ABSTRACT
This paper reports about a study to analyse the Internet diffusion in Italy by Italian firms, using domain names under the .it ccTLD as metric. The penetration rate calculated according to the number of companies is computed for highly dissimilar geographical areas (regions). A concentration analysis was performed in order to discover whether the geographical distribution of the Internet is less concentrated with respect to both the number of companies present in Italy and income level, suggesting a diffusive effect. Regression analysis was carried out using social, economic and infrastructure indicators. Results show that a “digital divide” exists in terms of geographical distribution (i.e., in macro-areas – Northern, Central, and Southern Italy - and at the regional level). In future we are going to carry out a research in order to compare the number of domains registered to businesses with that of domains registered in the non-profit sector.

Keywords
Domain names, Digital Divide, Internet Diffusion, Power laws, Zipf's law.

1. INTRODUCTION
According to the economics literature, the Internet is expanding very rapidly (Coffman, Odlyzko, 2001). Studies carried out by the Network User Association (NUA Ltd) estimated the worldwide on-line population in 1999 and in 2002. According to this society, in Europe the number of individuals on-line came to 190.91 million in 2002, compared to 47.15 million in 1999. Companies as well as individuals also turn to the Internet to exploit its communication potential. Today, information infrastructures are reaching out to the individual consumer, and telematic networks reduce the cost of communications. This statement agrees with the economics literature (Novak e Hoffman, 1996), which confirms that the Web is becoming a dynamic and personal means of communication.
According to other authors (Candi, d'Ignazio, Sabbioni, 2002) the spread of the Internet and the functions of electronic commerce will permit individual clients to choose from a wide array of products and reduce costs, selecting and buying goods directly from the source and allowing companies to sell while by passing traditional channels. Scandinavia, at 8.6%, leads the region with the highest percentage of on-line sales, usually computers and related products, travel, video and music, and books.
This situation could prove to be quite worrisome for traditional businesses, as emerges from a survey carried out by the Union Sindacati Agenti and representatives of the Commercio Italiani in November 2000. However, companies must adopt entirely new forms of commercial activity so that online sales will be successful.
The advantages for businesses provided by the Internet are not only linked to the sale of products and services (direct advantages) but can also be indirect (Hansons, 2000). For example, among the most important of these are reduced costs, image consolidation, greater customer loyalty, and a wider diffusion of products offered by the company. They are referred to as “indirect” since they do not lead directly to sales and do not generate immediate profits; however, they are important since they will probably be the greatest benefits offered to businesses by the Internet.
The gradual confirmation of the Internet as a means of communication also permits companies access data and a variety of other information; for example, it is possible to rapidly obtain information about the market in which one operates by visiting websites specialized in economic information or areas that furnish updates on laws, price changes, the appearance of any new operators in the field, fairs, competitive bidding, and other news of interest to operators. One can also identify the competition and analyze them by means of information published on company websites, etc.
Several metrics are available for measuring the diffusion of the Internet. The most suitable are the so-called endogenous metrics which can be “obtained in an automatic or semiautomatic way from the Internet itself” (Diaz-Picazo, 1999). These metrics have the unquestionable advantage of accuracy; according to the literature, the most frequently used are Internet hosts (based on hostcount procedures (see studies published by Internet Software Consortium or da Ripe) and second-level domain names (Naldi, 1997; Zook, 1999; Bauer, Berneand and Maitland, 2002). To measure the analysis of Internet diffusion in Italian companies, we have used the endogenous measure of second-level domain names registered under the ccTLD “.it.”, managed by the Institute of Informatics and Telematics of Pisa.

However, aside from the advantages offered by endogenous measures, there are a few disadvantages, since they tend in some cases to underestimate, and in others to overestimate, the phenomenon (Zook, 1999, 2000, 2001). Overestimation can occur when the number of hosts is used, often associated with more than one IP address, while if we consider the number of domains registered, more than one domain may be associated with the same registrant. Underestimation can occur because not all internet users register a domain name under their own ccTLD, and in many countries, the regulations allow foreign citizens to register under their own ccTLD.

In the case of hosts, underestimation may be due to the increasing presence of firewalls and private networks (Intranet) and the use of dynamic IP addresses, increasingly accompanied by new tools for accessing the Network (for example, mobile phones). In spite of these disadvantages, the numbers of hosts and Internet domains are the best means for analyzing Internet diffusion.

2. Methods
The Institute for Informatics and Telematics (IIT-CNR), which manages the “.it” ccTLD Registry, conducted a study to analyze the diffusion of Internet use in Italy. Data were extracted from databases of the registrations managed by the IIT-CNR, using automatic and semi-automatic procedures. Approximately 550,000 domain names have been analyzed and grouped into several categories (individuals, firms, universities, associations, public groups, foundations, committees, and other organizations) in order to identify the determinants of adoption and then of diffusion, for each category. A careful data cleaning procedure was followed. As of September 7, 2001, the database WHOIS contained 265,437 domains registered by companies, of which 3,503 were domains registered by companies with their legal headquarters in the various EU countries. The remaining 3,331 domains were not classified, since it was impossible to ascertain the area where these companies operate. A total of 258,603 domains registered by businesses were analyzed. Furthermore, to reduce the disadvantages previously described, regarding the overestimation of the analysis of Internet diffusion when using domain names as a measure, in this paper we will only consider one domain name for each business: in other words, if a company registered multiple domain names, we considered only the first one registered.

3. Results
3.1 Analysis of concentration
As the literature suggests, analysis shows that a technological divide exists among companies operating in specific geographic areas in Italy. As shown in the following Figure, the North is more likely than other areas to use the new technology. Of the domains registered by Italian firms, 59.20% is represented by the North (153,088 of 258,063); Central Italy attained a percentage of 23.51% (60,796 of 258,603), and finally the South showed a percentage of 17.29% (44,719 domain names registered, of a total 258,603).

Since Italy is divided into 20 administrative units called regions, in this article we have found it advisable to analyze internet diffusion by region. As seen in Figure 2 and 3, Lombardy, Alto Adige and Lazio (in that order) are the regions

---

2 http://www.ripe.net/statistics/hostcount.html
3 For example, Italy allows organizations and citizens of in the European Union countries to register under the “.it” ccTLD
4 Database Whois, Database of the state of the registrations, Database of the Letter of Assumption of Responsibility
showing the highest penetration rates (penetration rate is calculated by dividing the number of domains registered by companies by the number of companies in Italy, and multiplying the ratio by 100. At the macro-area level, the relative penetration rate recorded in North and Central Italy are higher compared to the South (7.23 and 7.07 respectively for every 100 companies, compared to 4.04 in the South.).

**Figure 2 - Penetration rates of domains in different regions**

![Figure 2](image)

**Figure 3 - Cumulate number of domains registered by firms**

![Figure 3](image)

The calculation of the Gini index and the construction of the Lorenz curve show that the concentration of domains is higher compared to the number of firms and total income. The Gini index, calculated according to the number of domains registered, was 0.557 compared to 0.468 computed according to the number of companies and 0.466 calculated according to total income.

**Lorenz curve for domains, income, number of firms**

![Lorenz curve](image)

This shows that, as the literature suggests (Hansons, 2000; Warschauer, 2001), the existing divide regarding Internet diffusion does not depend only on the number of companies and total income in a region but also on other, such as social and cultural, factors. However, it can be deduced that the wealthiest and most industrialized regions in Italy are more likely to adopt the new technology than are the other regions. The three leading regions, which register 45.8% of

---

5 The Gini index can assume values between 0 and 1, where 1 indicates a situation of maximum concentration and 0 indicates equidistribution.

6 This also emerges from a study conducted by the AMD Global Consumer Advisory Board (GCAB) which has examined the causes of the differences in Internet access and use.

the Internet domains, have a percentage of number of companies equivalent to 35.6%, and Pearson’s correlation between the number of domains registered by companies in the respective regions and total income was 0.969; instead, the correlation between the number of companies and registration of the number of domains was 0.971.

3.2. Determinants of adoption

According the OECD (1999), the digital divide is the distinction between “Who has and who does not have access to information (OECD)”. Several aspects of this phenomenon exist. It is possible to observe a digital divide among countries (international digital divide), among individuals or organizations and within a country at the local level (domestic digital divide). In order to analyze whether there is a domestic digital divide in Italy, we ran exploratory stepwise regressions using economic and social indicators at the regional level. This allowed us to examine the determinants of Internet diffusion, by singling out the factors leading to registration of domains by company. The dependent variable of our regression models is the penetration rate calculated as

Penetration Rate = (Number of domains registered by companies/number of companies) * 100;

We constructed three different models. The first two (M1, M2) are simple univariate regression models, analyzing the influence on Internet diffusion of two key economic factors: per-capita income and added value per employee. As expected, these two variables positively correlated with penetration rates per capita income and value added per employee (0.701 and 0.827 respectively). The economic literature, in fact, underlines that differences in income distribution play a crucial role in explaining differences in ITC access and utilization (Norris, 2000; Pohjola, 2002, Hansons, 2000 Warschauer, 2001). We included these crucial indicators in distinct models in order to address multicollinearity problems. Per capita income and added value per employee are very likely to be correlated with other social and economic indicators at a local level, generating distortions in the estimated coefficients. Results regarding M1 are reported in Table 1. Although it includes only a dependent variable, the model is quite powerful, and explains about 49% per cent of the variability in registrations among Italian regions. The coefficient of per capita income is highly significant and has the expected sign, showing a positive influence of per capita income on Internet diffusion at a local level.

$R^2 = 0.49$

<table>
<thead>
<tr>
<th>Model</th>
<th>Non-standardized coefficients</th>
<th>Standardized coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Standard Error</td>
<td>Beta</td>
<td>-6.68</td>
</tr>
<tr>
<td>1</td>
<td>(Costant) per capita income</td>
<td>0.000</td>
<td>0.000</td>
<td>0.701</td>
</tr>
</tbody>
</table>

Table 1: Stepwise regressions with per capita income as dependent variable.

Quite similar results are obtained in M2 (Table 2) including the added value per employees as dependent variable. The $R^2$ is even higher, stating that the efficiency of the productivity structure account for 68.5% of the variability in the Internet diffusion.

$R^2 = 0.68$

<table>
<thead>
<tr>
<th>Model</th>
<th>Non-standardized coefficients</th>
<th>Standardized coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Standard Error</td>
<td>Beta</td>
<td>-4.138</td>
</tr>
<tr>
<td>2</td>
<td>(Costant) added value employee</td>
<td>-10.836</td>
<td>2.619</td>
<td>0.827</td>
</tr>
</tbody>
</table>

Table 2: Stepwise regressions with per added value employee as dependent variable.

---

7 Pearson’s index can assume values ranging between –1 and +1. –1 indicates discordance between the variables examined, and this indicates that the positive variations of one coincide with the negative variations of the other while positive values indicate concordance.

8 The correlation is significant to 0.01.

As previously stated, and in agreement with the economics literature, the variation in Internet diffusion between regions may derive from other factors as well. In the following table (Table 3) we analyzed the stepwise regression, taking into consideration as dependent variable an economic factor (employees in the service sector); one related to education (number of college graduates); a socio-cultural variable (spending for theatrical and musical performances); one related to infrastructure (founds for telephony and telematics) and one relative to public spending (hydraulic works and electrical systems).

The Table shows that regions that spend considerable funds on musical and cultural activity are more likely to use the new technology (spending for theater has the second-highest Beta value compared to the other variables). As might be expected, the index of spending for telephone and telematics also plays an important role (the Beta is equivalent to 0.708). In fact as the literature proposes (Warschauer, 2001) one of the determining factors in internet diffusion is the presence of adequate network infrastructures.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Non-standardized coefficients</th>
<th>Standardized coefficients</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Service sector employees</td>
<td>1.719E-06</td>
<td>0.000</td>
<td>0.545</td>
<td>2.761</td>
</tr>
<tr>
<td>Number college graduates</td>
<td>0.000</td>
<td>0.000</td>
<td>0.548</td>
<td>2.703</td>
</tr>
<tr>
<td>Spending theater and music</td>
<td>2.834E-05</td>
<td>0.000</td>
<td>0.657</td>
<td>3.697</td>
</tr>
<tr>
<td>Founds for telephony and telematics</td>
<td>0.033</td>
<td>0.008</td>
<td>0.708</td>
<td>4.254</td>
</tr>
<tr>
<td>Hydraulic works and electrical systems</td>
<td>1.255E-05</td>
<td>0.000</td>
<td>0.543</td>
<td>2.743</td>
</tr>
</tbody>
</table>

Table 3: Stepwise regressions taking as independent variable the penetration rate of the companies

In brief, it is possible to conclude that regions with an efficient and service-oriented structure, a lively cultural scene, and a good educational level (greater number of college graduates) are more inclined to use the new technology and are the best candidates for a more active and interactive use of the Internet.

4. Further developments of the research

4.1 Comparison of Internet diffusion between profit-oriented and non-profit organizations

Our next goal in this ongoing project is to compare the results from this study with an analysis of Internet diffusion in the non-profit sector. Therefore we intend to verify:

1) Whether organizations that do not pursue monetary gain (i.e., foundations, committees and other organizations in the non-profit sector) take advantage of the new technology to the same degree as companies

2) Whether, using regression analysis, factors that significantly influence variability of Internet diffusion at the regional level in companies have an important role in recording indexes of penetration and domains also for non-profit organizations

3) Whether the registration of domains of companies is less concentrated compared to that of domains registered by non-profit organizations

4) Finally, predict by means of a temporal analysis between the various regions in registration of these organizations (either for-profit or non-profit) whether the divide will close.

4.2 The distribution of domain names

At present a growing body of literature is devoted to discovering and analyzing the regularities displayed by the Internet network (Pitkow, 1998). In particular, the presence of power laws (Blank, Solomon, 2000) for Internet-related phenomena is widely accepted. Power laws are discovered in the number of in and out links of a website (Barabasi, Albert, 1999; Adamic, Huberman, 2000), in the number of pages composing an Internet web site, and in the behavior of Internet surfers (Huberman, Pirolo, Pitkow, Lukose, 1998; Johansen, 2001).

Among the several types of power laws detected so far, Zipf’s law (Zipf, 1949) seems to play a central role. This law has been especially applied to the analysis of the distribution of the population of cities in a country or region (Gabaix, 1999; Krugman, 1996, 1998). Under this law, ordering cities according to their population and plotting the logarithm of
the population against the logarithm of the rank in a graph, a straight line whose slope is about –1 is obtained. Therefore, it turns out that the population of the first city in a country is approximately twice the population of the second city, three times that of the third one and so on. Moreover, this seems to be a broad general law and has been applied to linguistics (for the frequency of words in a text, Mandelbrot, 1965; Alexander, Sidorov, 2001), to the study of the intensity of earthquakes (Sonette et al., 1996) and to various fields in biology and physiology (Jorgensen, Mejer and Nielsen, 2001). In economics it has been applied to the distribution of company sizes (Axtell, 2001), measured in different ways (income, assets, number of employees) (Okuyama, Takayasu, Takayasu, 1999).

The idea is to investigate the presence of power laws for domains registered by firms in Italian cities. If a power law is discovered, the challenge will be to analyze how it is generated. However, at present there is no generally accepted theoretical foundation for this empirical regularity.

References
[16] Zook, M.A. (2000) Internet metrics: using host and domain counts to map the Internet Telecommunications Policy 24.6-7

TOP