IEEE Access®

Multidisciplinary : Rapid Review : Open Access Journal

World Influence of Infectious Diseases from Wikipedia Network Analysis

GUILLAUME ROLLIN[1], JOSÉ LAGES[1], AND DIMA L. SHEPELYANSKY[2]

[1]Institut UTINAM, Observatoire des Sciences de l'Univers THETA, CNRS, Université de Bourgogne Franche Comté, 25030 Besançon, France
[2]Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

Corresponding author: José Lages (e-mail: jose.lages@utinam.cnrs.fr).

ABSTRACT We consider the network of 5 416 537 articles of English Wikipedia extracted in 2017. Using the recent reduced Google matrix (REGOMAX) method we construct the reduced network of 230 articles (nodes) of infectious diseases and 195 articles of world countries. This method generates the reduced directed network between all 425 nodes taking into account all direct and indirect links with pathways via the huge global network. PageRank and CheiRank algorithms are used to determine the most influential diseases with the top PageRank diseases being Tuberculosis, HIV/AIDS and Malaria. From the reduced Google matrix we determine the sensitivity of world countries to specific diseases integrating their influence over all their history including the times of ancient Egyptian mummies. The obtained results are compared with the World Health Organization (WHO) data demonstrating that the Wikipedia network analysis provides reliable results with up to about 80 percent overlap between WHO and REGOMAX analyses.

INDEX TERMS Infectious diseases, Wikipedia, Markov processes, Ranking (statistics), Complex networks, Data mining.

I. INTRODUCTION

INFECTIOUS diseases account for about 1 in 4 deaths worldwide, including approximately two-thirds of all deaths among children younger than age 5 [1]. Thus the understanding of the world influence of infectious diseases is an important challenge. Here we apply the mathematical statistical methods originated from computer and network sciences using the PageRank and other Google matrix algorithms, developed at the early stage of search engines development [2], [3], and used since to analyze various complex networks (see e.g. [4] for a review and references therein). These methods are applied to English Wikipedia edition which is considered as a directed network generated by hyperlinks (citations) between articles (nodes). Nowadays, the free online encyclopedia supersedes old ones such as Encyclopaedia britannica [5] in volume and in quality of articles devoted to scientific topics [6]. For instance, Wikipedia articles devoted to biomolecules are actively maintained by scholars of the domain [7], [8]. The academic analysis of information contained by Wikipedia finds more and more applications as reviewed in [9], [10].

The Google matrix analysis, associated to the PageRank algorithm, initially invented by Brin and Page to efficiently rank pages of the World Wide Web [2], allows to probe the network of Wikipedia articles in order to measure the influence of every articles. The efficiency of this approach for Wikipedia networks has been demonstrated by ranking historical figures on a scale of 35 centuries of human history and by ranking world universities [11]–[15]. This approach produced also reliable results for the world trade during last 50 years reported by the UN COMTRADE database and other directed networks [16].

Recently, the reduced Google matrix method (REGOMAX) has been proposed using parallels with quantum scattering in nuclear physics, mesoscopic physics, and quantum chaos [17], [18]. This method allows to infer hidden interactions between a set of n_r nodes selected from a huge network taking into account all indirect pathways between these n_r nodes via the huge remaining part of the network. The efficient applications of this approach to the global biological molecular networks and their signaling pathways are demonstrated in [19]. In this case, the nodes of the complex network are the thousands of considered proteins, the links are the causal relations between proteins, and the n_r

nodes are e.g. proteins involved in a given signaling pathway. In addition to direct links (causal relations) among the n_r proteins obtained from literature, the REGOMAX method allows to find possible indirect interactions between the n_r proteins due to the embedding of the subnetwork of these n_r proteins into the global complex network of proteins. Also, the efficiency of REGOMAX has been demonstrated for the analyses of world terror networks [20], of geopolitical relations between countries [21], and of influence between world painters [22] from Wikipedia networks.

In this work we use REGOMAX method to investigate the world influence and importance of infectious diseases constructing the reduced Google matrix from English Wikipedia network with all infectious diseases and world countries listed there. This is an additional approach to the models of epidemic spreading on complex networks (see [23] for a survey and references therein) or to the estimates of infectious disease prevalence using Wikipedia usage [24].

The paper is constructed as follows: the data sets and methods are described in Section II, Results and Discussion are presented in Section III and Conclusions are given in Section IV; Appendix contains Tables 1, 2, 3, 4, 5; additional data are presented at [25].

II. DESCRIPTION OF DATA SETS AND METHODS
A. ENGLISH WIKIPEDIA EDITION NETWORK
We consider the network of articles extracted in May 2017 from the English language edition of Wikipedia. This complex network, available at [26] and already used in previous studies [13], [20]–[22], contains $N = 5\,416\,537$ articles (nodes) connected through $n_l = 122\,232\,932$ hyperlinks. From this data set we extract the $n_d = 230$ articles devoted to infectious diseases (see Tab. 1, Tab. 2) and the $n_c = 195$ articles devoted to countries (sovereign states, see Tab. 3). Both the list of infectious diseases and the list of sovereign states of 2017 are taken from Wikipedia [27], [28]. Thus the size of the reduced Google matrix is $n_r = n_d + n_c = 425$. This subset of n_r articles is embedded in the global Wikipedia network with N nodes. All data sets are available at [25].

B. GOOGLE MATRIX CONSTRUCTION
The construction of Google matrix G is described in detail in [2], [3], [16]. In short, the Google matrix G is constructed from the adjacency matrix A_{ij} with elements 1 if article (node) j points to article (node) i and zero otherwise. The Google matrix elements take the standard form $G_{ij} = \alpha S_{ij} + (1 - \alpha)/N$ [2], [3], [16], where S is the matrix of Markov transitions with elements $S_{ij} = A_{ij}/k_{out}(j)$. Here $k_{out}(j) = \sum_{i=1}^{N} A_{ij} \neq 0$ is the out-degree of node j (number of outgoing links) and $S_{ij} = 1/N$ if j has no outgoing links (dangling node). The parameter $0 < \alpha < 1$ is the damping factor. For a random surfer, jumping from one node to another, it determines the probability $(1-\alpha)$ to jump to any node; below we use the standard value $\alpha = 0.85$ [3].

The right eigenvector of G satisfies the equation $GP = \lambda P$ with the unit eigenvalue $\lambda = 1$. It gives the PageRank probabilities $P(j)$ to find a random surfer on a node j and has positive elements ($\sum_j P(j) = 1$). All nodes can be ordered by decreasing probability P numbered by PageRank index $K = 1, 2, ...N$ with a maximal probability at $K = 1$ and minimal at $K = N$. The numerical computation of $P(j)$ is efficiently done with the PageRank algorithm described in [2], [3].

It is also useful to consider the network with inverted direction of links. After inversion the Google matrix G^* is constructed within the same procedure with $G^*P^* = P^*$. This matrix has its own PageRank vector $P^*(j)$ called CheiRank [29] (see also [11], [16]). Its probability values can be again ordered in a decreasing order with CheiRank index K^* with highest P^* at $K^* = 1$ and smallest at $K^* = N$.

PageRank algorithm measures the influence of nodes, i.e. more a node is pointed by influential nodes, more it is influential. Conversely, CheiRank algorithm measures the communicative ability of nodes, i.e. more a node points towards communicative nodes, more it is communicative. On average, the high values of P (P^*) correspond to nodes with many ingoing (outgoing) links [16].

C. REDUCED GOOGLE MATRIX ANALYSIS
Reduced Google matrix is constructed for a selected subset of nodes (articles) following the method described in [17]–[19]. It is based on concepts of scattering theory used in different fields including mesoscopic and nuclear physics, and quantum chaos (see Refs. in [17]). It captures in a n_r-by-n_r Perron-Frobenius matrix the full contribution of direct and indirect interactions happening in the full Google matrix between the n_r nodes of interest. Also the PageRank probabilities of selected n_r nodes are the same as for the global network with N nodes, up to a constant multiplicative factor taking into account that the sum of PageRank probabilities over n_r nodes is unity. The elements of reduced matrix $G_R(i, j)$ can be interpreted as the probability for a random surfer starting at web-page j to arrive in web-page i using direct and indirect interactions. Indirect interactions refer to paths composed in part of web-pages different from the n_r ones of interest. The intermediate computation steps of G_R offer a decomposition of G_R into matrices that clearly distinguish direct from indirect interactions: $G_R = G_{rr} + G_{pr} + G_{qr}$ [17]. Here G_{rr} is given by the direct links between selected n_r nodes in the global G matrix with n nodes. In fact, G_{pr} is rather close to the matrix in which each column is given by the PageRank vector P_r, ensuring that PageRank probabilities of G_R are the same as for G (up to a constant multiplier). Thus G_{pr} doesn't provide much information about direct and indirect links between selected nodes. The component playing an interesting role is G_{qr}, which takes into account all indirect links between selected nodes appearing due to multiple paths via the global network nodes N (see [17]). The matrix $G_{qr} = G_{qrd} + G_{qrnd}$ has diagonal (G_{qrd}) and non-diagonal (G_{qrnd}) parts. Thus G_{qrnd} describes indirect interactions between nodes. The explicit formulas as well as the mathematical and numerical computation methods of all
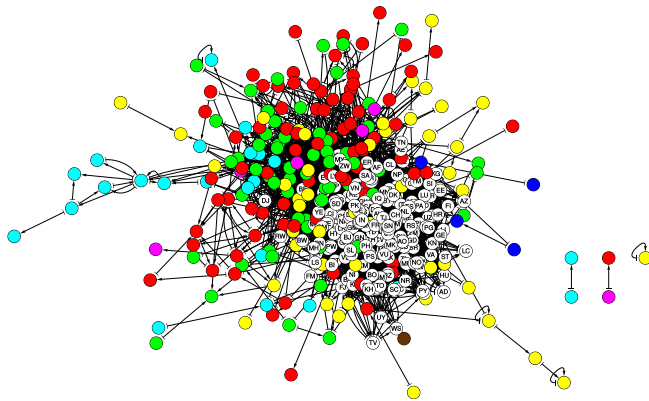
IEEE *Access*

FIGURE 1. Subnetwork of the 425 articles devoted to countries and infectious diseases in 2017 English Wikipedia. The bulk of the other Wikipedia articles is not shown. Articles devoted to countries are presented by empty nodes with country codes (see Tab. 3). Articles devoted to infectious diseases are presented by colored nodes with the following color code: bacterial diseases (red), viral diseases (green), parasitic diseases (yellow), fungal diseases (cyan), prionic diseases (blue), diseases with multiple origins (magenta), and other kind of diseases (brown) (see Tab. 1. Tab. 2). Network drawn with Cytoscape [30].
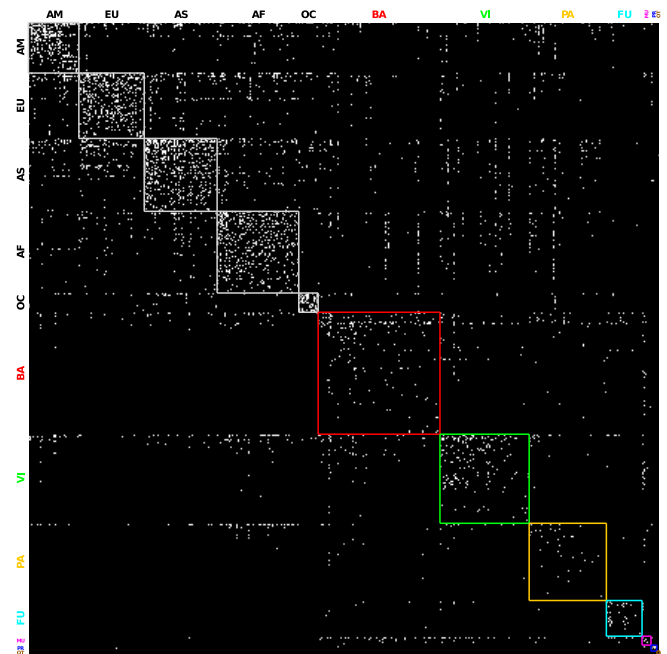

FIGURE 2. Adjacency matrix of the subnetwork of the 425 articles devoted to countries and infectious diseases in 2017 English Wikipedia. White (black) pixels represent existing (absent) links. In horizontal and vertical axis the articles are ordered by continent for countries (AM for Americas, EU for Europe, AS for Asia, AF for Africa, and OC for Oceania) and by disease type for diseases (BA for bacterial, VI for viral, PA for parasitic, FU for fungal, MU for multiple type, PR for prionic, and OT for other type). Inside each block of continent and disease type the articles are ordered according to PageRank algorithm (see Tabs. 1 and 3). From top left to bottom right: block diagonal elements corresponding to intra-continent links are delimited by light gray squares. Also block diagonal elements corresponding to intra-disease type links are delimited by squares with color contour. Off diagonal blocks indicates links between diseases of different type, links between countries of different continent, or links between diseases and countries.

three components of G_R are given in [17]–[19], [22].

After obtaining the matrix G_R and its components we can analyze the PageRank sensitivity in respect to specific links between n_r nodes. To measure the sensitivity of a country c to a disease d we change the matrix element $G_R(d \rightarrow c)$ by a factor $(1+\delta)$ with $\delta \ll 1$, we renormalize to unity the sum of the column elements associated with disease d, and we compute the logarithmic derivative of PageRank probability $P(c)$ associated to country c: $D(d \rightarrow c, c) = d \ln P(c)/d\delta$ (diagonal sensitivity). It is also possible to consider the nondiagonal (or indirect) sensitivity $D(d \rightarrow c, c') = d \ln P(c')/d\delta$ when the variation is done for the link from d to c and the derivative of PageRank probability is computed for another country c'. This approach was already used in [20], [21] showing its efficiency.

III. RESULTS AND DISCUSSION

All calculations have been performed using FORTRAN codes compiled and ran on a computer with Intel Xeon processor E5-2630 v4 25M Cache, 2.20 GHz, with 64GB RDRAM. Unless otherwise stated figures have been prepared using gnuplot.

A. NETWORK OF DIRECT LINKS

For the reduced Google matrix analysis we have $n_r = 425$ selected nodes of countries (195) and infectious diseases (230). The diseases are attributed to 7 groups corresponding to the standard disease types as it is given in Tab. 1, Tab. 2. These n_r nodes constitute a subnetwork embedded in the huge global English Wikipedia network with more than 5 million nodes. This subnetwork is shown in Fig. 1 which has been generated with Cytoscape software [30]. In Fig. 1 black arrow links represent the nonzero elements of adjacency matrix between the selected n_r nodes. The image

of this adjacency matrix is shown in Fig. 2 where white pixels depicted a link between two nodes. In this picture, nodes are ordered with respect to the PageRank order in each subgroup: American countries, European countries, Asian countries, African countries, Oceanian countries, bacterial diseases, viral disease, parasitic diseases, fungal diseases, multiple origins diseases, prionic diseases and other kind of disease origins. There are visibly more links inside subgroups but links between groups are also significant. Fig. 1 gives us the global view of network of direct links, shown in Fig. 2, corresponding to the component G_{rr} of the reduced Google matrix. We see that countries are located in the central part of the network of Fig. 1 since they have many ingoing links. While it is useful to have such a global view it is clear that it does not take into account the indirect links appearing between n_r nodes due to pathways via the complementary network part with a huge number of nodes $N - n_r \simeq N$. The indirect links emerging between n_r from this indirect pathways are analyzed in the frame of REGOMAX method below.
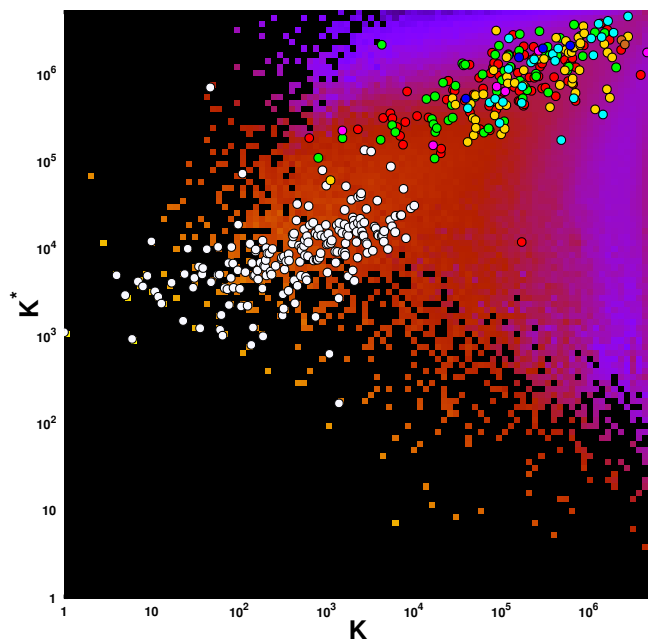
FIGURE 3. Density of articles of English Wikipedia 2017 on PageRank $K -$ CheiRank K^* plane. Data are averaged over a 100×100 grid spanning the $(\log_{10} K, \log_{10} K^*) \in [0, \log_{10} N] \times [0, \log_{10} N]$ domain. Density of articles ranges from very low density (purple tiles) to very high density (bright yellow tiles). The cells without articles are represented by black tiles. The superimposed white (colored) circles give the positions of countries (infectious diseases) listed in Tab. 3 (Tab. 1). The color code for infectious diseases is the same as in Fig. 1.

B. PAGERANK AND CHEIRANK OF THE REDUCED NETWORK NODES

At first we compute the PageRank and CheiRank probabilities for the global network with N nodes attributing to each node PageRank and CheiRank indexes K and K^*. For selected n_r nodes the results of PageRank are shown in Tab. 2. As usual (see [11], [16]), the countries are taking the top PageRank positions with US, France, Germany, etc at $K = 1, 2, 3$, etc as shown in Tab. 3. In the list of n_r nodes the infectious diseases start to appear from $K = 106$ with Tuberculosis (Tab. 2). If we consider only infectious diseases ordered by their disease PageRank index K_d then we obtain at the top Tuberculosis, HIV/AIDS, Malaria, Pneumonia, Smallpox at first positions with $K_d = 1, 2, 3, 4, 5$ (see Tab. 2). It is clear that PageRank order gives at the top positions severe infectious diseases which are (were for Smallpox) very broadly spread worldwide.

In Fig. 3 we show the location of selected n_r nodes on the global (K, K^*) plane of density of Wikipedia articles (see details of this representation in [11], [16]). Here the positions of countries are shown by white circles and diseases by color circles. The countries are taking the top positions since they have many ingoing links from variety of other articles. The infectious diseases are located on higher values of K, K^* even if some diseases are overlapping with the end list of countries (see Tab. 2).

All $n_r = 425$ selected articles can be ordered by their local PageRank and CheiRank indexes K_r and K_r^* which
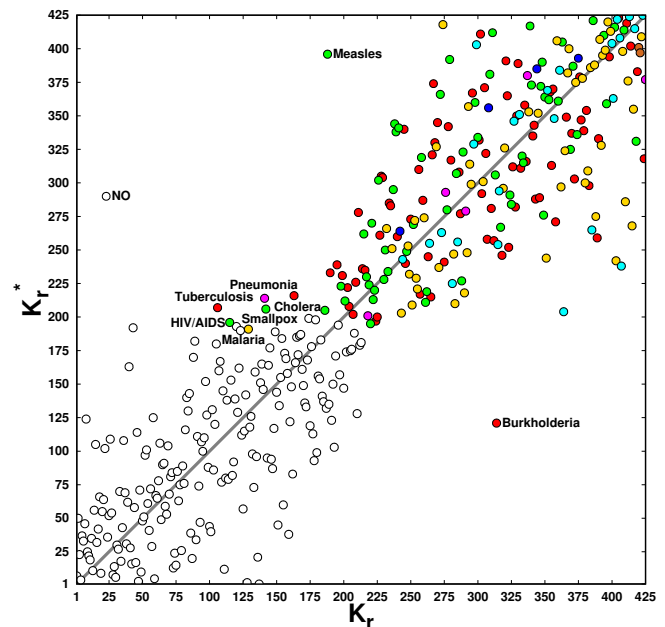
FIGURE 4. Distribution of the articles devoted to infectious diseases (colored circles) and to countries (white circles) in the PageRank $K_r -$ CheiRank K_r^* plane. The color code for infectious diseases is the same as in Fig. 1.

range from 1 to $n_r = 425$. Their distribution in the local PageRank-CheiRank plane is shown in Fig. 4. As discussed previously, countries are at the top K_r, K_r^* positions. The names of top PageRank diseases are marked on the figure. The most communicative articles of infectious diseases are those with top K_r^* positions. Thus the top CheiRank disease is Burkholderia due to many outgoing links present in this article. The next ones are Malaria and HIV/AIDS.

C. REDUCED GOOGLE MATRIX

To study further the selected subset of 425 nodes we use the reduced Google matrix approach and compute numerically G_R and its three components G_{pr}, G_{rr}, G_{qr}. It is convenient to characterize each component by its weight defined as the sum of all elements divided by the matrix size n_r. By definition we have the weight $W_R = 1$ for G_R and we obtain weights $W_{pr} = 0.91021$, $W_{rr} = 0.04715$, $W_{qr} = 0.04264$ (with nondiagonal weight $W_{qrnd} = 0.02667$) respectively for G_{pr}, G_{rr}, G_{qr} (G_{qrnd}). The weight of G_{pr} is significantly larger than others but this matrix is close to the matrix composed from equal columns where the column is the PageRank vector (see also discussions in [18]–[20]). Due to this reason the components G_{rr} and G_{qr} provide an important information about interactions of nodes. Since the weights of these two components are approximately equal we see that the direct and indirect (hidden) links have a comparable contribution.

As an illustration we show in Fig. 5 a close up on African countries and viral diseases sectors of the full $n_r \times n_r$ reduced Google matrix G_R is shown (there are 55 African countries and 60 viral diseases shown in fig5). Detailed presentations

IEEE *Access*

FIGURE 5. Reduced Google matrix G_R (top left panel) and three of its components, G_{qrnd} (top right panel), G_{rr} (bottom left panel), and $G_{rr} + G_{qrnd}$, associated to the subnetwork of articles (bottom right) devoted to countries and infectious diseases in 2017 English Wikipedia. For the sake of clarity, here we show only matrix entries corresponding to the subset of articles devoted to African countries and viral diseases (see [25] for the full subnetwork constituted by the $n_r = 425$ articles devoted to all countries and all infectious diseases). Here lines and columns corresponding to African countries (viral diseases) are ordered as in Tab. 3 (Tab. 1). Horizontal and vertical tics on close-ups are placed every 10 entries.

of the G_R matrix components for the complete subset of $n_r = 425$ countries and infectious diseases are given in [25]. In Fig. 5, G_R and its components are composed of diagonal blocks corresponding to country → country and disease → disease effective links, and off-diagonal blocks corresponding to disease → country (upper off-diagonal block) and country → disease (lower off-diagonal block) effective links.

D. FRIENDSHIP NETWORK OF NODES

We use the matrix of direct and indirect transition $G_{rr} + G_{qr}$ to determine the proximity relations between 230 nodes corresponding to diseases and to all $n_r = 425$ nodes of diseases and countries. We call this the friendship networks being shown in Fig. 6. For each of 7 disease groups (see Tab. 2) we take a group leader as a disease with highest PageRank probability inside the group. Then on each step (level) we take 2 best friends define them as those nodes to which a leader has two highest transition matrix elements of $G_{rr} + G_{qr}$. This gives us the second level of nodes below the 7 leaders. After that we generate the third level keeping again two better friends of the nodes of second level (those with highest transition probabilities). This algorithm is repeated until no new friends are found and the algorithm stops. In this way we obtain the network of 17 infectious diseases shown in the top panel of Fig. 6. The full arrows show the proximity links between disease nodes. The red arrows mark links with dominant contribution of G_{qr} indirect transitions while the black ones mark the links with dominance of G_{rr}

FIGURE 6. Infectious diseases friendship network. Top panel: we consider the set of top PageRank infectious disease for each of the seven type of infectious diseases, i.e., tuberculosis for bacterial type, HIV/AIDS for virus type, malaria for parasitic type, candidiasis for fungal type, pneumonia for multiple type, Creutzfeldt–Jakob for prionic type, and Desmodesmus for other type (see Tab. 1). From each one of these top PageRank diseases (placed along the main grey circle) we determine the two best linked diseases in $G_{rr} + G_{qr}$, e.g., from tuberculosis the two best linked diseases are leprosy and HIV/AIDS. If not already present in the reduced network, we add these best linked diseases along secondary circles centered on the previous diseases. Then from each one of the newly added diseases we determine the two best linked diseases, and so on. At the fourth iterations of this process no new diseases can be added. The arrows represent the links between diseases (1st iteration: plain line; 2nd iteration: dashed line; 3rd iteration: dashed–dotted line; 4th iteration: dotted line). Black arrows correspond to links existing in the adjacency matrix, red arrows are purely hidden links absent from adjacency matrix but present in G_{qr} component of the reduced Google matrix G_R. The color code for infectious diseases is the same as in Fig. 1. Bottom panel: same reduced network as in the top panel but at each iteration also the best linked countries are determined. At each iteration no new links are determined from the newly added countries. Countries are represented by ring shape nodes with red color for countries from Americas, gold for African countries, cyan for Asian countries, blue for European countries, and orange for Oceanian countries. Network drawn with Cytoscape [30].

direct transitions. Full arrows are for transitions from group leaders to nodes of second level, etc (see Fig. 6 caption for details). The obtained network is drawn with the Cytoscape software [30].

In Fig. 6 top panel, four of the leader diseases are well connected; nodes corresponding to Tuberculosis (bacterial disease), HIV/AIDS (viral disease), Malaria (parasitic disease), Pneumonia (multiple origin disease) have 6 or more degrees. Nodes of Creutzfeldt-Jakob disease (prion disease) and of Desmodesmus (other origin disease) are more isolated. Focusing on first level friendship links (solid arrows), we retrieve several well known interactions between infectious diseases such as:

- the Tuberculosis–HIV/AIDS syndemic (see e.g. [31], [32]) which is here represented by a closed loop between the two diseases,
- the interaction between HIV/AIDS and Syphilis (see e.g. [33]) which is a typical example of syndemic between AIDS and sexually transmitted diseases,
- the Candidiasis interaction with HIV/AIDS, since the former is a very common opportunistic fungal infection for patients with HIV/AIDS (see e.g. [34]), and the Candidiasis interaction with Sepsis[1] since e.g. invasive Candidiasis which in some rare cases can lead to fulminant sepsis with an associated mortality exceeding 70% (see e.g. [35]),
- the Pneumonia to Sepsis interaction or the Malaria to Sepsis interaction, the first interaction reflects the fact that Sepsis is one of the possible complications of Pneumonia, the second interaction reflects that symptoms of Malaria resemble to those of Sepsis [36],
- the closed loop interaction between Tuberculosis and Leprosy reflecting that these two diseases are caused by two different species of mycobacteria (see e.g. [37]),
- the relation between Pneumonia and Tuberculosis, two severe pulmonary diseases (see e.g. [38]),
- the Creutzfeldt–Jakob disease pointing to Pneumonia since patient infected by this prion disease develop a fatal Pneumonia due to impaired coughing reflexes (see e.g. [39]),
- the closed loop interaction between Kuru and Creutzfeldt–Jakob diseases since these two diseases are representatives of transmissible spongiform encephalopathies (see e.g. [40]).

Taking into account also the other 2nd to 4th friendship levels, peculiar features appear such as:

- the cluster of bacterial diseases Tuberculosis–Leprosy–Syphilis; since there were a confusion between Leprosy and Syphilis in diagnosis before XXth century (see e.g. [41], [42]), and false positives can occur with Tuberculosis for patients with Syphilis (see e.g. [43]),
- the mosquito diseases cluster grouping Malaria, Yellow fever, Dengue fever and West Nile fever,

- the Meningitis–Sepsis closed loop since the Sepsis is usually developed at early stage by patient with Meningitis (see e.g. [44]).

Red arrows in Fig. 6 indicate pure indirect links between infectious diseases: Desmodesmus and Malaria are both waterborne diseases (see e.g. [45]), Desmodesmus and HIV/AIDS are related by a Wikipedia page devoted to immunocompetence (see e.g. [46]), and Kuru in Papua New Guinea Foré language possibly means to shiver from cold (see e.g. [47]).

From the above analysis we observe that the wiring between infectious diseases is meaningful guaranteeing that information encoded in the reduced Google matrix G_R, and more precisely in its $G_{rr} + G_{qr}$ component, is reliable, and can be used to infer possible relations between infectious diseases and any other subjects contained in Wikipedia such as e.g. countries, drugs, proteins, etc.

In the bottom panel of Fig. 6 we analyze the proximity between diseases of top panel with the world countries. Thus we add the two better "friend" countries being those to which a given disease has most strong matrix elements in $G_{rr} + G_{qr}$ (there is no next iterations for country nodes). The friend countries (or proximity countries) are Egypt and Swaziland for Tuberculosis; Cameroon and Cote d'Ivoire for HIV/AIVS; Peru and Thailand for Malaria; Liberia and Uganda for Pneumonia; United Kingdom (UK)[2] and Papua New Guinea for Creutzfeldt-Jakob disease and others. These strong links from an infectious disease to a given country well correspond to known events involving a disease and a country, like e.g. UK and Creutzfeldt-Jakob disease. We will see this in a more direct way using the sensitivity analysis presented in the next subsection.

E. WORLD COUNTRY SENSITIVITY TO INFECTIOUS DISEASES

We also perform analysis of the sensitivity $D(d \to c, c')$ of country node c' to the variation of the link $d \to c$, where d denotes a disease node and c a country node. The diagonal sensitivity $D(d \to c, c)$ of world countries to Tuberculosis and HIV/AIDS are shown in Fig. 7. The most sensitive countries to Tuberculosis are Swaziland (SZ), Egypt (EG) and New Zealand (NZ). Indeed, in 2007 SZ had the highest estimated incidence rate of Tuberculosis as it is described in the corresponding Wikipedia article. Egypt also appears in this article since tubercular decay has been found in the spine of Egyptian mummy kept in the British Museum. NZ is present in this article since this country had a relatively successful effort to eradicate bovine tuberculosis. Thus Tuberculosis has direct links to these three countries (in agreement with two close country friends shown in the network of Fig. 6) that results in their high sensitivity to this disease. Of course, the origins of this sensitivity are different for SZ, EG, NZ. Thus Wikipedia network integrates all historical events related to Tuberculosis including ancient Egyptian mummy and recent years of high incidence rate

---

[1]Even if most of the Sepsis are bacterial, it can also be fungal or viral
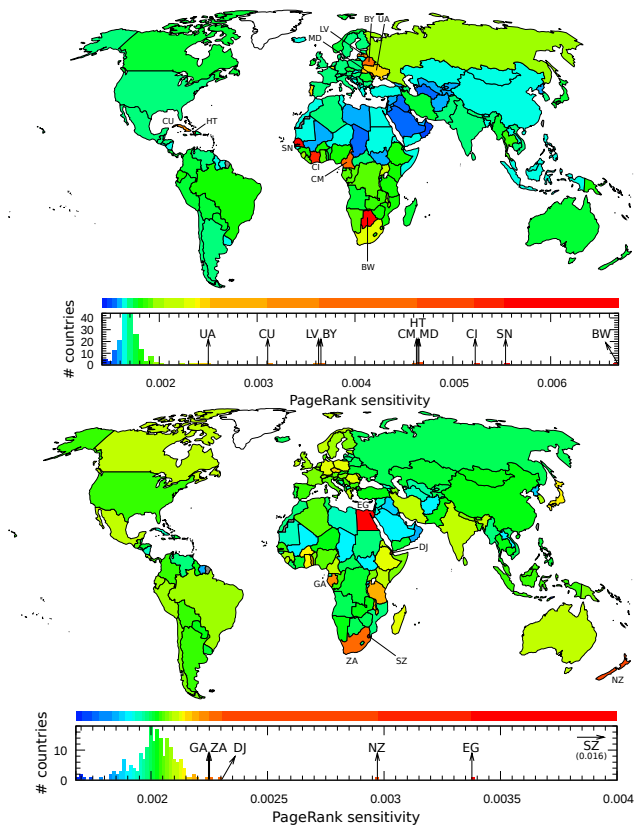
[2]Noted GB in Fig. 6 and Tab. 3.

FIGURE 7. Country PageRank sensitivity to the variation of the reduced Google matrix HIV→country link (upper panel) and to the variation of the reduced Google matrix Tuberculosis→country link (bottom panel). The color categories are obtained using the Jenks natural breaks classification method [48]. We use the blank world map svg file available at [49].

FIGURE 8. Country PageRank sensitivity to the variation of the reduced Google matrix HIV → USA link (top panel) and to the variation of the reduced Google matrix HIV → Cameroon link (bottom panel). The color categories are obtained using the Jenks natural breaks classification method [48]. We use the blank world map svg file available at [49].

in SZ. It can be discussed how important are these rather different types of links between disease and counties. Of course, a simplified network view cannot take into account all richness of historical events and describe them by a few number of links. However this approach provides a reliable global view of the interactions and dependencies between a disease and world countries.

The sensitivity of countries to HIV/AIDS is shown in the top panel of Fig. 7. The most sensitive countries are Botswana (BW), Senegal (SN) and Cote d'Ivoire (CI). This happens since HIV/AIDS article directly points that estimated life expectancy in BW dropped from 65 to 35 years in 2006; SN and CI appears since the closest relative of HIV-2 exists in monkey living in coastal West Africa from SN to CI. The friendship network in Fig. 6 indeed marks the countries close to HIV/AIDS as CI and Cameroon (CM). The sensitivity map in Fig. 7 also shows that CM has high sensitivity to HIV/AIDS since HIV-1 appears to have originated in southern Cameroon.

The case of two diseases considered in Fig. 7 demonstrates that the REGOMAX approach is able to reliably determine the sensitivity of world countries to infectious diseases taking into account their relations on a scale of about 3 thousands of years.
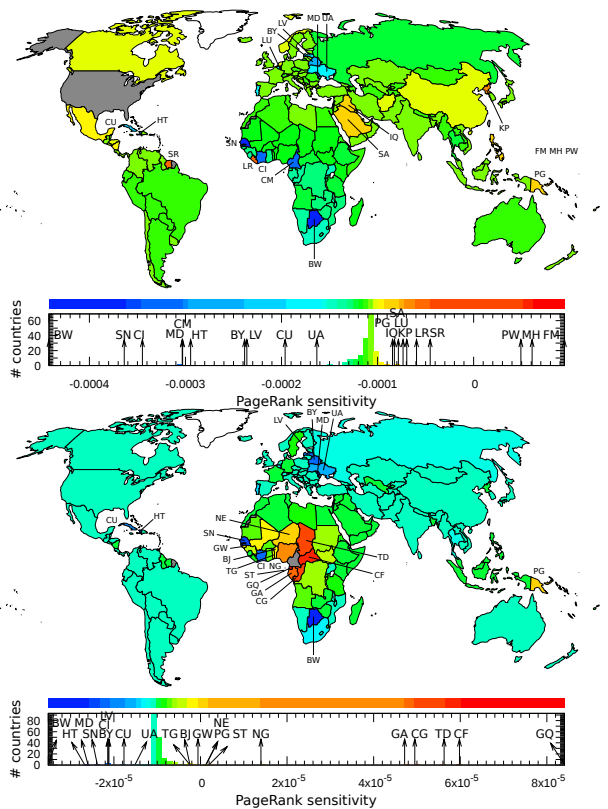
A part of the sensitivity relations between diseases and countries can be visible from the friendship network as those in the bottom panel of Fig. 6. However, the REGOMAX approach can handle also indirect sensitivity $D(d \rightarrow c, c')$ which is rather hard to be directly extracted from the friendship network. The examples of indirect sensitivity are shown in Fig. 8. Thus the variation of link from HIV/AIDS to Cameroon (CM) (Fig. 8 bottom panel) mainly affects Equatorial Guinea (GQ), Central African Republic (CF) and Chad (TD). The variation of link HIV/AIDS to USA (Fig. 8 top panel) produces the strongest sensitivity for Federal States of Micronesia (FM), Marshall Islands (MH) and Rwanda (RW). These countries are not present in the Wikipedia article HIV/AIDS and the obtained sensitivity emerges from a complex network interconnections between HIV/AIDS, USA (or Cameroon) to these countries. Thus the REGOMAX analysis allows to recover all network complexity of direct and indirect interactions between nodes.

F. COMPARISON OF REGOMAX AND WHO RESULTS

It is important to compare the results of REGOMAX analysis with those of World Health Organization (WHO) or other sources on number of infected people. With this aim we extract from WHO reports [38] the number of Tuberculosis
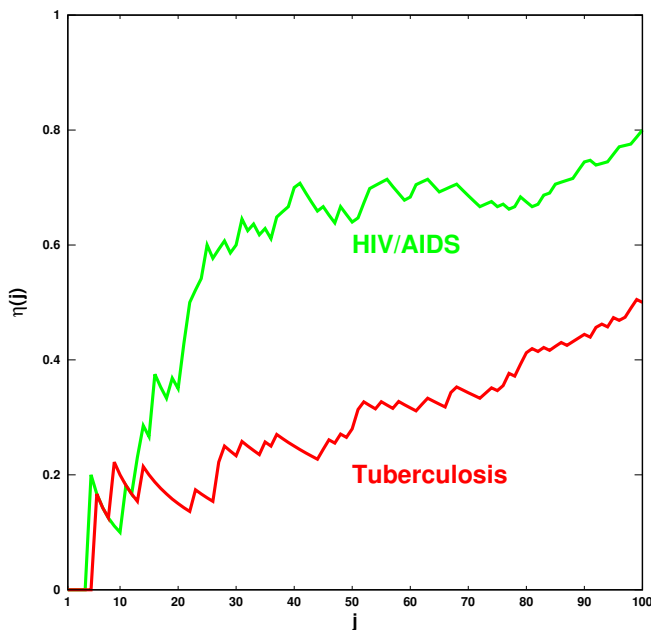
FIGURE 9. Overlap $\eta(j) = j_c/j$ between the ranking of countries obtained from the REGOMAX PageRank sensitivity computation and the ranking of countries obtained from WHO data (Tab. 4 for Tuberculosis in 2016 and Tab. 5 for HIV/AIDS in 2017). Here, j_c is the number of common countries in the top j of the two rankings. Green curve is for the HIV/AIDS disease, and red curve is for the Tuberculosis disease.
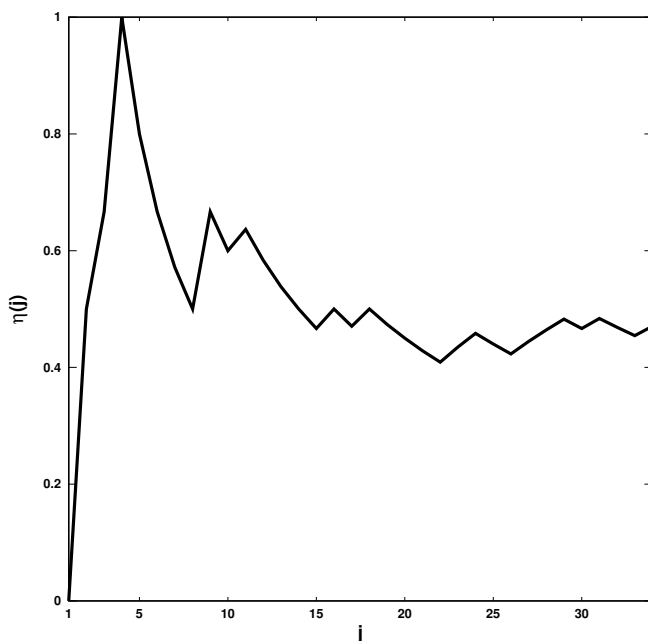

FIGURE 11. Maps of the maximal sensitivity of diseases. Each country is colored according to the disease giving the maximal sensitivity: Tuberculosis (blue), HIV/AIDS (red), Malaria (cyan), Diarrheal diseases (magenta), Dengue, Meningitis, Hepatitis. Top panel shows the maximal sensitivity for the 7 diseases, and bottom panel for 6 of these diseases (Tuberculosis is not taken into account). Gray color indicates countries for which the maximal sensitivity to these diseases is less than 1/8 (1/4) of the greatest maximal sensitivity for any country. Note that Dengue, Meningitis and Hepatitis never gives the maximal sensitivity for any country. We use the blank world map svg file available at [49].

FIGURE 10. Overlap $\eta(j) = j_c/j$ between the PageRank of the diseases and the diseases ranking based on the estimated number of deaths during the year 2010 (Tab. 6). Here, j_c is the number of common countries in the top j of the two rankings. The estimated numbers of deaths caused by each disease during the year 2010 are extracted from [50].
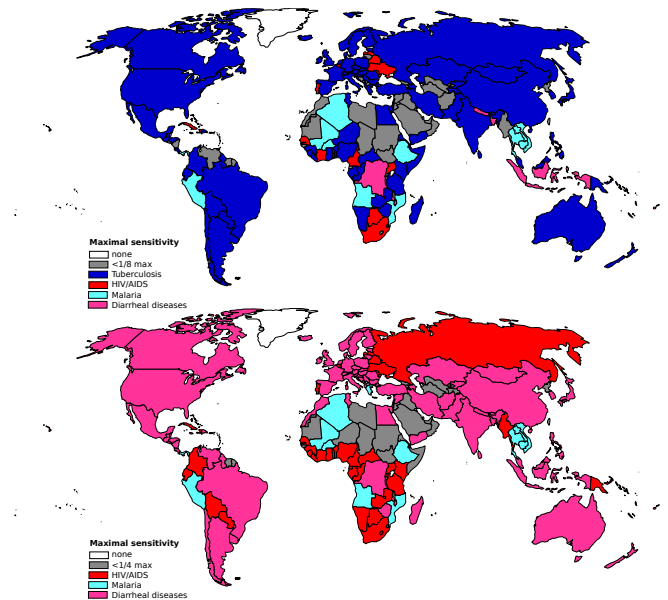
incidences per 100000 population of a given country in 2016, and the number of new HIV/AIDS infections per 1000 uninfected population in 2017. The ranking of countries by the number of incidences is presented in Tab. 4 for Tuberculosis and the ranking of countries by new infections is presented in Tab. 5 for HIV/AIDS. We also analyze Global Burden of Disease (GBD) Study data [50] for the estimated number of deaths caused by each disease in 2010. The resulting ranking of diseases is given in Tab. 6.

We compare these official WHO ranking results with those obtained from REGOMAX analysis. Thus we determine the ranking of countries by their sensitivity to Tuberculosis and to HIV/AIDS for top 100 countries (these ranking lists are given in [25]). The overlap $\eta(j)$ of these REGOMAX rankings with those of WHO from Tab. 4 and Tab. 5 are shown in Fig. 9. We obtain the overlap of 50% for Tuberculosis and 79% for HIV/AIDS for the top 100 countries. These numbers are comparable with overlaps obtained for top 100 historical figures found from Wikipedia and historical analysis (see [12]) and for top 100 world universities determined by Wikipedia and Shanghai ranking (see [13], [15]).

Another comparison is presented in Fig. 10. Here we take the infectious diseases ordered by their PageRank index and compare them with the ranking list of diseases ordered by the estimated number of deaths caused by them in 2010 (see Tab. 6). The obtained overlap is shown in Fig. 10 with an overall of 100% for top 4 deadliest diseases (which from Tab. 6 are 1 HIV/AIDS, 2 Pneumonia, 3 Tuberculosis, 4

Malaria) and 46% for the whole list of 34 considered diseases. In addition the REGOMAX analysis allows to determine the world map of countries with the highest sensitivity to the top list of 7 diseases. Such maps are shown in Fig. 11. We see the world dominance of Tuberculosis (top panel) while after it (or in its absence) we find the world dominance of HIV/AIDS (bottom panel). The geographical influence of Malaria and Diarrheal diseases are also well visible.

Thus the performed comparison with WHO data shows that the Google matrix analysis of Wikipedia network provides us a reliable information about world importance and influence of infectious diseases.

G. DISCUSSION

Reduced networks of infectious diseases, as displayed in Fig. 6, allow to efficiently summarize main direct and indirect (hidden) interactions between diseases and between diseases and countries. This method could be applied to e.g. rare diseases to possibly infer relations between them; in comparison with infectious diseases, we would expect many hidden links as rare diseases form a less coherent and less studied group of diseases. We can imagine many other applications such as e.g. the representation of the interactions between cancers and drugs.

According to Global Burden of Disease organization [50] raw data extracted all around the world from vital registrations, censuses, hospitals, police records, etc, is generally incomplete and only estimations and trends can be obtained by applying statistical models, such as e.g. CODEm to estimate worldwide causes of death [51]. Conversely, Wikipedia forms an open coherent knowledge database, from which we can determine the global influence of infectious diseases. We believe that our reduced Google matrix method applied to worldwide causes of deaths could constitute a complementary method to those usually used in GBD Study [50].

The REGOMAX method could be applied to periodically taken snapshots of Wikipedia networks in order to determine over a time window global sensitivities to infectious diseases. Although possibly interesting information could result from such study, it is not clear it could allow estimation of near real time prevalence as it is the case for methods based on Wikipedia usage [24]; indeed editing latency could be too slow to consider editors activity as a probe of infection peaks.

IV. CONCLUSIONS

In this work we presented the reduced Google matrix (or REGOMAX) analysis of world influence of infectious diseases from the English Wikipedia network of 2017. This method allows to take into account all direct and indirect links between the selected nodes of countries and diseases. The importance of diseases is determined by their PageRank probabilities. The REGOMAX analysis allows to establish the network of proximity (friendship) relations between the diseases and countries. The sensitivity of world countries to a specific disease is determined as well as the influence of link variation between a disease and a country on other countries.

The comparison with the WHO data confirms the reliability of REGOMAX results applied to Wikipedia network.

REFERENCES

[1] NIH. (2018) Infectious diseases. National Institutes of Health. Accessed Aug. 2018. [Online]. Available: https://www.nih.gov/about-nih/what-we-do/nih-turning-discovery-into-health/infectious-diseases

[2] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," Computer Networks and ISDN Systems, vol. 30, no. 1, pp. 107 – 117, 1998, proceedings of the Seventh International World Wide Web Conference. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016975529800110X

[3] A. N. Langville and C. D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, 2012.

[4] D. Gleich, "Pagerank beyond the web," SIAM Review, vol. 57, no. 3, pp. 321–363, 2015. [Online]. Available: https://doi.org/10.1137/140976649

[5] Encyclopaedia Britannica. (2018) Accessed Aug. 2018. [Online]. Available: http://www.britannica.com

[6] J. Giles, "Internet encyclopaedias go head to head," Nature, vol. 438, pp. 900–901, Dec. 2005. [Online]. Available: https://doi.org/10.1038/438900a

[7] D. Butler, "Publish in wikipedia or perish," Nature, dec 2008. [Online]. Available: https://doi.org/10.1038/news.2008.1312

[8] E. Callaway, "No rest for the bio-wikis," Nature, vol. 468, no. 7322, pp. 359–360, nov 2010. [Online]. Available: https://doi.org/10.1038/468359a

[9] J. M. Reagle Jr., Good Faith Collaboration: The Culture of Wikipedia (History and Foundations of Information Science). The MIT Press, 2012.

[10] F. Å. Nielsen, "Wikipedia research and tools: Review and comments," SSRN Electronic Journal, 2012. [Online]. Available: https://doi.org/10.2139/ssrn.2129874

[11] A. O. Zhirov, O. V. Zhirov, and D. L. Shepelyansky, "Two-dimensional ranking of wikipedia articles," The European Physical Journal B, vol. 77, no. 4, pp. 523–531, Oct 2010. [Online]. Available: https://doi.org/10.1140/epjb/e2010-10500-7

[12] Y.-H. Eom, P. Aragón, D. Laniado, A. Kaltenbrunner, S. Vigna, and D. L. Shepelyansky, "Interactions of cultures and top people of wikipedia from ranking of 24 language editions," PLOS ONE, vol. 10, no. 3, pp. 1–27, 03 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0114825

[13] J. Lages, A. Patt, and D. L. Shepelyansky, "Wikipedia ranking of world universities," The European Physical Journal B, vol. 89, no. 3, mar 2016. [Online]. Available: https://doi.org/10.1140/epjb/e2016-60922-0

[14] G. Katz and L. Rokach, "Wikiometrics: a wikipedia based ranking system," World Wide Web, vol. 20, no. 6, pp. 1153–1177, Nov 2017. [Online]. Available: https://doi.org/10.1007/s11280-016-0427-8

[15] C. Coquidé, J. Lages, and D. L. Shepelyansky, "World influence and interactions of universities from wikipedia networks," Eur. Phys. J. B, vol. 92, no. 1, p. 3, 2019. [Online]. Available: https://doi.org/10.1140/epjb/e2018-90532-7

[16] L. Ermann, K. M. Frahm, and D. L. Shepelyansky, "Google matrix analysis of directed networks," Rev. Mod. Phys., vol. 87, pp. 1261–1310, Nov 2015. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.87.1261

[17] K. M. Frahm and D. L. Shepelyansky, "Reduced Google matrix," arXiv, vol. arXiv:1602.02394, Feb. 2016. [Online]. Available: https://arxiv.org/abs/1602.02394

[18] K. M. Frahm, K. Jaffrès-Runser, and D. L. Shepelyansky, "Wikipedia mining of hidden links between political leaders," The European Physical Journal B, vol. 89, no. 12, p. 269, Dec 2016. [Online]. Available: https://doi.org/10.1140/epjb/e2016-70526-3

[19] J. Lages, D. L. Shepelyansky, and A. Zinovyev, "Inferring hidden causal relations between pathway members using reduced google matrix of directed biological networks," PLOS ONE, vol. 13, no. 1, pp. 1–28, 01 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0190812

[20] S. El Zant, K. M. Frahm, K. Jaffrès-Runser, and D. L. Shepelyansky, "Analysis of world terror networks from the reduced google matrix of wikipedia," The European Physical Journal B, vol. 91, no. 1, p. 7, Jan 2018. [Online]. Available: https://doi.org/10.1140/epjb/e2017-80570-0

[21] S. El Zant, K. Jaffrès-Runser, and D. L. Shepelyansky, "Capturing the influence of geopolitical ties from wikipedia with reduced google matrix," PLOS ONE, vol. 13, no. 8, pp. 1–31, 08 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0201397

[22] S. E. Zant, K. Jaffrès-Runser, K. M. Frahm, and D. L. Shepelyansky, "Interactions and influence of world painters from the reduced google

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI
10.1109/ACCESS.2019.2899339, IEEE Access

IEEE Access

G.Rollin *et al.*: World influence of infectious diseases

matrix of wikipedia networks," IEEE Access, vol. 6, pp. 47 735–47 750, 2018.

[23] W. Wang, M. Tang, H. E. Stanley, and L. A. Braunstein, "Unification of theoretical approaches for epidemic spreading on complex networks," Reports on Progress in Physics, vol. 80, no. 3, p. 036603, 2017. [Online]. Available: http://stacks.iop.org/0034-4885/80/i=3/a=036603

[24] D. J. McIver and J. S. Brownstein, "Wikipedia usage estimates prevalence of influenza-like illness in the united states in near real-time," PLOS Computational Biology, vol. 10, no. 4, pp. 1–8, 04 2014. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1003581

[25] Wiki4InfectiousDiseases. (2018) Wikipedia network of infectious diseases. Accessed Sep. 2018. [Online]. Available: http://perso.utinam.cnrs.fr/ lages/datasets/Wiki4InfectiousDiseases/

[26] K. Frahm and D. Shepelyansky. (2017) Wikipedia networks of 24 editions of 2017. Accessed Jul. 2018. [Online]. Available: http://www.quantware.ups-tlse.fr/QWLIB/24wiki2017

[27] Wikipedia. (2018) List of infectious diseases. Wikipedia. Accessed Jul. 2018. [Online]. Available: https://en.wikipedia.org/wiki/List_of_infectious_diseases

[28] ——. (2017) List of sovereign states in 2017. Wikipedia. Accessed Jul. 2018. [Online]. Available: https://en.wikipedia.org/wiki/List_of_sovereign_states_in_2017

[29] A. D. Chepelianskii, "Towards physical laws for software architecture," arXiv, vol. arXiv:1003.5455, 2010. [Online]. Available: http://arxiv.org/abs/1003.5455

[30] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," Genome Research, vol. 13, no. 11, pp. 2498–2504, 2003. [Online]. Available: http://genome.cshlp.org/content/13/11/2498.abstract

[31] A. Pawlowski, M. Jansson, M. Skÿld, M. E. Rottenberg, and G. KÃd'llenius, "Tuberculosis and hiv co-infection," PLOS Pathogens, vol. 8, no. 2, pp. 1–7, 02 2012. [Online]. Available: https://doi.org/10.1371/journal.ppat.1002464

[32] C. K. Kwan and J. D. Ernst, "Hiv and tuberculosis: a deadly human syndemic," Clinical Microbiology Reviews, vol. 24, no. 2, pp. 351–376, 2011. [Online]. Available: https://cmr.asm.org/content/24/2/351

[33] G. Karp, F. Schlaeffer, A. Jotkowitz, and K. Riesenberg, "Syphilis and hiv co-infection," European Journal of Internal Medicine, vol. 20, no. 1, pp. 9 – 13, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0953620508001301

[34] D. Armstrong-James, G. Meintjes, and G. D. Brown, "A neglected epidemic: fungal infections in hiv/aids," Trends in Microbiology, vol. 22, no. 3, pp. 120 – 127, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966842X1400002X

[35] P. G. Pappas, M. S. Lionakis, M. C. Arendrup, L. Ostrosky-Zeichner, and B. J. Kullberg, "Invasive candidiasis," Nature Reviews Disease Primers, vol. 4, pp. 18 026 EP –, May 2018, primer. [Online]. Available: http://dx.doi.org/10.1038/nrdp.2018.26

[36] M. A. Auma, M. J. Siedner, D. Nyehangane, A. Nalusaji, M. Nakaye, J. Mwanga-Amumpaire, R. Muhindo, L. A. Wilson, Y. Boum, and C. C. Moore, "Malaria is an uncommon cause of adult sepsis in south-western uganda," Malaria Journal, vol. 12, no. 1, p. 146, 2013. [Online]. Available: https://doi.org/10.1186/1475-2875-12-146

[37] Wikipedia. (2018) Mycobacterium. Wikipedia. Accessed Sep. 2018. [Online]. Available: https://en.wikipedia.org/wiki/Mycobacterium

[38] WHO. (2018) Global health observatory data repository. World Health Organization. Accessed Aug. 2018. [Online]. Available: http://apps.who.int/gho/data/node.home

[39] A. Al Balushi, M. W. Meeks, G. Hayat, and J. Kafaie, "Creutzfeldt–jakob disease: Analysis of four cases," Frontiers in Neurology, vol. 7, p. 138, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2016.00138

[40] B. Sikorska and P. P. Liberski, Human Prion Diseases: From Kuru to Variant Creutzfeldt-Jakob Disease. Dordrecht: Springer Netherlands, 2012, pp. 457–496. [Online]. Available: https://doi.org/10.1007/978-94-007-5416-4_17

[41] "Leprosy and syphilis," The Lancet, vol. 135, no. 3487, p. 1436, 1890, originally published as Volume 1, Issue 3487. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673602192002

[42] "Syphilis and leprosy," Journal of the American Medical Association, vol. XXXIII, no. 17, pp. 1048–1049, 1899. [Online]. Available: + http://dx.doi.org/10.1001/jama.1899.02450690052009

[43] A. L. Shane, Red Book: 2006 Report of the Committee on Infectious Diseases, 27th Edition. CDC, 2006, vol. 12, no. 12. [Online]. Available: http://wwwnc.cdc.gov/eid/article/12/12/06-1045

[44] CDC. (2018) Bacterial meningitis. Centers for Disease Control and Prevention. Accessed Sep. 2018. [Online]. Available: https://www.cdc.gov/meningitis/bacterial.html

[45] Wikipedia. (2018) Waterborne diseases. Wikipedia. Accessed Sep. 2018. [Online]. Available: https://en.wikipedia.org/wiki/Waterborne_diseases

[46] ——. (2018) Immunocompetence. Wikipedia. Accessed Sep. 2018. [Online]. Available: https://en.wikipedia.org/wiki/Immunocompetence

[47] P. P. Liberski and J. W. Ironside, "Chapter 23 - prion diseases," in Neurobiology of Brain Disorders, M. J. Zigmond, L. P. Rowland, and J. T. Coyle, Eds. San Diego: Academic Press, 2015, pp. 356 – 374. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123982704000239

[48] Wikipedia. (2018) Jenks natural breaks optimization. Wikipedia. Accessed Jun. 2018. [Online]. Available: https://en.wikipedia.org/wiki/Jenks_natural_breaks_optimization

[49] simplemaps. (2018) Blank world map. simplemaps. Accessed Nov. 2018. [Online]. Available: https://simplemaps.com/static/demos/resources/svg-library/svgs/world.svg

[50] Global Burden of Disease organisation, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010," The Lancet, vol. 380, no. 9859, pp. 2095 – 2128, 2012. [Online]. Available: https://doi.org/10.1016/S0140-6736(12)61728-0

[51] K. J. Foreman, R. Lozano, A. D. Lopez, and C. J. Murray, "Modeling causes of death: an integrated approach using codem," Population Health Metrics, vol. 10, no. 1, p. 1, Jan 2012. [Online]. Available: https://doi.org/10.1186/1478-7954-10-1

.

APPENDIX A TABLES

TABLE 1. List of infectious diseases ordered by type (82 of bacterial type, 60 of viral type, 52 of parasitic type, 24 of fungal type, 6 of multiple types, and 2 of other types) then by PageRank of the corresponding article in English edition of Wikipedia [26]. The $n_d = 230$ elements of this list have been extracted from the *list of infectious diseases* article in 2017 English Wikipedia [25], [27].

Infectious diseases

#	Bacterial type	#	Bacterial type (Cont.)	#	Viral type (Cont.)	#	Parasitic type (Cont.)
1	Tuberculosis	61	Prevotella	38	CCHF[e]	37	Anaplasmosis
2	Cholera	62	Ehrlichiosis	39	Caliciviridae	38	Baylisascaris
3	Typhoid fever	63	Vibrio vulnificus	40	PML[f]	39	Fasciolopsiasis
4	Syphilis	64	Actinomycosis	41	Rift Valley fever	40	Head lice infestation
5	Bubonic plague	65	Ureaplasma urealyticum	42	Lassa fever	41	Angiostrongyliasis
6	Leprosy	66	Melioidosis	43	Astrovirus	42	Isosporiasis
7	Sepsis	67	Granuloma inguinale	44	Fifth disease	43	Diphyllobothriasis
8	Meningitis	68	Buruli ulcer	45	VEE[g]	44	Cyclosporiasis
9	Plague (disease)	69	Yersinia pseudotuberculosis	46	Molluscum contagiosum	45	Balantidiasis
10	Typhus	70	Kingella kingae	47	MERS[h]	46	Blastocystosis
11	Diphtheria	71	H. granulocytic anaplasmosis	48	Monkeypox	47	Dientamoebiasis
12	Anthrax	72	Pneumococcal inf.	49	Hand, foot, and mouth disease	48	Gnathostomiasis
13	Helicobacter pylori	73	Nocardiosis	50	Bolivian hemorrhagic fever	49	Capillariasis
14	Lyme disease	74	Bartonellosis	51	H. metapneumovirus	50	Metagonimiasis
15	Staphylococcus	75	Group B streptococcal inf.	52	HFRS[i]	51	Pediculosis corporis
16	Haemophilus influenzae	76	Pasteurellosis	53	LCM[j]	52	Free-living Amoebozoa inf.
17	Tetanus	77	Carrion's disease	54	BK virus		**fungal type**
18	Gonorrhea	78	Yersiniosis	55	Colorado tick fever	1	Candidiasis
19	Pertussis	79	Mycoplasma pneumonia	56	Argentine hemorrhagic fever	2	Histoplasmosis
20	Botulism	80	HME[a]	57	Heartland virus	3	Athlete's foot
21	Chlamydia inf.	81	Ehrlichiosis ewingii inf.	58	Venezuelan hemorrhagic fever	4	Coccidioidomycosis
22	Rickettsia	82	A. haemolyticum[b]	59	Brazilian hemorrhagic fever	5	Chytridiomycosis
23	Brucellosis		**Viral type**	60	H. bocavirus	6	Pneumocystis pneumonia
24	Leptospirosis	1	HIV/AIDS		**Parasitic type**	7	Cryptococcosis
25	Pelvic inflammatory disease	2	Smallpox	1	Malaria	8	Onychomycosis
26	Cellulitis	3	Influenza	2	Schistosomiasis	9	Aspergillosis
27	Legionnaires' disease	4	Measles	3	African trypanosomiasis	10	Tinea versicolor
28	Clostridium difficile inf.	5	Yellow fever	4	Toxoplasmosis	11	Tinea cruris
29	Rocky Mountain spotted fever	6	Poliomyelitis	5	Onchocerciasis	12	Paracoccidioidomycosis
30	Enterococcus	7	SARS[c]	6	Scabies	13	Tinea corporis
31	Bacterial vaginosis	8	Hepatitis C	7	Chagas disease	14	Blastomycosis
32	Bacterial pneumonia	9	Hepatitis B	8	Filariasis	15	Zygomycosis
33	Epidemic typhus	10	Ebola virus disease	9	Hookworm inf.	16	Tinea capitis
34	Trachoma	11	Common cold	10	Leishmaniasis	17	Tinea manuum
35	Salmonellosis	12	Rabies	11	Trichinosis	18	Sporotrichosis
36	Tularemia	13	Dengue fever	12	Giardiasis	19	Tinea barbae
37	Kawasaki disease	14	Chickenpox	13	Amoebiasis	20	Tinea nigra
38	Bacteroides	15	H. papillomavirus inf.	14	Trichomoniasis	21	Chromoblastomycosis
39	Acinetobacter	16	West Nile fever	15	Cryptosporidiosis	22	White piedra
40	Chlamydophila pneumoniae	17	Herpes simplex	16	Myiasis	23	Black piedra
41	Q fever	18	Rubella	17	Dracunculiasis	24	Geotrichosis
42	Shigellosis	19	Hepatitis A	18	Lymphatic filariasis		**Multiple types**
43	Gas gangrene	20	Cytomegalovirus	19	Echinococcosis	1	Pneumonia
44	Bacillus cereus	21	Mumps	20	Cysticercosis	2	Foodborne illness
45	Group A streptococcal inf.	22	Infectious mononucleosis	21	Babesiosis	3	Hemolytic-uremic syndrome
46	Meningococcal disease	23	Shingles	22	Pinworm inf.	4	Keratitis
47	Burkholderia	24	Norovirus	23	Ascariasis	5	Neonatal conjunctivitis
48	Relapsing fever	25	Rotavirus	24	Strongyloidiasis	6	Mycetoma
49	Glanders	26	Hantavirus	25	Toxocariasis		**Prionic type**
50	Psittacosis	27	H. respiratory syncytial virus	26	Anisakis	1	Creutzfeldt–Jakob disease
51	Listeriosis	28	Marburg virus	27	Fasciolosis	2	Kuru (disease)
52	Rickettsialpox	29	Rhinovirus	28	Paragonimiasis	3	Fatal familial insomnia
53	Campylobacteriosis	30	Enterovirus	29	Cutaneous larva migrans	4	GSS[k]
54	Naegleriasis	31	Chikungunya	30	Trichuriasis		**Other types**
55	Murine typhus	32	Hepatitis D	31	Hymenolepiasis	1	Desmodesmus
56	Fusobacterium	33	Viral pneumonia	32	Clonorchiasis	2	Rhinosporidiosis
57	Chancroid	34	Hepatitis E	33	Pediculosis pubis		
58	Cat-scratch disease	35	SSPE[d]	34	Microsporidiosis		
59	Staphylococcal inf.	36	Roseola	35	Opisthorchiasis		
60	Vibrio parahaemolyticus	37	H. parainfluenza viruses	36	Taeniasis		

Abbreviations H. and inf. stand for Human and infection. [a] HME: Human monocytotropic ehrlichiosis. [b] A. haemolyticum: Arcanobacterium haemolyticum. [c] SARS: Severe acute respiratory syndrome. [d] SSPE: Subacute sclerosing panencephalitis. [e] CCHF: Crimean–Congo hemorrhagic fever. [f] PML: Progressive multifocal leukoencephalopathy. [g] VEE: Venezuelan equine encephalitis virus. [h] MERS: Middle East respiratory syndrome. [i] HFRS: Hantavirus hemorrhagic fever with renal syndrome. [j] LCM: Lymphocytic choriomeningitis. [k] GSS: Gerstmann–Sträussler–Scheinker syndrome.

TABLE 2. Ranking of infectious diseases and countries in English Wikipedia 2017 according to PageRank algorithm. The color code distinguishes type of infectious diseases: **bacterial**, **viral**, **parasitic**, **fungal**, **prionic**, **multiple origin**, and **other origin**.

Rank	Disease or country	Rank	Disease or country	Rank	Disease or country	Rank	Disease or country
1	United States	228	Haemophilus influenzae	294	Cysticercosis	360	Bolivian hemorrhagic fever
...		229	Tetanus	295	Babesiosis	361	Blastomycosis
105	Sudan	230	H. papillomavirus inf.	296	Bacteroides	362	Cutaneous larva migrans
106	Tuberculosis	231	West Nile fever	297	Pneumocystis pneumonia	363	H. metapneumovirus
107	Uganda	232	Schistosomiasis	298	Viral pneumonia	364	Zygomycosis
...		233	Herpes simplex	299	Cryptococcosis	365	Trichuriasis
114	Somalia	234	Gonorrhea	300	Hepatitis E	366	Granuloma inguinale
115	HIV/AIDS	235	Pertussis	301	Acinetobacter	367	Hymenolepiasis
116	Ivory Coast	236	African trypanosomiasis	302	Chlamydophila pneumoniae	368	Clonorchiasis
...		237	Rubella	303	Q fever	369	HFRS[g]
128	Fiji	238	Hepatitis A	304	Pinworm inf.	370	Buruli ulcer
129	Malaria	239	Cytomegalovirus	305	Shigellosis	371	LCM[h]
130	Mali	240	Botulism	306	Gas gangrene	372	Yersinia pseudotuberculosis
...		241	Mumps	307	Bacillus cereus	373	Pediculosis pubis
140	Oman	242	Creutzfeldt–Jakob disease	308	Kuru (disease)	374	BK virus
141	Pneumonia	243	Toxoplasmosis	309	SSPE[b]	375	GSS[i]
142	Smallpox	244	Candidiasis	310	Group A streptococcal inf.	376	Kingella kingae
143	Suriname	245	Chlamydia inf.	311	Roseola	377	H. granulocytic anaplasmosis
...		246	Rickettsia	312	Meningococcal disease	378	Microsporidiosis
162	Malawi	247	Infectious mononucleosis	313	H. parainfluenza viruses	379	Pneumococcal inf.
163	Cholera	248	Onchocerciasis	314	Burkholderia	380	Opisthorchiasis
164	Togo	249	Scabies	315	Onychomycosis	381	Nocardiosis
...		250	Brucellosis	316	Aspergillosis	382	Taeniasis
185	San Marino	251	Chagas disease	317	CCHF[c]	383	Bartonellosis
186	Influenza	252	Shingles	318	Relapsing fever	384	Anaplasmosis
187	Saint Lucia	253	Filariasis	319	Ascariasis	385	Tinea capitis
188	Measles	254	Hookworm inf.	320	Strongyloidiasis	386	Colorado tick fever
189	Palau	255	Leishmaniasis	321	Glanders	387	Baylisascaris
190	Typhoid fever	256	Leptospirosis	322	Psittacosis	388	Fasciolopsiasis
191	Marshall Islands	257	Pelvic inflammatory disease	323	Listeriosis	389	Group B streptococcal inf.
192	Equatorial Guinea	258	Norovirus	324	Caliciviridae	390	Pasteurellosis
193	Dominica	259	Cellulitis	325	PML[d]	391	Head lice infestation
194	Guinea-Bissau	260	Trichinosis	326	Rickettsialpox	392	Angiostrongyliasis
195	Syphilis	261	Rotavirus	327	Tinea versicolor	393	Isosporiasis
196	Comoros	262	Hantavirus	328	Campylobacteriosis	394	Argentine hemorrhagic fever
197	Djibouti	263	Legionnaires' disease	329	Naegleriasis	395	Diphyllobothriasis
198	Yellow fever	264	Histoplasmosis	330	Murine typhus	396	Heartland virus
199	Bubonic plague	265	Clostridium difficile inf.	331	Tinea cruris	397	Cyclosporiasis
200	Fed. States of Micronesia	266	Rocky Mountain spotted fever	332	Fusobacterium	398	Carrion's disease
201	Poliomyelitis	267	Enterococcus	333	Rift Valley fever	399	Balantidiasis
202	Tuvalu	268	Bacterial vaginosis	334	Lassa fever	400	Tinea manuum
203	Leprosy	269	Giardiasis	335	Chancroid	401	Sporotrichosis
204	Sepsis	270	Bacterial pneumonia	336	Cat-scratch disease	402	Venezuelan hemorrhagic fever
205	Nauru	271	Amoebiasis	337	Neonatal conjunctivitis	403	Blastocystosis
206	St. Vincent & Grenadines	272	H. respiratory syncytial virus	338	Toxocariasis	404	Tinea barbae
207	Meningitis	273	Athlete's foot	339	Astrovirus	405	Yersiniosis
208	Kiribati	274	Trichomoniasis	340	Fifth disease	406	Tinea nigra
209	Plague (disease)	275	Epidemic typhus	341	Staphylococcal inf.	407	Chromoblastomycosis
210	Saint Kitts and Nevis	276	Hemolytic-uremic syndrome	342	Vibrio parahaemolyticus	408	Dientamoebiasis
211	Typhus	277	Marburg virus	343	Prevotella	409	Brazilian hemorrhagic fever
212	Antigua and Barbuda	278	Trachoma	344	Fatal familial insomnia	410	Gnathostomiasis
213	São Tomé & Príncipe	279	Rhinovirus	345	Anisakis	411	Mycoplasma pneumonia
214	Diphtheria	280	Salmonellosis	346	Ehrlichiosis	412	Capillariasis
215	SARS[a]	281	Coccidioidomycosis	347	VEE[e]	413	White piedra
216	Anthrax	282	Cryptosporidiosis	348	Molluscum contagiosum	414	HME[j]
217	Hepatitis C	283	Myiasis	349	MERS[f]	415	Metagonimiasis
218	Foodborne illness	284	Enterovirus	350	Monkeypox	416	Pediculosis corporis
219	Hepatitis B	285	Chytridiomycosis	351	Fasciolosis	417	Black piedra
220	Ebola virus disease	286	Tularemia	352	Paracoccidioidomycosis	418	H. bocavirus
221	Common cold	287	Kawasaki disease	353	Hand, foot, and mouth disease	419	Ehrlichiosis ewingii inf.
222	Rabies	288	Chikungunya	354	Vibrio vulnificus	420	Desmodesmus
223	Dengue fever	289	Hepatitis D	355	Actinomycosis	421	Rhinosporidiosis
224	Helicobacter pylori	290	Dracunculiasis	356	Ureaplasma urealyticum	422	Free-living Amoebozoa inf.
225	Lyme disease	291	Keratitis	357	Tinea corporis	423	Geotrichosis
226	Chickenpox	292	Lymphatic filariasis	358	Melioidosis	424	A. haemolyticum[k]
227	Staphylococcus	293	Echinococcosis	359	Paragonimiasis	425	Mycetoma

Abbreviations H. and inf. stand for Human and infection. [a] SARS: Severe acute respiratory syndrome. [b] SSPE: Subacute sclerosing panencephalitis. [c] CCHF: Crimean–Congo hemorrhagic fever. [d] PML: Progressive multifocal leukoencephalopathy. [e] VEE: Venezuelan equine encephalitis virus. [f] MERS: Middle East respiratory syndrome. [g] HFRS: Hantavirus hemorrhagic fever with renal syndrome. [h] LCM: Lymphocytic choriomeningitis. [i] GSS: Gerstmann–Sträussler–Scheinker syndrome. [j] HME: Human monocytotropic ehrlichiosis. [k] A. haemolyticum: Arcanobacterium haemolyticum.

TABLE 3. List of countries ordered by PageRank of the corresponding article in English Wikipedia 2017. Here the countries correspond to the $n_c = 195$ sovereign states listed in [28]. The country code (CC) is given for each country.

Rank	Country	CC	Rank	Country	CC	Rank	Country	CC
1	United States	US	66	Morocco	MA	131	Kyrgyzstan	KG
2	France	FR	67	Cuba	CU	132	Guyana	GY
3	Germany	DE	68	Algeria	DZ	133	Trinidad and Tobago	TT
4	United Kingdom	GB	69	Bosnia and Herzegovina	BA	134	Mauritius	MU
5	Iran	IR	70	Ecuador	EC	135	Tajikistan	TJ
6	India	IN	71	Saudi Arabia	SA	136	Monaco	MC
7	Canada	CA	72	Lithuania	LT	137	Oman	OM
8	Australia	AU	73	Iceland	IS	138	Suriname	SR
9	China	CN	74	Bolivia	BO	139	Liberia	LR
10	Italy	IT	75	Tanzania	TZ	140	Solomon Islands	SB
11	Japan	JP	76	Ethiopia	ET	141	Sierra Leone	SL
12	Russia	RU	77	Democratic Republic of the Congo	CD	142	Bhutan	BT
13	Brazil	BR	78	Madagascar	MG	143	The Bahamas	BS
14	Spain	ES	79	Armenia	AM	144	Vatican City	VA
15	Netherlands	NL	80	Lebanon	LB	145	Bahrain	BH
16	Poland	PL	81	Cyprus	CY	146	Barbados	BB
17	Sweden	SE	82	Kazakhstan	KZ	147	Botswana	BW
18	Mexico	MX	83	Georgia (country)	GE	148	Rwanda	RW
19	Turkey	TR	84	Latvia	LV	149	Turkmenistan	TM
20	Romania	RO	85	Panama	PA	150	Benin	BJ
21	New Zealand	NZ	86	Belarus	BY	151	Niger	NE
22	South Africa	ZA	87	Albania	AL	152	Gabon	GA
23	Norway	NO	88	Papua New Guinea	PG	153	Brunei	BN
24	Switzerland	CH	89	Luxembourg	LU	154	Belize	BZ
25	Philippines	PH	90	Ghana	GH	155	Guinea	GN
26	Austria	AT	91	United Arab Emirates	AE	156	Chad	TD
27	Belgium	BE	92	Uruguay	UY	157	Malawi	MW
28	Pakistan	PK	93	North Korea	KP	158	Togo	TG
29	Argentina	AR	94	Yemen	YE	159	Liechtenstein	LI
30	Indonesia	ID	95	Costa Rica	CR	160	Samoa	WS
31	Greece	GR	96	Malta	MT	161	Burundi	BI
32	Denmark	DK	97	Tunisia	TN	162	South Sudan	SS
33	South Korea	KR	98	Jamaica	JM	163	Republic of the Congo	CG
34	Israel	IL	99	Zimbabwe	ZW	164	East Timor	TL
35	Hungary	HU	100	Cambodia	KH	165	Cape Verde	CV
36	Finland	FI	101	Cameroon	CM	166	Eritrea	ER
37	Egypt	EG	102	Mongolia	MN	167	Mauritania	MR
38	Portugal	PT	103	Burkina Faso	BF	168	Central African Republic	CF
39	Ukraine	UA	104	Jordan	JO	169	Maldives	MV
40	Sri Lanka	LK	105	Sudan	SD	170	Tonga	TO
41	Czech Republic	CZ	106	Uganda	UG	171	Andorra	AD
42	Malaysia	MY	107	Republic of Macedonia	MK	172	Vanuatu	VU
43	Peru	PE	108	Guatemala	GT	173	State of Palestine	PS
44	Thailand	TH	109	Libya	LY	174	Lesotho	LS
45	Colombia	CO	110	Dominican Republic	DO	175	The Gambia	GM
46	Bulgaria	BG	111	Haiti	HT	176	Swaziland	SZ
47	Chile	CL	112	Moldova	MD	177	Seychelles	SC
48	Republic of Ireland	IE	113	Somalia	SO	178	Grenada	GD
49	Singapore	SG	114	Ivory Coast	CI	179	San Marino	SM
50	Serbia	RS	115	Namibia	NA	180	Saint Lucia	LC
51	Azerbaijan	AZ	116	Paraguay	PY	181	Palau	PW
52	Vietnam	VN	117	Angola	AO	182	Marshall Islands	MH
53	Nepal	NP	118	Uzbekistan	UZ	183	Equatorial Guinea	GQ
54	Estonia	EE	119	Montenegro	ME	184	Dominica	DM
55	Croatia	HR	120	Kuwait	KW	185	Guinea-Bissau	GW
56	Nigeria	NG	121	Laos	LA	186	Comoros	KM
57	Afghanistan	AF	122	Mozambique	MZ	187	Djibouti	DJ
58	Iraq	IQ	123	Nicaragua	NI	188	Federated States of Micronesia	FM
59	Bangladesh	BD	124	Qatar	QA	189	Tuvalu	TV
60	Syria	SY	125	Senegal	SN	190	Nauru	NR
61	Myanmar	MM	126	Fiji	FJ	191	Saint Vincent and the Grenadines	VC
62	Kenya	KE	127	Mali	ML	192	Kiribati	KI
63	Slovakia	SK	128	Honduras	HN	193	Saint Kitts and Nevis	KN
64	Venezuela	VE	129	Zambia	ZM	194	Antigua and Barbuda	AG
65	Slovenia	SI	130	El Salvador	SV	195	São Tomé and Príncipe	ST

TABLE 4. List of countries ordered by Tuberculosis incidences in 2016 (per 100000 population per year; taken from World Health Organization [38])

Rank	Country	Incidence	Rank	Country	Incidence	Rank	Country	Incidence
1	South Africa	781.0	65	Mauritania	102.0	129	Portugal	20.0
2	Lesotho	724.0	66	Moldova	101.0	130	Serbia	19.0
3	Kiribati	566.0	67	São Tomé and Príncipe	99.0	131	Turkey	18.0
4	Philippines	554.0	68	Niger	93.0	132	Trinidad and Tobago	18.0
5	Mozambique	551.0	69	Guyana	93.0	133	Poland	18.0
6	North Korea	513.0	70	Malaysia	92.0	134	Montenegro	16.0
7	East Timor	498.0	71	Georgia (country)	92.0	135	Japan	16.0
8	Gabon	485.0	72	Ukraine	87.0	136	Estonia	16.0
9	Namibia	446.0	73	Tajikistan	85.0	137	Chile	16.0
10	Papua New Guinea	432.0	74	Solomon Islands	84.0	138	Albania	16.0
11	Marshall Islands	422.0	75	Sudan	82.0	139	Seychelles	15.0
12	Central African Republic	407.0	76	South Korea	77.0	140	Iran	14.0
13	Swaziland	398.0	77	Uzbekistan	76.0	141	Egypt	14.0
14	Indonesia	391.0	78	Romania	74.0	142	Malta	13.0
15	Republic of the Congo	378.0	79	Eritrea	74.0	143	Lebanon	12.0
16	Zambia	376.0	80	Algeria	70.0	144	Croatia	12.0
17	Guinea-Bissau	374.0	81	Kazakhstan	67.0	145	Bahrain	12.0
18	Angola	370.0	82	Russia	66.0	146	Spain	10.0
19	Myanmar	361.0	83	Brunei	66.0	147	Saudi Arabia	10.0
20	Kenya	348.0	84	Azerbaijan	66.0	148	Belgium	10.0
21	Cambodia	345.0	85	Sri Lanka	65.0	149	United Kingdom	9.9
22	Djibouti	335.0	86	China	64.0	150	Costa Rica	9.5
23	Botswana	326.0	87	Turkmenistan	60.0	151	Oman	9.0
24	Democratic Republic of the Congo	323.0	88	El Salvador	60.0	152	Hungary	8.8
25	Liberia	308.0	89	Dominican Republic	60.0	153	Tonga	8.6
26	Sierra Leone	304.0	90	Fiji	59.0	154	Sweden	8.2
27	Tanzania	287.0	91	Benin	59.0	155	Austria	8.2
28	Somalia	270.0	92	Vanuatu	56.0	156	Germany	8.1
29	Pakistan	268.0	93	Mali	56.0	157	Switzerland	7.8
30	Madagascar	237.0	94	Panama	55.0	158	Dominica	7.8
31	Bangladesh	221.0	95	Lithuania	53.0	159	Samoa	7.7
32	Nigeria	219.0	96	Belarus	52.0	160	France	7.7
33	India	211.0	97	Singapore	51.0	161	New Zealand	7.3
34	Zimbabwe	208.0	98	Burkina Faso	51.0	162	Republic of Ireland	7.1
35	Tuvalu	207.0	99	Rwanda	50.0	163	Cuba	6.9
36	Cameroon	203.0	100	Ecuador	50.0	164	Slovenia	6.5
37	Uganda	201.0	101	Maldives	49.0	165	Grenada	6.4
38	Afghanistan	189.0	102	Yemen	48.0	166	Saint Vincent and the Grenadines	6.3
39	Haiti	188.0	103	Nicaragua	48.0	167	Norway	6.1
40	Mongolia	183.0	104	Togo	46.0	168	Italy	6.1
41	Equatorial Guinea	181.0	105	Armenia	44.0	169	Denmark	6.1
42	Bhutan	178.0	106	Iraq	43.0	170	Australia	6.1
43	Federated States of Micronesia	177.0	107	Paraguay	42.0	171	Andorra	6.0
44	Ethiopia	177.0	108	Brazil	42.0	172	Slovakia	5.9
45	Guinea	176.0	109	Libya	40.0	173	Netherlands	5.9
46	Laos	175.0	110	Honduras	40.0	174	Luxembourg	5.8
47	The Gambia	174.0	111	Tunisia	38.0	175	Jordan	5.6
48	Thailand	172.0	112	Belize	38.0	176	Cyprus	5.6
49	Malawi	159.0	113	Latvia	37.0	177	Canada	5.2
50	Ghana	156.0	114	Comoros	35.0	178	Czech Republic	5.0
51	Nepal	154.0	115	Venezuela	32.0	179	Finland	4.7
52	Ivory Coast	153.0	116	Colombia	32.0	180	Jamaica	4.5
53	Chad	153.0	117	Bosnia and Herzegovina	32.0	181	Greece	4.4
54	South Sudan	146.0	118	Uruguay	29.0	182	Israel	3.5
55	Kyrgyzstan	145.0	119	Bulgaria	27.0	183	Antigua and Barbuda	3.4
56	Senegal	140.0	120	Suriname	26.0	184	United States	3.1
57	Cape Verde	137.0	121	The Bahamas	26.0	185	Iceland	2.1
58	Vietnam	133.0	122	Kuwait	24.0	186	Saint Lucia	1.9
59	Palau	123.0	123	Guatemala	24.0	187	Barbados	1.2
60	Burundi	118.0	124	Argentina	24.0	188	United Arab Emirates	0.79
61	Peru	117.0	125	Qatar	23.0	189	San Marino	0.0
62	Bolivia	114.0	126	Mexico	22.0	190	Saint Kitts and Nevis	0.0
63	Nauru	112.0	127	Mauritius	22.0	191	Monaco	0.0
64	Morocco	103.0	128	Syria	21.0			

TABLE 5. List of countries ordered by new HIV/AIDS infections in 2017 (per 1000 uninfected population; taken from World Health Organization [38])

Rank	Country	New infections	Rank	Country	New infections	Rank	Country	New infections
1	Lesotho	12.68	44	Ethiopia	0.33	87	Azerbaijan	0.1
2	Swaziland	9.37	45	Cape Verde	0.31	88	Spain	0.09
3	South Africa	5.58	46	Trinidad and Tobago	0.29	89	Peru	0.09
4	Botswana	5.52	47	Cuba	0.29	90	Niger	0.09
5	Namibia	4.37	48	Georgia (country)	0.28	91	Lithuania	0.09
6	Zambia	4.08	49	Chile	0.28	92	France	0.09
7	Mozambique	3.63	50	Dominican Republic	0.24	93	Armenia	0.09
8	Zimbabwe	3.03	51	Brazil	0.24	94	Senegal	0.08
9	Equatorial Guinea	2.71	52	Latvia	0.23	95	Albania	0.08
10	Malawi	2.29	53	Myanmar	0.22	96	Sweden	0.06
11	Central African Republic	1.8	54	Venezuela	0.21	97	Nicaragua	0.06
12	Republic of the Congo	1.65	55	Paraguay	0.2	98	Malta	0.06
13	Uganda	1.5	56	Burundi	0.2	99	Italy	0.06
14	Kenya	1.46	57	Malaysia	0.19	100	Republic of Ireland	0.06
15	Cameroon	1.39	58	Indonesia	0.19	101	Iran	0.06
16	South Sudan	1.35	59	Costa Rica	0.19	102	India	0.06
17	Nigeria	1.23	60	Burkina Faso	0.19	103	Australia	0.05
18	Tanzania	1.19	61	Madagascar	0.18	104	Yemen	0.04
19	Angola	0.94	62	Luxembourg	0.18	105	Romania	0.04
20	Gabon	0.92	63	Guatemala	0.18	106	Czech Republic	0.04
21	Sierra Leone	0.86	64	Somalia	0.17	107	Cambodia	0.04
22	Ivory Coast	0.86	65	Democratic Republic of the Congo	0.17	108	Bahrain	0.04
23	Ghana	0.78	66	Kazakhstan	0.16	109	Tunisia	0.03
24	Haiti	0.77	67	El Salvador	0.16	110	Sri Lanka	0.03
25	Guyana	0.77	68	Uruguay	0.15	111	Slovenia	0.03
26	Belize	0.75	69	Tajikistan	0.15	112	Serbia	0.03
27	Guinea-Bissau	0.72	70	Eritrea	0.15	113	Netherlands	0.03
28	Rwanda	0.7	71	Sudan	0.13	114	Nepal	0.03
29	Guinea	0.67	72	Kyrgyzstan	0.13	115	Morocco	0.03
30	Liberia	0.66	73	Argentina	0.13	116	Bulgaria	0.03
31	The Gambia	0.65	74	Vietnam	0.12	117	Afghanistan	0.03
32	Jamaica	0.63	75	Mauritania	0.12	118	Slovakia	0.02
33	Suriname	0.62	76	Fiji	0.12	119	Saudi Arabia	0.02
34	Togo	0.59	77	Ecuador	0.12	120	Qatar	0.02
35	Djibouti	0.58	78	Colombia	0.12	121	Lebanon	0.02
36	Barbados	0.58	79	Philippines	0.11	122	Kuwait	0.02
37	Ukraine	0.38	80	Montenegro	0.11	123	Egypt	0.02
38	Moldova	0.38	81	Honduras	0.11	124	Croatia	0.02
39	Papua New Guinea	0.37	82	Thailand	0.1	125	Algeria	0.02
40	Panama	0.34	83	Pakistan	0.1	126	Mongolia	0.01
41	Chad	0.34	84	Mexico	0.1	127	Jordan	<0.01
42	Benin	0.34	85	Laos	0.1	128	Comoros	<0.01
43	Mali	0.33	86	Bolivia	0.1	129	Bangladesh	<0.01

TABLE 6. List of infectious diseases ordered by the estimated number of worldwide deaths caused in 2010. Data are taken from Global Burden of Disease Study 2010 [50].

Rank	Diseases	Number of deaths per year ($\times 10^3$)
1	HIV/AIDS	1465.4
2	Pneumonia	1460.7
3	Tuberculosis	1196.0
4	Malaria	1169.5
5	Meningitis	422.9
6	Rotavirus	250.9
7	Typhoid fever	190.2
8	Hepatitis B	132.2
9	Measles	125.4
10	Shigellosis	122.8
11	Syphilis	113.3
12	Campylobacteriosis	109.7
13	Hepatitis A	102.8
14	Cryptosporidiosis	99.8
15	Pertussis	81.4
16	Salmonellosis	81.3
17	Tetanus	61.3
18	Cholera	58.1
19	Hepatitis E	56.6
20	Amoebiasis	55.5
21	Leishmaniasis	51.6
22	Rabies	26.4
23	Hepatitis C	16.0
24	Dengue fever	14.7
25	Schistosomiasis	11.7
26	Chagas disease	10.3
27	African trypanosomiasis	9.1
28	Chickenpox	6.8
29	Diphtheria	2.9
30	Ascariasis	2.7
31	Chlamydia infection	1.2
32	Echinococcosis	1.2
33	Cysticercosis	1.2
34	Gonorrhea	0.9

GUILLAUME ROLLIN received a Master (M.Sc.) degree (2012) and a Ph.D. degree (2015) in theorical physics from Franche-Comté University, Besançon, France. His thesis was devoted to the chaotic dynamics in the restricted three body problem. He received a Master (M.A.) degree (2018) in Epistemology, Logic and History of Science from Lorraine University, Nancy, France. Currently, he is Post-Doctoral resercher at the theorical physics and astrophysics team of the UTINAM institute at Bourgogne Franche-Comté University (UBFC). His fields of research are dynamical chaos, restricted three body problem, Google matrix of directed networks and epistemology.

JOSÉ LAGES received a DEA (M.Sc.) degree in theoretical physics from University of Montpellier, France (1998), and a Ph.D. degree in theoretical physics from University Paul Sabatier, Toulouse, France (2001). Then, he was Post-Doctoral researcher at the Condensed Matter Theory group of the Center for Fundamental Interactions, Instituto Superior Técnico, Lisbon, at Ames Laboratory, Iowa State University, USA, and at the Laboratoire de Physique Théorique, University Paul Sabatier, Toulouse, France. Since 2005, he is Associate Professor at Bourgogne Franche-Comté University (UBFC). He is currently working at UTINAM institute, UBFC, Besançon. His fields of research are electronic transport in disordered systems, quantum computing, Google matrix of directed networks, and dynamical chaos in Solar System and galaxies.

DIMA L. SHEPELYANKY received a Diploma degree in physics from Novosibirsk State University in 1978, the Ph.D. and Russian Doctor degrees from the Institute of Nuclear Physics of Russian Academy of Sciences, Novosibirsk, Russia, in 1982 and 1989, respectively. He has been a Researcher with CNRS Toulouse, France, since 1991, and the Directeur de Recherche with the Laboratoire de Physique Théorique du CNRS, Toulouse, since 1994. His research fields include classical and quantum chaos, nonlinear waves, dark matter dynamics, atom ionization in strong fields, electronic transport, quantum computing, Markov chains, and Google matrix of directed networks.

• • •