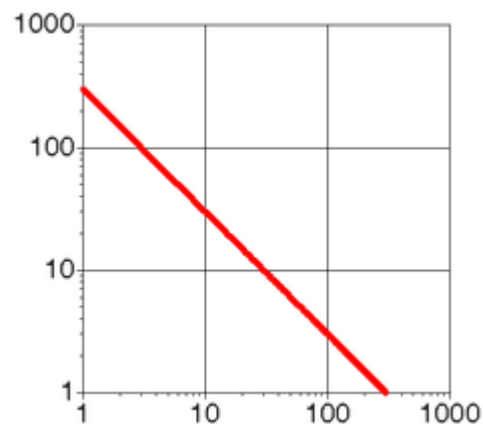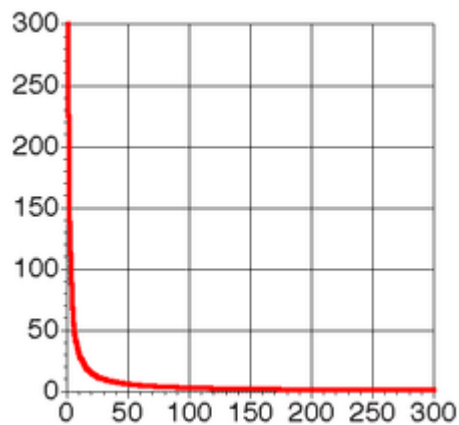# blogosphere
quantitative elements

Joannès Vermorel

Tuesday, October 17th 2006

*(disclaimer: I am not a specialist)*

Blog demographics

- October 2006
    - 1.1 billion web users.
    - 1.6 million blog post per day.
    - 0.1% of daily posters.
- Zipf curve phenomenon: $f(n) = a / n^s$

Blog demographics

- Can be interpreted as the 990-9-1 rule (lurker, occasional and regular).

- !!! Blog observations can be **super-biased**.

- An insignificant user percentage "rules".

- The **drooping tails**: there is not enough content on the web!

Blog perception

Joannès Vermorel
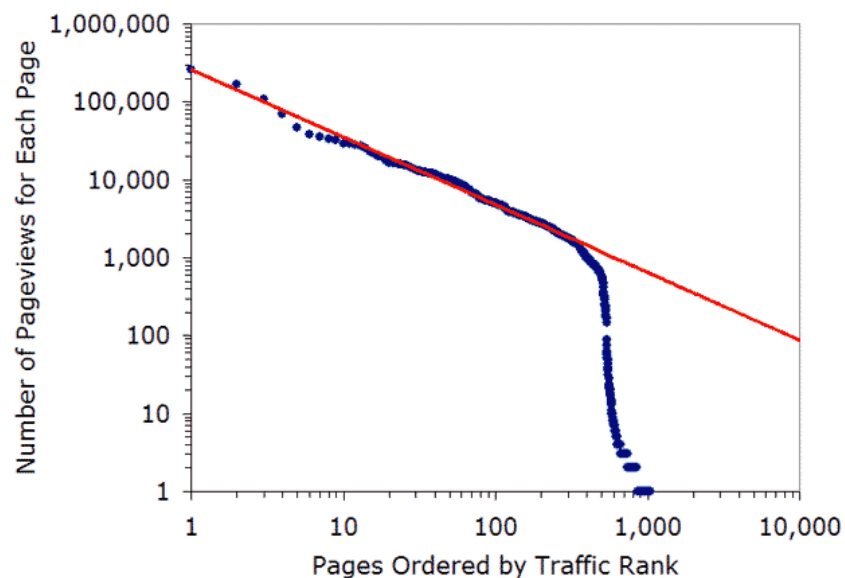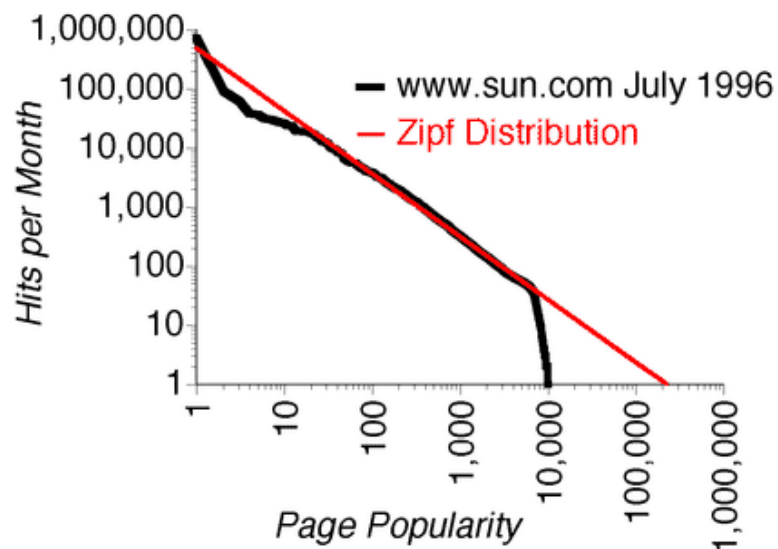
Blog perception

- Dominant reading pattern
  1. → movement.
  2. → movement.
  3. ↓ ↓ movement.

- Why?
  – Average design on the web.
  – Users follow mentally the average design.

- Conclusions
  – Two first paragraphs must carry the information.
 – Subheads, paragraphs and bullet points must be information-carrying words.

Joannes Vermorel

Blog navigation

- The navigation within the page follows the Fitts' Law
 $$T = a + b \log 2(D/W+1)$$
- The **user performance** (reading/navigating) follows an *approximate* Gaussian distribution.
- Very strong individual inequalities.

- Conclusions
  - Fast people are not that fast (physical limits)
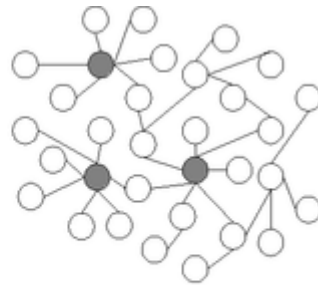 - Slow people can be really (really) slow: disable users
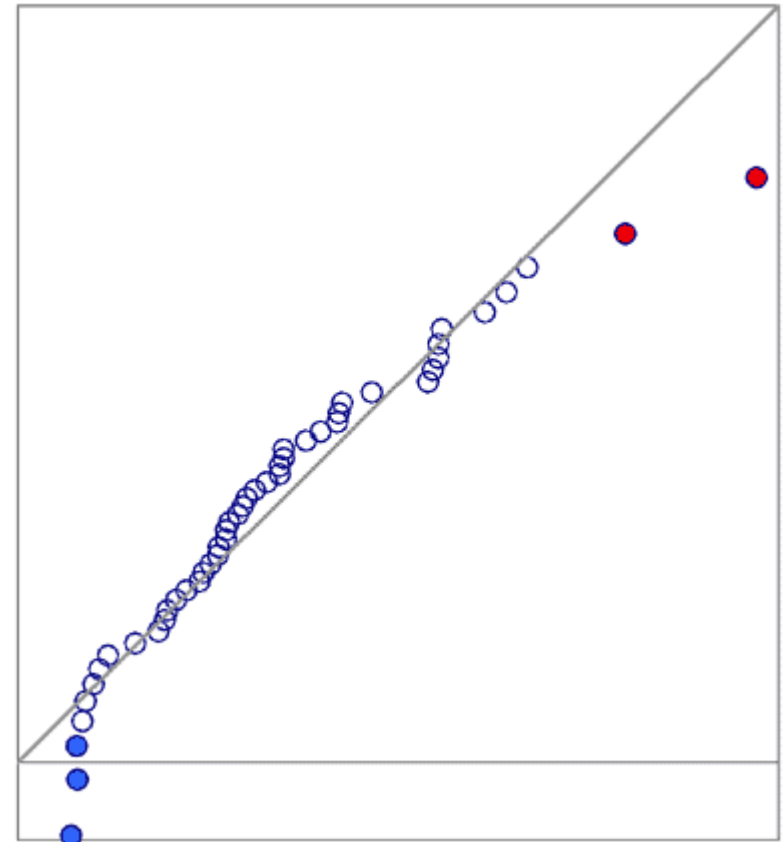
X: measured time Y: expected time
(assuming a Gaussian distribution)
The diagram is omitting total failures

Blog networks

- Network is defined by hyperlinks.

- Scale-free network
  - Power law connectivity
    $P(k) = a\,k^{-b}$
  - Very "Human" (like all self-organized networks)
  - Nearly optimal networks for information diffusion.
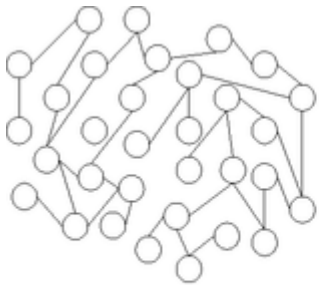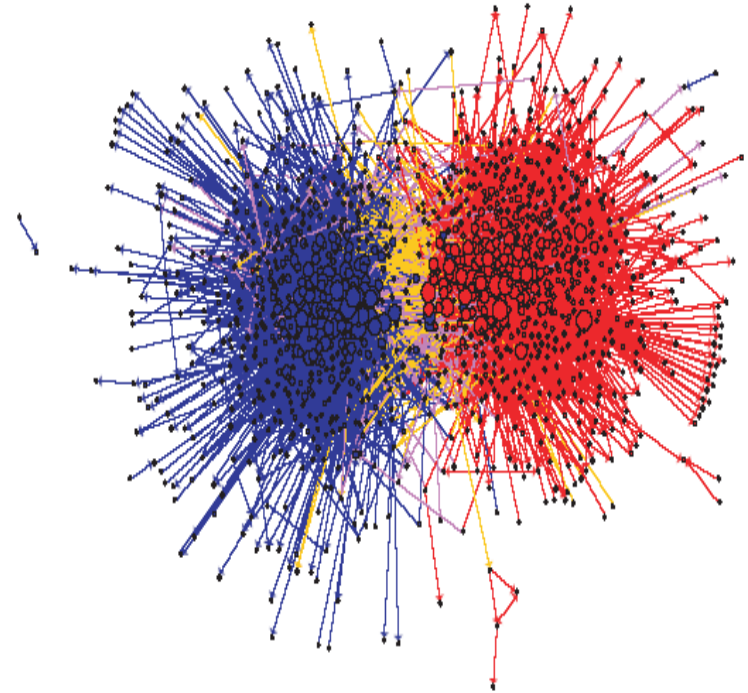

(a) Random network   (b) Scale-free network

Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

Blog networks

- Hyperlinks aren't everything.
- Other criterions include
  - Betweenness
  - Centrality (closeness, degree, eigenvector)
  - Centralization
  - Cohesion
  - Density
  - Radiality
  - Reach
  - Structural holes
  - …
- Mathematical tools → discover communities and complex social networks.

Joannes Vermorel

Signal to Noise ratio in blogs

- Noise is overwhelming
  - Splogs (+1M)
  - Zombie comments
  - Moronic comments
  - Flame wars
  - Godwin Law
  - Blog spam
  - …

- Not-so-intuitive ideas
  - Niche content can be of very high quality.
  - Junk content doesn't matter as long as your filters are powerful enough.
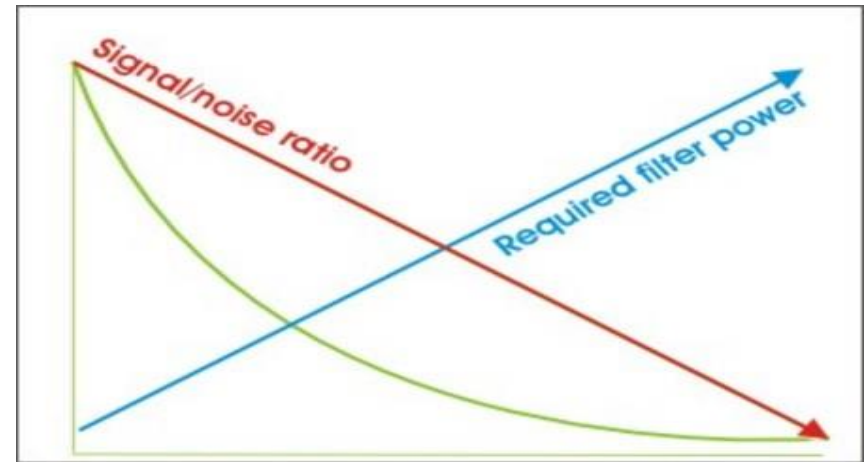
Signal is present in every part of the tail.

Conclusions

- Blogs are not a "mass" phenomenon (but it's definitively large).

- New methods to "study" blogs
  - Perception focused vs. the actual content.
  - Large networks study vs. case-based study.
  - Noise is an issue vs. Access is an issue.

- Still, a whole new set of paradigms are emerging.

Joannes Vermorel

References

- Behind the scene in the Blogsphere: http://www.umassd.edu/cmr/studies/cmrblogstudy.pdf
- Blog perception: http://www.useit.com/alertbox/reading_pattern.html
- Blog demographics: http://www.useit.com/alertbox/participation_inequality.html
- Bring tha noize! http://longtail.typepad.com/the_long_tail/2005/06/more_on_signals.html
- Democrat vs Republican graphs: http://www.willisms.com/
- Fit's Law: http://en.wikipedia.org/wiki/Fitts'_law
- Godwin Law: http://fr.wikipedia.org/wiki/Loi_de_Godwin
- Splogs: http://www.splogspot.com
- Usuability performance: http://www.useit.com/alertbox/outlier_performance.html
- Scale free graphs: http://en.wikipedia.org/wiki/Scale-free_network
- Zipf's Law: http://en.wikipedia.org/wiki/Zipf's_law

Joannès Vermorel