A Web Metrics of the Universities Mutual Impact: G-Factor revisited

8th International Conference on Webometrics, Informetrics & Scientomtrics & 13th COLLNET Meeting 2012

Seoul, October 23-26, 2012

Dmytro Filchenko

Ph.D, senior lecturer

Director of the Center for Webometrics & Web Marketing, Sumy State University, Ukraine

dm.filchenko@linet.sumdu.edu.ua

SUMY STATE UNIVERSITY
Welcomes cooperation

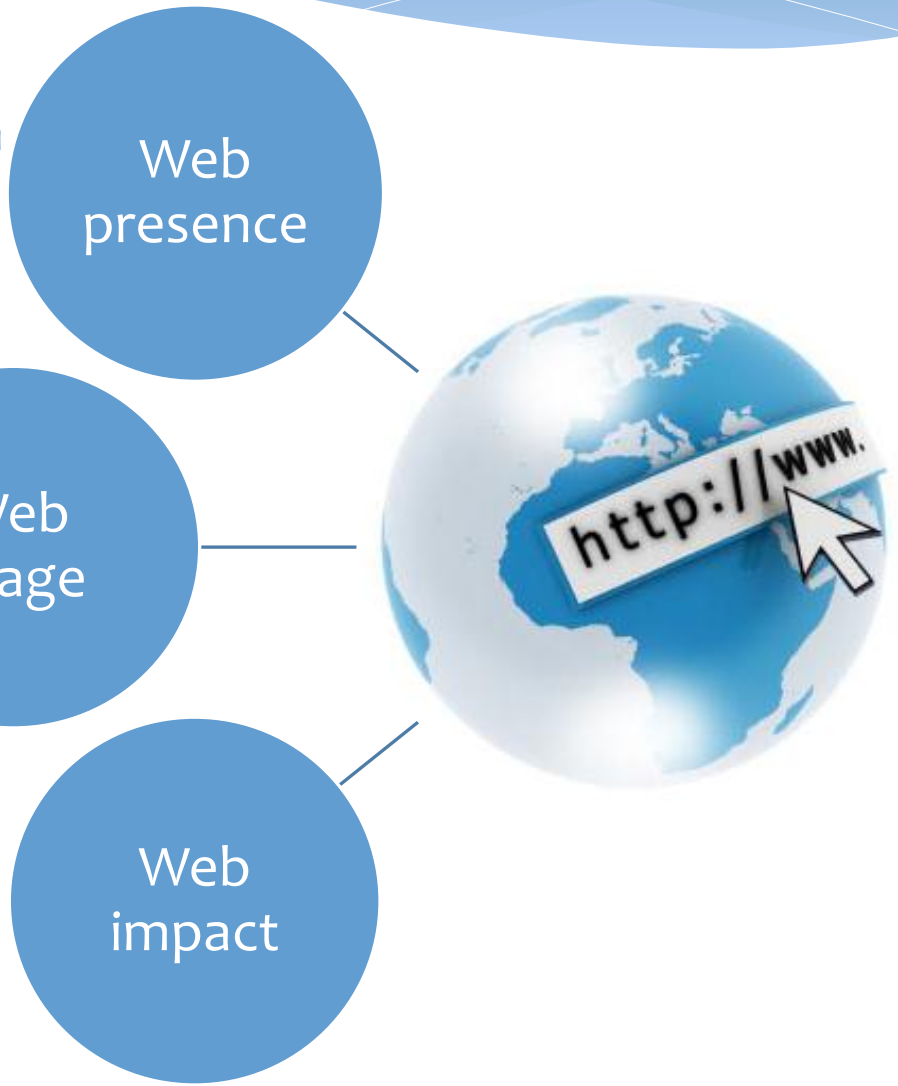
>>> web.sumdu.edu.ua

Web metrics of university performance

- **Metrics of indexation in search engines (web content mining techniques)**

- **Web traffic metrics (web usage mining techniques)**

- **Metrics of citation in Web (web structure mining techniques)**



University Webometric Models

Project	Concept of Web Metrics	Reliability
Ranking Web	Web presence metrics (50%) **Web impact metrics** (35%) Non-web impact metrics (15%)	Open methodology, contemporary (2004-current time)
G-Factor	**Web impact metrics** (100%)	Open methodology, outdated (2006)
Mike Tung's Ranking	Web presence metrics (adjustable weight) **Web impact metrics** (adjustable weight)	Open methodology, outdated (2006)
4INTERNATIONAL COLLEGES & UNIVERSITIES	**Web impact metrics** (unknown weight) Web usage metrics (unknown weight)	Methodology is closed, contemporary (2006-2012)
EduRoute	Web presence metrics (30%) **Web impact metrics** (70%)	Methodology is closed, contemporary (2009-2012)

Major universities impact web metrics

Number of backlinks (WR, EduRoute, Mike Tung's models)

- number of hyperlinks directed to the university web domain from all other web domains

Number of referring domains (WR, 4ICU models)

- number of all other web domains which have hyperlinks directed to the university web domain

Google's PageRank (4ICU model)

- the impact of the university web domain is determined by the impact of all other web domains that have a hyperlink to the university website

G-Factor metrics (G-factor model)

- number of hyperlinks directed to the university web domain from web domains of other universities

Why to introduce a new metrics?

One can see that:

- universities impact web metrics (*IM*) are all functions of universities backlinks (*BL*) and/or referred domains (RD):

$$IM = f(BL, RD)$$

- functions like

$$IM = BL, IM = RD \text{ or } IM = RD * RD$$

and etc. are too general and vulnerable to be robust estimators of universities impact in Web

- classical G-Factor metrics is also vulnerable (the possibility of universities colluding)
- Google's PageRank is robust, but (a) it is determined on the whole sample of webdomain indexed by Google and (b) it is given as a natural number from 1 to 10

New metrics introduction

The model

- G-Factor model
- Google's PageRank model

Let us mix them together?

Data sources

- Google does not provide backlinks no more
- nor Yahoo! SiteExplorer

Where to get the backlinks data?

'G' in G-Factor was actually stated for 'Google' as a major data source. We are going to find a new data source. But want to ground on Google's PageRank model. So let us call a new metrics an 'extended G-Factor'.

Extended G-Factor: the model

Google's Page Rank model:

$$\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}}(\alpha \mathrm{A}' + (\mathbf{1} - \alpha)\mathbf{e}\mathbf{e}^{\mathrm{T}}/n) \quad (1)$$

- n is the number of universities in the sample
- $\boldsymbol{\pi} \in [0, 1]$ is the column of n Page Ranks
- $\mathbf{e}$ is nx1 vector of 1's
- T is a matrix transposition symbol
- $\mathbf{A}$ is stochastic adjacency matrix for a given web graph $G = (U, H)$

$$a_{ij} = \begin{cases} \dfrac{1}{O_i}, \text{if } (i,j) \in H, \\ 0 \ - \text{otherwise.} \end{cases} \quad (2)$$

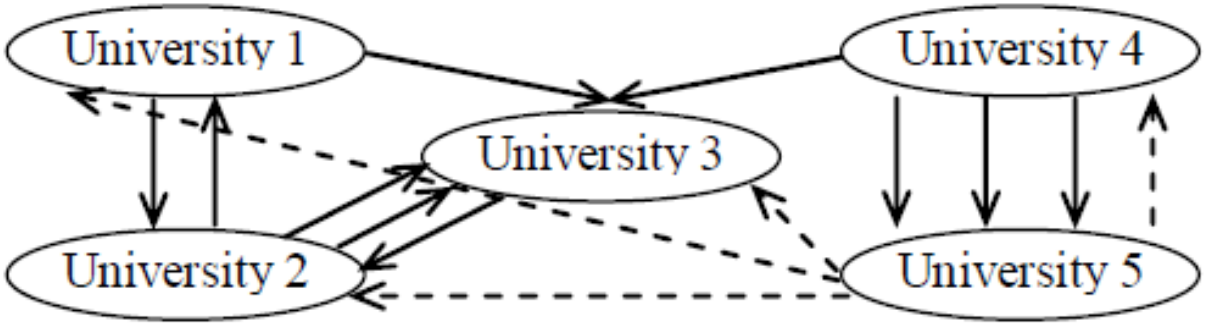
- α is a parameter, which sensitivity may be represented by the following inequality:

$$\left|\frac{d\pi_i(\alpha)}{d\alpha}\right| \leq \frac{1}{1-\alpha} \quad \forall\ \alpha{=}1,n \quad (3)$$

Extended G-Factor: the model (continued)

G-Factor model:

- O_i is the number of hyperlinks from the i-th university web domain to all the rest university web domain in the sample U



Reasoning for backlinks provider

- ✓ Majestic SEO has the largest index of its kind publicly available (it was compared by us to Yahoo! SiteExplorer and other contemporary site exploring tool)
- ✓ Index updates several times a day
- ✓ User-friendly user interface
- ✓ APIs are available
- ✓ Powerful own metrics (Citation & Trust Flow)
- ✓ Variety of analytical tools (Keyword Checker, Clique Hunter, Comparator etc.)
- ✓ The largest university rankings (WR, 4ICU) use Majestic SEO as a backlinks provider as well

Algorithm of extended G-Factor

Step 1
- Extracting raw data (list of backlinks) for each of *n* web domains from the sample by Majestic SEO APIs

Step 2
- Text parsing & raw data analysis

Step 3
- Computation of adjacency matrix A' for a given web graph

Step 4
- Computation of *n* Page Ranks (for example, by power iteration technique)

Step 5
- Results validation & dissemination

Numerical experiment and outcomes

Sample: 324 Ukrainian educational institutions (WR directory, January 2012)

- High interaction level between universities by specialization
- High interaction level between large universities and small specialized schools
- Bias towards regional interaction
- The hyperlink are mostly not specific (universities are not in favor of citing each other)

Benefits of extended G-Factor

- Create an incentive for universities to explore the content of each other via mutual citation
- Content citation results in content enhancement
- Content exploring means mutual information retrieval and its evaluation (peer-review)
- All this actually are prerequisites of starting collaboration

Future dissemination of extended G-Factor

Sumy Web Rank: The First Ukrainian Ranking Portal (ranking.sumdu.edu.ua)

Integration in Ukrainian national university rankings

Expansion on cross-counties samples

Integration in global rankings

감사합니다!
Thank you!
Дякую!

Dmytro Filchenko

Ph.D, prof. assistant

dm.filchenko@linet.sumdu.edu.ua

Center for Webometrics & Web Marketing,
Sumy State University, Ukraine

SUMY STATE UNIVERSITY
Welcomes cooperation

>>> www.sumdu.edu.ua