Interactions of pharmaceutical companies with world countries, cancers and rare diseases from Wikipedia network analysis

Guillaume Rollin[1], José Lages[1], Tatiana S. Serebriyskaya[2] Dima L. Shepelyansky[3]

1 Institut UTINAM, CNRS, UMR 6213, OSU THETA, Université de Bourgogne Franche-Comté, Besançon, France
2 Laboratory for Translational Research and Personalized Medicine, Moscow Institute of Physics and Technology, Moscow, Russia
3 Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

guillaume.rollin@utinam.cnrs.fr
jose.lages@utinam.cnrs.fr
ts.serebriyskaya@gmail.com
dima@irsamc.ups-tlse.fr

Abstract

Using English Wikipedia network of more than 5 million articles we analyze interactions and interlinks between 34 largest pharmaceutical companies, 195 world countries, 47 rare renal diseases and 37 types of cancer. The recently developed algorithm of reduced Google matrix (REGOMAX) allows to take into account direct Markov transitions between these articles but also all indirect ones generated by the pathways between these articles via the global Wikipedia network. Thus this approach provides a compact description of interactions between these articles that allows to determine the friendship networks between articles, the PageRank sensitivity of countries to pharmaceutical companies and rare renal diseases. We also show that the top pharmaceutical companies of Wikipedia PageRank are not those of the top list of market capitalization.

Introduction

The improvement of human health and its treatment from various diseases is the vital task of human society [1]. The creation of efficient medicaments and drugs is now mainly controlled by large biotechnology and pharmaceutical companies. The 34 world largest companies are listed in Wikipedia [2]. The analysis of interactions between these companies and their influence to various diseases is an important but not easy task. Here we develop the data mining approach to this task using the directed network of articles of English Wikipedia, dated by May 2017, generated by citation links between articles. At present Wikipedia accumulates a huge amount of human knowledge exceeding the one of Encyclopedia Britanica in volume and accuracy of articles devoted to scientific topics [3]. Scientific articles are actively maintained as it is shown by the example of articles on biomolecules [4]. The academic research and analysis of Wikipedia information is growing with the development of new tools and methods as reviewed in [5]. The quality of Wikipedia articles is improving with time as it is shown by the analysis reported in [6].

At present there is a variety of methods for the analysis and the characterization of complex networks (see e.g. [7]). The most used of them is the PageRank algorithm invented by Brin and Page in 1998 for ranking the World Wide Web (WWW) sites [8]. The detailed descriptions of the algorithm and the mathematical properties of the related Google matrix are given in [9] while a variety of applications of the Google matrix to real directed networks are described in [10]. The applications of Google matrix methods to Wikipedia networks of 24 language editions allowed to obtain a reliable ranking of historical figures over 15 centuries of human history [11] and of world universities [12].

Recently the reduced Google matrix (REGOMAX) algorithm has been proposed for the analysis of the effective interactions and links between a selected subset of nodes of interest embedded in a much larger global network [13]. The REGOMAX algorithm originates from the scattering theory of nuclear and mesoscopic physics and field of quantum chaos. Its efficiency has been demonstrated for Wikipedia networks recovering effective interactions between politicians [14], geopolitical relations and links between world countries [15], universities [16] and banks [17].

The applications of REGOMAX approach were also performed for networks of protein-protein interactions [18], and for Wikipedia networks to determine influence of infectious diseases [19], drugs and cancers [20]. Here we extend this approach to analyze the global influence and their mutual interactions of the 34 world largest biotechnology and pharmaceutical companies [2]. We use the Wikipedia network with 5 416 537 articles and establish sensitivities of 195 countries to these companies. We also construct the interaction networks between these companies and 47 rare renal diseases and 37 types of cancer. The number of people suffered from cancer was about 32.6 million worldwide in 2012 with annual increase 6-9 million people [21]. The number of people suffered from rare diseases reaches 350 million globally [22]. Despite cancer and 80% of rare diseases are genetic diseases, development of drugs for these two groups of disease meets some different challenges. One of fundamental challenge of drug development for rare diseases is their huge number and diversity. It's known more than 7 000 rare diseases [22] whereas number of known cancer types is quite limited. Big number and diversity rare diseases leads to the problems of drug development because small number of patients for individual disease, the logistics involved in reaching widely dispersed patients, the lack of validated biomarkers and surrogate end-points, and limited clinical expertise and expert centers [23]. Representation of information about cancer and rare diseases in Wikipedia clearly illustrates difference between cancers and rare diseases. Most of cancer types listed on site of National Cancer Institute [24] are represented in Wikipedia. In the same time only 8.5% and 15% of phenotype described in Orphanet and OMIM correspondingly have topics in Wikipedia. For our calculation we use only 47 rare renal diseases listed in [25] and represented on Wikipedia.

We hope that our results will allow to obtain a better understanding of the pharmaceutical companies influence on these diseases and to obtain a specialization profiles of pharmaceutical companies.

The paper is constructed as follows: the mathematical methods are described in Section Methods and data are overviewed in Section Datasets. The two following Sections are devoted to the presentation of Results and of the Discussion.

Methods

Google matrix construction

The detailed description of Google matrix construction and its properties are given in [9, 10] so that here we provide only a brief description for reader convenience.

The Google matrix G is built from the adjacency matrix A_{ij} with elements being 1 if article (node) j points to article (node) i and zero otherwise. The matrix elements have the standard form $G_{ij} = \alpha S_{ij} + (1-\alpha)/N$ [8–10], where S is the matrix describing Markov transitions with elements $S_{ij} = A_{ij}/k_{out}(j)$ if the out-degree of node j is $k_{out}(j) = \sum_{i=1}^{N} A_{ij} \neq 0$, or $S_{ij} = 1/N$ if j has no outgoing links (dangling node), i.e., $k_{out}(j) = 0$. Here $0 < \alpha < 1$ is the damping factor and below we use its standard value $\alpha = 0.85$ [9]. We note that for the range $0.5 \leq \alpha \leq 0.95$ the results are not sensitive to α [9, 10]. Thus, for a random surfer, jumping from one node to another, the probability to jump to any node is $(1-\alpha)$.

The leading right eigenvector P of G, called the PageRank vector, is the solution of the equation $GP = \lambda P$ for the unit eigenvalue $\lambda = 1$. According to the Perron-Frobenius theorem the PageRank vector components P_j give probabilities to find a random surfer on a node j, therefore $\sum_j P_j = 1$ [9]. We order all nodes by decreasing probability P_j 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 probabilities is done efficiently with the PageRank iteration algorithm described in [8, 9].

It is also useful to consider the original network with inverted direction of links. After inversion the Google matrix G^* is constructed via the same procedure with $G^* P^* = P^*$. The matrix G^* has its own PageRank vector P^* called CheiRank [26, 27] (see also [10]). Its components are also probabilities that can be again ordered in a decreasing order with CheiRank index K^* with highest component of P^* at $K^* = 1$ and smallest at $K^* = N$. On average, the high values of P (P^*) correspond to nodes with many ingoing (outgoing) links [9, 10]. Thus PageRank stress the article importance on the network while CheiRank highlights the communicative properties of articles.

Reduced Google matrix

The REGOMAX algorithm was proposed in [13] and it is described in detail in [14–16, 18]. Here we give the main elements of this method using the notations of [14].

This algorithm computes efficiently a *reduced Google matrix* of size $N_r \times N_r$ that captures the full contributions of direct and indirect pathways happening in the full Google matrix between N_r nodes of interest. For the selected N_r nodes their PageRank probabilities 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 the N_r nodes is unity. The computation of G_{R} determines also a decomposition of G_{R} into matrix components that clearly distinguish direct from indirect interactions, $G_{\mathrm{R}} = G_{rr} + G_{\mathrm{pr}} + G_{\mathrm{qr}}$ [14]. Here G_{rr} is given by the direct links between selected N_r nodes in the global matrix G with N nodes. In fact, G_{pr} is rather close to the matrix in which each column is equal to the PageRank vector P_r. This ensures that PageRank probabilities of G_{R} are the same as for G (up to a constant multiplier). Due to that G_{pr} doesn't provide much information about direct and indirect links between selected nodes.

Thus the most interesting and nontrivial role is played by G_{qr}. It takes into account all indirect links between selected nodes emerging due to multiple pathways via the global network nodes N (see [13, 14]). The matrix $G_{\mathrm{qr}} = G_{\mathrm{qrd}} + G_{\mathrm{qrnd}}$ has diagonal (G_{qrd}) and non-diagonal (G_{qrnd}) parts with G_{qrnd} including indirect interactions between nodes. The exact formulas for the three components of G_{R} are given in [13, 14].

A useful additional characteristic provided by G_{R} matrix is the sensitivity of the PageRank probability to the variation of a specific link between a pair of nodes chosen among the N_r nodes of interest. The useful results obtained with this method have been demonstrated in [15–17, 19, 20]. Thus, in our case, the sensitivity of a country c to a specific pharmaceutical company ph is determined by a change of matrix element

$G_{\mathrm{R}\ c,ph}$ by a factor $(1+\delta)$ with $\delta \ll 1$ and renormalization to unity of the sum of the column matrix elements associated with pharmaceutical company ph. Then the sensitivity is defined by the logarithmic derivative of PageRank probability $P(c)$ associated to country c, $D(ph \to c, c) = d\ln P(c)/d\delta$ (diagonal sensitivity). In a similar way we determine the sensitivity of countries to a rare renal disease rd: $D(rd \to c, c) = d\ln P(c)/d\delta$.

Table 1. List of the 34 largest pharmaceutical companies ranked by the relative PageRank index K_r of their corresponding articles in Wikipedia. The K_{LMC} index gives the ranking by the largest market capitalization since 2000 [2] and the K_{MC} index gives the ranking by market capitalization in 2017 [2].

K_r	K_{LMC}	K_{MC}	Company	K	K_{LMC}	K_{MC}	Company
1	2	3	Pfizer	18	18	18	Biogen
2	8	15	GSK[a]	19	28	28	Mylan
3	14	10	Bayer	20	7	13	Gilead
4	1	1	J&J[b]	21	21	21	Shire
5	3	4	Novartis	22	23	23	Takeda
6	6	6	Merck	23	29	27	Astellas
7	16	14	Lilly	24	32	29	Daiichi Sankyo
8	4	2	Roche	25	5	5	AbbVie
9	19	16	AstraZeneca	26	22	22	Regeneron
10	10	9	Sanofi	27	33	31	Eisai
11	15	11	BMS[c]	28	20	19	Stryker
12	13	12	Abbott	29	34	33	BioMarin
13	24	24	Illumina	30	27	26	Zoetis
14	11	8	Amgen	31	25	25	Vertex
15	9	7	Novo Nordisk	32	26	30	Alexion
16	12	20	Allergan	33	31	34	Perrigo
17	17	17	Celgene	34	30	32	Incyte

[a]GSK: GlaxoSmithKline, [b]J&J: Johnson & Johnson, [c]BMS: Bristol-Myers Squibb.

Datasets

We consider the English Wikipedia edition collected in May 2017 with $N = 5416537$ articles (nodes) and $N_l = 122232932$ hyperlinks between articles. This network has been considered also in [16, 17, 19, 20]. For the REGOMAX analysis we select $N_c = 195$ world countries (see the list and PageRank order in [19, 20]), $N_{ph} = 34$ of the largest pharmaceutical companies (see Table 1), $N_{rd} = 47$ rare renal diseases (see Table 2), and $N_{cr} = 37$ types of cancer listed in [20]. Thus in total we consider $N_r = N_c + N_{ph} + N_{rd} + N_{cr} = 313$ Wikipedia articles as nodes of interest. Compared to the studies reported in [20] the new nodes correspond to $N_{ph} = 34$ pharmaceutical companies and $N_{rd} = 47$ rare renal diseases. The relative PageRank order of the $N_{ph} = 34$ companies is given in Table 1, and of the $N_{rd} = 47$ rare renal diseases in Table 2.

In Table 2, the 47 renal rare diseases can be grouped in five categories to which we have for convenience associated a color: congenital abnormalities of the kidney and urinary tract (red), glomerular diseases (blue), renal tubular diseases and metabolic diseases (gold), nephrolithiasis (cyan), and ciliopathies and nephronophthisis (green) in manner they were classified in [25]. From 166 genetic disorders of renal growth,

Table 2. List of 47 rare renal diseases ranked by the relative PageRank index K_r of their corresponding articles in Wikipedia. The relative CheiRank index, K_r^*, is also given. The list of rare renal diseases is splitted in 5 categories with the following color code: 🟥 Congenital abnormalities of the kidney and urinary tract, 🟦 Glomerular diseases, 🟧 Renal tubular diseases and metabolic diseases, 🔷 Nephrolithiasis and 🟩 Ciliopathies and Nephronophthisis.

K_r	K_r^*		Rare renal disease	Short name
1	6	🟦	Alport syndrome	Alport
2	1	🟩	Bardet–Biedl syndrome	Bardet–Biedl
3	7	🟦	Fabry disease	Fabry
4	2	🟥	Kallmann syndrome	KS
5	4	🟧	Renal tubular acidosis	RTA
6	5	🔷	Cystinuria	Cystinuria
7	14	🟥	Renal agenesis	Renal agenesis
8	16	🟩	Nephronophthisis	Nephronophthisis
9	33	🟧	X-linked hypophosphatemia	XLH
10	8	🟧	Bartter syndrome	Bartter
11	41	🔷	Xanthinuria	Xanthinuria
12	3	🟧	Oculocerebrorenal syndrome	Oculocerebrorenal
13	17	🟦	Nail–patella syndrome	Nail-patella
14	45	🟦	Familial renal amyloidosis	Amyloidosis
15	9	🟧	Episodic ataxia	EA
16	24	🟧	Gitelman syndrome	Gitelman
17	21	🟧	Medullary cystic kidney disease	MCKD
18	32	🟧	Familial hypocalciuric hypercalcemia	FHH
19	42	🟧	Renal glycosuria	Glycosuria
20	40	🟧	Fanconi-Bickel syndrome	Fanconi-Bickel
21	23	🟧	Liddle's syndrome	Liddle
22	31	🟧	Hypomagnesemia with secondary hypocalcemia	HSH
23	13	🟥	Fraser syndrome	Fraser
24	36	🟦	WAGR syndrome	WAGR
25	19	🟥	Branchio-oto-renal syndrome	BOR
26	26	🟥	Townes–Brocks syndrome	Townes–Brocks
27	44	🟧	Autosomal dominant hypophosphatemic rickets	ADHR
28	18	🟩	Orofaciodigital syndrome 1	Orofaciodigital
29	28	🟦	Denys–Drash syndrome	DDS
30	10	🟥	Perlman syndrome	Perlman
31	11	🟥	Simpson–Golabi–Behmel syndrome	SGBS
32	12	🟩	Caroli disease	Caroli
33	25	🟦	Congenital nephrotic syndrome	GNS
34	39	🟦	Fechtner syndrome	Fechtner
35	27	🟧	Abderhalden–Kaufmann–Lignac syndrome	AKL
36	22	🟧	Lysinuric protein intolerance	LPI
37	47	🟩	Majewski's polydactyly syndrome	Majewski
38	20	🟥	Urofacial syndrome	Urofacial
39	15	🟥	Barakat syndrome	Barakat
40	35	🟧	Dicarboxylic aminoaciduria	DCBXA
41	38	🟥	MODY 5	MODY 5
42	30	🟥	Papillorenal syndrome	Papillorenal
43	43	🟧	OCRL	OCRL
44	37	🟦	COQ9	COQ9
45	29	🟩	Juvenile nephronophthisis	J. nephronophthisis
46	46	🟩	Renal-hepatic-pancreatic dysplasia	RHPD
47	34	🟩	Conorenal syndrome	Conorenal

structure renal function listed in article for 47 phenotypes (28%) only corresponding articles were found on Wikipedia.

Results

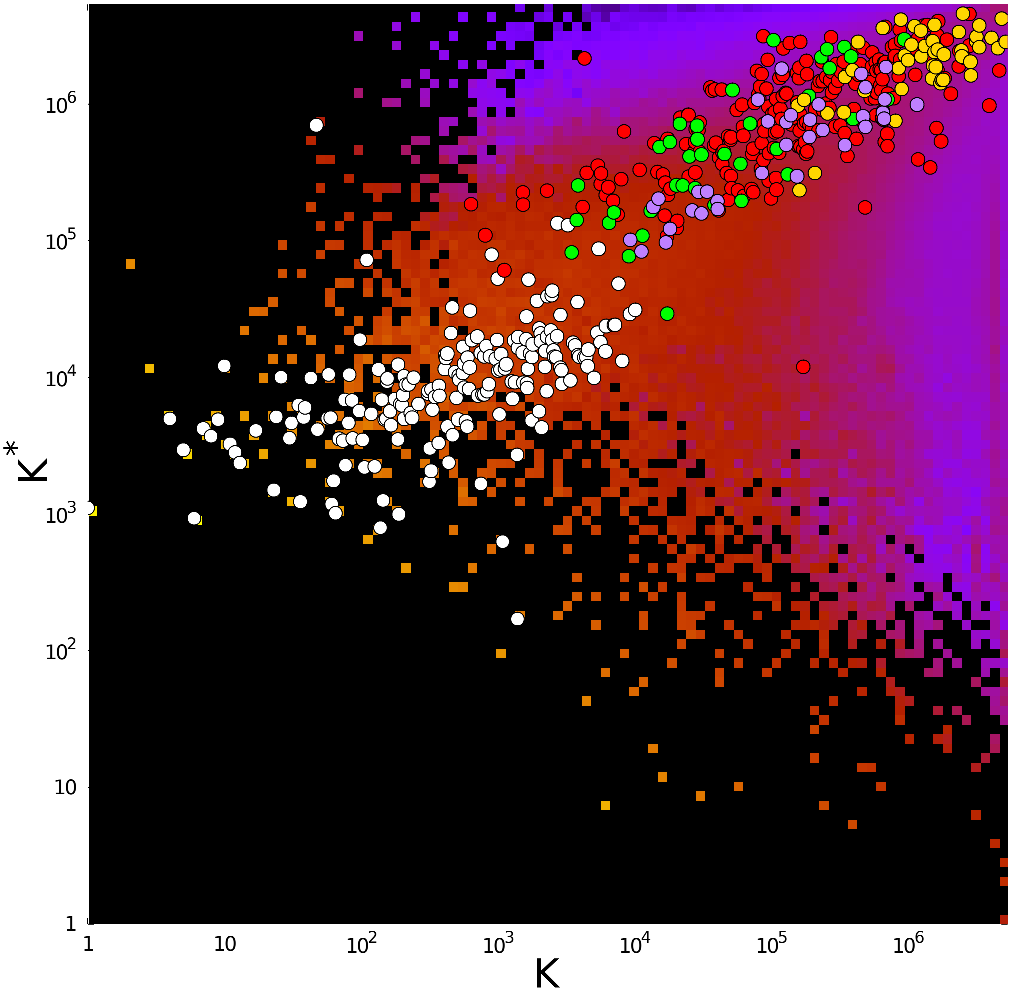

Fig 1. Density of Wikipedia articles in the PageRank-CheiRank plane (K, K^*). Data are averaged over a 100×100 grid for $(\log_{10} K, \log_{10} K^*)$ spanning the domain $[0, \log_{10} N] \times [0, \log_{10} N]$. Density of articles ranges from very low density (purple tiles) to very high density (bright yellow tiles). The absence of article is represented by black tiles. The superimposed white circles give the positions of articles devoted to 195 countries, the red circles are the positions of articles devoted to 230 kind of infectious diseases studied in [19], the green circles are the positions of 37 articles of cancers studied in [20], the gold circles mark the positions of 47 articles of rare renal diseases and the purple circles give the positions of 34 articles of pharmaceutical companies.

PageRank vs CheiRank distributions

In Fig. 1 we show the coarse-grained distribution of all N Wikipedia articles on PageRank-CheiRank plane (K, K^*) in logarithmic scale. On this plane we plot positions of 195 countries (white circles), 37 cancer types (green circles), 47 rare renal diseases (gold circles), 34 pharmaceutical companies (purple circles) and we also add the 230 infectious diseases (red circles) studied in [19]. As usual countries take the top PageRank positions (see also [10]), then we have the groups of infectious diseases and cancers. The group of rare diseases starts approximately at the middle of the group of cancers that corresponds to the fact that these 47 diseases are rare. The head of parmaceutical companies group is a bit behind the head of the group of cancers. In the global PageRank list the top 3 of each group are $K = 1$ US, $K = 2$ France, $K = 3$ Germany for countries, $K = 639$ tuberculosis, $K = 810$ HIV/AIDS, $K = 1116$ malaria for infectious diseases [19], $K = 3478$ lung cancer, $K = 3788$ breast cancer, $K = 3871$ leukemia for cancers [20], $K = 9345$ Pfizer, $K = 11290$ GlaxoSmithKline, $K = 13737$ Bayer for pharmaceutical companies, $K = 156963$ Alport syndrome, $K = 161731$ Bardet–Biedl syndrome, $K = 174780$ Fabry disease for rare renal diseases.

In Table 1 we give the PageRank relative rank K_r list of 34 pharmaceutical companies. We also give in this table the rank K_{LMC} of these companies from the largest market capitalization ranking [2] and the rank K_{MC} from the market capitalization ranking in 2017 [2].

The distribution of companies on the rank plane (K_r, K_{LMC}) is shown in Fig. 2. The richest company Johnson & Johnson (J&J) is only at the 4th position in PageRank $(K_r = 4)$. This shows that its public influence in Wikipedia is well behind its first position in market capitalization. We can also note that AbbVie, spin-off of Abbott laboratories, which performed the 5th most important market capitalization since 2000 is with $K_r = 25$ among the less influential of the considered pharmaceutical companies. In contrast Pfizer is the most influential company on Wikipedia with $K_r = 1$ and $K_{LMC} = 2$. A significant influence in Wikipedia is found for pharmaceutical companies with a top K_r position and a somewhat worse K_{LMC}, such companies are GlaxoSmithKline (GSK) with $K_r = 2$ and $K_{LMC} = 8$ and Bayer with $K_r = 3$ and $K_{LMC} = 14$. As a summary, if the influence of the pharmaceutical companies were strictly proportional to their richness, the companies would lie on the diagonal $(K_r = K_{LMC})$, thus here companies above (below) the diagonal have an excess (lack) of influence in Wikipedia in comparison with their market capitalization.

The overlap $\eta(j)$ between the two ranking lists K_r and K_{LMC} (see Table 1) and also between the two ranking lists K_r and K_{MC} (see Table 1) are shown in Fig. 3. We see that for the first three positions the overlap is only $\eta = 1/3$. This shows that the most rich pharmaceutical companies still can improve significantly their public visibility at Wikipedia. At present Wikipedia web site is at the 5th position among the most visited web sites of the world [28] and the improvement of Wikipedia article content can bring significantly better world visibility of a company, being free of charge. Nevertheless 7 of the top 10 most influential pharmaceutical companies in Wikipedia are also among the top 10 largest market capitalization since 2000 and among the top 10 market capitalization in 2017.

The distribution of 47 articles of rare renal diseases on the PageRank-CheiRank plane if shown in Fig. 4. The top three PageRank positions are taken by Alport syndrome, Bardet-Biedl syndrome, Fabry disease $(K_r = 1, 2, 3)$ so that these are most influential rare renal diseases. The top three CheiRank positions are Bardet-Biedl syndrome, Kallmann syndrome, Oculocerebrorenal syndrome $(K_r^* = 1, 2, 3)$ thus being the most communicative articles of among the 47 articles devoted to rare renal diseases.

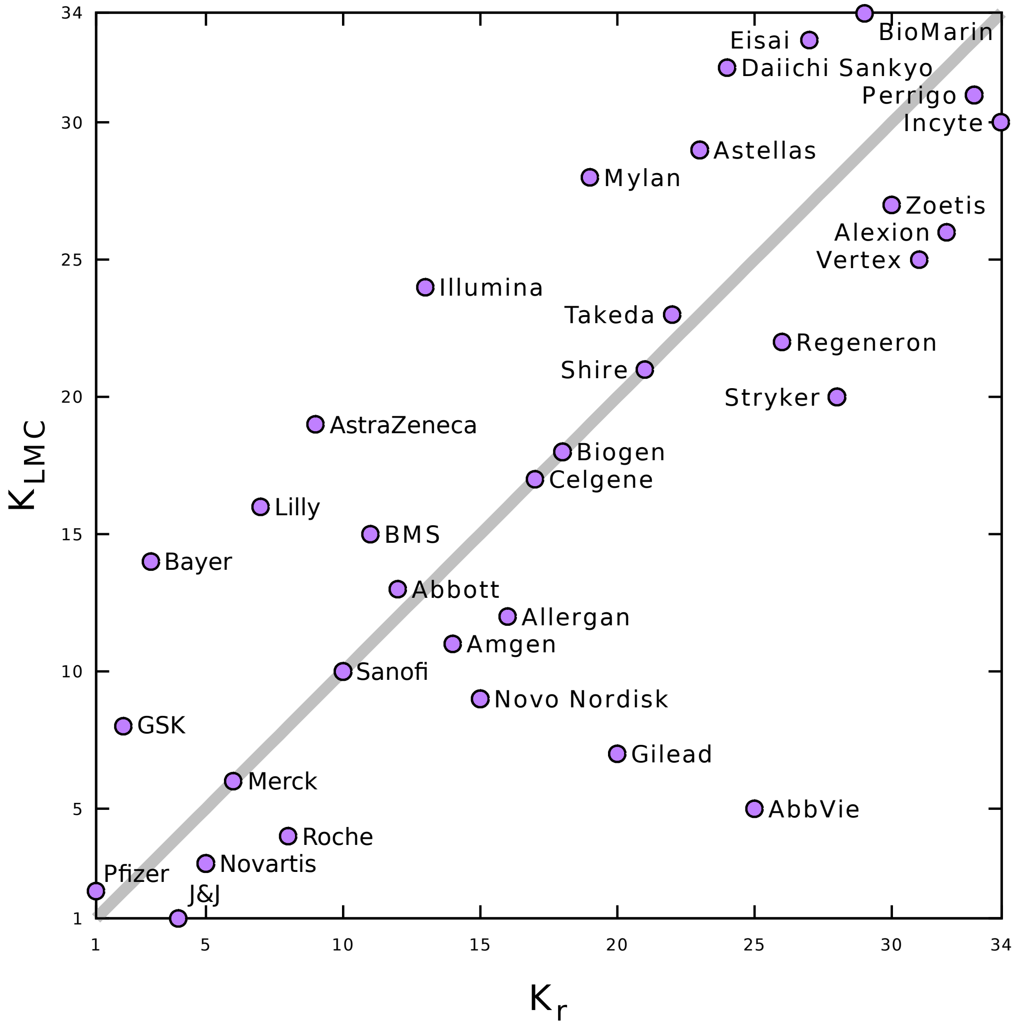

Fig 2. Distribution of pharmaceutical companies ranked by the largest market capitalization index, K_{LMC}, and by the relative PageRank index, K_r, of their article in Wikipedia. See rankings in Table 1.

Example of reduced Google matrix $G_{\rm R}$

We will consider Wikipedia articles (nodes of interest) among the 313 selected articles (nodes) including countries ($N_c = 195$), pharmaceutical companies ($N_{ph} = 34$), cancers ($N_{cr} = 37$) and rare renal diseases ($N_{rd} = 47$). Since the interactions between countries and cancers have been analyzed in [20], we show below other interactions such as between pharmaceutical companies and countries, cancers, rare renal diseases, and between rare renal diseases and countries. In Fig. 5 we show the 81 × 81 reduced Google matrix $G_{\rm R}$ (top left panel) and its G_{rr} (top right panel), $G_{\rm pr}$ (bottom left panel), $G_{\rm qr}$ (bottom right panel) components associated to pharmaceutical companies and rare renal diseases. Focusing on $G_{\rm R}$ (Fig. 5 top left panel), the first 34 × 34 block diagonal sub-matrix (delimited by the purple square) gives the effective directed links between pharmaceutical companies. The remaining 47 × 47 block diagonal sub-matrix (sub-divided in five block diagonal sub-matrices delimited by colored square; the rare renal diseases color code is given in Table 2) gives the effective directed links between rare renal diseases. The 47 × 34 (34 × 47) rectangular sub-matrix gives the effective

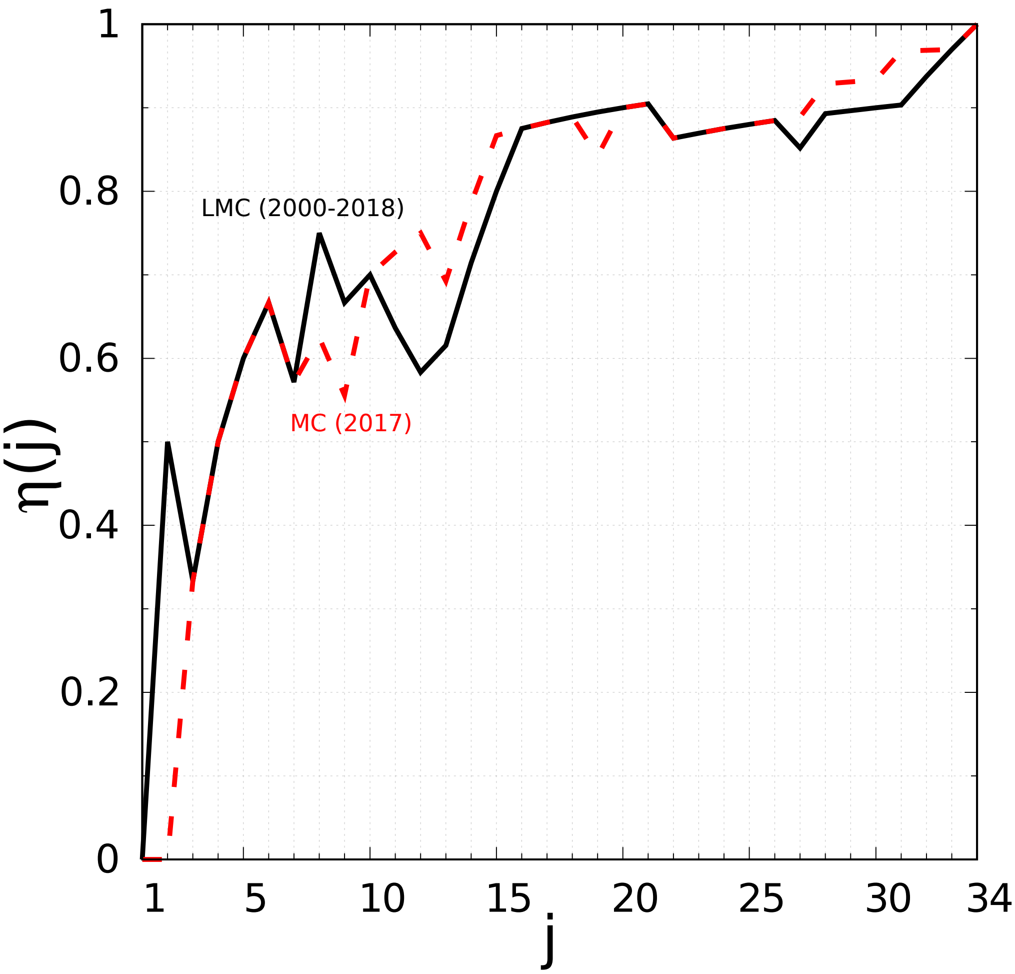

Fig 3. Overlap between the PageRanking of pharmaceutical companies articles in Wikipedia and the ranking of pharmaceutical companies by market capitalizations. The overlap function is $\eta(j) = j_{ph}/j$ where j_{ph} is the number of common pharmaceutical companies in the top j of two lists: the PageRanking of pharmaceutical companies articles in Wikipedia (see first column of Tab. 1) and the list of pharmaceutical companies ranked by market capitalization in 2017 (red curve, see third column in Tab. 1) and by the largest market capitalization since 2000 (black curve, see second column Tab. 1).

directed links from (pointing to) pharmaceutical companies to (from) rare renal diseases. We clearly see that the G_{pr} components gives us back information about the global PageRank, indeed each column are similar. We retrieve the fact that Wikipedia articles about top pharmaceutical companies are better PageRanked that those concerning rare renal diseases. The 22 most influential companies have a global PageRank index from $K \sim 10^4$ to $K \sim 10^5$ (see Fig. 1), whereas the most influential rare renal disease have a PageRank index above $K \sim 10^5$. This is also clearly seen in the G_{pr} matrix picture (Fig. 5, bottom left panel). The direct links are visible on the G_{rr} matrix (Fig. 5 top right panel) which gives a picture of the adjacency matrix A. We observe that direct links are denser in the block diagonal of pharmaceutical companies than in the rare renal diseases one. In the pharmaceutical companies sector we have many direct links with almost every Wikipedia articles devoted to a pharmaceutical company pointing

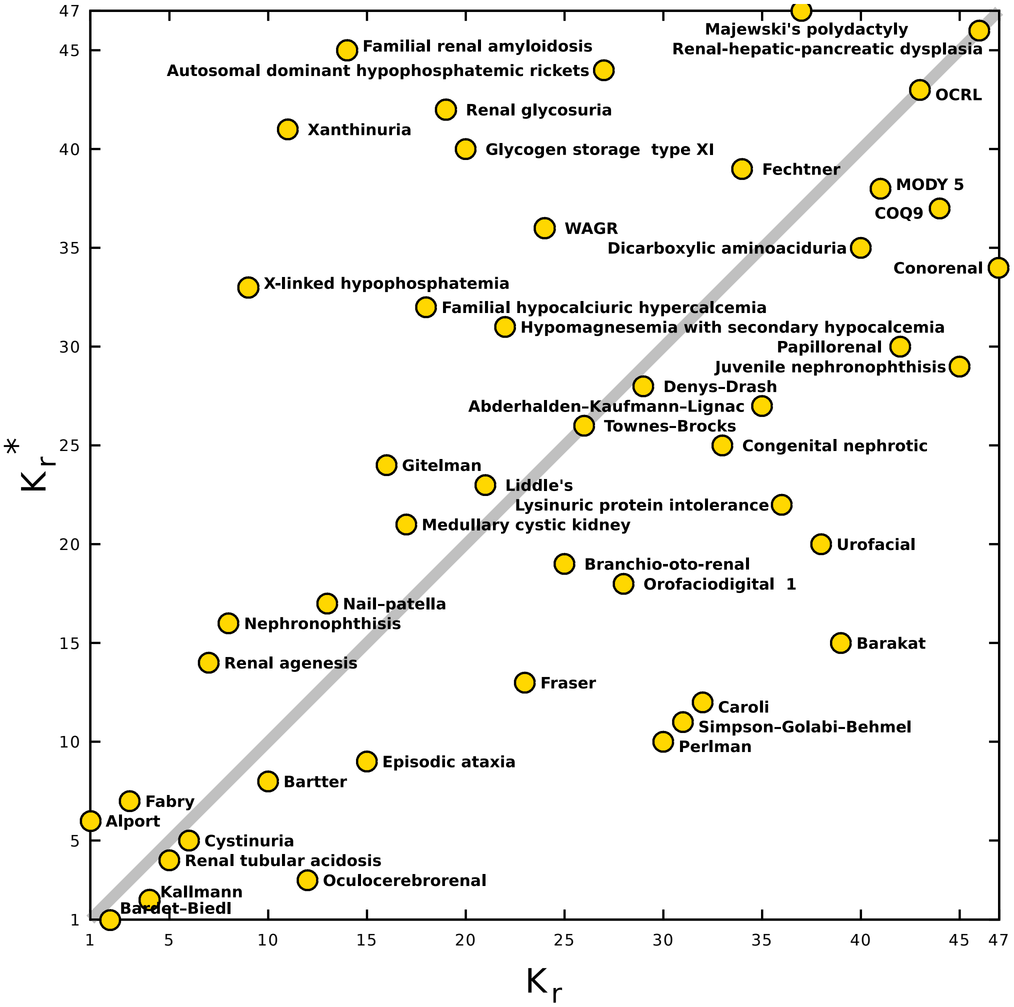

Fig 4. Distribution of $N_{rd} = 47$ articles of rare renal diseases on the plane of relative PageRank-CheiRank indexes (K_r, K_r^*); positions in the plane are given by golden circles with short disease names.

toward at least another one (with exceptions of Alexion, BioMarin, Novo Nordisk). This is due to the very competitive economical industry where drug department acquisitions from a company to another is routine. Still from Fig. 5 top right panel, from the rare renal diseases sector we observe that the few direct links concern rare renal diseases belonging to the same category (pixels enclosed in the red, blue, gold cyan and green squares). The exceptions are nephronophthisis, which belongs to the ciliopathies category, and medullary cystic kidney disease, which belongs to the renal tubular diseases and metabolic diseases category; each one these diseases directly point to the other. In fact these two diseases share similar morphological and clinical features (see, e.g., [29]) which explains reciprocal direct links between these diseases. We also observe that none of the Wikipedia articles devoted to the largest market capitalization pharmaceutical companies directly cites one of the rare renal diseases. This observation highlights the orphan status of these diseases. The only direct link existing between pharmaceutical companies and rare renal diseases is Fabry disease pointing toward Shire which manufactures Replagal, a dedicated drug for Fabry disease [30]. Finally, the G_{qr} matrix component (Fig. 5 bottom right panel) gives indirect links, some of them being

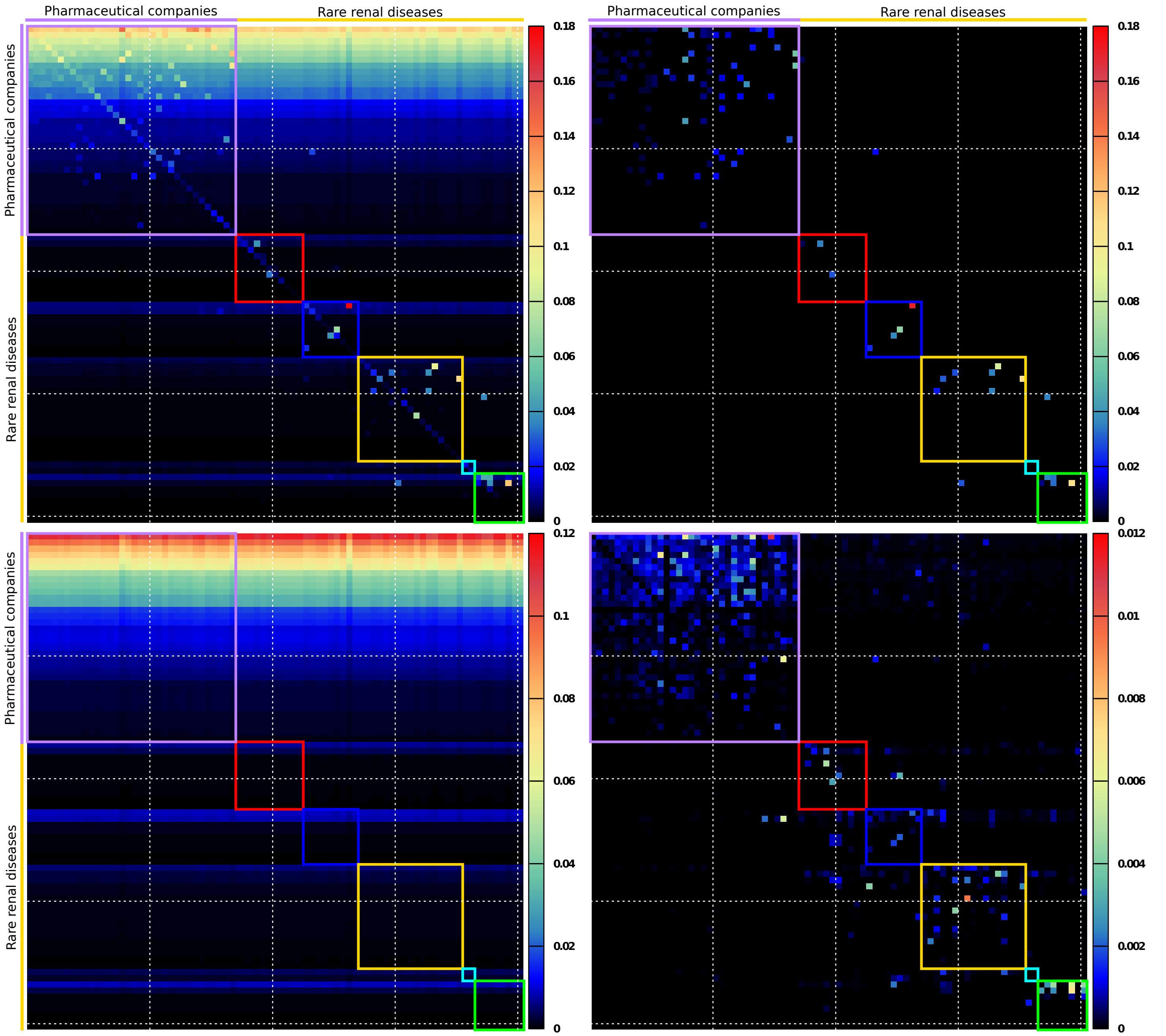

Fig 5. Reduced Google matrix G_{R} of pharmaceutical companies and rare renal diseases. We show the reduced Google matrix G_{R} (top left panel) and its three components G_{rr} (top right panel), G_{pr} (bottom left panel), and G_{qrnd} (bottom right panel). Each "pixel" represents a matrix entry which the amplitude is given by a color. Color bars gives the corresponds between matrix entry amplitudes and colors. For each 81 × 81 matrix the first 34 entries correspond to pharmaceutical companies (ordered as in Table 1) and the other 47 entries correspond to rare renal diseases (ordered by categories then by PageRank order inside each category, see Table 2). The first 34 × 34 block diagonal sub-matrix (purple square) corresponds to directed interactions between pharmaceutical companies. The other five smallest block diagonal sub-matrices correspond to directed interactions between rare renal diseases belonging to one of the five categories defined in Table 2. The colors of the squares correspond to color categories given in Table 2. For the sake of visibility horizontal and vertical white dashed lines are drawn after every 20 entries.

purely hidden links since they do not appear in the G_{rr} matrix component (Fig. 5 top right panel). The G_{qr} matrix component reveals many hidden links between pharmaceutical companies which are certainly due to the complex economic entanglements of these companies. The two most intense hidden links appear from Alexion to Shire and from Abbott to Johnson & Johnson. Many indirect pathways through the entire Wikipedia network contribute to these hidden links. E.g., one of the shortest path going from Alexion to Shire is the Alexion → Ludwig N. Hantson → Shire link [31]. Indeed, Ludwig N. Hantson has been named Alexion Pharmaceuticals CEO in March 2017 [32], and he was CEO of Baxalta prior its acquisition by Shire in June 2016 [33]. Also the indirect link Abbott → Advanced Medical Optics → Johnson & Johnson contributes to the Abbott → Johnson & Johnson hidden link. Indeed Advanced Medical Optics, also known as Abbott Medical Optics, has been acquired by Johnson & Johnson in February 2017 [34]. Concerning the rare renal diseases sector the most intense hidden link among diseases belonging to the same category are: Familial hypocalciuric hypercalcemia → Gitelman syndrome as both are hypocalciuric diseases [35], the inversed link Gitelman syndrome → Familial hypocalciuric hypercalcemia, Juvenile nephronophthisis → Bardet–Biedl syndrome as nephronophthisis and Bardet–Biedl syndrome typical cited examples of ciliopathies, Townes–Brocks syndrome → Branchio-oto-renal syndrome as both of these syndrome imply sensorineural hearing loss [36]. Between diseases belonging to different category we observe the notable hidden links: Alport syndrome → Oculocerebrorenal syndrome (the inverted hidden link is also present but less intense) as both are examples of genetic cause of cataract in childhood or early life [37], Denys–Drash syndrome → Perlman syndrome both leading to a high risk of Wilms' tumor [38, 39]. The most interesting hidden links are those appearing between pharmaceutical companies and rare renal diseases. The most intense one is Alexion → Fabry disease. Indeed Alexion Pharmaceuticals acquired in May 2015 Synageva BioPharma Corp. which is a company dedicated to rare diseases and in particular to Fabry disease [40]. Also with a somewhat less intensity, we observe hidden links from BioMarin and Vertex to Fabry disease. Here indirect pathways through Wikipedia going from Alexion, BioMarin, or Vertex to Fabry diseases are not evident to find as it is the case for hidden links only between pharmaceutical companies or between rare renal diseases.

Network structures

As the G_{pr} matrix component of the $N_r \times N_r$ reduced Google matrix G_{R} gives trivial results already known from a direct PageRank analysis of the $N \times N$ global Google matrix, we use the $G_{rr} + G_{\text{qr}}$ matrix component to infer an effective reduced network between possibly the N_r nodes of interest.

First we build the reduced network of interactions between the $N_r = N_{ph} + N_c$ nodes of interest constituted by the $N_{ph} = 34$ pharmaceutical companies and the $N_c = 195$ countries (see Fig. 6). The reduced network construction rules are described in the caption of Fig. 6. At the first level we take the top 5 companies in PageRank list (Table 1) and trace the links from a company to two countries and two other companies providing these links have the largest column matrix elements in $G_{rr} + G_{\text{qr}}$. Then the process is repeated to other levels from the newly added companies. The process is stopped when no new companies can be added. We obtain a compact reduced network of 15 pharmaceutical companies over the 34 initially selected, and 12 of these constitute the top 12 of the PageRank list (Table 1). The 3 remaining companies are nevertheless in the top 9 of the largest market capitalization since 2000.

From this network, following direct links (black arrows), i.e., hyperlinks from Wikipedia articles, we see, e.g., that for Bayer the closest companies are Sanofi and Roche and the two country friends are Germany, where its main office is located, and

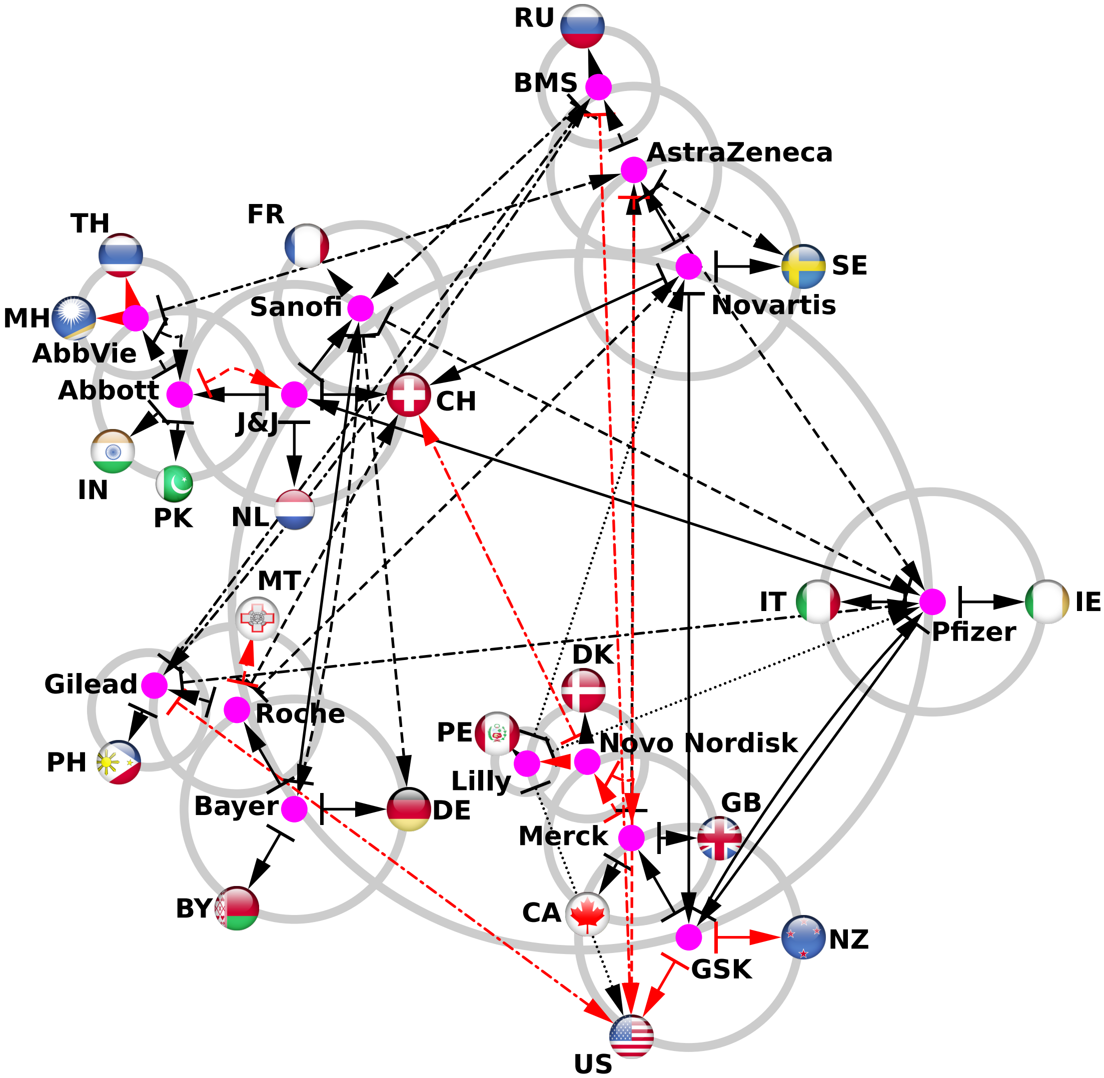

Fig 6. Reduced network of pharmaceutical companies with the addition of their best connected countries. We consider the first five pharmaceutical companies from the Wikipedia PageRank list (see Tab. 1). Each one of these companies are represented by purple circles (●) placed along the main grey circle. From these five most influential pharmaceutical in Wikipedia, we determine the two best connected companies, i.e., for a pharmaceutical company ph, we determine the two companies ph_1 and ph_2 giving the highest $(G_{rr} + G_{\mathrm{qr}})_{ph_{1\text{ or }2},ph}$ values. If not already present in the network, we add these best connected companies along secondary circles centered on the previous companies. Also from the initial five pharmaceutical companies we determine the two best connected countries, i.e., for a company ph, we determine the two countries c_1 and c_2 giving the highest $(G_{rr} + G_{\mathrm{qr}})_{c_{1\text{ or }2},ph}$ values. From the newly added pharmaceutical of the first iteration, we determine the two best connected pharmaceutical companies and the two best connected countries. This constitutes the second iteration, an so on. At the fourth iteration of this process, no new pharmaceutical companies are added, and consequently the network construction process stops. The links obtained at the first iteration are represented by plain line arrows, at the second iteration by dashed line arrows, at the third iteration by dashed-dotted line arrows, and at the fourth iteration by dotted line arrows. Black color arrows correspond to links existing in the adjacency matrix (direct hyperlinks in Wikipedia), and red color arrows are purely hidden links absent from the adjacency matrix but present in G_{qr} component of the reduced Google matrix G_{R}. The obtained network is drawn with the Cytoscape software [41]. Countries are marked by their ISO 3166-1 alpha-2 codes.

Belarus. These two links are direct ones from Bayer page in Wikipedia. Here Belarus is cited as an example of country from the Commonwealth of Independent States (CIS) where over-the-counter drug business was developing; for that purpose Bayer acquired in June 2008 Sagmel, Inc. who was already implanted in CIS [42, 43]. For Pfizer the two company friends are GlaxoSmithKline and Johnson & Johnson while two country friends are Italy and Ireland. All these direct links are easily explainable looking through hyperlinks inside Wikipedia articles. Direct links between companies testifies from major transactions between them reported in devoted Wikipedia articles. These direct links gives a compact picture of pharmaceutical companies relationships. The non obvious hidden links (red arrows in Fig. 6) are potentially more interesting. Here, most of the hidden links points to US, highlight the fact that most of the pharmaceutical companies are American.

The reduced network of interactions between the $N_{ph} = 34$ pharmaceutical companies and the $N_{cr} = 37$ cancer types is shown in Fig. 7. At each level, it shows two closest company friends and two cancer types. The construction rules are the same as in Fig. 6. The structure of the reduced network between pharmaceutical companies remains the same as in Fig. 6. The reduced network gives the most strong links from a company to the related types of cancer. We observe a clear polarization toward breast cancer, since 10 of the 15 pharmaceutical companies preferentially point to breast cancer, the Wikipedia article of which is the second most influential article among the 37 articles devoted to cancer types [20]. The second and third most connected cancer type are leukemia and lung cancer which are preferentially pointed by 4 and 3 pharmaceutical companies respectively. Although the (list lung cancer, breast cancer, and leukemia) constitutes the top 3 of the most influential cancer types in Wikipedia [20], we observe the peculiar interest to breast cancer from pharmaceutical companies. Looking with more details, Pfizer and GlaxoSmithKline are mostly linked to breast and leukemia cancers; Bayer to breast and thyroid cancers. Johnson & Johnson has most strong links to leukemia and ovarian cancers. For Novartis the most strong links are to melanoma and breast cancers. Links to cancers for other companies and also well visible in Fig. 7. We argue that our REGOMAX approach allows to determine from Wikipedia the main orientations of pharmaceutical companies in their treatments of cancers. From Fig. 7 we also see that in addition to direct links (black arrows) the indirect links (red arrows) play very important role.

The reduced network of pharmaceutical companies and rare renal diseases is shown in Fig. 8. Its construction rules are the same as for Figs. 6 & 7. Again the network structure of the pharmaceutical companies remains the same as in Figs. 6 & 7. Since 15 pharmaceutical companies are presents, 30 preferentially connected rare renal diseases could have potentially emerged from this reduced network, but in fact only 10 rare renal diseases are presents. From this network we directly see the main orientations of a company to specific rare renal diseases. We observe a strong polarization toward Fabry disease ($K_r = 3$ in Table 2) since 12 of the 15 pharmaceutical companies point to Fabry disease. The second and third most connected rare renal diseases are Alport syndrome ($K_r = 1$ in Table 2) with 5 ingoing links and Kallmann syndrome ($K_r = 4$ in Table 2) with 3 ingoing links. Looking with more details, Pfizer has the most strong links to Kallmann syndrome and Fabry disease. GlaxoSmithKline is linked to Fabry disease and medullary cystic kidney disease while bayer is oriented to renal agenesis and familial renal amyloidosis disease. For Johnson & Johnson the main orientations are Alport syndrome and Fabry disease. Novartis is mainly linked to renal tubular acidosis and Fabry disease. We observe that none of the articles devoted to the largest market capitalization pharmaceutical companies cite any of the rare renal diseases, thus all the connections from pharmaceutical companies to rare renal diseases are indirect hidden links (red arrows in Fig. 8), confirming indeed the effective status of orphan diseases.

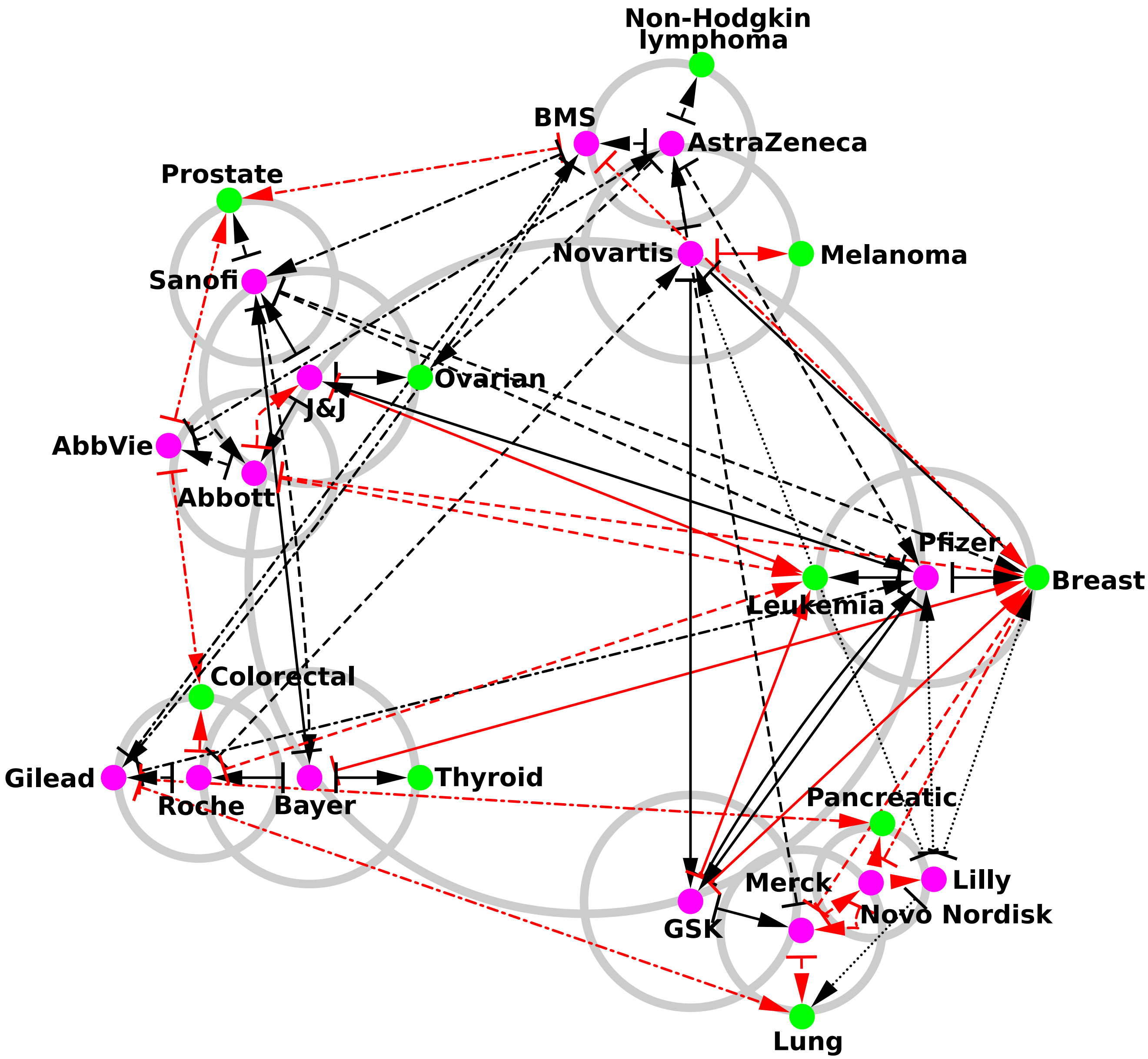

Fig 7. Reduced network of pharmaceutical companies with the addition of their best connected cancers. The construction algorithm is the same as the one used to generate Fig. 6 excepting that we replace at each iteration the two best connected countries by the two best connected cancers. Pharmaceutical companies are represented by purple circles (●) and cancers by green circles (●).

We also present the friendship network between companies and infectious diseases studied in [19] in SupInfo Fig. S2.

Sensitivity of countries to pharmaceutical companies

To see the global influence of a specific pharmaceutical company on world countries we use the PageRank sensitivity D described in Section Reduced Google matrix.

In Fig. 9 we show the sensitivity of world countries to two companies, Pfizer and Bayer. The two most sensitive countries for Pfizer are Ireland and Italy in correlation

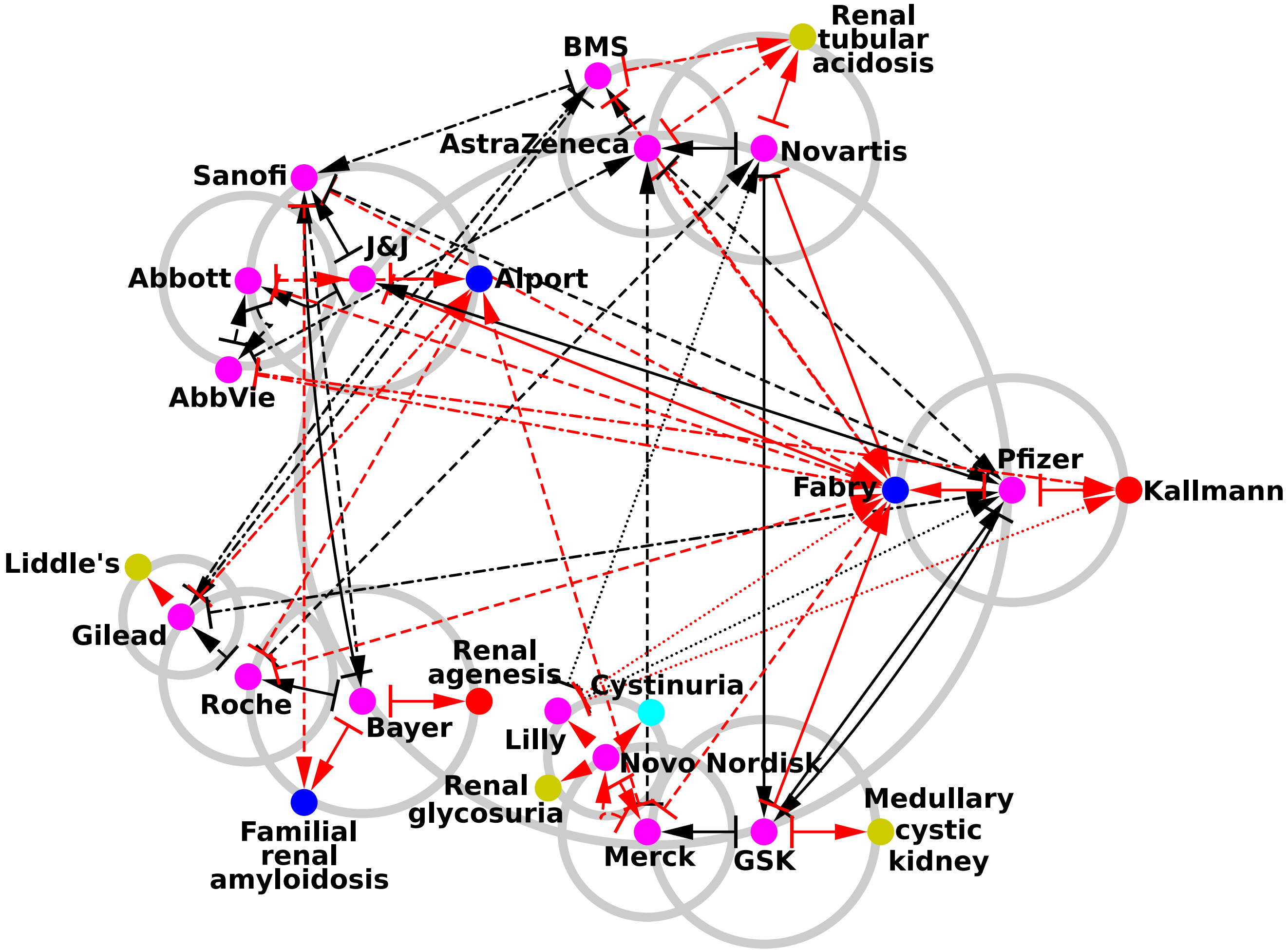

Fig 8. Reduced network of pharmaceutical companies with the addition of their best connected rare renal diseases. The construction algorithm is the same as the one used to generate Fig. 6 excepting that we replace at each iteration the two best connected countries by the two best connected rare renal diseases. Pharmaceutical companies are represented by purple circles (●) and rare renal diseases by red circles (●) for congenital abnormalities of the kidney and urinary tract, blue circles (●) for glomerular diseases, gold circles (●) for renal tubular diseases and metabolic diseases, cyan circles (●) for nephrolithiasis, and green circles (●) for ciliopathies.

with the most strong direct links shown in Fig. 6. Italy is present here for historical reasons as calcium citrate used to produce citric acid were supplied to Pfizer by Italy until a shortage caused by World War I which forced Pfizer chemists to develop fermentation technology to obtain citric acid from sugar using a fungus, this technology were then used massively to produce antibiotic penicillin during World War II [45]. In 2016, Pfizer tempted to acquire the Allergan, an Irish–tax registered pharmaceutical company, in order to create the world's largest drugmaker and to relocate its headquarters in Ireland to decrease taxes. The deal was called-off after US Treasury adopted new anti tax inversion rules [46]. The next most Pfizer impacted countries are Canada, Australia, and New Zealand. There is no direct link between Pfizer and these countries. Among the shortest indirect paths we have, e.g., Pfizer → Terre Haute → Canada, Pfizer → Wyeth → Australia, and Pfizer → Helen Clark → Judicial Committee of the Privy Council → New Zealand. Pfizer is present in Canada since

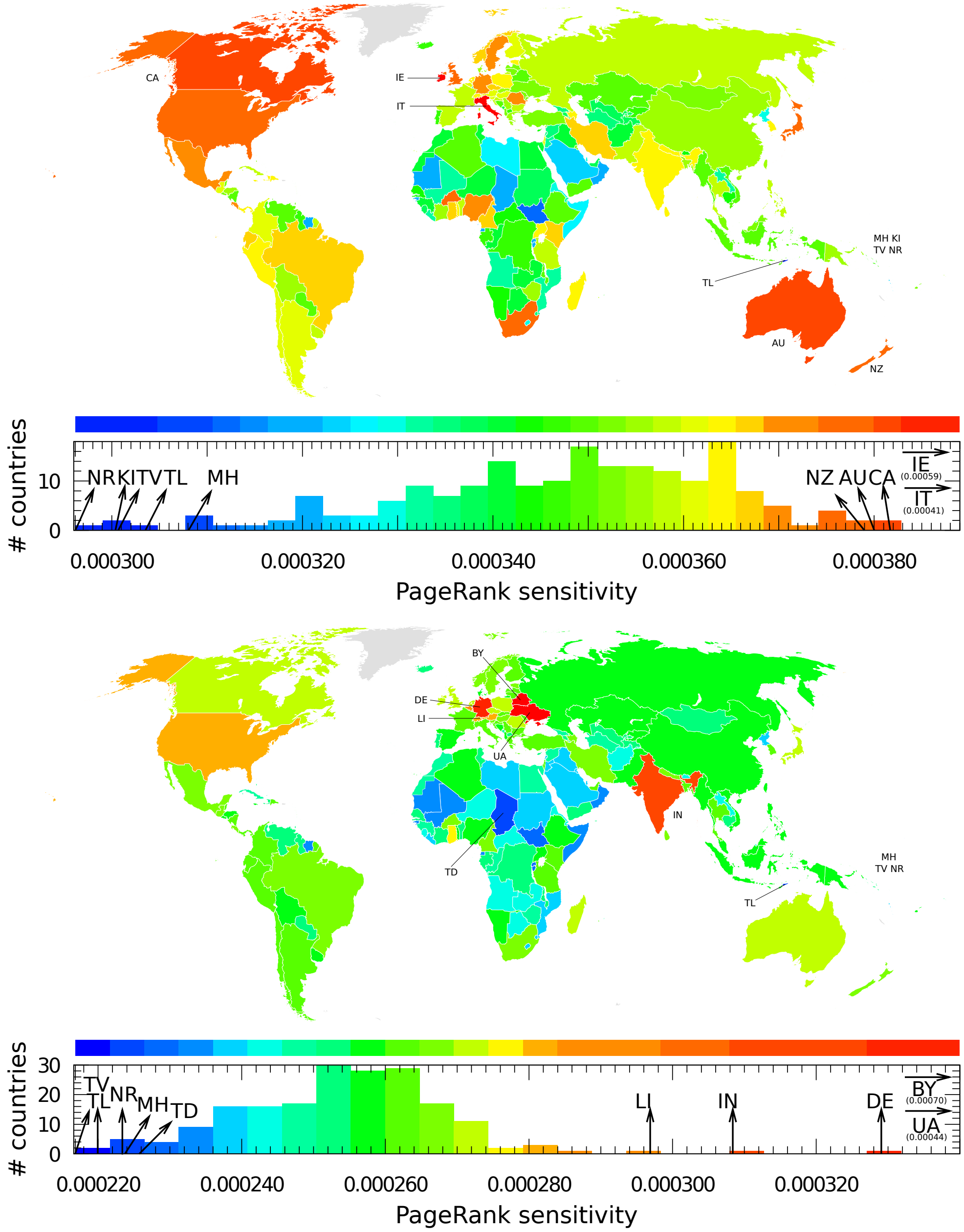

Fig 9. Sensitivity of countries to Pfizer company (top panel) and to Bayer company (bottom panel). A country c is colored according to its diagonal PageRank sensitivity $D(ph \to c, c)$, where ph is the pharmaceutical company. Color categories are obtained using the Jenks natural breaks classification method [44].

fifties, and in 2012 Pfizer's Canadian division was recognized as one of the 15 best places to work in Canada [47]. In 2009, Pfizer acquired Wyeth which supplied a pneumococcal vaccine approved for young children in Australia [48]. Wikileaks revealed that in 1990 Pfizer was lobbying in the US against New Zealand considering its drug buying rules as restrictive; Helen Clark was at that time New Zealand Health minister [49]. For Bayer the most sensitive countries are Belarus, Ukraine and Germany. Indeed, the article Bayer has direct links to Belarus and Ukraine. There is also direct link to Germany and many indirect links to it since the company is located in Germany. It is possible that Germany is less sensitive compared to Ukraine and Belarus since its PageRank probability is significantly higher. We retrieve here Belarus and Ukraine, members of CIS, as it is explained in Section Network structures. Also India appears as it host a part of Bayer Business Services [50].

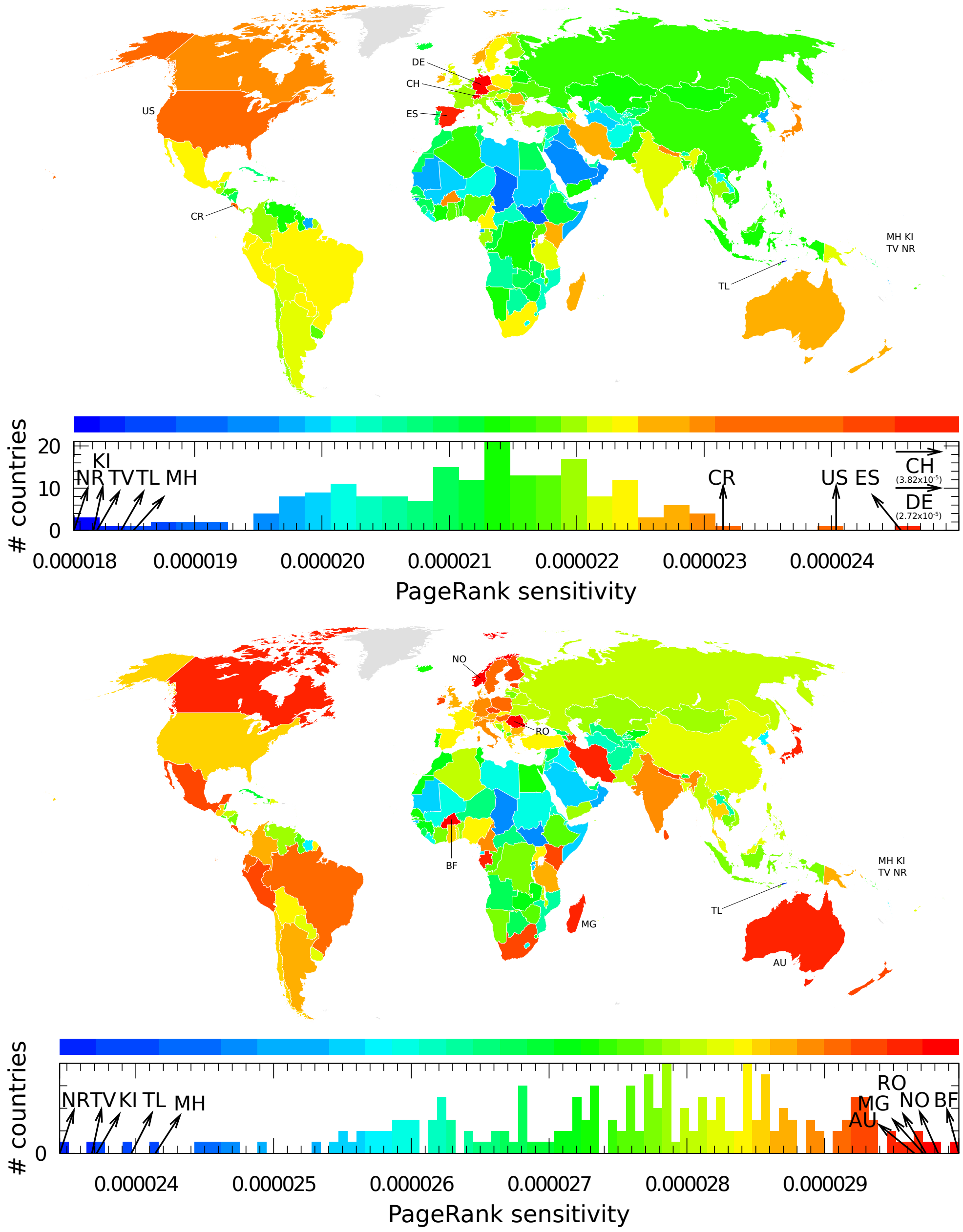

Fig 10. Sensitivity of countries to rare renal diseases; here to Kallmann syndrome (top panel) and to Bardet–Biedl syndrome (bottom panel). A country c is colored according to its diagonal PageRank sensitivity $D(rd \to c, c)$, where rd is the rare renal disease. Color categories are obtained using the Jenks natural breaks classification method [44].

Sensitivity of countries to rare renal diseases

In Fig. 10 we present the PageRank sensitivity of countries to Kallmann syndrome and Bardet–Biedl syndrome. For Kallmann syndrome the most sensitive countries are Switzerland, Germany and Spain. From Kallmann syndrome many paths converge toward Germany, we have, e.g., the direct link Kallmann syndrome → Germany, and, e.g., the indirect link Kallmann syndrome → Franz Josef Kallmann → Germany. Franz Josef Kallmann who gave the first description of this disease was a German born geneticist. The Wikipedia article devoted to Kallmann syndrome points to Switzerland trough a direct link and an indirect link, Kallmann syndrome → Lausanne University Hospital → Switzerland. The Lausanne University Hospital is cited as one of the main research research facility on this disease [51]. Also the Spanish doctor Aureliano Maestre de San Juan noted the link between anosmia and hypogonadism (Kallmann syndrome → Aureliano Maestre de San Juan → Spain). Country sensitivity appears due to the combination of direct and indirect links captured by the REGOMAX algorithm.

For Bardet–Biedl syndrome the most sensitive countries are Burkina Faso, Norway and Romania. Here there are no direct links to these countries from the article of Bardet–Biedl syndrome [52]. The indirect link to Romania appears since Arthur Biedl was born in today Romania. Indirect paths leading to the other most sensitive countries are difficult to find.

We also present the sensitivity of pharmaceutical companies to cancers and rare renal disease respectively in Fig. S3 and Fig. S4 in Section Supplementary Information.

Discussion

We use the reduced Google matrix (REGOMAX) algorithm for analysis of English Wikipedia network with more than 5 million articles. The analysis is focused on 195 world countries, 34 largest biotechnology and pharmaceutical companies, 47 rare renal diseases and 37 types of cancer. The algorithm allows to construct the reduced Google matrix of these entries taking into account the direct and indirect links between them. While the direct links are directly present in the global Wikipedia network the indirect links are obtained with the REGOMAX summing the contributions between entries from all pathways connecting them via the global network. With the reduced Google matrix we determine the interaction networks between companies and world countries, companies and rare renal disease, companies and cancers. From the sensitivity of PageRank probabilities we obtain the influence of specific companies on world countries. This approach also provides the sensitivity of world countries to specific rare renal diseases. We obtain the most influential and communicative pharmaceutical companies showing that the top PageRank positions belong to the companies which are not at all at the top list of market capitalization. We argue that the improvement of Wikipedia articles of specific pharmaceutical companies can increase their world wide visibility without significant additional investments. Our study shows that the knowledge accumulated at Wikipedia can be efficiently analyzed by the REGOMAX algorithm determining the effective interactions between specific Wikipedia articles being of interest for researchers.

Acknowledgments

This work was supported in part by the Pogramme Investissements d'Avenir ANR-11-IDEX-0002-02, reference ANR-10-LABX-0037-NEXT (project THETRACOM). This work was also supported in part by the Programme Investissements d'Avenir

ANR-15-IDEX-0003, ISITE-BFC (project GNETWORKS) and by Bourgogne Franche-Comté region (project APEX).

Supplementary Information

In Fig. S1 we present in local PageRank K_r – CheiRank K_r^* indexes plane the set of selected Wikipedia articles described in Section Datasets. In Fig. S2 we show the friendship network of interactions between pharmaceutical companies and infections diseases studied in [19].

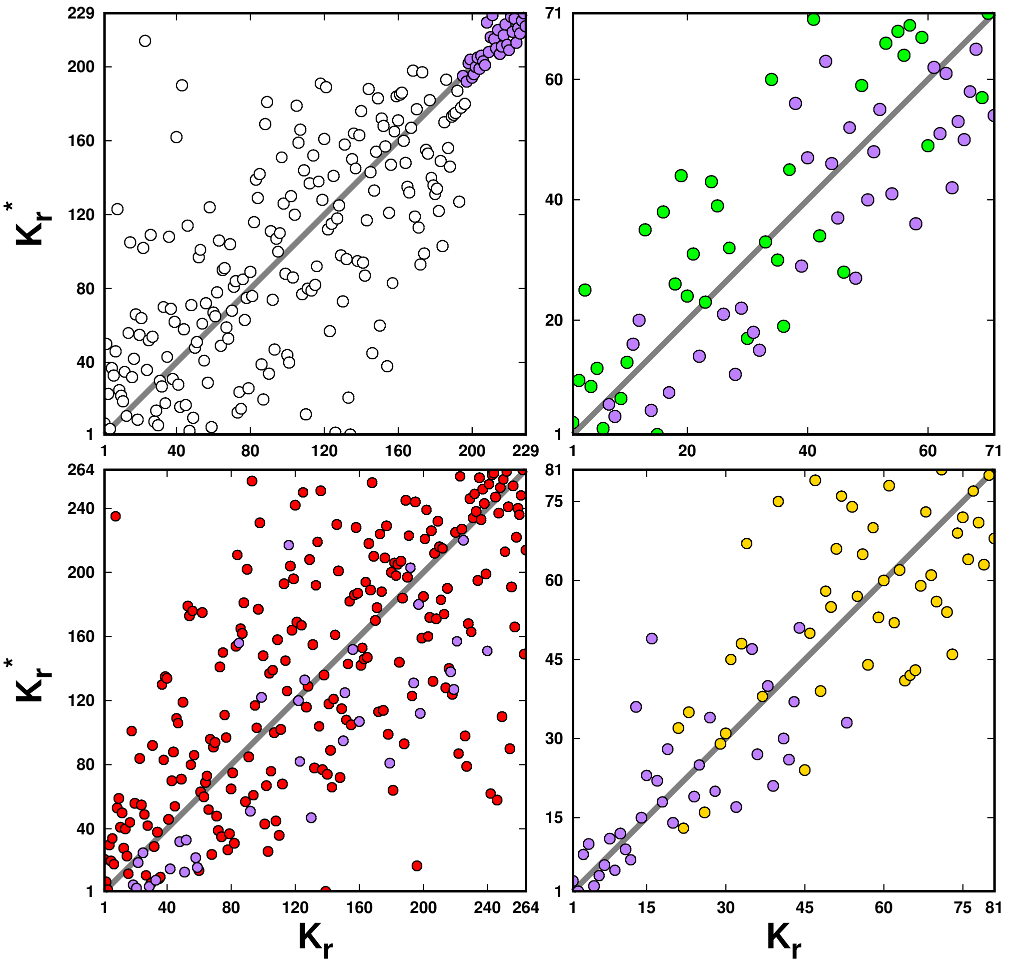

Fig S1. Distribution of the May 2017 English Wikipedia articles in the relative PageRank K_r – CheiRank K_r^* indexes plane for pharmaceutical companies and countries (top left panel), and cancer types (top right panel), and infectious diseases (bottom left panel), and rare renal diseases (bottom right panel). The $N_{ph} = 34$ pharmaceutical companies are represented by purple circles (●), the $N_c = 195$ countries by white circles (○), the $N_{cr} = 37$ cancer types by green circles (●), the $N_d = 230$ infectious diseases by red circles (●), and the $N_{rd} = 47$ rare renal diseases by gold circles (●).

We also show that the REGOMAX analysis allows to determine the inter-sensitivity of pharmaceutical companies to cancers and rare renal diseases. We show the sensitivity of 34 pharmaceutical companies to 37 cancers and 47 rare renal diseases (and vise versa)

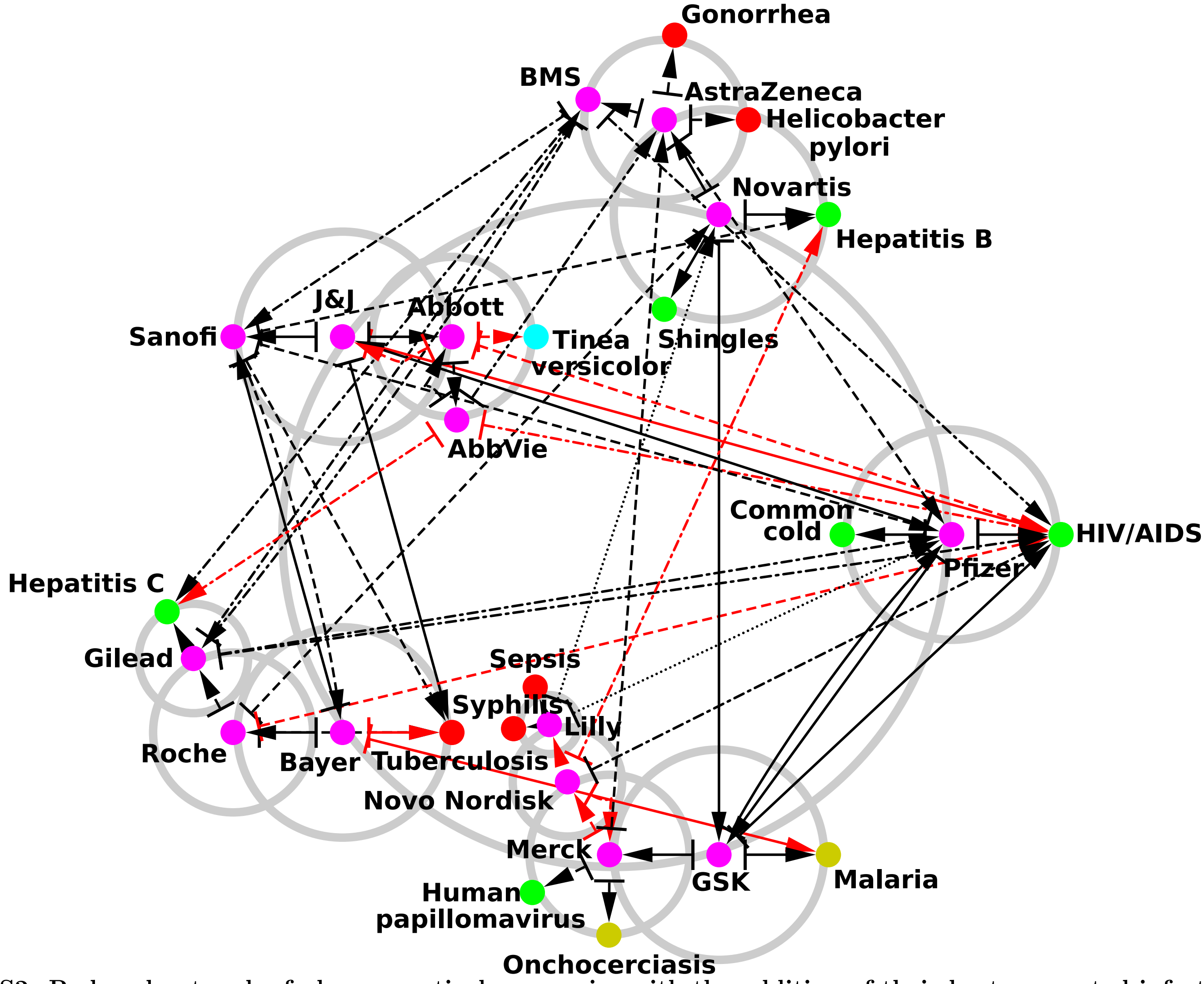

Fig S2. Reduced network of pharmaceutical companies with the addition of their best connected infectious diseases. The construction algorithm is the same as the one used to generate Fig. 6 excepting that we replace at each iteration the two best connected countries by the two best connected infectious diseases. Pharmaceutical companies are represented by purple circles (●) and infectious diseases by red circles (●) for bacterial type diseases, green circles (●) for viral type diseases, gold circles (●) for parasitic type diseases, and cyan circles (●) for fungal type diseases. The list of infectious diseases is available in [19].

in Fig. S3 and Fig. S4 respectively. We note the very strong matrix element in Fig. S4 (red square) between Shire and Fabry disease appears due to the drug Replagal which is a treatment for this disease and which is produced by Shire company.

References

1. World Health Organization. World Health Statistics; 2018. Available from: https://www.who.int/gho/publications/world_health_statistics/.

2. Wikipedia contributors. List of largest biotechnology and pharmaceutical companies — Wikipedia, The Free Encyclopedia; 2019. Available from:

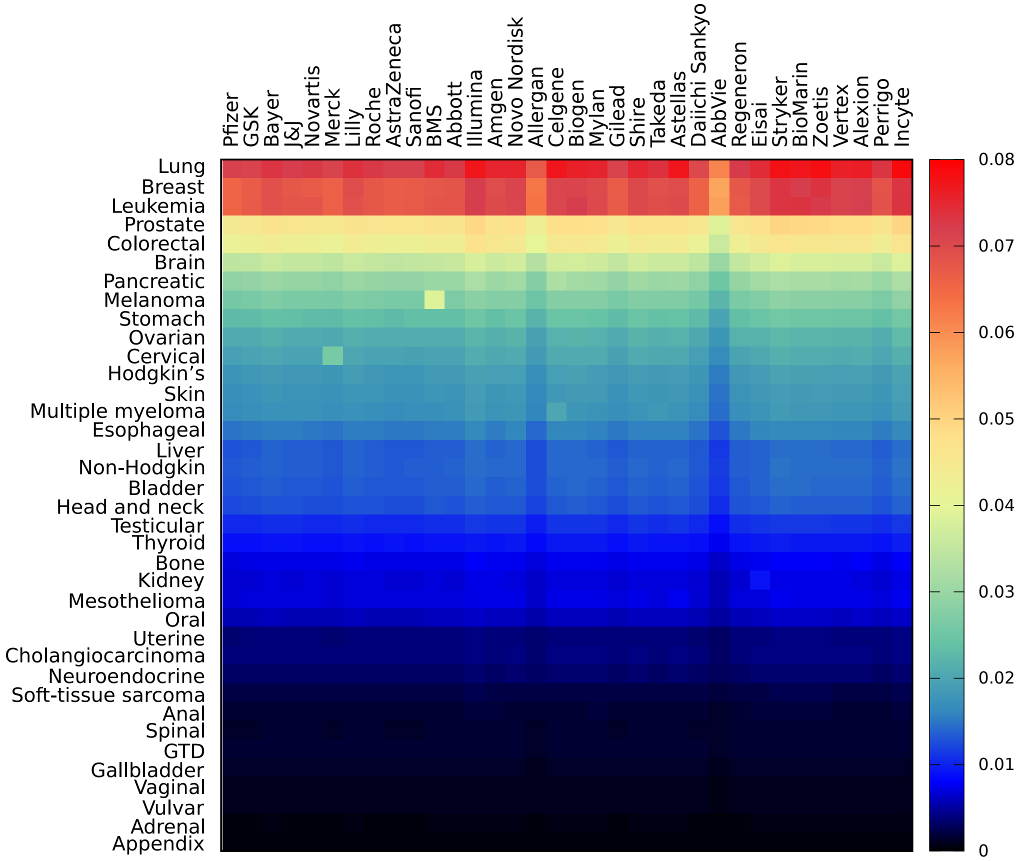

Fig S3. Sensitivity of pharmaceutical companies to cancers. Horizontal (vertical) entries represents pharmaceutical companies (cancer types). The acronym GTD stands for gestational trophoblastic disease.

https://en.wikipedia.org/w/index.php?title=List_of_largest_biotechnology_and_pharmaceutical_companies&oldid=884151571.

3. Giles J. Internet encyclopaedias go head to head. Nature. 2005;438:900–901. doi:10.1038/438900a.

4. Callaway E. No rest for the bio-wikis. Nature. 2010;468(7322):359–360. doi:10.1038/468359a.

5. Nielsen FÅ. Wikipedia Research and Tools: Review and Comments. SSRN Electronic Journal. 2012;doi:10.2139/ssrn.2129874.

6. Lewoniewski W, Wecel K, Abramowicz W. Relative Quality and Popularity Evaluation of Multilingual Wikipedia Articles. Informatics. 2017;4(4):43. doi:10.3390/informatics4040043.

7. Dorogovtsev S. Lectures on Complex Networks. Oxford University Press; 2010.

8. Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. 1998;30(1):107 – 117. doi:10.1016/S0169-7552(98)00110-X.

9. Langville AN, Meyer CD. Google’s PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press; 2012.

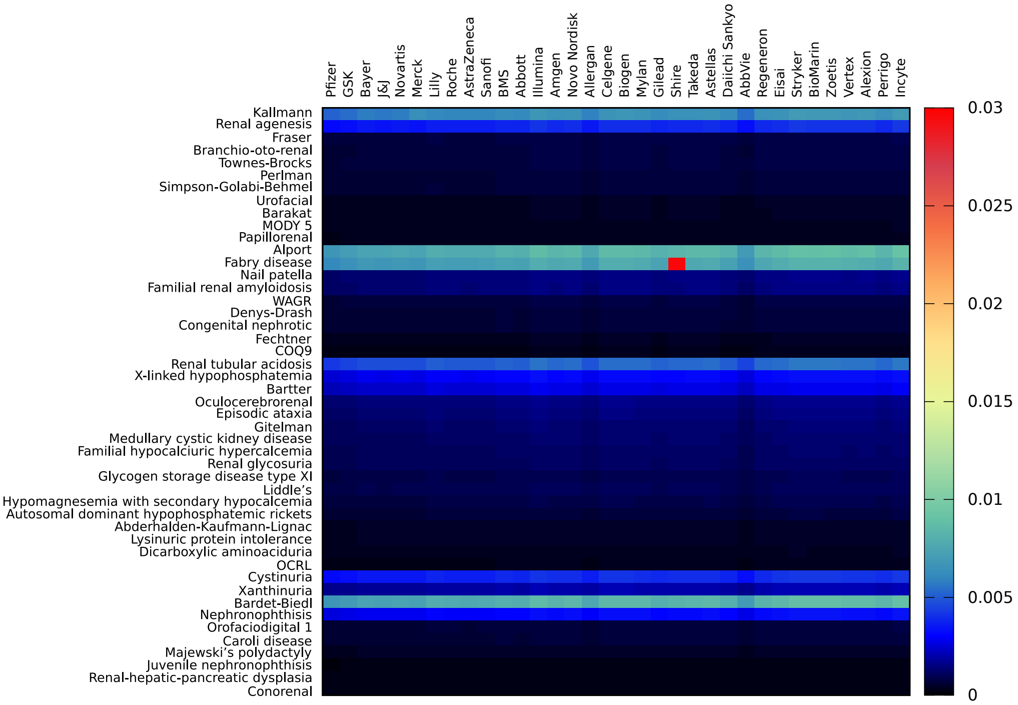

Fig S4. Sensitivity of pharmaceutical companies to rare renal diseases. Horizontal (vertical) entries represents pharmaceutical companies (rare renal diseases).

10. Ermann L, Frahm KM, Shepelyansky DL. Google matrix analysis of directed networks. Rev Mod Phys. 2015;87:1261–1310. doi:10.1103/RevModPhys.87.1261.

11. Eom YH, Aragón P, Laniado D, Kaltenbrunner A, Vigna S, Shepelyansky DL. Interactions of Cultures and Top People of Wikipedia from Ranking of 24 Language Editions. PLOS ONE. 2015;10(3):1–27. doi:10.1371/journal.pone.0114825.

12. Lages J, Patt A, Shepelyansky DL. Wikipedia ranking of world universities. The European Physical Journal B. 2016;89(3). doi:10.1140/epjb/e2016-60922-0.

13. Frahm KM, Shepelyansky DL. Reduced Google matrix. arXiv. 2016;arXiv:1602.02394.

14. Frahm KM, Jaffrès-Runser K, Shepelyansky DL. Wikipedia mining of hidden links between political leaders. The European Physical Journal B. 2016;89(12):269. doi:10.1140/epjb/e2016-70526-3.

15. El Zant S, Jaffrès-Runser K, Shepelyansky DL. Capturing the influence of geopolitical ties from Wikipedia with reduced Google matrix. PLOS ONE. 2018;13(8):1–31. doi:10.1371/journal.pone.0201397.

16. Coquidé C, Lages J, Shepelyansky DL. World influence and interactions of universities from Wikipedia networks. The European Physical Journal B. 2019;92(1):3. doi:10.1140/epjb/e2018-90532-7.

17. Demidov D, Frahm KM, Shepelyansky DL. What is the central bank of Wikipedia? arXiv e-prints. 2019; p. arXiv:1902.07920.

18. Lages J, Shepelyansky DL, Zinovyev A. Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks. PLOS ONE. 2018;13(1):1–28. doi:10.1371/journal.pone.0190812.

19. Rollin G, Lages J, Shepelyansky DL. World Influence of Infectious Diseases From Wikipedia Network Analysis. IEEE Access. 2019;7:26073–26087. doi:10.1109/ACCESS.2019.2899339.

20. Rollin G, Lages J, Shepelyansky D. Wikipedia network analysis of cancer interactions and world influence. bioRxiv. 2019;doi:10.1101/527879.

21. International Agency for Research on Cancer, World Health Organization. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 v1.0; 2012. Available from: http://publications.iarc.fr/Databases/Iarc-Cancerbases/GLOBOCAN-2012-Estimated-Cancer-Incidence-Mortality-And-Prevalence-Worldwi 0-2012.

22. Global Genes. RARE Facts; 2019. Available from: https://globalgenes.org/rare-facts/.

23. Kaplan W, Wirtz VJ, Mantel-Teeuwisse A, Stolk P, Duthey B, Laing R. Priority Medicines for Europe and the World 2013 Update; 2013. Available from: https://www.who.int/medicines/areas/priority_medicines/Ch6_19Rare.pdf.

24. Institute NC. Cancer types; 2019. Available from: https://www.cancer.gov/types.

25. Devuyst O, Knoers NVAM, Remuzzi G, Schaefer F. Rare inherited kidney diseases: challenges, opportunities, and perspectives. The Lancet. 2014;383:1844–1859. doi:10.1016/S0140-6736(14)60659-0.

26. Chepelianskii AD. Towards physical laws for software architecture. arXiv e-prints. 2010; p. arXiv:1003.5455.

27. Zhirov AO, Zhirov OV, Shepelyansky DL. Two-dimensional ranking of Wikipedia articles. The European Physical Journal B. 2010;77(4):523–531. doi:10.1140/epjb/e2010-10500-7.

28. Alexa. World top visited web sites; 2019. Available from: https://www.alexa.com/topsites.

29. Scolari F, Ghiggeri G. Nephronophthisis-Medullary Cystic Kidney Disease: From Bedside to Bench and Back Again. Saudi Journal of Kidney Diseases and Transplantation. 2003;14(3):316–327.

30. Tsuboi K, Yamamoto H. Clinical observation of patients with Fabry disease after switching from agalsidase beta (Fabrazyme) to agalsidase alfa (Replagal). Genetics In Medicine. 2012;14:779. doi:10.1038/gim.2012.39.

31. Wikipedia contributors. Ludwig N. Hantson — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Ludwig_N._Hantson&oldid=773365464.

32. Bloomberg. Ludwig N. Hantson: Executive Profile and Biography; 2019. Available from: https://www.bloomberg.com/research/stocks/people/person.asp?personId=12703772&privcapId=347983.

33. Wikipedia contributors. Baxalta — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Baxalta&oldid=768715356.

34. Wikipedia contributors. Abbott Medical Optics — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Abbott_Medical_Optics&oldid=767847481.

35. Wikipedia contributors. Hypocalciuria — Wikipedia, The Free Encyclopedia; 2016. Available from: https://en.wikipedia.org/w/index.php?title=Hypocalciuria&oldid=749626627.

36. Wikipedia contributors. Sensorineural hearing loss — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Sensorineural_hearing_loss&oldid=770911856.

37. Wikipedia contributors. Cataract — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Cataract&oldid=777851128.

38. Astuti D, Morris MR, Cooper WN, Staals RHJ, Wake NC, Fews GA, et al. Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. Nature Genetics. 2012;44:277–284. doi:10.1038/ng.1071.

39. Drash A, Sherman F, Hartmann WH, Blizzard RM. A syndrome of pseudohermaphroditism, Wilms' tumor, hypertension, and degenerative renal disease. The Journal of Pediatrics. 1970;76(4):585 – 593. doi:https://doi.org/10.1016/S0022-3476(70)80409-7.

40. Alexion. Alexion to Acquire Synageva to Strengthen Global Leadership in Developing and Commercializing Transformative Therapies for Patients with Devastating and Rare Diseases; 2019. Available from: https://news.alexion.com/node/524/all/2006/all.

41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003;13(11):2498–2504. doi:10.1101/gr.1239303.

42. Bayer. Bayer HealthCare successfully completes acquisition of Sagmel, Inc.'s OTC Business; 2019. Available from: https://www.investor.bayer.com/en/nc/news/archive/investor-news-2008/investor-news-2008/bayer-healthcare-successfully-completes-acquisition-of-sagmel-incs-otc-bu

43. Bayer. Bayer HealthCare to acquire OTC Business of Sagmel, Inc.; 2019. Available from: https://www.investor.bayer.de/en/nc/news/archive/investor-news-2008/investor-news-2008/bayer-healthcare-to-acquire-otc-business-of-sagmel-inc/.

44. Wikipedia contributors. Jenks natural breaks optimization — Wikipedia, The Free Encyclopedia; 2018. Available from: https://en.wikipedia.org/w/index.php?title=Jenks_natural_breaks_optimization&oldid=842502988.

45. ACS Chemistry for Life. Pfizer's work on penicillin for World War II becomes a National Historic Chemical Landmark; 2019. Available from: https://www.acs.org/content/acs/en/pressroom/newsreleases/2008/june/

pfizers-work-on-penicillin-for-world-war-ii-becomes-a-national-historic-c html.

46. The Washington Post. Pfizer, Allergan call off $160 billion merger after U.S. moves to block inversions; 2019. Available from: http://wapo.st/1VahSX7?tid=ss_mail&utm_term=.2d3472c4dcf9.

47. Wikipedia contributors. Pfizer — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Pfizer&oldid=777843272.

48. Wikipedia contributors. Wyeth — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Wyeth&oldid=777064151.

49. New Zealand Herald. WikiLeaks: Drug firms tried to ditch Clark; 2019. Available from: https://www.nzherald.co.nz/pharmaceuticals/news/article.cfm?c_id=278&objectid=10695239.

50. Wikipedia contributors. Bayer — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Bayer&oldid=776044035.

51. Wikipedia contributors. Kallmann syndrome — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Kallmann_syndrome&oldid=776085509.

52. Wikipedia contributors. Bardet–Biedl syndrome — Wikipedia, The Free Encyclopedia; 2017. Available from: https://en.wikipedia.org/w/index.php?title=Bardet%E2%80%93Biedl_syndrome&oldid=775800611.