Typosquatting
The "Curse" of Popularity

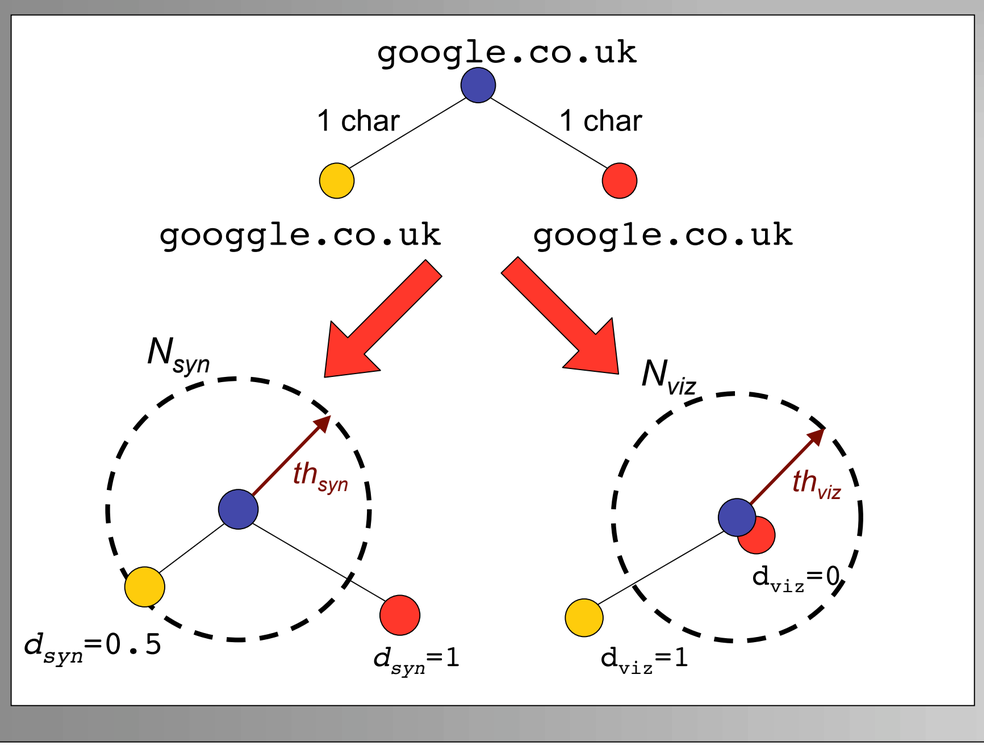
A. Linari, *F. Mitchell, D. Duce, S. Morris*
{**alinari**, frmitchell, daduce}@brookes.ac.uk, stephen.morris@nominet.org.uk

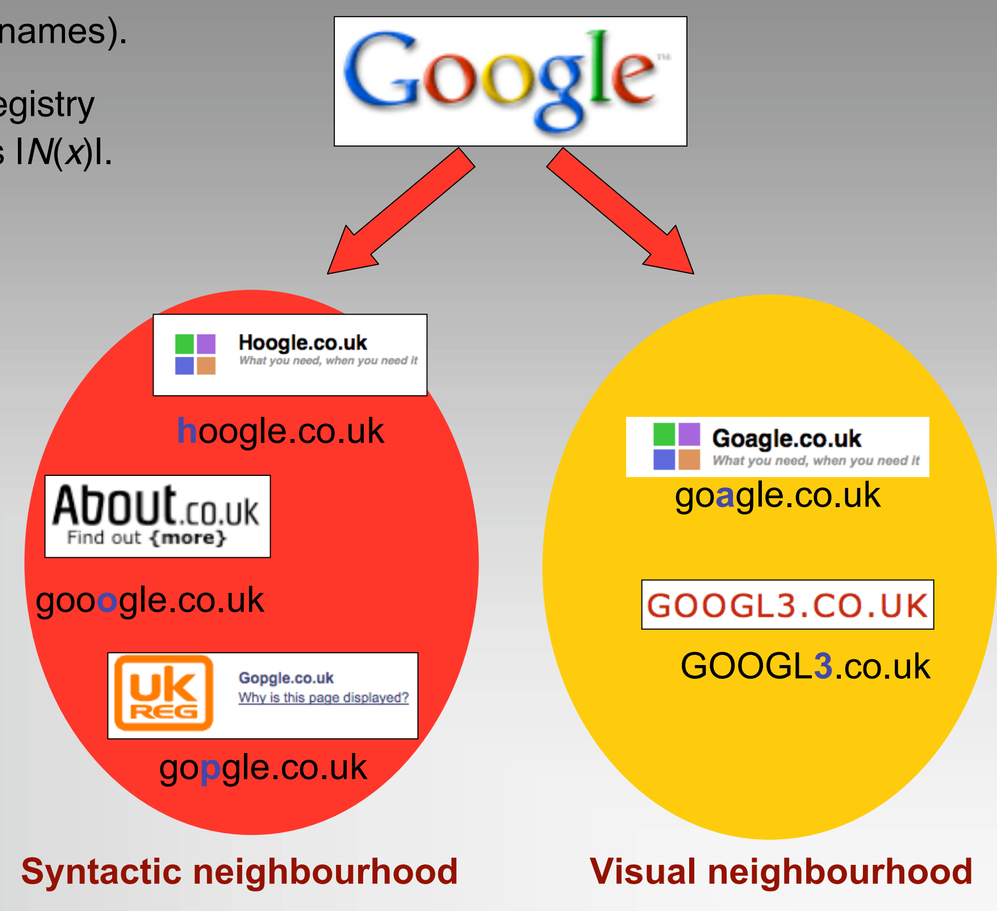
INTRODUCTION

Typosquatting is the practice of registering a domain name with the intent to confuse it with the name of a trademark or a famous domain name (**trademark infringement** on domain names).

The **neighbourhood $N_d(x)$** of a domain name x is the set of all domain names in the registry whose distance $d()$ from x is lower than a threshold th_d and we denote its cardinality as $|N(x)|$.

A **distance function $d()$** is a transformation in the space of domain names.

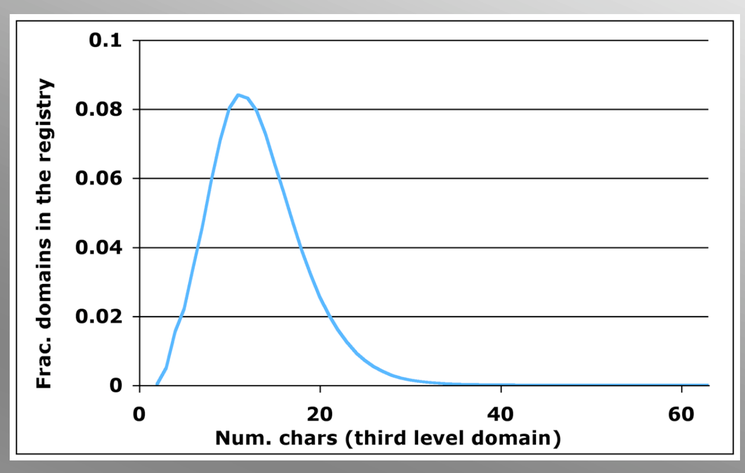




Syntactic neighbourhood **Visual neighbourhood**

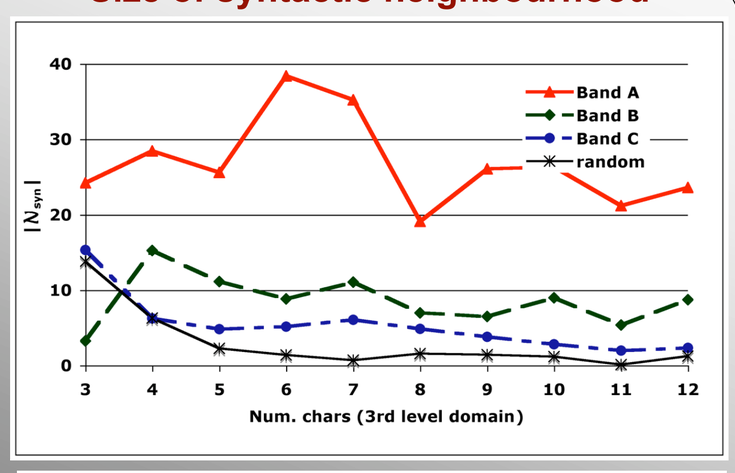
EXPERIMENTS

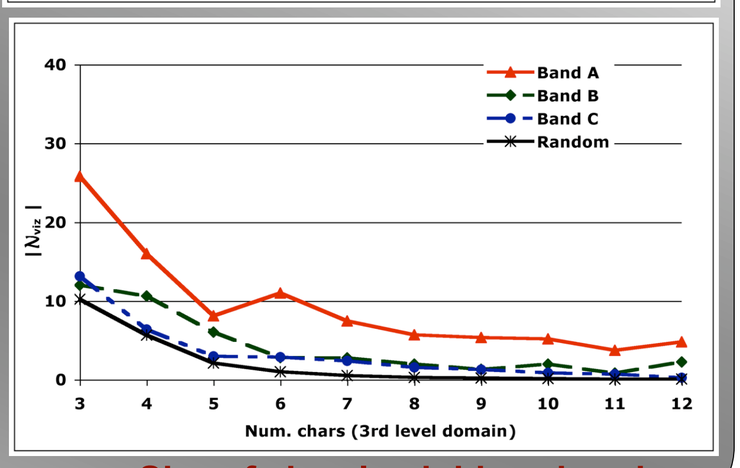
- Analysed the **'.co.uk' registry** (March 2008)
  - **Length** refers to the third-level label
- Bands of popularity (source NetCraft.com)
  - **Band A**: domains with ranking in [1,100]
  - **Band B**: domains with ranking in [101,500]
  - **Band C**: domains with ranking in [501,1000]
  - **Random**: 100 randomly chosen domains



Distribution of lengths of '.co.uk' domains

Size of syntactic neighbourhood





Size of visual neighbourhood

MAIN RESULTS

- Reliability of the results increase with the length of domain names
  - We suggest to not consider the statistics for 3-4 character domain names
- Syntactic and visual neighbourhood
  - High correlation between **popularity** and **likelihood of being typosquatted**
- Syntactic neighbourhood does not dependent on *length*(x)

DISCUSSION

- Social networks and Web 2.0 applications rely on a loose concept of **digital identities** based on simple identifiers
  - Which are the risks associated to typo-squatting in these scenarios?
  - How would they benefit from a cluster-based analysis?
- We should move towards a concept of digital identity is associated to a **cluster in some identifier space**
  - The cluster comprises a **primary identifier** and **secondary identifiers** that are mapped to it

REFERENCES

[1] A. Banerjee, D. Barman, M. Faloutsos, and L. N. Bhuyan. *Cyber-fraud is one typo away*. In INFOCOM, 2008.
[2] P. E. Black. *Visual similarity of top-level domains*. http://hissa.nist.gov/black/GTLD/ (valid Mar. 2009).
[3] McAfee. *What's in a name: The state of typo-squatting 2007*. http://www.siteadvisor.com/studies/typo_squatters_nov2007.html (valid Mar. 2009).
[4] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels. *Strider typo-patrol: Discovery and analysis of systematic typo-squatting*. In SRUTI, 2006.