Relationship Between Web Links and Trade

Ricardo Baeza-Yates
Yahoo! Research
Barcelona, Spain and Santiago, Chile

Carlos Castillo
Università di Roma "La Sapienza"
Rome, Italy

Summary

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

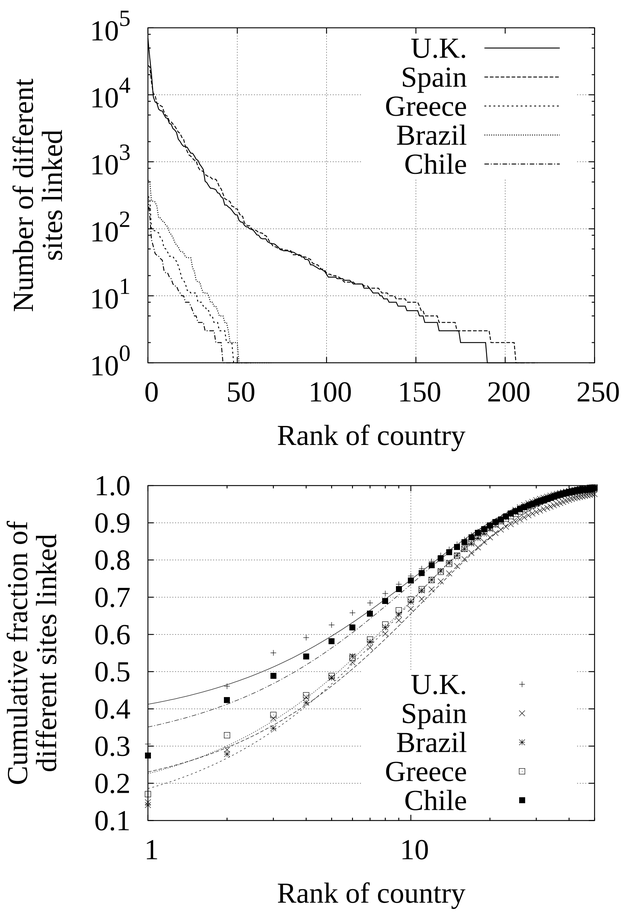
The World Wide Web can be seen as a directed graph in which each page is a node and each hyperlink between two pages is an edge. This graph can be naturally partitioned into different hosts: if we collapse all pages in the same host to a single node, keeping their links to pages in other hosts, we obtain a new graph called the **Hostgraph** [Dill et al., 2002]. This graph is studied in [Bharat et al., 2001], in which hosts are grouped into country-code top-level domains; they observe geographical connections between the most linked countries.

We are interested in relating Web links with trade, studying the following Web samples:

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

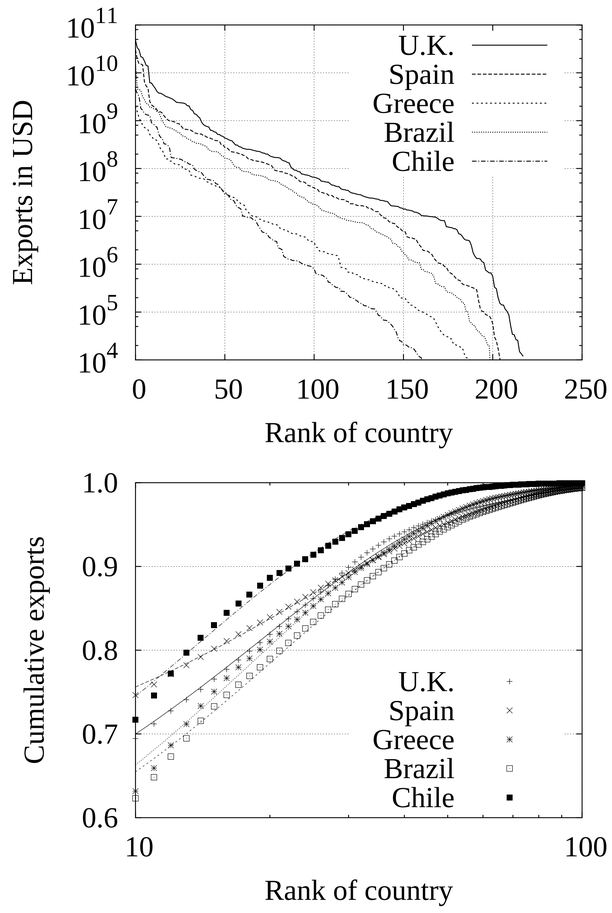
Link data

We obtained the number of links to pages in other countries; they are shown in the figure below. 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 $\lambda = 0.10 \pm 0.01$.



Trade data

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:



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.

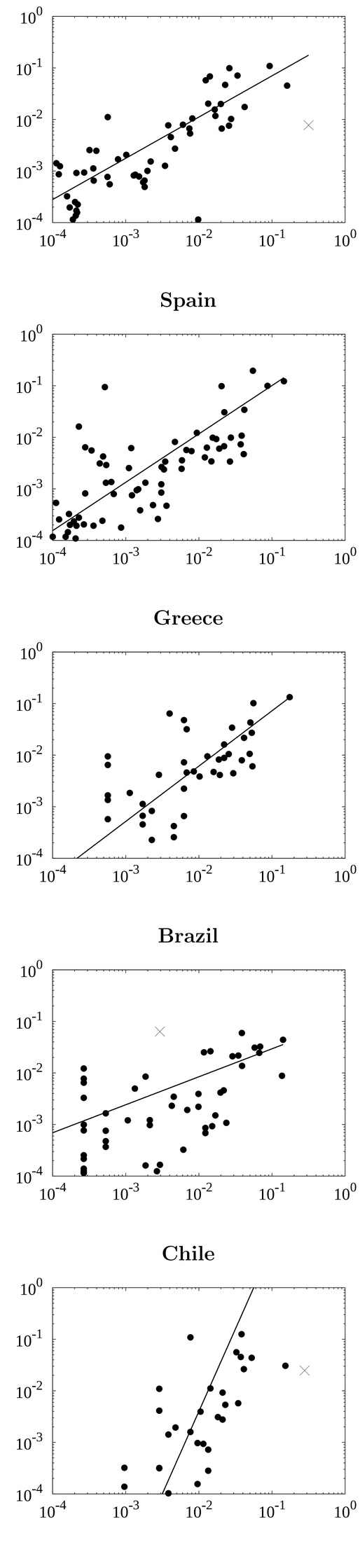
Correlation

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 the graphs on the right. We include in the calculation only pairs of countries where the trade and link is more than 10^{-3} of the total, as lower than that threshold, the data becomes very noisy.

We also measured the correlation of the ranked lists of links of each country with the ranked lists of trade to every country in our collection. In general, a country's Web links are more similar to that country's trade than to other countries.

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.

Each dot represents a country
X-axis is the fraction of out-links to that country.
Y-axis is the fraction of exports to that country.



References

[Bharat et al., 2001] Bharat, K., Chang, B. W., Henzinger, M., and Ruhl, M. (2001). Who links to whom: Mining linkage between web sites. In *International Conference on Data Mining (ICDM)*, pages 51–58, San Jose, California, USA. IEEE CS.

[Dill et al., 2002] Dill, S., Kumar, R., Mccurley, K. S., Rajagopalan, S., Sivakumar, D., and Tomkins, A. (2002). Self-similarity in the web. *ACM Trans. Inter. Tech.*, 2(3):205–223.