Relationship Between Web Links and Trade

Ricardo Baeza-Yates
Yahoo! Research Barcelona
Barcelona, Spain
ricardo@baeza.cl

Carlos Castillo
Università di Roma "La Sapienza"
Rome, Italy
castillo@dis.uniroma1.it

ABSTRACT

We report on observations on Web characterization studies that suggest that the amount of Web links among sites under different country-code top-level domains is related to the amount of trade between the corresponding countries.

Categories and Subject Descriptors: J.4 [Social and Behavioral Sciences]: Economics; H.4.m [Information Systems]: Miscellaneous

General Terms: Measurement

Keywords: World trade graph; National Web domains

1. WEB LINKS AND TRADE

The World Wide Web can be seen as a directed graph in which each page is a node and each hyperlink between two pages is an edge. This graph can be naturally partitioned into different hosts: if we collapse all pages in the same host to a single node, keeping their links to pages in other hosts, we obtain a new graph called the **Hostgraph** [3]. This graph is studied extensively in [2], in which hosts are grouped into country-code top-level domains; they observe geographical connections between the most linked countries.

We focus on the relationship between Web links and commercial trade, using several Web collections obtained using Web crawls between 2002 and 2004 [1]. Table 1 shows the characteristics of them; for each collection, we show the number of external links to *country-code top-level domains*, excluding `.com`, `.net`, `.org`, `.biz` and other generic domains.

Table 1: Characteristics of the collections studied.

Country	Pages [millions]	Number of external links	
		Total	Different sites
U.K.	18.5	1,857,948	229,731
Spain	16.2	2,785,377	184,754
Greece	3.7	11,004	1,798
Brazil	4.7	478,446	3,794
Chile	3.3	7,368	1,061

We obtained the number of links to pages in other countries; they are shown in Figure 1. A good model for the number of links is an exponential distribution (with CDF $\lambda e^{-\lambda x}$). The parameters for the fit in different countries are quite similar, with $\lambda = 0.10 \pm 0.01$.
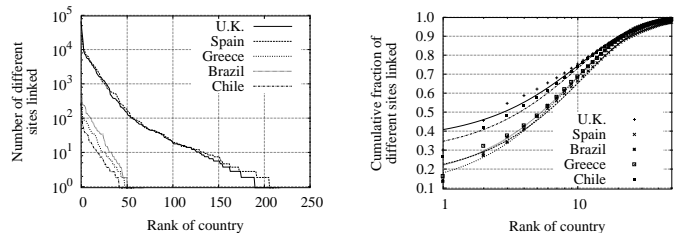
Figure 1: Distribution of links to sites outside the country.

For the data about commercial trade, we use the Commodity Trade Database (COMTRADE) of the United Nations Statistics Division (available online at `http://unstats.un.org/unsd/comtrade/`). The distribution of the exports to other countries from our collection is shown in Figure 2.
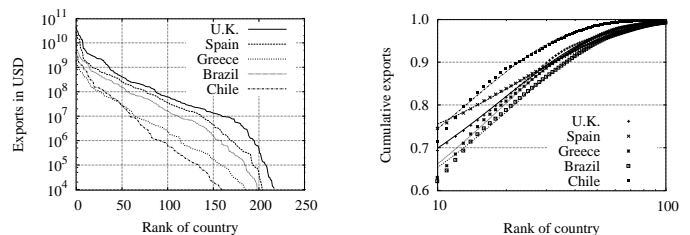

Figure 2: Distribution of exports between countries.

In [4] the authors considered an unweighted version of this data, and used a graph considering that two countries are linked if their volume of trade is above a certain threshold. They found that this graph exhibit scale-free properties. In our case, except for the first few countries (roughly 10) at the beginning, which appear to follow a power-law, the behavior for most of the trade partners is roughly an exponential distribution with parameter $\lambda = 0.05 \pm 0.01$ for the exports, and $\lambda = 0.06 \pm 0.01$ for the imports, which means that in these countries the exports are slightly more diversified than imports. The variations in the parameter λ depend on the diversification of the trade of the country. Chile has the smaller diversification in this sample, and Spain the larger.

2. CORRELATION

We found that there is a relationship between the number of links to pages in other countries and the amount of trade with those countries, as shown in Figure 3. We include in the calculation only pairs of countries where the trade and link
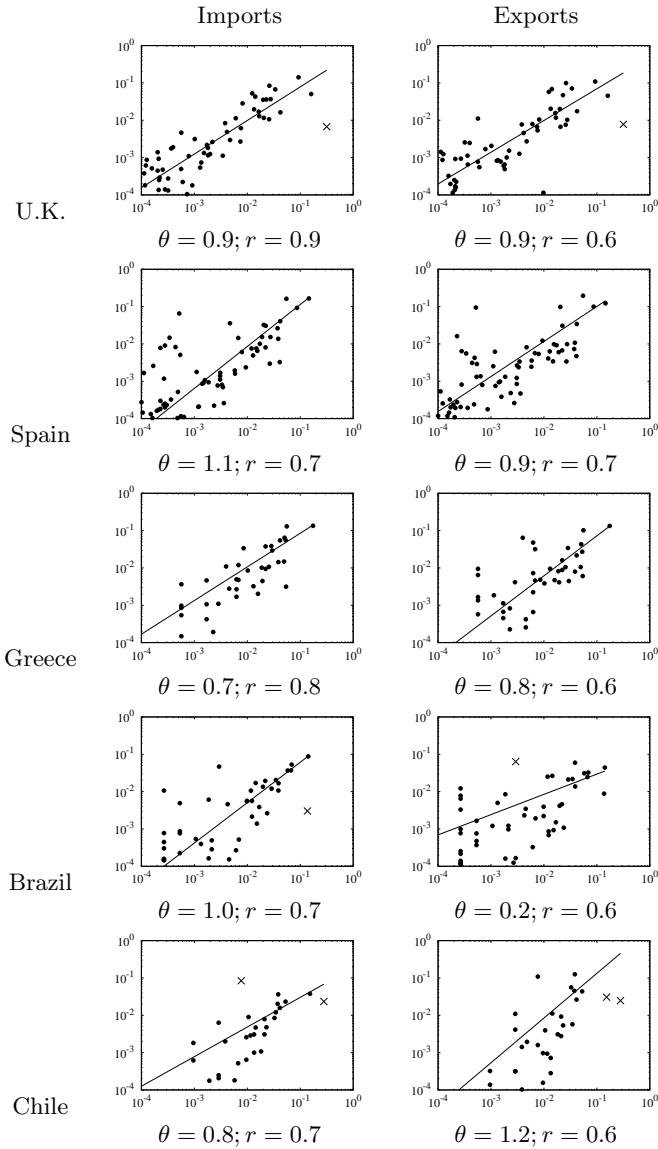
Figure 3: Relationship between links and trade, each dot represents a country; the x-axis are links, and the y-axis the amount of imports (left) or exports (rights) from/to that country. The fit exponent θ and Pearson's correlation r are shown.

Table 2: Comparison of cross-similarities between Web links and exports. The values in **boldface** should be the higher in their rows, but they are sometimes lower than other values, which are shown underlined. Top: total number of links, bottom: number of different sites

Total number of links	U.K. Exp.	Spain Exp.	Greece Exp.	Brazil Exp.	Chile Exp.
U.K.	**0.23**	0.23	<u>0.28</u>	0.15	0.20
Spain	0.20	**0.24**	0.24	0.14	0.18
Greece	0.42	0.47	**0.49**	0.48	0.46
Brazil	0.26	<u>0.34</u>	0.31	**0.32**	0.32
Chile	0.36	0.43	0.45	0.48	**0.50**

Different sites	U.K. Exp.	Spain Exp.	Greece Exp.	Brazil Exp.	Chile Exp.
U.K.	**0.23**	0.23	<u>0.27</u>	0.15	0.19
Spain	0.22	**0.25**	<u>0.26</u>	0.15	0.17
Greece	0.45	<u>0.51</u>	**0.48**	0.47	<u>0.51</u>
Brazil	<u>0.37</u>	<u>0.47</u>	<u>0.44</u>	**0.31**	<u>0.35</u>
Chile	0.43	0.49	0.53	0.53	**0.55**

confusion (between different countries) than using the number of different sites. In the latter case, the sample of Brazil seems too small to give meaningful results. For Chile, Greece, and Spain, the results are better as the countries are more related to themselves than to other countries; this metric tends to put the United Kingdom and Spain closer to Greece than to themselves, this suggests that there might be a relationship between the ranked lists of trade partners of the U.K., Spain and Greece.

Preliminary results suggest that the ordering of trade partners is indeed strongly correlated to geographical distance and cultural ties, and we are currently analyzing this relationship. So far, our results show that the high-level structure of Web links among country-code domains is clearly related to commercial trade, and to the best of our knowledge this relationship had never been depicted in the past.

Acknowledgements

3. REFERENCES

[1] R. Baeza-Yates, C. Castillo, and E. Efthimiadis. Characterization of national Web domains. Technical report, Universitat Pompeu Fabra, July 2005.
[2] K. Bharat, B. W. Chang, M. Henzinger, and M. Ruhl. Who links to whom: Mining linkage between web sites. In *ICDM*, pp. 51–58, San Jose, California, USA, 2001. IEEE CS.
[3] S. Dill, R. Kumar, K. S. Mccurley, S. Rajagopalan, D. Sivakumar, and A. Tomkins. Self-similarity in the web. *ACM Trans. Inter. Tech.*, 2(3):205–223, 2002.
[4] A. M. Serrano and M. Boguna. Topology of the world trade web. *Physical Review E*, 68(1):015101+, July 2003.

is more than 10^{-3} of the total, as lower than that threshold, the data becomes very noisy. We also have removed 1 or 2 outliers from some graphs to improve the fit, they are marked with a cross in their graphs.

One explanation for this correlation is that the Web captures economical relationships, another is that the correlations observed are just recovering link and trade "popularity" of the countries, this is, the receivers of the larger amount of links and trade will always be the same countries, no matter which collection we are observing. To test this idea, we measured the correlation of the ranked lists of links of each country with the ranked lists of trade to every country in our collection. It is expected that a country's links will be more similar to that country's trade than to other countries. The results are mixed, as shown in Table 2.

Using the total number of links tends to generate less