Text-mining of the IGF2014 Day 2: An Overview

We talked about many different things on Day 2...

IGF2014 Day 2

Figure 1a. IGF2014, Day 2: the most frequently used words and phrases.

... and our talk was quite complicated:

IGF2014 Day 2
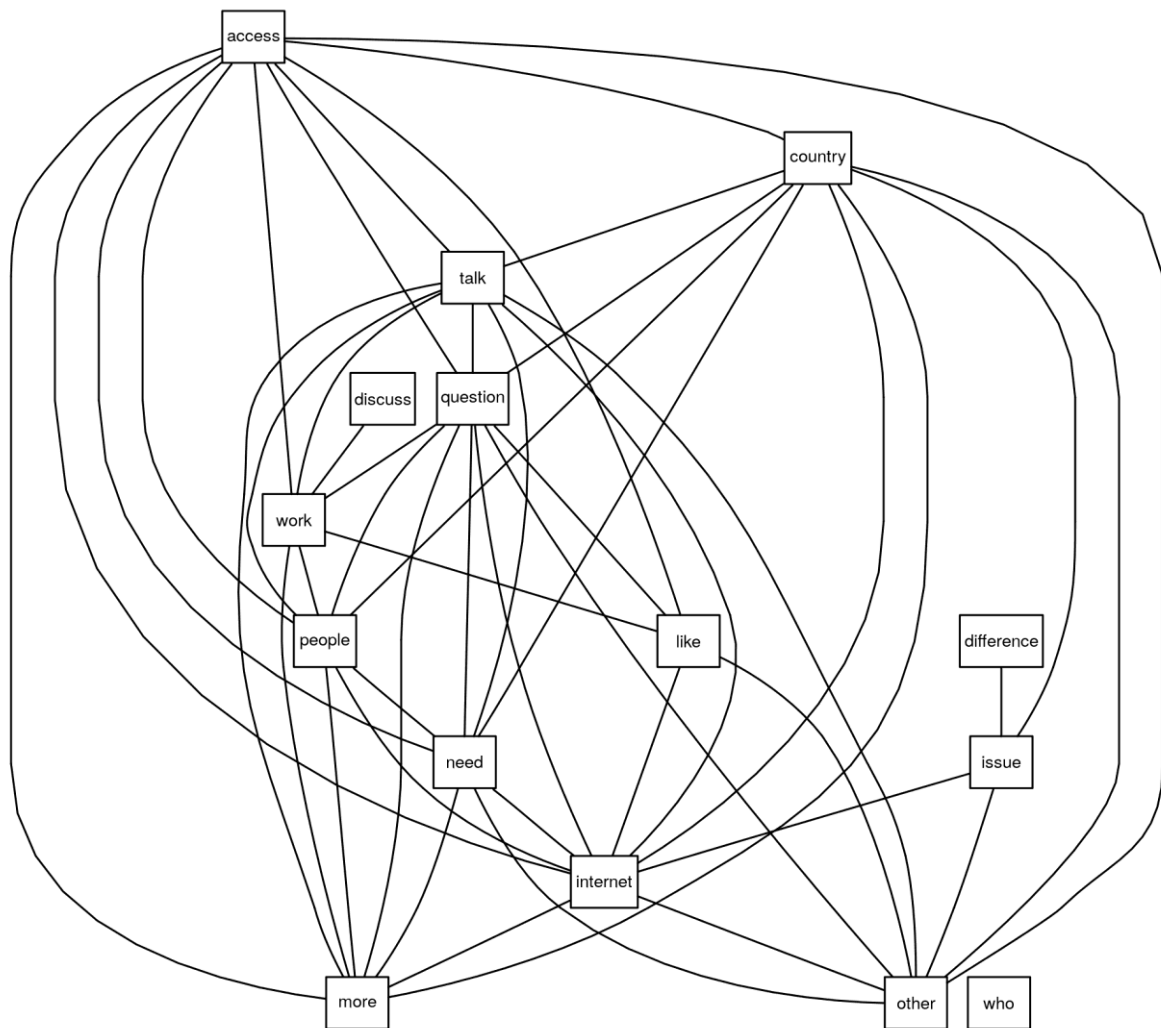## associations >= 0.5

Figure 1b. IGF2014, Day 2: Associations between 15 most frequently used words and phrases.

However, when we asked our computing machinery to address only IG-specific words and phrases, interesting structures emerged:
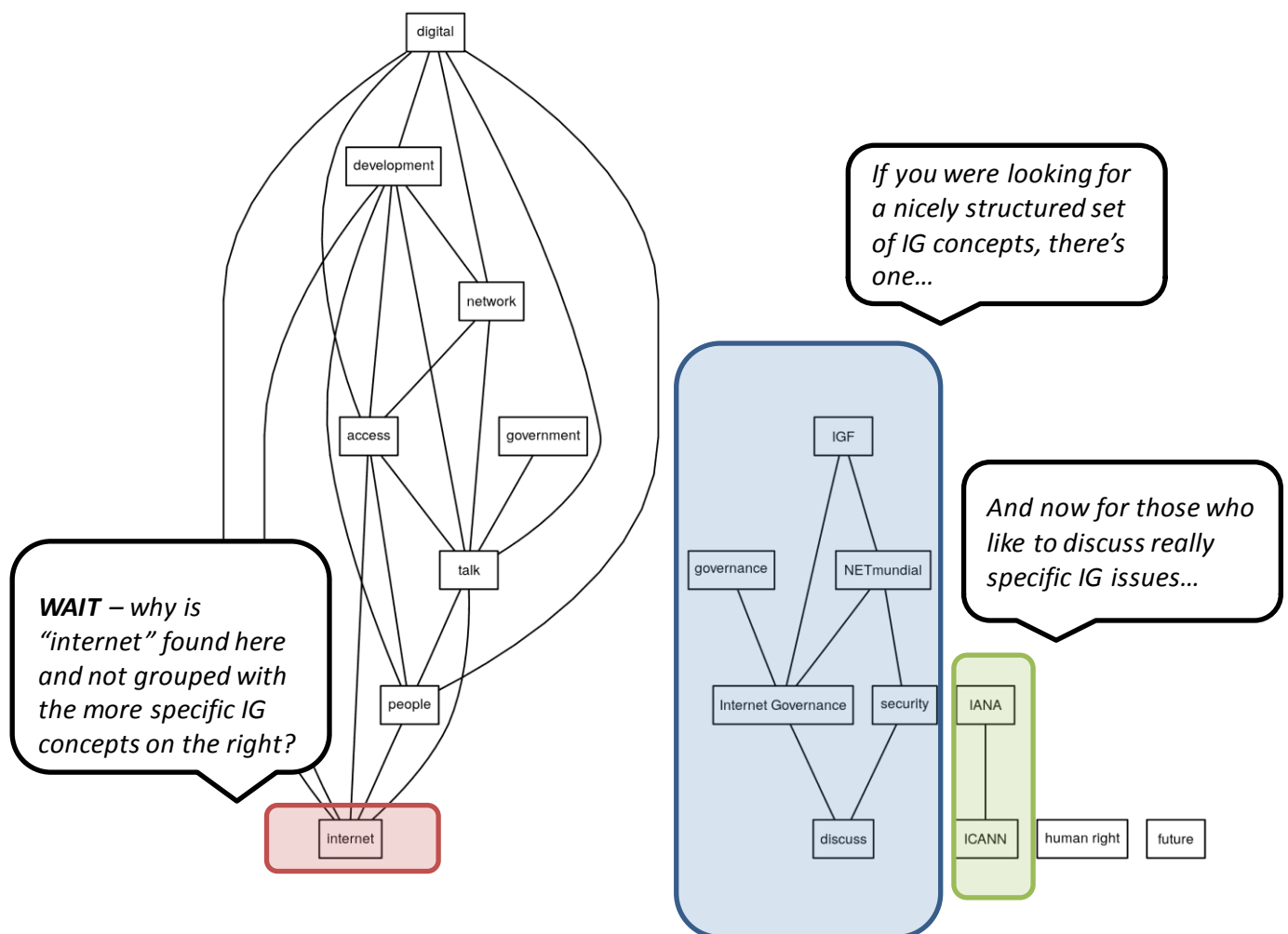

Figure 2. IGF2014, Day 2: associations between some IG specific words and phrases.

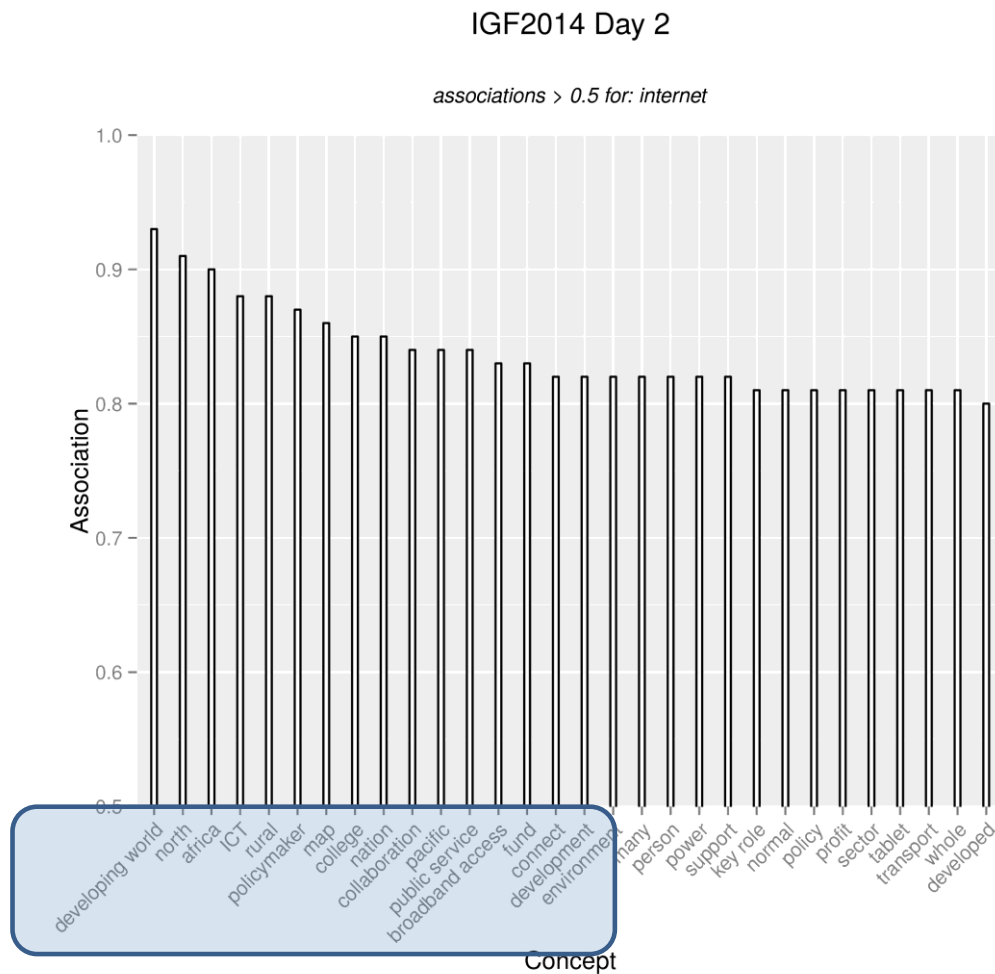Let's have a look at how the keyword **Internet** was associated with other words and phrases in Day 2:

Figure 3a. IGF2014, Day 2: Associations for **Internet.** Interesting: (a) several geographical references are present; (b) the 'development discourse' is present; (c) many non-IG, non-technical words are present.

And what about the phrase **Internet Governance Forum,** certainly among the central players in the IG discourse?
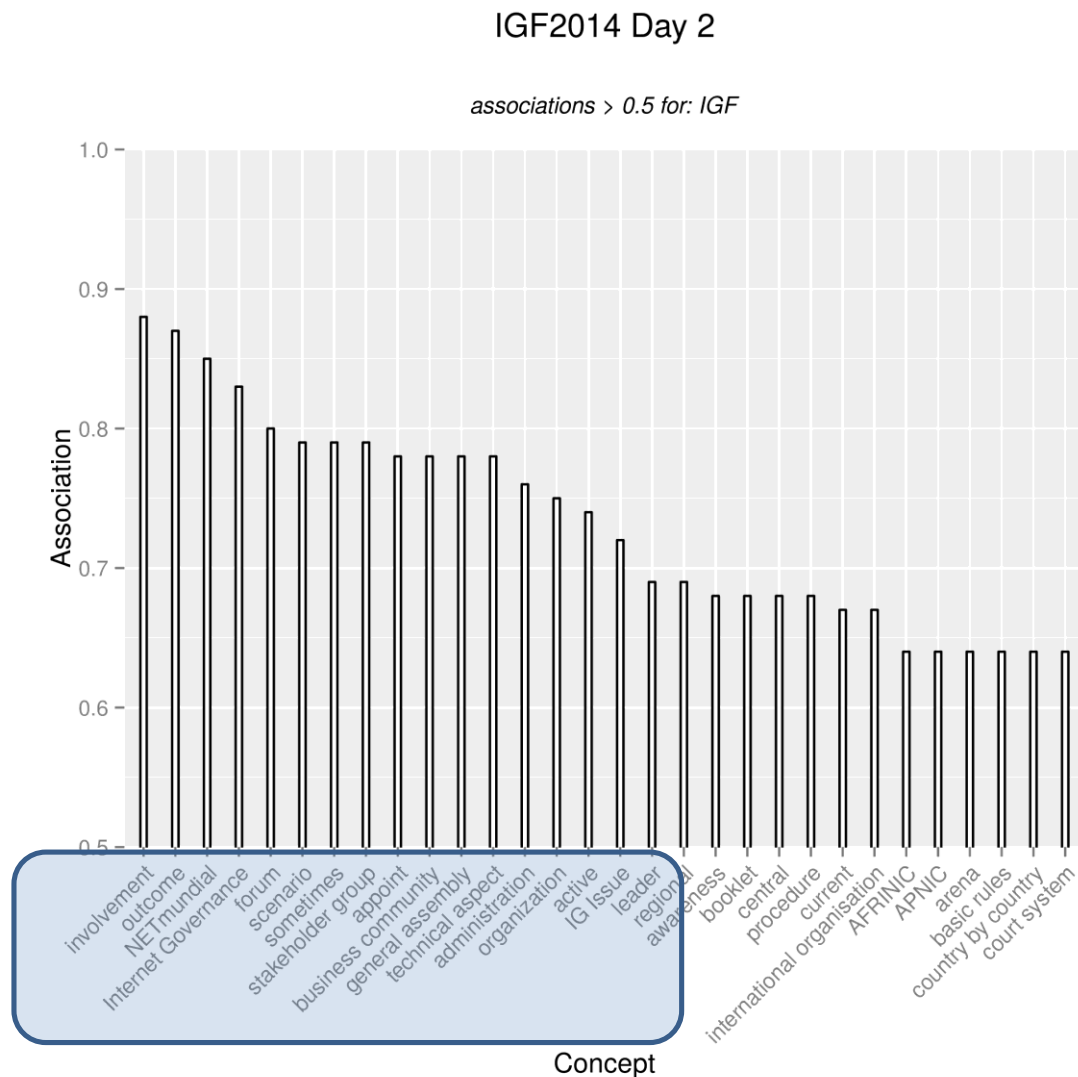

IGF2014 Day 2

associations > 0.5 for: IGF

Figure 3b. IGF2014, Day 2: Associations for **IGF.**

Judging by the highest number of associates for the abbreviation **IGF**, it seems as if the IGF talk is beginning to be about itself: *involvement, outcome, NETmundial, IG, forum, scenario, stakeholder group, appointments, businesses, general assembly, technical aspect, administration, organization, active, IG issue, leader, etc.*

After inspecting the associations for **Internet Governance**, a similar pattern emerged; the strongest associates were *IG issue, governance process, W3C, involvement, forum, IGF, organisation, IG research, scenario, internal, outcome, parliament, stakeholder group, etc.*
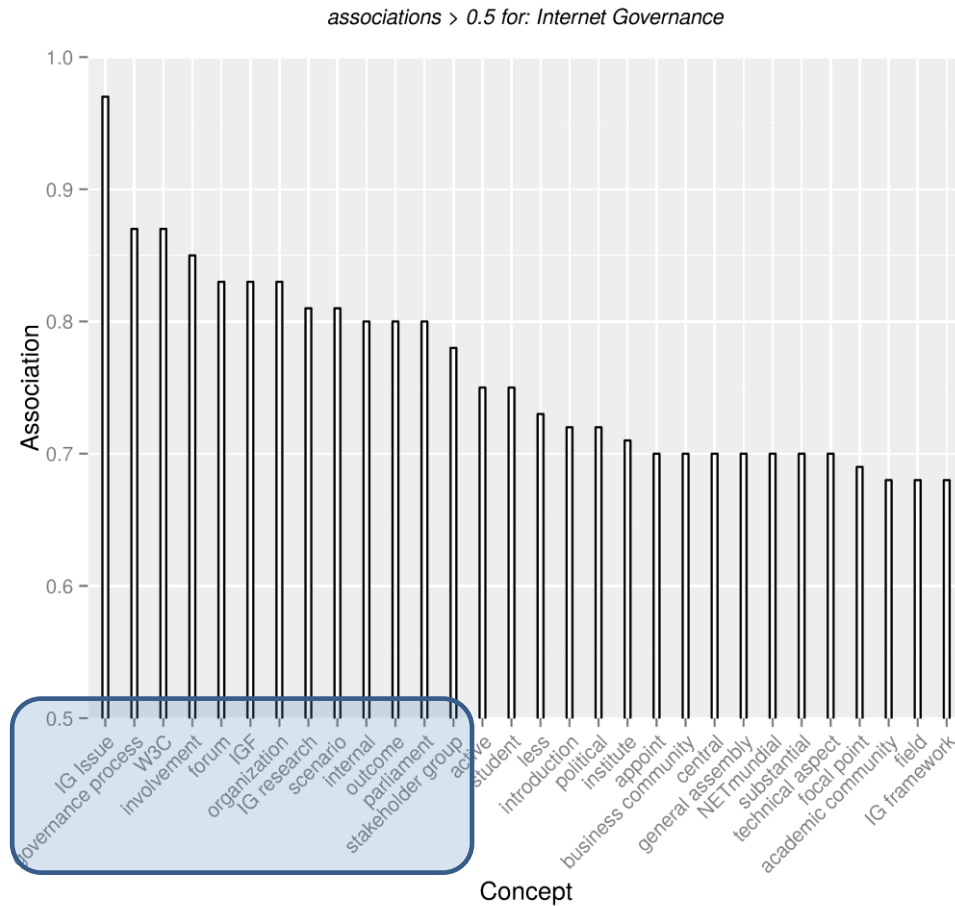

Figure 3c. IGF2014, Day 2: Associations for **Internet governance.**

During the IGF2014 Opening Session, the word with the highest association with the word Internet was: *people*, followed by *developing*, and *broadband access*.
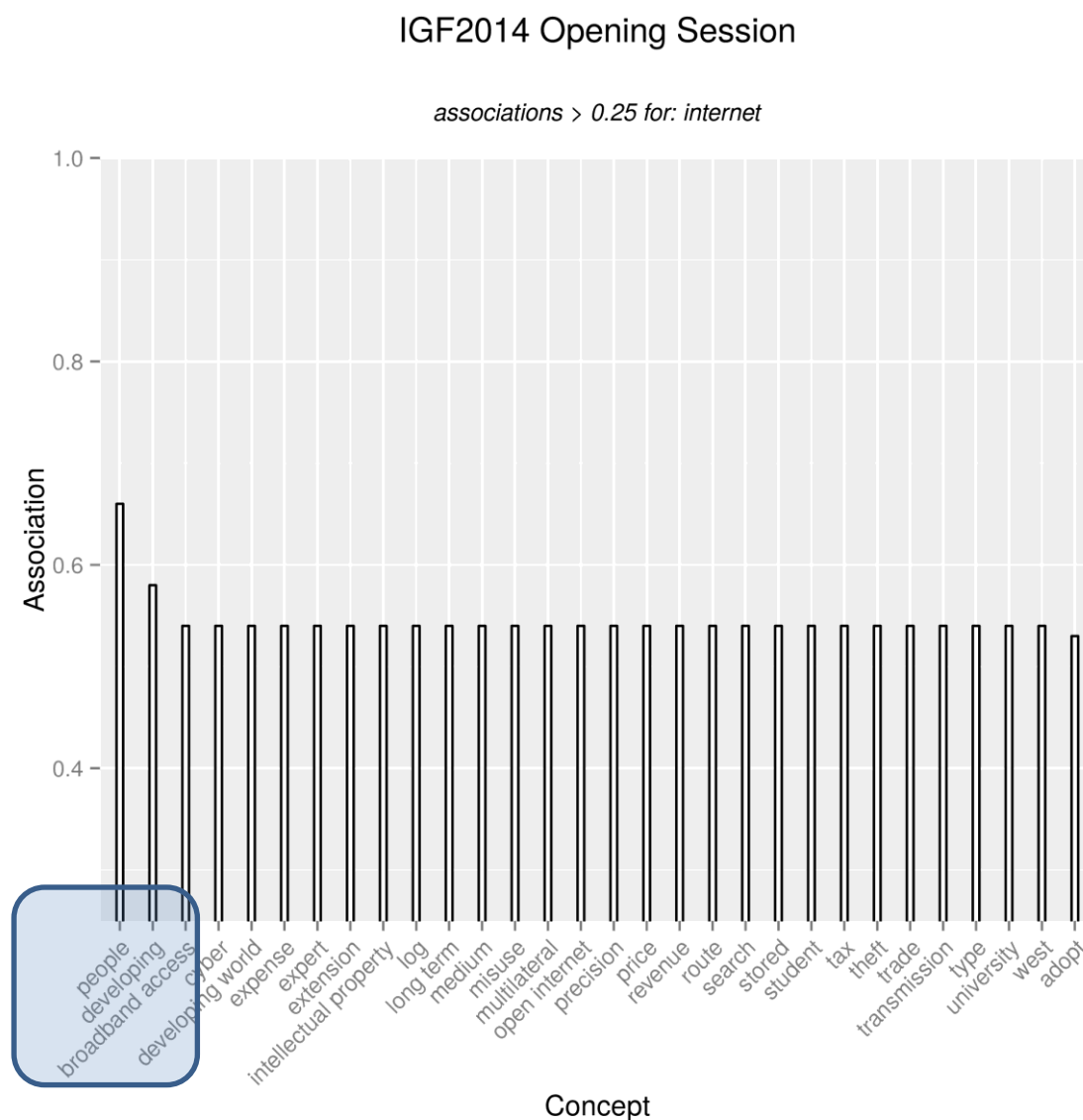

Figure 4. IGF2014, Opening Session: Associations for **Internet.**

In other words, 'IGF talk' and 'IG talk' are becoming somewhat self-referential, encompassing highly associated and highly specific words and phrases, while 'Internet talk' is more referential, encompassing associations with somewhat simpler, more concrete words and phrases.

The two – IGF/IG discourse and Internet discourse – must go together. Moreover, they must fit. That is what the IGF, as a discussion forum, is meant to be about: allowing participants to exercise abstract schemas, specific knowledge, complex policy-related concepts, and the finesse of diplomatic language, while never making them lose sight of the real, referential world of the Internet and information society needs and developments.

In the scope of contemporary text-mining methods and approaches, the one we have taken here is rather classic. Its simplicity, compared to the resulting, rather straightforward, empirical discovery of two conceptual realms that govern the IG discourse, illustrates only the nearest frontiers of contemporary data science applied to the study of governance processes. Stay with us for further results and analyses.

References

Text mining procedures used to produce Figures 1 – 4 were developed in R, using and extending the functionality of the tm() package.

The R Project for Statistical Computing
http://www.r-project.org/

tm: Text Mining Package
http://cran.r-project.org/web/packages/tm/index.html

Feinerer, I., Hornik, H. & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, Vol. 25, Issue 5, Mar 2008.
http://www.jstatsoft.org/v25/i05/paper

Statistical analysis and R programming:

Goran S. Milovanović, PhD
DiploFoundation, *Belgrade*
Twitter: @GSMilovanovic