia	\$ 5.491,00	FMI	2011
Grécia	\$ 26.294,00	FMI	2011
Guatemala	\$ 507,00	FMI	2011
Guiana	\$ 7.466,00	FMI	2011
Guiné	\$ 1.083,00	FMI	2011
Holanda	\$ 42.183,00	FMI	2011
Honduras	\$ 4.345,00	FMI	2011
Hong Kong	\$ 46.502,00	Banco Mundial	2010
Hungria	\$ 19.591,00	FMI	2011
Iémen	\$ 2.307,00	FMI	2011

Economia	PIB per capita (Dólar internacional)	Fonte	Ano
Índia	\$ 3.694,00	FMI	2011
Indonésia	\$ 4.666,00	FMI	2011
Irã	\$ 13.053,00	FMI	2011
Irlanda	\$ 39.639,00	FMI	2011
Islândia	\$ 38.061,00	FMI	2011
Israel	\$ 30.975,00	FMI	2011
Itália	\$ 30.464,00	FMI	2011
Jamaica	\$ 9.029,00	FMI	2011
Japão	\$ 3.474,00	FMI	2011
Jordânia	\$ 59,00	FMI	2011
Lao P.D.R.	\$ 2.659,00	FMI	2011
Letônia	\$ 15.662,00	FMI	2011
Líbano	\$ 15.523,00	FMI	2011
Lituânia	\$ 18.856,00	FMI	2011
Luxemburgo	\$ 80.119,00	FMI	2011
Macau, China	\$ 63.681,00	Banco Mundial	2010
Macedônia	\$ 10.367,00	FMI	2011
Madagáscar	\$ 934,00	FMI	2011
Malásia	\$ 15.568,00	FMI	2011
Maldivas	\$ 8.731,00	FMI	2011
Mali	\$ 1.128,00	FMI	2011
Malta	\$ 25.428,00	FMI	2011
Marrocos	\$ 5.052,00	FMI	2011
Maurício	\$ 14.954,00	FMI	2011
Mauritânia	\$ 2.179,00	FMI	2011
México	\$ 1.461,00	FMI	2011
Moçambique	\$ 1.085,00	FMI	2011
Moldávia	\$ 3.373,00	FMI	2011
Mongólia	\$ 4.744,00	FMI	2011
Montenegro	\$ 11.545,00	FMI	2011
Namíbia	\$ 7.363,00	FMI	2011
Nepal	\$ 1.328,00	FMI	2011
Nicarágua	\$ 3.206,00	FMI	2011
Níger	\$ 771,00	FMI	2011
Nigéria	\$ 2.578,00	FMI	2011
Noruega	\$ 53.471,00	FMI	2011
Nova Zelândia	\$ 27.668,00	FMI	2011
Oman	\$ 26.519,00	FMI	2011
Panamá	\$ 14.097,00	FMI	2011
Papua Nova Guiné	\$ 2.532,00	FMI	2011
Paquistão	\$ 2.787,00	FMI	2011
Paraguai	\$ 5.413,00	FMI	2011
Peru	\$ 10.062,00	FMI	2011
Polónia	\$ 20.334,00	FMI	2011
Portugal	\$ 23.361,00	FMI	2011
Quênia	\$ 1.746,00	FMI	2011
Quirguistão	\$ 2.372,00	FMI	2011
Reino Unido	\$ 3.609,00	FMI	2011
Rep. Dominicana	\$ 9.287,00	FMI	2011
República Checa	\$ 27.062,00	FMI	2011
República Eslovaca	\$ 23.304,00	FMI	2011

Economia	PIB per capita (Dólar internacional)	Fonte	Ano
Romênia	\$ 12.476,00	FMI	2011
Ruanda	\$ 1.341,00	FMI	2011
Rússia	\$ 16.736,00	FMI	2011
Senegal	\$ 1.871,00	FMI	2011
Sérvia	\$ 10.642,00	FMI	2011
Seychelles	\$ 24.726,00	FMI	2011
Síria	\$ 5.041,00	FMI	2010
Sri Lanka	\$ 5.674,00	FMI	2011
Suazilândia	\$ 5.302,00	FMI	2011
Suécia	\$ 40.394,00	FMI	2011
Suíça	\$ 4.337,00	FMI	2011
Suriname	\$ 9.475,00	FMI	2011
Tailândia	\$ 9.396,00	FMI	2011
Tanzânia	\$ 1.515,00	FMI	2011
Togo	\$ 899,00	FMI	2011
Trinidad & Tobago	\$ 20.053,00	FMI	2011
Tunísia	\$ 9.478,00	FMI	2011
Turcomenistão	\$ 7.846,00	FMI	2011
Turquia	\$ 14.517,00	FMI	2011
Ucrânia	\$ 7.233,00	FMI	2011
Uganda	\$ 1.317,00	FMI	2011
Uruguai	\$ 15.113,00	FMI	2011
Uzbequistão	\$ 3.302,00	FMI	2011
Venezuela	\$ 12.568,00	FMI	2011
Vietnã	\$ 3.359,00	FMI	2011
Zâmbia	\$ 1.611,00	FMI	2011
Zimbábue	\$ 487,00	FMI	2011

Apêndice 4 – PIB per capita dos estados e regiões do Brasil

Estado	PIB per capita (R\$)
ACRE	R\$ 10.322,25
ALAGOAS	R\$ 6.498,30
AMAPA	R\$ 11.412,81
AMAZONAS	R\$ 14.121,32
BAHIA	R\$ 9.044,71
CEARA	R\$ 7.423,96
DISTRITO FEDERAL	R\$ 48.714,91
ESPIRITO SANTO	R\$ 18.490,96
GOIAS	R\$ 13.953,02
MARANHAO	R\$ 6.045,54
MATO GROSSO	R\$ 18.435,06
MATO GROSSO DO SUL	R\$ 14.880,48
MINAS GERAIS	R\$ 13.838,99
PARA	R\$ 7.590,63
PARAIBA	R\$ 7.357,40
PARANA	R\$ 17.171,58
PERNAMBUCO	R\$ 8.597,74
PIAUI	R\$ 5.844,33
RIO DE JANEIRO	R\$ 21.347,69
RIO GRANDE DO NORTE	R\$ 8.589,98
RIO GRANDE DO SUL	R\$ 19.102,54
RONDONIA	R\$ 12.995,77
RORAIMA	R\$ 12.817,00
SANTA CATARINA	R\$ 20.489,60
SAO PAULO	R\$ 25.306,86
SERGIPE	R\$ 9.452,81
TOCANTINS	R\$ 10.892,33

Região	PIB per capita (R\$)
Centro-Oeste	R\$ 21.600,40
Nordeste	R\$ 7.888,65
Norte	R\$ 10.262,69
Sudeste	R\$ 21.390,42
Sul	R\$ 18.664,30

FONTE: IPEA (2009)

Apêndice 5 –Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso da internet

P(INTERNET = Sim | ...)

P(INTERNET = Sim | ...)

Apêndice 6 – Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso do celular

P(USOU CELULAR = Sim | ...)

P(USOU CELULAR = Sim | ...)

Apêndice 7 – Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso do comércio eletrônico

P(eCOMMERCE = Sim | ...)

P(eCOMMERCE = Sim | ...)

Apêndice 8 – Diagramas de barra com as probabilidades associadas ao estado de cada uma das variáveis do modelo de uso do internet banking

P(BANKING = Sim | ...)

P(BANKING = Sim | ...)