

Maps of influence and interactions of infectious and cancer diseases from Wikipedia networks

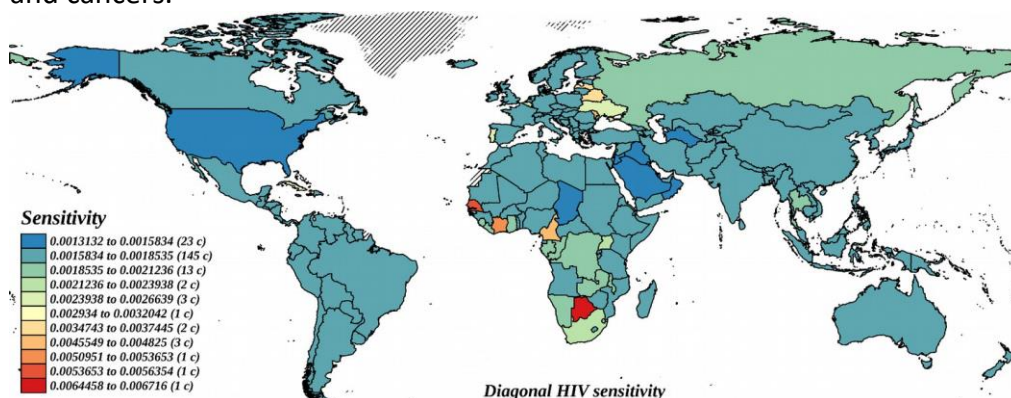
Guillaume Rollin¹, Jose Lages¹, Dima L Shepelyansky²

Institut UTINAM -Observatoire de Besançon, Université de Franche-Comté, Besançon, France, Université de Toulouse, UPS, CNRS, Laboratoire de Physique Théorique, Toulouse, France²

We use the Google matrix analysis of the English Wikipedia articles network to infer influence of diseases on countries and to infer interactions between diseases and drugs. Nowadays, the free online encyclopedia supersedes old ones such as Encyclopedia Britanica in volume and in quality of articles devoted to scientific topics [1]. For instance, articles devoted to biomolecules are actively maintained by scholars of the domain [2,3]. The Google matrix analysis, associated to the PageRank algorithm [4] initially invented by Sergey Brin and Larry Page to efficiently rank pages of the WWW, allows to probe the network of Wikipedia articles in order to measure the influence of every articles. Recently, using parallels with quantum scattering in nuclear physics, mesoscopic physics, and quantum chaos, we have suggested a novel methodology, called Googlomics [5], for the structural analysis of directed biological networks using spectral analysis of their Google matrices. Moreover we used the new reduced Google matrix method which allows to infer hidden interactions between a set of nodes selected from a huge network. We successfully applied this method for the regulatory biological networks and demonstrate how its computation allows inferring hidden causal relations between the members of a signaling pathway or a functionally related group of genes [5].

Here we study diseases through their entries in the English Wikipedia edition. In particular we focus:

- on the set of articles devoted to infectious diseases and the set of articles devoted to countries, in order to measure the influence of different diseases on different countries. Also the reduced network of infectious diseases is built showing direct and hidden relations between diseases,
- on the set of articles devoted to cancer types and the set of articles devoted to drugs for cancer treatment, in order to possibly measure hidden interactions between drugs and cancers.



PageRank sensitivity of countries to the variation of the HIV→country link in 2017 English Wikipedia.

References:

- [1] J. Giles, Internet encyclopaedias go head to head, *Nature* 438, 900–901 (2005)
- [2] D. Butler, Publish in Wikipedia or perish, published online 16 December 2008, *Nature*
- [3] E. Callaway, No rest for the bio-wikis, *Nature* 468, 359-360 (2010)

- [4] S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* 30, 107 (1998)
- [5] J. Lages, D. L. Shepelyansky, A. Zinovyev, Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks, *PLoS ONE* 13(1): e0190812 (2018)