

Article

Fibrosis protein-protein interactions from Google matrix analysis of MetaCore network

Ekaterina Kotelnikova¹, Klaus M. Frahm², Dima L. Shepelyansky^{2,*} and Oksana Kunduzova^{3,4}

¹ Clarivate Analytics, Barcelona 08025, Spain

² Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

³ National Institute of Health and Medical Research (INSERM) U1048, 31432, Toulouse, Cedex 4, France

⁴ Université de Toulouse, UPS, 31062 Toulouse, France

* Correspondence: dima@irsamc.ups-tlse.fr; Tel. +33-56155-60-68 (D.L.S.)

Received: October 20, 2021; ...

Abstract: Protein-protein interactions is a longstanding challenge in cardiac remodeling processes and heart failure. Here we use the MetaCore network and the Google matrix algorithms for prediction of protein-protein interactions dictating cardiac fibrosis, a primary causes of end-stage heart failure. The developed algorithms allow to identify interactions between key proteins and predict new actors orchestrating fibroblast activation linked to fibrosis in mouse and human tissues. These data hold great promise for uncovering new therapeutic targets to limit myocardial fibrosis.

Keywords: fibrosis; Markov chains; Google matrix; directed networks; protein-protein interactions

1. Introduction

Cardiovascular disease, a class of diseases that impact cardiovascular system, is responsible for 31% of all deaths and remains the leading cause of mortality worldwide [1]. Myocardial fibrosis is a central element of cardiac remodeling that leads to human failure and death [2]. Myocardial fibrosis results from uncontrolled fibroblast activity and excessive extracellular matrix deposition [2]. Although a number of factors have been implicated in orchestrating the fibrotic response, tissue fibrosis is dominated by a central mediator: transforming growth factor- β (TGF- β) [3]. Sustained TGF- β production leads to a continuous cycle of growth factor signaling and dysregulated matrix turnover. However, despite intensive research, the factors that orchestrate fibrosis are still poorly understood and as a result, effective strategies for reversing fibrosis are lacking [2,4]. Considering the complex heterogeneity of fibrosis, research strategy on a system-level understanding of the disease using mathematical modeling approaches is a driving force to dissect the complex processes involved in fibrotic disorders. Recently, we have reproduced the classic hallmarks of aberrant cardiac fibroblast activation leading to fibrosis including high collagen production and deposition [5]. Tools designed in this work, with RNA sequence data sets, enable analyses to help generate hypotheses about a gene's function in activated fibroblasts.

The present work is based on the recent clinical results presented in [5] which allowed to determine the protein response detected in fibrosis tissue of mouse and human as a feedback on TGF protein stimulation, which is known to play an important role [3]. These experiments allowed to determine proteins with most positive and most negative response. We use the proteins with top 20 positive and top 20 negative strongest response. Their names are given in Table 1 marked by indexes $K_u = 1, 2, \dots, 20$; $K_d = 1, 2, \dots, 20$. These proteins are ordered monotonically from the strongest $K_u = 1$ to weakest $K_u = 20$ positive responses; the same monotonic ordering is done by modulus of negative response with strongest $K_d = 1$ to weakest $K_d = 20$ responses. In addition there are 4 proteins TGF- β with indexes $K_t = 1, 2, 3, 4$ used in experiments [5]. These 44 proteins form the internal selected fibrosis group. For the analysis of protein-protein interactions (PPI) characterizing fibrosis we add a group of 10 external proteins with indexes as $K_x = 1, 2, \dots, 10$ (their choice will be explained below). Thus in

total we have the PPI fibrosis network with 54 proteins (nodes). They are ordered by their global index $K_g = 1, 2, \dots, 54$ as it is shown in Table 1 (first 4 K_t , then 20 K_u , 20 K_d and 10 K_x).

To analyze the properties of this PPI fibrosis network we use the developed commercial MetaCore network database of Clarivate [6]. This network database has been shown to be useful for analysis of various specific biological problems (see e.g. [7,8]). At present the MetaCore network has $N = 40079$ nodes with $N_\ell = 292191$ links (without self connections) with on average $n_\ell = N_\ell/N \approx 7.3$ links per node [9]. The nodes are given mainly by proteins but also there are certain molecules and molecular clusters catalyzing the interactions with proteins. This MetaCore PPI network is directed and non-weighted. Also its network links mark the bi-functional nature of interactions leading to the activation or the inhibition of one protein by another one. For some nodes link action is neutral or unknown.

For the investigation of fibrosis PPI network we use the Google matrix algorithms developed for the analysis of the World Wide Web [10,11] and other directed networks like Wikipedia networks, world trade networks and others (see review [12]). Such an approach to network characterization is based on the concept of Markov chains invented by Markov in an article published 1906 in the proceeding of the Kazan University [13].

The important method for analysis of directed networks is the reduced Google matrix (REGOMAX) algorithm developed and described in detail in [14,15]. The REGOMAX algorithm has been applied to PPI networks of SIGNOR database as reported in [16,17]. However, the number of nodes in the SIGNOR database is by a factor ten smaller than the number of nodes of the MetaCore network and its use can only be considered as a test bed for of the numerical algorithms and its conceptual base. A first description of the statistical properties of the global MetaCore network, including PageRank, CheiRank and REGOMAX characteristics, was presented in [9]. However, this work only represents a statistical study of the MetaCore network without any applications to a concrete biological problem. In this work we apply the REGOMAX analysis to the specific biological problem of fibrosis.

The important feature of the REGOMAX algorithm is that it constructs the Google matrix of a selected subset of nodes $N_r \ll N$ (here we have $N_r = 54$) taking into account not only direct links between these N_r nodes but also all indirect pathways connecting them via the global MetaCore network of much larger size N . The efficiency the REGOMAX approach was demonstrated for various applications concerning the Wikipedia and world trade networks [18–21] and we also expect that this method will provide useful and new insights in the context of fibrosis protein-protein interactions using the MetaCore network.

The paper constructing as follows: Section 2 describes the data sets and Google matrix algorithms, Section 3 presents the obtained results of the reduced Google matrix and sensitivity analysis for the particular group of 54 proteins (of Table 1) we consider here and Section 4 provides the discussion of the results and the conclusion; In the Appendix we provide additional Figures and a simple analytical estimate for the sensitivity matrix to which we refer in the main part of the work; more detail numerical data files obtained from the Google matrix computations are available at [22].

2. Data sets and methods

Here we describe the data sets, construction rules and algorithms of Google matrix.

2.1. Network data sets

The global MetaCore PPI network contains $N = 40079$ nodes with $N_\ell = 292191$ links (without self connections). The number of activation/inhibition links is $N_{\ell_+}/N_{\ell_-} = 65157/49321 \approx 1.3$ and the number of neutral links is $N_{\ell_n} = N - N_{\ell_+} - N_{\ell_-} = 177713$. Here we mainly present the results without taking into account the bi-functional nature of links. However, a part of the results takes into account this bi-functionality of links using the Ising Google matrix approach described in [9,17]. The subset of selected $N_r = 54$ fibrosis proteins (nodes) is given in Table 1; these nodes are represented

by 4 TGF- β proteins/nodes ($K_t = 1, 2, 3, 4$), 20 “up-proteins” ($K_u = 1, \dots, 20$), 20 “down-proteins” ($K_d = 1, \dots, 20$), both obtained from experiments [5] (as described above) and 10 new “X-proteins” (or “X-nodes”; $K_x = 1, \dots, 10$) whose selection is explained later. The TGF- β 4 nodes correspond to different isoforms of this protein.

The Google matrix approach used in this work is explained in detail in [10–12] and the related REGONAX algorithm is described in [9,14,15,17]. Below we present a short description of these methods following mainly the presentation given in [9], keeping the same notations.

Table 1. Table of the subset of $N_r = 54$ selected fibrosis proteins (nodes). Here K_g represents the global index of this group, $K_{t,u,d,x}$ represent the index of the four subgroups of 4 TFG- β proteins, 20 up-proteins, 20 down-proteins and 10 additional X-proteins; K (K^*) represents the local PageRank (CheiRank) index inside this group and K_M (K_M^*) represents the PageRank (CheiRank) index for the global MetaCore network of $N = 40079$ nodes; the last column gives the associated protein names.

K_g	$K_{t,u,d,x}$	K	K^*	K_M	K_M^*	Protein
1	$K_t = 1$	30	37	10780	26299	TGF- β 0
2	$K_t = 2$	9	14	235	5690	TGF- β 1
3	$K_t = 3$	13	33	968	25073	TGF- β 2
4	$K_t = 4$	20	45	4726	29508	TGF- β 3
5	$K_u = 1$	46	35	28737	25928	ADAMTS16
6	$K_u = 2$	17	34	3478	25137	FGF21
7	$K_u = 3$	52	39	40048	28152	TNFSF18
8	$K_u = 4$	16	26	2467	19160	ACAN
9	$K_u = 5$	14	31	1489	24511	RPH3A
10	$K_u = 6$	42	46	26600	29559	ADAMTS8
11	$K_u = 7$	51	47	34769	39960	MEGF6
12	$K_u = 8$	40	38	26295	27326	SV2B
13	$K_u = 9$	44	48	27111	36021	C1QTNF3
14	$K_u = 10$	50	49	34616	39841	ANO4
15	$K_u = 11$	32	24	12696	16566	IL11
16	$K_u = 12$	43	30	26624	23640	CDH10
17	$K_u = 13$	26	50	7263	30243	HTR2B
18	$K_u = 14$	19	16	4647	6551	LAMA1
19	$K_u = 15$	28	36	8342	26295	LAMA1
20	$K_u = 16$	18	17	4021	8252	RAPGEF4
21	$K_u = 17$	48	51	29945	36964	DNER
22	$K_u = 18$	36	18	22159	8569	GALNT3
23	$K_u = 19$	47	23	29145	15531	ACSBG1
24	$K_u = 20$	37	20	24786	8735	OLFM2
25	$K_d = 1$	35	40	19039	28262	CLEC3B
26	$K_d = 2$	41	41	26477	28290	SCARA5
27	$K_d = 3$	39	22	26109	11185	SLC10A6
28	$K_d = 4$	24	44	6360	29204	CXCL5
29	$K_d = 5$	33	19	14952	8729	MYOC
30	$K_d = 6$	22	28	5961	22288	IFITM1
31	$K_d = 7$	21	13	5599	4483	ANGPTL4
32	$K_d = 8$	38	25	25538	17434	SELENBP1
33	$K_d = 9$	34	52	18938	33179	FMO1
34	$K_d = 10$	49	53	34080	39427	GPR88
35	$K_d = 11$	23	27	6276	22141	HMGCS2
36	$K_d = 12$	53	43	37060	28328	LGI2
37	$K_d = 13$	29	11	9162	2485	PTN
38	$K_d = 14$	11	15	513	5974	ADORA2A
39	$K_d = 15$	27	29	7789	22652	GFRA1
40	$K_d = 16$	25	21	6718	8844	IL1R2
41	$K_d = 17$	54	42	35446	28306	IL1R2
42	$K_d = 18$	31	12	12148	3444	PEG10
43	$K_d = 19$	45	54	27829	36195	FMO2
44	$K_d = 20$	15	32	1973	24994	COX4I2
45	$K_x = 1$	1	4	3	13	β -catenin
46	$K_x = 2$	2	1	4	6	p53
47	$K_x = 3$	3	2	11	10	ESR1
48	$K_x = 4$	4	5	13	25	STAT3
49	$K_x = 5$	5	3	22	11	RelA
50	$K_x = 6$	6	6	38	82	PPAR- γ
51	$K_x = 7$	7	8	111	767	IKK- β
52	$K_x = 8$	8	7	179	198	SNAIL1
53	$K_x = 9$	10	9	237	1520	MMP-14
54	$K_x = 10$	12	10	578	2123	Flotillin-1

2.2. Google matrix construction, PageRank and CheiRank

First, we construct the Google matrix G of the MetaCore network for the simple case where the bi-functional nature of links is neglected. Furthermore, the directed links are non weighted. First one defines an adjacency matrix with elements A_{ij} being equal to 1 if node j points to node i and equal to 0 otherwise. In the next step, the stochastic matrix S describing the node-to-node Markov transitions is obtained by normalizing each column sum of the matrix A elements to unity. For dangling nodes j corresponding to zero columns of A , i.e. $A_{ij} = 0$ for all nodes i , the corresponding elements of S are defined by $S_{ij} = 1/N$. The stochastic matrix S describes a Markov process on the network: a random surfer jumps from node j to node i with the probability S_{ij} , therefore following the directed links. The column sum normalization $\sum_i S_{ij} = 1$ ensures the conservation of probability. The elements of the Google matrix G are then defined by the standard form

$$G_{ij} = \alpha S_{ij} + (1 - \alpha)/N \quad (1)$$

where $\alpha = 0.85$ is the usual damping factor [10,11]. The Google matrix is also column sum normalized and now the random surfer jumps on the network in accordance to the stochastic matrix S with a probability α and with a complementary probability $(1 - \alpha)$, to an arbitrary random node of the network. The damping factor allows to escape from possible isolated communities and ensures that the Markov process converges for long times rather quickly to a uniform stationary probability distribution. The latter is given by the PageRank vector P which is the right eigenvector of the Google matrix G corresponding to the leading eigenvalue, here $\lambda = 1$. The corresponding eigenvalue equation is then $GP = P$. According to the Perron-Frobenius theorem, the PageRank vector P has positive elements and their sum is normalized to unity. The PageRank vector elements $P(j)$ gives the probability to find the random surfer on the node j at the stationary state of the Markov process. Thus, all the nodes can be ranked by monotonically by decreasing PageRank probability. The PageRank index $K(j)$ gives the rank of the node j with the highest (lowest) PageRank probability $P(j)$ corresponding to $K(j) = 1$ ($K(j) = N$). The PageRank probability $P(j)$ is proportional, on average, to the number of ingoing links pointing to node j . However, it also takes into account the “importance” (i.e. PageRank probability) of the nodes having a direct link to j .

It is also useful to consider a network obtained by the inversion of all link directions. For this inverted network, the corresponding Google matrix is denoted G^* and the corresponding PageRank vector, called the CheiRank vector P^* , is defined such as $G^*P^* = P^*$. A detailed statistical analysis of the CheiRank vector can be found in [23,24] (see also [12]). Similarly to the PageRank vector, the CheiRank probability $P^*(j)$ is proportional, on average, to the number of outgoing links going out from node j . The CheiRank index $K^*(j)$ is also defined as the rank of the node j according to decreasing values of the CheiRank probability $P^*(j)$.

2.3. Reduced Google matrix

The concept of the REGOMAX algorithm was introduced in [14] and a detailed description of the first applications to groups of political leaders having articles in Wikipedia networks (different language editions) can be found in [15]. This algorithm determines effective interactions between a selected subset of N_r nodes enclosed in a global network of size $N \gg N_r$. These interactions are determined taking into account direct and all indirect transitions between N_r nodes via all the other $N_s = N - N_r$ nodes of the global network. We note that quite often in certain network analyses only direct links of a subset of elected N_r nodes are taking into account and their indirect interactions via the global network are omitted thus clearly missing the important interactions.

On a mathematical level the REGOMAX approach uses ideas similar to those of the Schur complement in linear algebra (see e.g. [25]) and quantum chaotic scattering in the field of quantum chaos and mesoscopic physics (see e.g. [26,27]). The Schur complement was introduced by Issai Schur in 1917 (see history in [25]). In the context of Markov chains this approach was discussed in [28]. It

is clear that the Schur complement attracted a lot of studies from far 1917. However, there are new elements, developed in [14,15], related to a specific matrix decomposition of the Schur complement which allows to understand its new features and to compute efficiently (numerically) the three related matrix components in the framework of the reduced Google matrix approach for very large networks (e.g. $N \sim 5 \times 10^6$ as for English Wikipedia).

We write the full Google matrix G of the global network in the block form

$$G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix} \quad (2)$$

where the label “r” refers to the nodes of the reduced network, i.e. the subset of N_r nodes, and “s” to the other $N_s = N - N_r$ nodes which form the complementary network acting as an effective “scattering network”. The reduced Google matrix G_R acts on the subset of N_r nodes, has the size $N_r \times N_r$. It is defined by

$$G_R P_r = P_r \quad (3)$$

Here P_r is a vector of size N_r , its components are the normalized PageRank probabilities of the N_r nodes, $P_r(j) = P(j) / \sum_{i=1}^{N_r} P(i)$. The REGOMAX approach allows to find an effective Google matrix for the subset of N_r nodes keeping fixed the relative ranking probabilities between these nodes. The reduced Google matrix G_R has the form [14,15]

$$G_R = G_{rr} + G_{rs}(\mathbf{1} - G_{ss})^{-1}G_{sr} \quad (4)$$

Furthermore it satisfies the relation (3) and it is also column sum normalized. The reduced Google matrix G_R can be represented as the sum of three components [14,15]:

$$G_R = G_{rr} + G_{pr} + G_{qr} \quad (5)$$

Here, the first component, G_{rr} , corresponds to the direct transitions between the N_r nodes; the second component, G_{pr} , is a matrix of rank 1 with all the columns being proportional (actually approximately equal to the reduced PageRank vector P_r); the third component, G_{qr} , describes all the “interesting indirect pathways” passing through the global network of G matrix. Without going into the details we mention here that mathematically (and also numerically) G_{pr} is obtained from (4) by extracting the contribution of the leading eigenvector of G_{ss} (which is very close to the PageRank of the complementary scattering network of N_s nodes) whose eigenvalue is close to unity but it is *not exactly* unity since G_{ss} is not column normalized and there is small escape probability from the N_s scattering nodes to the selected subset with N_r nodes. This eigenvector dominates therefore the matrix inverse in (4) and its contribution produces the rank 1 matrix G_{pr} and the remaining contributions of the other eigenvectors of G_{ss} to the matrix inverse provide the matrix G_{qr} which can be efficiently computed by a rapid convergent matrix series (see [14,15] for details). This point is crucial since it allows for a highly efficient numerical evaluation of all three components of G_R also for the case where a direct numerical computation of the matrix inverse of $(\mathbf{1} - G_{ss})$ is not possible due to very large values of N (note G_{ss} has the size $N_s \times N_s$ with $N_s \approx N \gg N_r$). While G_{pr} , being typically numerical dominant, has a very simple rank 1 structure the matrix G_{qr} contains the most nontrivial information related to indirect hidden transitions. Actually, mathematically both components G_{pr} and G_{qr} arise from indirect pathways through the scattering nodes (represented by the matrix inverse term in (4)) but G_{pr} can be viewed as a uniform background generated by the long-time limit (i.e. the leading eigenvector of G_{ss}) of the effective process in the complementary scattering network. The component G_{qr} gives the deviations from this background and in the following when we speak of “contributions from indirect pathways” we refer essentially to the contributions of G_{qr} . It is possible that certain matrix elements of G_{qr} are negative and if this happens this is also an important information since it indicates a reduction

from the uniform background for certain links (matrix elements of G_R , G_{rr} and G_{pr} are always positive due to mathematical reasons).

Furthermore, we also define the matrix $G_{qr}^{(nd)}$ which is obtained from the matrix G_{qr} by putting its diagonal elements to zero (these elements correspond to indirect self-interactions of nodes). We consider that this matrix contains the most interesting link information, direct links and “relevant” indirect links describing the deviations from the uniform background due to G_{pr} . The contribution of each component is characterized by their weights W_R , W_{pr} , W_{rr} , W_{qr} ($W_{qr}^{(nd)}$) respectively for G_R , G_{pr} , G_{rr} , G_{qr} ($G_{qr}^{(nd)}$). The weight of a matrix is given by the sum of all the matrix elements divided by its size N_r ($W_R = 1$ due to the column sum normalization of G_R). Examples of interesting applications and studies of reduced Google matrices associated to various directed networks are described in [16–19].

2.4. Bi-functional Ising MetaCore network

To take into account the bi-functional nature (activation and inhibition) of MetaCore links, we use the approach proposed in [17] with the construction of a larger network where each node is split into two new nodes with labels (+) and (-). These two nodes can be viewed as two Ising-spin components associated to the activation and the inhibition of the corresponding protein. In the construction of the doubled “Ising” network of proteins, each elements of the initial adjacency matrix is replaced by one of the following 2×2 matrices

$$\sigma_+ = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad \sigma_0 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (6)$$

where σ_+ applies to “activation” links, σ_- to “inhibition” links, and σ_0 when the nature of the interaction is “unknown” or “neutral”. For the rare cases of multiple interactions between two proteins, we use the sum of the corresponding σ -matrices which increases the weight of the adjacency matrix elements. Once the “Ising” adjacency matrix is obtained, the corresponding Google matrix is constructed in the usual way as described above. The doubled Ising MetaCore network corresponds to $N_I = 80158$ nodes and $N_{I,\ell} = 939808$ links given by the non-zero entries of the used σ -matrices.

Now, the PageRank vector associated to this doubled Ising network has two components $P_+(j)$ and $P_-(j)$ for every node j of the simple network. Due to the particular structure of the σ -matrices (6), one can show analytically the exact identity, $P(j) = P_+(j) + P_-(j)$, where $P(j)$ is the PageRank of the initial single PPI network [17]. The numerical verification shows that the identity $P(j) = P_+(j) + P_-(j)$ holds up to the numerical precision $\sim 10^{-13}$.

As in [17], we characterize each node by its PageRank “magnetization” given by

$$M(j) = \frac{P_+(j) - P_-(j)}{P_+(j) + P_-(j)}. \quad (7)$$

By definition, we have $-1 \leq M(j) \leq 1$. Nodes with positive M are mainly activated nodes and those with negative M are mainly inhibited nodes.

In this work the results are mainly presented for the simple network without taking into account the bi-functional nature of links. However, for an illustration we also present some results for the bi-functional network, keeping for further studies a more detailed analysis of this case.

2.5. Sensitivity derivative

The reduced Google matrix G_R of the fibrosis network describes effective interactions between N_r nodes taking into account all direct and indirect pathways via the global MetaCore network.

As in [9], we determine the sensitivity of PageRank probabilities with respect to a small variation of the matrix elements of G_R . The PageRank sensitivity of the node j with respect to a small variation of the link $b \rightarrow a$ is defined as

$$D_{(b \rightarrow a)}(j) = \frac{1}{P_r(j)} \left. \frac{dP_{r\epsilon}(j)}{d\epsilon} \right|_{\epsilon=0} = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon P_r(j)} [P_{r\epsilon}(j) - P_r(j)] . \quad (8)$$

Here, for fixed values of a and b , $P_{r\epsilon}(j)$ is the PageRank vector computed from a perturbed matrix $G_{R\epsilon}$ where the elements are defined by $G_{R\epsilon}(a, b) = G_R(a, b)(1 + \epsilon) / [1 + \epsilon G_R(a, b)]$; $G_{R\epsilon}(c, b) = G_R(c, b) / [1 + \epsilon G_R(a, b)]$ if $c \neq a$ and $G_{R\epsilon}(c, d) = G_R(c, d)$ if $d \neq b$ and for arbitrary c (including $c = a$). In other words the element $G_R(a, b)$, corresponding to the transition $b \rightarrow a$, is enhanced/multiplied with $(1 + \epsilon)$ and then the column b is resum-normlized by multiplying it with the factor $1/[1 + \epsilon G_R(a, b)]$ and all other columns $d \neq b$ are not modified. We use here an efficient algorithm described in [29] to evaluate the derivative in (8) exactly without usage of finite differences (see also the Appendix for some details on this and other related points). In the following, we consider the case where $j = a$ and we define the “sensitivity matrix” as $D_{ab} = D_{(b \rightarrow a)}(a)$. It turns out from the numerical computations that for the cases considered here all values of D_{ab} are positive: $D_{ab} > 0$ which can also be analytically understood as explained in the Appendix.

2.6. Determination of external X-proteins

From the experimental results of [5] we have 44 nodes of our selected subset (see first 44 rows of Table 1). Of course, the interactions between these nodes are very important but it is also important to determine how these 44 fibrosis proteins are influenced by external nodes. To find the most important and influential external nodes we take 5 top up- and 5 down-proteins with $K_u = 1, \dots, 5$ and $K_d = 1, \dots, 5$ from Table 1. Then we determine all external nodes having direct ingoing 134 links to one of these 5 + 5 fibrosis proteins. There are 122 such proteins. The first 44 proteins of Table 1 together with these 122 external proteins (ordered by their PageRank index) constitute an intermediary group of size 166 for which we compute first the reduced Google matrix by (4) and which we note as $G_R^{(166)}$ and from this the associated sensitivity matrix $D_{ab}^{(166)}$ (8) (with $j = a$; see also Figure A3). Then we compute the sum of sensitivities $D_s^{(5+5)}(b) = \sum_{a=5}^9 D_{ab}^{(166)} + \sum_{a=25}^{29} D_{ab}^{(166)}$ (a -sum over top 5 up- and top 5 down-proteins) for $b = 45, \dots, 166$ (new external proteins). Then we select the top 10 external proteins b with highest values of $D_s^{(5+5)}(b)$. In the following we call this new subgroup the subgroup of X-proteins (or X-nodes). They are given in the last 10 rows of Table 1 (for $K_g = 45, \dots, 54$ and $K_x = 1, \dots, 10$). We mention that these 10 X-proteins have index values of (1, 2, 3, 4, 6, 8, 10, 15, 27) with respect to the initial list of 122 external proteins (which were already PageRank ordered). It turns out that this procedure selects automatically 10 external nodes which have approximately the strongest PageRank values. This can be understood by the fact that the matrix $D_{ab}^{(166)}$ is roughly proportional to $P(b)$ except for a small number cells with strong peak values (see also Figure A3 and the Appendix for a theoretical explanation). In this way we obtain the full subset of 54 fibrosis proteins given in Table 1. The REGOMAX analysis is performed for these 54 fibrosis proteins and unless stated otherwise all results for G_R , D_{ab} etc. refer to this group of 54 proteins.

3. Results

In this Section we present the results of Google matrix analysis of fibrosis protein-protein interactions.

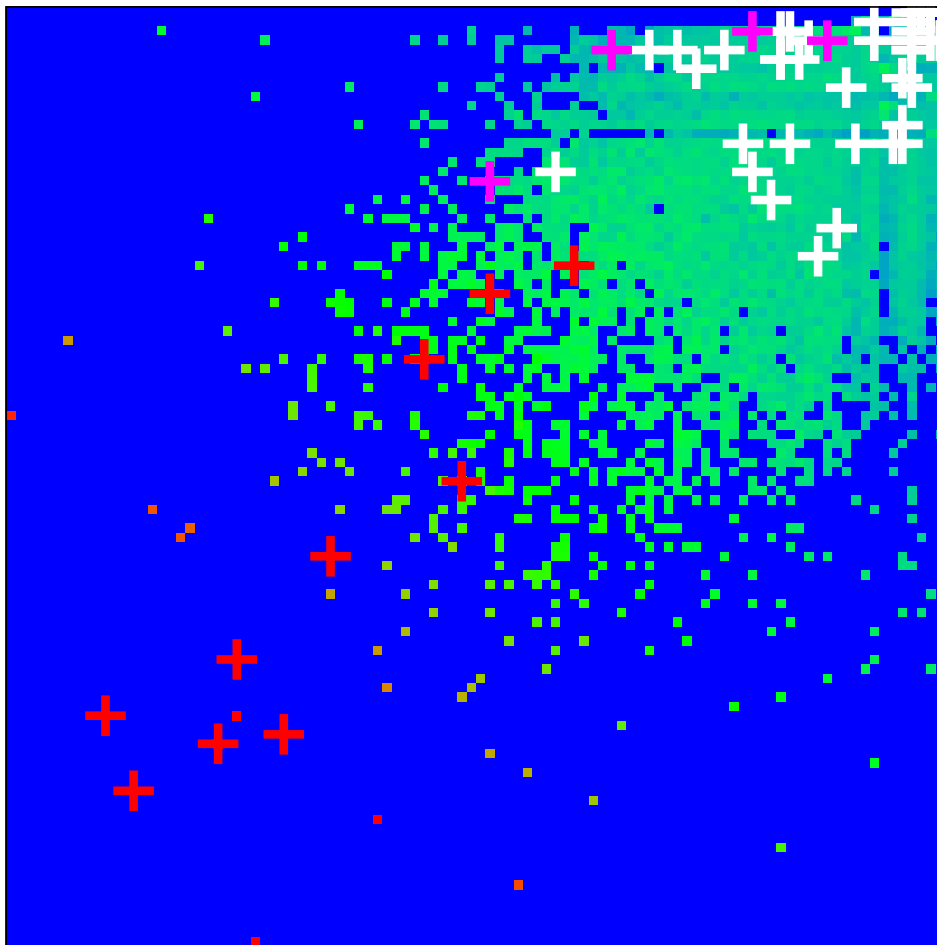


Figure 1. Density of nodes $W(K_M, K_M^*)$ on PageRank-CheiRank plane (K_M, K_M^*) averaged over 100×100 logarithmically equidistant grids for $0 \leq \ln K_M, \ln K_M^* \leq \ln N$, the density is averaged over all nodes inside each cell of the grid, the normalization condition is $\sum_{K_M, K_M^*} W(K_M, K_M^*) = 1$. Color varies from blue at zero value to red at maximal density value. In order to increase the visibility large density values have been reduced to (saturated at) $1/16$ of the actual maximum density and typical green cells correspond to density values of $\sim 1/2^8$ of the (reduced) maximum density. The x -axis corresponds to $\ln K_M$ and the y -axis to $\ln K_M^*$ with $K_M (K_M^*)$ being the global PageRank (CheiRank) index for the full MetaCore network. The crosses mark the positions of the 54 proteins of Table 1 with colors: red for the X-proteins, pink for the TGF- β subgroup and white for the up- and down-protein subgroups.

3.1. Fibrosis proteins on PageRank-CheiRank plane

As in [9] we determine the density distribution of all proteins of the MetaCore network on the PageRank-CheiRank plane of logarithms $(\ln K, \ln K^*)$ of indexes (K, K^*) which is shown in Figure 1. The whole plane is divided on 100×100 logarithmically equidistant cells and the density is defined as the number of proteins in a given cell divided by a total possible nodes in a given cell (this approach is discussed in more detail e.g. in [24]). The highest density is located at top indexes K, K^* but in this region there is a relatively small number of proteins. The positions of fibrosis proteins of Table 1 are marked by crosses of 3 colors: red for 10 external X-proteins ($K_x = 1, \dots, 10$), pink for 4 TGF- β proteins ($K_t = 1, 2, 3, 4$) and white for the 40 up- and down-proteins ($K_u, K_d = 1, \dots, 20$). We see that X-proteins have highest rank positions; 2 of the TGF- β proteins approximately follow after K_x values of PageRank and 2 others have significantly lower K -rank positions (positions in K^* -rank are rather low); proteins K_u and K_d have on average rather low rank positions (very large K, K^* values). Therefore the X-proteins have the highest network influence and communicativity (small K, K^* values).

The presentation of Figure 1 uses the global MetaCore rank index values (in the following these values are noted as K_M, K_M^* ; see also Table 1). For the selected subset of 54 fibrosis proteins we note their local rank indexes in this group as K, K^* which are also given in Table 1. The distribution of these 54 local rank indexes on the PageRank-Cheirank plane of size 54×54 is given in Appendix Figure A1.

3.2. Reduced Google matrix of fibrosis

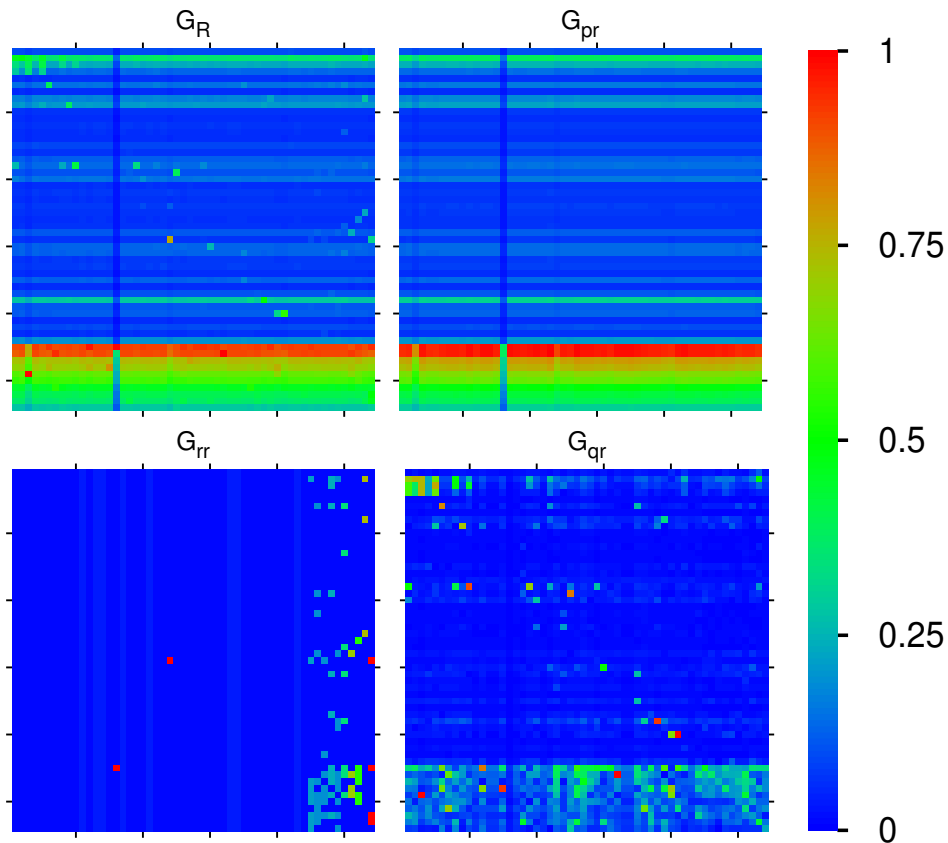


Figure 2. Color density plots of the matrix components $G_R, G_{pr}, G_{rr}, G_{qr}$ for the group of Table 1; the x -axis corresponds to the first (row) index (increasing values of K_g from top to down) and the y -axis corresponds to the second (column) index of the matrix (increasing values of K_g from left to right). The outside ticks indicate multiples of 10 of K_g . The numbers in the color bar correspond to $\sqrt{|g|/g_{\max}}$ with g being the value of the matrix element and g_{\max} being the maximum value. In order to increase the visibility for the cases of G_R, G_{rr}, G_{qr} the maximum value has been reduced (saturated) to the value of the 3rd largest value of g for each case and the cells corresponding to the 1st and 2nd largest values are reduced to the saturation value. In particular $G_R(45, 15)$ ($G_R(46, 13)$) has been reduced from 0.876387 (0.297512) to $G_R(49, 3) = 0.208777$; $G_{rr}(45, 16)$ ($G_{rr}(29, 24)$) has been reduced from 0.850004 (0.121432) to $G_{rr}(29, 54) = 0.019322$ (same 3rd value also for the other three cells in column 54); $G_{qr}(49, 3)$ ($G_{qr}(40, 41)$) has been reduced from 0.240629 (0.062024) to $G_{qr}(46, 32) = 0.041108$. For the matrix G_{qr} there are some negative values and here we show their absolute values (see text).

The reduced Google matrix G_R of 54 fibrosis proteins and its 3 matrix components G_{pr}, G_{rr}, G_{qr} are shown in Figure 2. The weights of these matrices are: $W_{pr} = 0.9522$, $W_{rr} = 0.0228$, $W_{qr} = 0.0250$, ($W_{qr}^{(nd)} = 0.0211$) and $W_R = 1$ (due to the column sum normalization of G_R). Thus the weight of G_{pr} is significantly higher as compared to the two other components. This behavior is quite typical and was also observed for Wikipedia networks (see e.g. [15,18,19]). The physical reason of this is that G_{pr} is obtained from the contribution of the leading eigenvector of the matrix G_{ss} whose eigenvalue is close to unity and dominates numerically the matrix inverse in (4) (see also the discussion in the last

section and [14,15] for details). Furthermore G_{pr} has a very simple structure since it is of rank one, i.e. all columns are exact multiples of the first column. Furthermore, these columns are approximately equal to the local PageRank vector. Therefore the component G_{pr} does not provide any new interesting information about possible interactions other than that it trivially reproduces the PageRank vector.

Numerically G_R is dominated by G_{pr} (with its high weight $W_{pr} = 0.9521$). However, the other two components give us important additional information about direct interactions between the 54 fibrosis proteins (G_{rr}), and, even more important, about all indirect interactions (G_{qr}) between these proteins via the global MetaCore network performing an effective summation over all indirect pathways (see [14,15] for details). The weights of the components of G_{rr} and G_{qr} are comparable. We also see that nearly all direct transitions visible in G_{rr} are from X-proteins to other proteins (all subgroups) which is not astonishing due to the selection rule that any X-node must have at least one direct link to the first 5 top- or first 5 up-proteins and also due to the fact that they have rather high PageRank but also CheiRank positions (according to Table 1, Figure 1 and Appendix Figure A1). Since the PageRank probabilities are higher for X-proteins (see Figure 1) there are rather strong transitions to these X-proteins well visible for G_R , G_{pr} and to a lesser extent also in G_{qr} . We note that the component G_{qr} has a small number of non-vanishing diagonal matrix elements which appear due to the possibility that a pathway over the global MetaCore network can return to an initial protein.

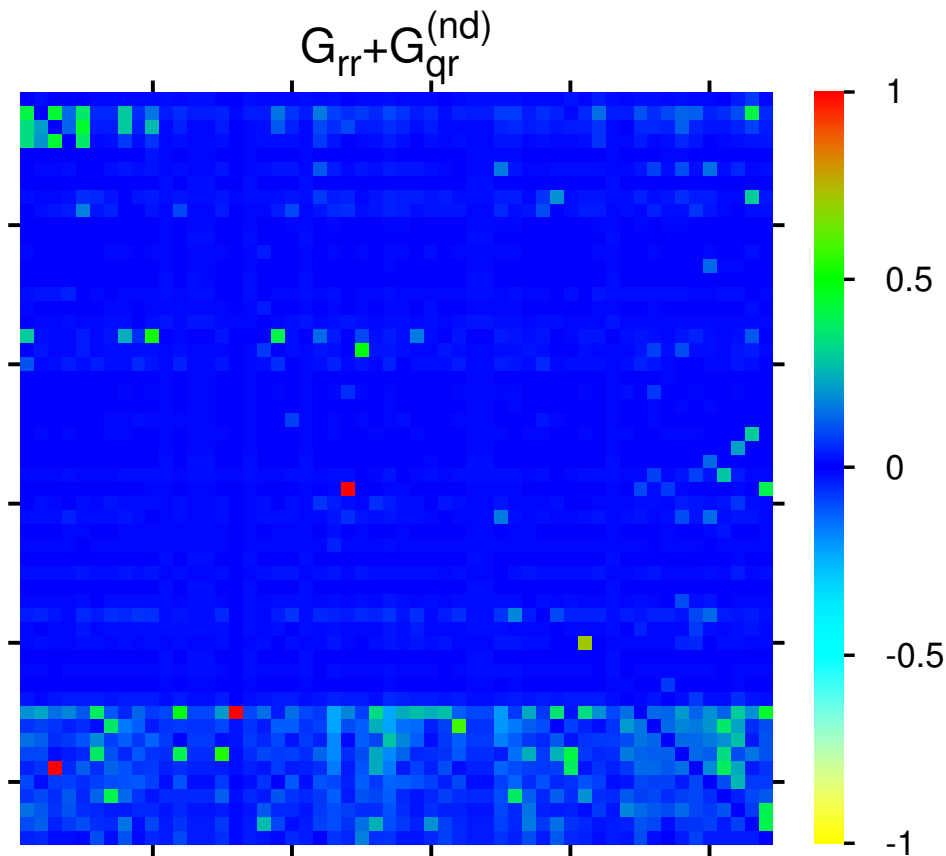


Figure 3. Color density plot $G_{rr} + G_{qr}^{(nd)}$ for the group of Table 1 (same plot style and color bar as Figure 2). The matrix element at (45, 16) ((49, 3)) has been reduced from 0.849861 (0.240632) to the value 0.121433 at (29, 24); a few matrix elements of $G_{rr} + G_{qr}^{(nd)}$ have negative values visible as cyan color (see text).

It should be noted that a few matrix elements of G_{qr} have negative values. Such a situation has been already found for other directed networks, e.g. Wikipedia networks studied in [15]. To be more precise for G_{qr} and $G_{rr} + G_{qr}^{(nd)}$ there are about 340 out of 2916 negative values ($\approx 11\%$). Most of them are very small. However, there are 10 values between -0.00668 and -0.00334 for both matrices

corresponding to 5-10% of the red-color saturation value used for G_{qr} . However, in Figure 2 only the modulus of matrix elements is shown in order to have a uniform style for all components (the 10 strongest negative values of G_{qr} correspond to green color with color bar values of 0.3 to 0.4 and after taking the modulus). Of course, the matrix elements of G_R , G_{Tr} and G_{Pr} are always positive due to strict mathematical properties.

Figure 3 shows the effective matrix of transitions for direct links and relevant indirect pathways (without self interactions) which is obtained as the sum of the two components $G_{Tr} + G_{qr}^{(nd)}$. There are also some cells with cyan color for negative matrix elements (corresponding to -0.3 to -0.2 in units of the color bar for the strongest 10 negative values). Most links are due to the interactions from K_x to K_t, K_u, K_d proteins but there are also some other significant transitions between the other members of the group of 54 proteins.

3.3. Network diagrams of fibrosis interactions

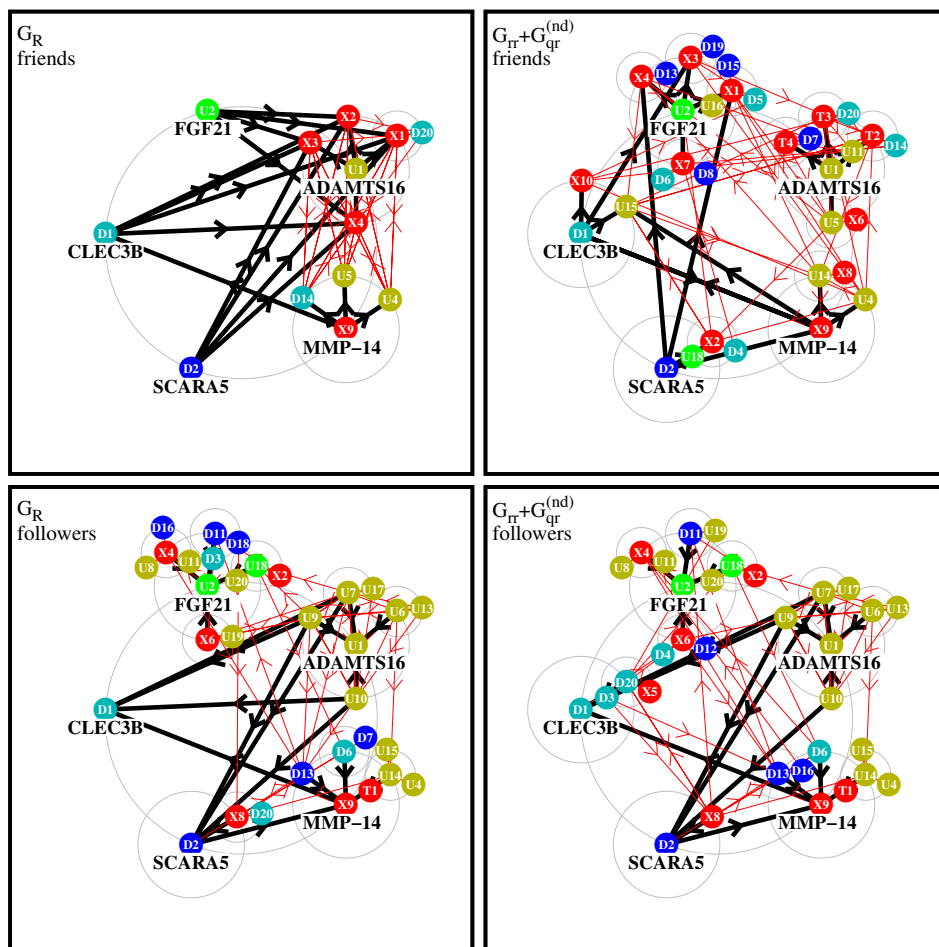


Figure 4. Effective friend and follower networks generated from G_R and $G_{Tr} + G_{qr}^{(nd)}$. Starting from 5 top nodes the 4 strongest friends/followers for each initial node are selected and links are shown by thick black arrows. For each selected new node further 4 strongest friends/followers are selected and corresponding new links are shown by thin red arrows. In this procedure the direct links between two nodes belonging both to one of the two subgroups of X-proteins or TGF- β proteins are not taken into account. The node labels T_j, U_j, D_j, X_j (with j being an integer value) correspond the local subgroup index $K_t = j, K_u = j, K_d = j$ or $K_x = j$ respectively which are given in Table 1. Further details about precise selection rules of links, top nodes and color attribution are given in the text.

In this section we discuss two types of effective networks (of most important PPI links) obtained from the two matrices G_R and $G_{rr} + G_{qr}^{(nd)}$, the latter containing the “interesting” links without the uniform background generated by the component G_{pr} (and without self-interactions). We remind that the value of a matrix element $g(a, b)$ (with g being either G_R or $G_{rr} + G_{qr}^{(nd)}$) corresponds to the strength of the link $b \rightarrow a$. If this value is sufficiently high we say that a is a “friend” of b and b is a “follower” of a . This distinction allows to construct for each matrix two types of effective networks by choosing a few number of “top nodes” and adding a certain number of the strongest friends (or followers) according to the values of $|g(a, b)|$ and repeating this procedure for a modest number of depth levels.

In Figure 4, we show four graphical representations of such effective networks for the two cases of friend or follower networks and the two matrices G_R and $G_{rr} + G_{qr}^{(nd)}$ visible in Figures 2 and 3. In these figures and the remainder of this subsection we use the short notations Tj, Uj, Dj or Xj for a protein/node where $j = 1, 2, \dots$ is the integer value of the subgroup index K_t, K_u, K_d or K_x respectively with real protein names given in Table 1.

To construct the effective network for a matrix component g (with g being either G_R or $G_{rr} + G_{qr}^{(nd)}$) we first choose 5 initial top nodes/proteins corresponding to $U1, U2$ (ADAMTS16, FGF21), $D1, D2$ (CLEC3B, SCARA5) and $X9$ (MMP-14). $U1, U2$ ($D1, D2$) have the strongest positive (negative) TGF- β response observed experimentally in [5]. The node corresponding to $X9$ (MMP-14) produces the strongest sensitivity D_{ab} (among those elements D_{ab} where a is an up- or down-protein and b is a TGF- β or X-protein; see next subsection for details on this). These 5 proteins form the set of level-0 nodes which are placed on a large circle.

We attribute the color red to the combined subgroups of 10 external X-proteins ($K_x = 1, \dots, 10$) and 4 TGF- β proteins ($K_t = 1, 2, 3, 4$) The transitions inside this red group are not taken into account since we are mainly interested in the influence of this group on the other up- and down-proteins. We attribute two colors to the up-proteins (olive green to $U1$, green to $U2$) and two colors to the down-proteins (cyan to $D1$, blue to $D2$). Inside the group of up-proteins, we attribute the color olive green to a protein Uj if Uj is a stronger follower of $U1$ than of $U2$ with respect to $g = G_{rr} + G_{qr}^{(nd)}$, i.e. if $g(K_u = 1, K_u = j) > g(K_u = 2, K_u = j)$ and green otherwise. In other words, we compare the strength of the links $Uj \rightarrow U1$ and $Uj \rightarrow U2$ to determine if Uj has the color olive green of $U1$ or green of $U2$. In a similar way, comparing the strength of the two links from a Dj protein to either $D1$ or $D2$, we attribute the two colors cyan and blue to down-proteins. This attribution rule, using the strongest followers with respect to $G_{rr} + G_{qr}^{(nd)}$ of the two top nodes inside a subgroup, ensures that for all colors there is a considerable number of proteins and it is the same for all four network diagrams (both matrices and both friend/follower cases).

For each of the 5 level-0 proteins, noted a , we first search the 4 strongest friends (followers), noted b , with largest value of $|g(b, a)|$ (or $|g(a, b)|$) corresponding the strongest link $a \rightarrow b$ (or $b \rightarrow a$), where the matrix g is either G_R or $G_{rr} + G_{qr}^{(nd)}$. The new nodes b (if not yet present in the set of level-0 nodes) nodes form the set of new level-1 nodes and they are placed on medium size circles of level 1 around the corresponding “parent” node a of level-0. The links between the nodes a and b are drawn as thick black arrows with direction $a \rightarrow b$ ($b \rightarrow a$) for the friend (follower) case. If a node b already belongs to the set of level-0 nodes we also draw a thick black arrow but using its already existing position on the initial large circle. If a node b has several parent nodes a we place it only on one medium circle preferably around a parent node of the same color if possible.

This procedure is repeated once: for each level-1 protein we determine the 4 strongest level-2 friends (or followers) which are placed on smaller circles of level 2 around the corresponding level-1 protein provided that they are not yet present in the former sets of level-0 or level-1 proteins. The links corresponding to this stage are drawn as thin red arrows with the same directions as in the first stage (we also draw thin arrows for a selected nodes who were already previously selected and using their former positions). As already mentioned above, links where *both* proteins (a and b) belong to

the combined set of X- and TGF- β proteins are not taken into account (otherwise they would strongly dominate these diagrams). We limit our-self to two stages of the procedure (i.e three levels of nodes) because otherwise the diagrams would require still smaller circles and many nodes would be hidden by former nodes. We note for the friend- G_R diagram a further third stage would not add any new nodes since the strongest friends of level-2 are already in the network. For the other cases additional further stages would only add a few number of new nodes with a quite rapid saturation of the network at some limit level where no new nodes are selected.

Figure 4 shows diagrams of level-2 networks for the cases of friend (top row) and follower (bottom row) diagrams and the two matrices $g = G_R$ (left column) or $g = G_{rr} + G_{qr}^{(nd)}$ (right column). Concerning the two cases of $g = G_{rr} + G_{qr}^{(nd)}$ about 15% of the shown arrows correspond to negative values of the matrix element of g (link strength is determined by the modulus of the matrix element).

For the friend network of G_R there is a dominance of links (black arrows) $U1, U2, D1, D2 \rightarrow X_j$ for certain X-proteins X_j which can be understood by the fact most X_j proteins have a significantly higher PageRank probabilities than the other proteins. Furthermore, the total number of nodes in this diagram is quite small because the strongest friends of level-1 nodes ($X1, X2, X3, U4, U5, D14$) are mostly other level-1 nodes and there is only one new level-2 node ($D20$). This diagram is obviously dominated by the uniform background (of the component G_{pr} contributing to G_R) which tends to select mostly the “same new friends” at each level.

For the friend case of $G_{rr} + G_{qr}^{(nd)}$ the network structure is significantly richer, since here the global PageRank transitions (due to the uniform background of G_{pr}) do not play a role. The group around $U1$ includes $T2, T3, T4$. Thus we see a formation of groups of friends around $U1$ and especially $U2$ with many friends, smaller groups of friends appear around $D1, D2$ and $X9$.

For the follower network of $G_{rr} + G_{qr}^{(nd)}$, the largest groups of followers are again formed around $U1, U2$. In the group around $U1$ we have only other up-proteins while in the group around $U2$ we have up-, down- and X-proteins. The third group around $X9$ is composed of several up- and down-proteins as well as one TGF- β protein ($T1$) on level 2. The fourth group around $D1$ includes $D3, D20$ and $X5$ but there are also two other followers $U7, U9$ but which are placed on the $U1$ -circle. The fifth group around $D2$ includes only $X8$ (on its own circle) and $U7, U9, U10$ from the $U1$ -circle.

The follower network of G_R matrix has a similar structure, since for followers the contribution of G_{pr} is not so significant such that several links of followers of G_R and $G_{rr} + G_{qr}^{(nd)}$ are similar.

It should be noted that the few negative matrix elements of G_{qr} have a modest impact on the network diagrams of $G_{rr} + G_{qr}^{(nd)}$ ($\sim 15\%$ of links and only one stage 1 link for the friend case).

These network diagrams allow to obtain a qualitative graphical view on the most significant fibrosis PPI interactions from a friend or a follower point of view.

We note that in principle it is possible to choose another initial set of 5 proteins at level-0. In Appendix Figure A2, we show the network diagrams for the modified level-0 set: $D1, D2, U9, U18$ and $X9$. Here the 4 up- and down-proteins have the highest sensitivity with respect to X-proteins (see next section). Some features are quite similar as to the first case: the friend diagram of G_R has only a modest number of nodes with a domination of X-proteins and generally the groups associated to the two up-top nodes appear somewhat larger than the groups for the two down-top nodes.

3.4. Sensitivity of fibrosis proteins

In addition to the matrix components $G_R, G_{pr}, G_{rr}, G_{qr}$ and the network diagrams (of G_R and $G_{rr} + G_{qr}^{(nd)}$) it is also important to analyze the sensitivity matrix D_{ab} defined previously in (7). This matrix D_{ab} gives the sensitivity of a protein a with respect to a small variation of the transition matrix element of G_R from protein b to a on the basis of logarithmic derivative of the PageRank probability (see subsection 2.5 and also the Appendix for more technical details on this).

As described previously (see subsection 2.6), we first compute the sensitivity matrix $D_{ab}^{(166)}$ associated to $G_R^{(166)}$ being the reduced Google matrix for a larger intermediary subset containing the

44 TGF- β , up- and down-proteins and further 122 external proteins having direct links (of the full MetaCore network) to the first 5 up- ($K_u = 1, \dots, 5$) and the first 5 down-proteins ($K_d = 1, \dots, 5$). This matrix is shown in Appendix Figure A3.

Then from the set of 122 external proteins we select the 10 proteins b with the largest effective sensitivity given by the sum $D_s^{(5+5)}(b) = \sum_{a=5}^9 D_{ab}^{(166)} + \sum_{a=25}^{29} D_{ab}^{(166)}$ (see Subsection 2.6) which form the group of 10 X-proteins. The 44 TGF- β , up- and down-proteins together with these 10 X-proteins form our main group of 54 proteins given Table 1 and for which have presented results of the reduced Google matrix in the last subsections.

The sensitivity matrix D_{ab} of size 54×54 for this main group is shown in Figure 5 with zoomed parts visible in Figure 6.

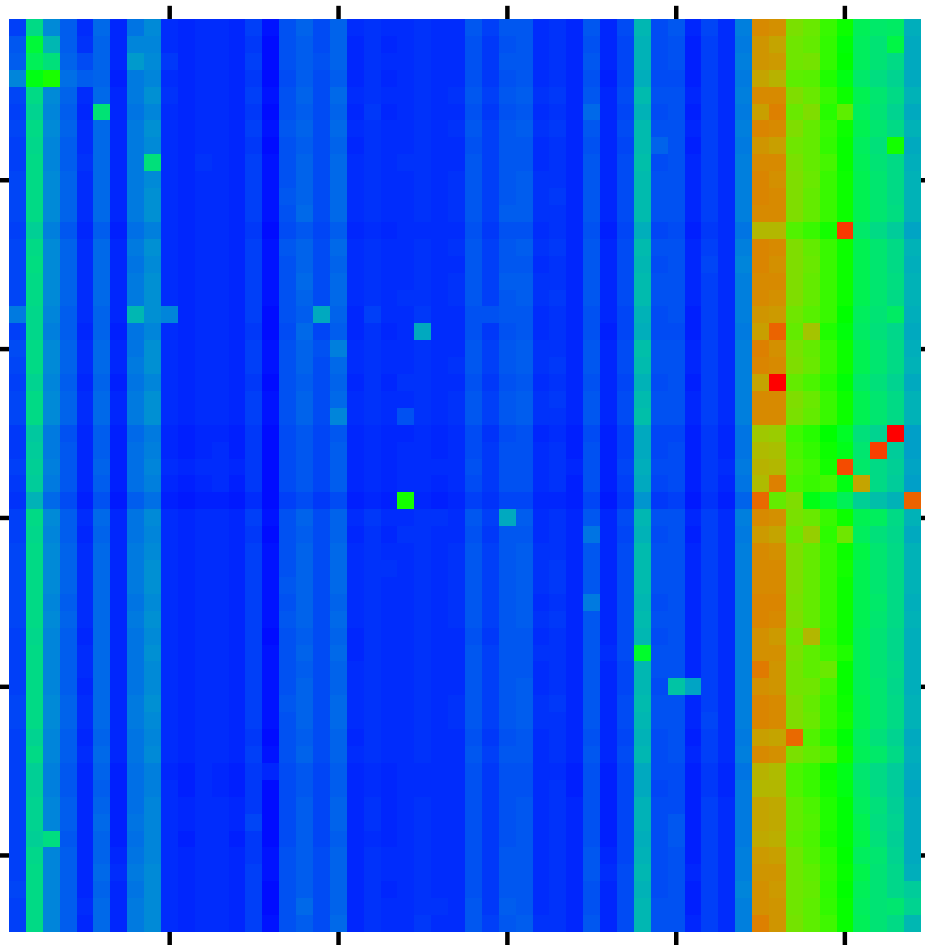


Figure 5. Color density plot of the sensitivity matrix D_{ab} of fibrosis proteins of Table 1; the axes and colors are defined as in Figure 2 (without saturation); the strongest top 40 sensitivity values are given in Table 2.

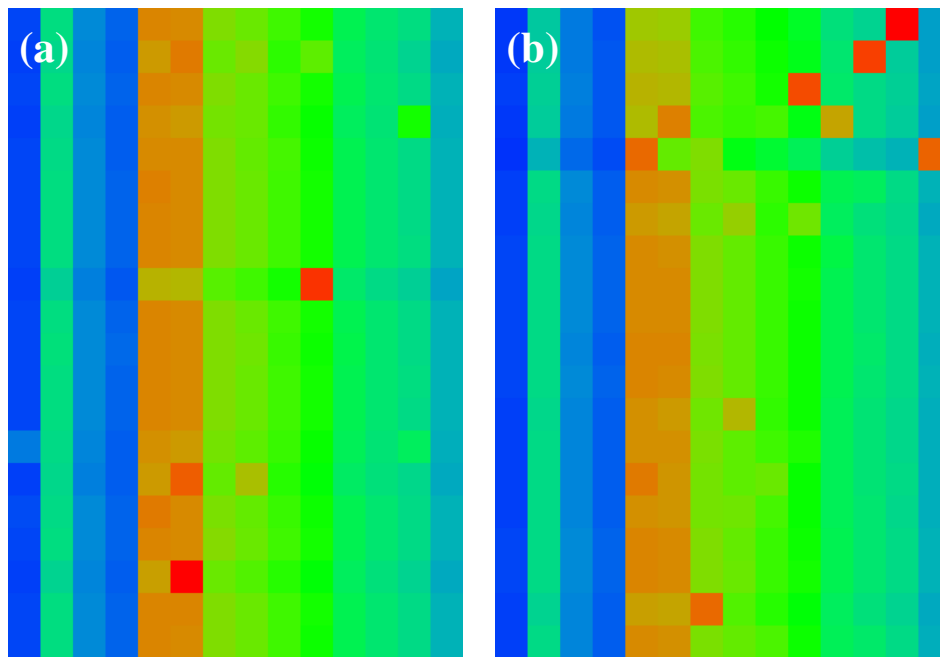


Figure 6. Zoomed parts of sensitivity matrix D_{ab} of Figure 5. Both panels show a selected sub-region of Figure 5 with the index a (vertical axis from top to down) belonging to the set of up-nodes ($a = 5, \dots, 24$ in panel (a)) or down-nodes ($a = 25, \dots, 44$ in panel (b)) and the index b (horizontal axis from left to right) corresponds for both panels to the 4 nodes of the TGF- β subgroup ($b = K_t = 1, \dots, 4$ for 4 left columns in each panel) and the 10 nodes of the X-proteins ($b = 45, \dots, 54$ or $K_x = 1, \dots, 10$ for 10 right columns in each panel).

Table 2. List of 40 top protein pairs (a, b) with strongest sensitivity matrix element D_{ab} with a belonging to the subgroups of up- or down-proteins and b belonging to the subgroups of TGF- β and X-proteins. The first column gives the ranking index K_s of D_{ab} matrix elements ordered by a decreasing value, the 2nd to 4th columns provide the $K_g, K_{u,d}$ indexes and the name of the protein (a), the 5th to 7th columns provide the $K_g, K_{t,x}$ indexes and the name of the protein (b), and the 8th column shows the value of D_{ab} . See also Figure 5 which shows a color density plot for all matrix elements D_{ab} and Table 1 for the list of considered proteins. An ordered list of all 560 values of sensitivity influence values D_{ab} of TGF- β or X-proteins (for “ b ”) on up-/down proteins (for “ a ”) is available at [22].

K_s	$K_g(a)$	$K_{u,d}(a)$	Protein(a)	$K_g(b)$	$K_{t,x}(b)$	Protein(b)	D_{ab}
1	25	$K_d = 1$	CLEC3B	53	$K_x = 9$	MMP-14	0.263109
2	22	$K_u = 18$	GALNT3	46	$K_x = 2$	p53	0.259298
3	13	$K_u = 9$	C1QTNF3	50	$K_x = 6$	PPAR- γ	0.225877
4	26	$K_d = 2$	SCARA5	52	$K_x = 8$	SNAIL1	0.219938
5	27	$K_d = 3$	SLC10A6	50	$K_x = 6$	PPAR- γ	0.214345
6	29	$K_d = 5$	MYOC	54	$K_x = 10$	Flotillin-1	0.200157
7	19	$K_u = 15$	LAMA1	46	$K_x = 2$	p53	0.199892
8	43	$K_d = 19$	FMO2	47	$K_x = 3$	ESR1	0.196550
9	29	$K_d = 5$	MYOC	45	$K_x = 1$	β -catenin	0.196394
10	39	$K_d = 15$	GFRA1	45	$K_x = 1$	β -catenin	0.184019
11	6	$K_u = 2$	FGF21	46	$K_x = 2$	p53	0.182339
12	20	$K_u = 16$	RAPGEF4	45	$K_x = 1$	β -catenin	0.182303
13	28	$K_d = 4$	CXCL5	46	$K_x = 2$	p53	0.181444
14	10	$K_u = 6$	ADAMTS8	45	$K_x = 1$	β -catenin	0.177848
15	42	$K_d = 18$	PEG10	45	$K_x = 1$	β -catenin	0.177726
16	35	$K_d = 11$	HMGCS2	45	$K_x = 1$	β -catenin	0.177443
17	15	$K_u = 11$	IL11	45	$K_x = 1$	β -catenin	0.177227
18	35	$K_d = 11$	HMGCS2	46	$K_x = 2$	p53	0.176906
19	21	$K_u = 17$	DNER	45	$K_x = 1$	β -catenin	0.176820
20	11	$K_u = 7$	MEGF6	45	$K_x = 1$	β -catenin	0.176612
21	36	$K_d = 12$	LGI2	45	$K_x = 1$	β -catenin	0.176606
22	7	$K_u = 3$	TNFSF18	45	$K_x = 1$	β -catenin	0.176603
23	41	$K_d = 17$	IL1R2	45	$K_x = 1$	β -catenin	0.176598
24	14	$K_u = 10$	ANO4	45	$K_x = 1$	β -catenin	0.176556
25	34	$K_d = 10$	GPR88	45	$K_x = 1$	β -catenin	0.176432
26	23	$K_u = 19$	ACSBG1	45	$K_x = 1$	β -catenin	0.176323
27	5	$K_u = 1$	ADAMTS16	45	$K_x = 1$	β -catenin	0.176315
28	12	$K_u = 8$	SV2B	45	$K_x = 1$	β -catenin	0.176264
29	17	$K_u = 13$	HTR2B	45	$K_x = 1$	β -catenin	0.176197
30	16	$K_u = 12$	CDH10	45	$K_x = 1$	β -catenin	0.176192
31	24	$K_u = 20$	OLFM2	45	$K_x = 1$	β -catenin	0.176038
32	32	$K_d = 8$	SELENBP1	45	$K_x = 1$	β -catenin	0.175939
33	33	$K_d = 9$	FMO1	45	$K_x = 1$	β -catenin	0.175776
34	33	$K_d = 9$	FMO1	46	$K_x = 2$	p53	0.175367
35	30	$K_d = 6$	IFITM1	45	$K_x = 1$	β -catenin	0.175056
36	44	$K_d = 20$	COX4I2	45	$K_x = 1$	β -catenin	0.174371
37	23	$K_u = 19$	ACSBG1	46	$K_x = 2$	p53	0.174167
38	34	$K_d = 10$	GPR88	46	$K_x = 2$	p53	0.173893
39	5	$K_u = 1$	ADAMTS16	46	$K_x = 2$	p53	0.173822
40	14	$K_u = 10$	ANO4	46	$K_x = 2$	p53	0.173770

The list of all 560 sensitivity matrix values D_{ab} with a belonging to the subgroups of up- or down-proteins and b belonging to the subgroups of TGF- β and X-proteins is available at [22]. The strongest 40 D_{ab} values of this list are shown in Table 2. Among the top 3 pairs we find that the protein MMP-14 gives the top sensitivity (influence) on the protein CLEC3B ($D_{ab} = 0.263109$); next is the protein p53 giving the sensitivity ($D_{ab} = 0.259298$) on the protein GALNT3, and the third place is for the sensitivity of C1QTNF3 from PPAR- γ ($D_{ab} = 0.225877$).

We mention that the appearance of MMP-14 ($K_x = 9$) at the top position of Table 2 is the reason why we selected this protein as one of the 5 top nodes in the net diagrams discussed in the last subsection. For the net diagrams shown in Figure 4, the other four top nodes were simply chosen as the first two up- ($K_u = 1, 2$) and down-proteins ($K_d = 1, 2$). However, for the net diagrams shown in Appendix Figure A2 also the two top up- and down-nodes were chosen by the criterion of top positions in Table 2 resulting in $K_u = 9, 18$ and $K_d = 1, 2$.

We have also computed the effective TGF- β sensitivity on up- or down-proteins (noted a) defined by the sum $D_s^{(\text{TGF}-\beta)}(a) = \sum_{b=1}^4 D_{ab}$. Ordering these values in decreasing order we obtain the ranking index $K_s^{(\text{TGF}-\beta)} = 1, \dots, 40$ whose dependence on K_u and K_d is visible Appendix Figure A4. We see that for the up-proteins we have 14 ranking values located at $K_s^{(\text{TGF}-\beta)} \leq 20$ and for the down-proteins only 6 values at $K_s^{(\text{TGF}-\beta)} \leq 20$ (with 3 values at $K_s^{(\text{TGF}-\beta)} = 18, 19, 20$). This shows that the overall influence of TGF- β proteins is somewhat stronger on the up-proteins as compared to the down-proteins.

However, we mention that the different values of $D_s^{(\text{TGF}-\beta)}(a)$ used to determine this ranking have only modest size variations in the interval 0.0250 to 0.0465 with most values between 0.040 and 0.043. Furthermore, in global the external X-proteins have a much higher influence (on up- and down-proteins) than the TGF- β proteins. For instance in Table 2 the TGF- β proteins do not appear at all (in the three “ b ” columns) and in the full list of 560 entries the first appearance of a TGF- β protein is at the ranking position $K_s = 319$.

Both of these points can be explained by the approximate expression $D_{ab} \approx [1 - P_r(a)]P_r(b) \approx P_r(b)$ which is derived in the appendix for a simplified model of a rank 1 G_R matrix but which also holds approximately for arbitrary G_R matrices due to the strong numerical weight of the rank 1 component G_{pr} . This behavior is also confirmed, for a “uniform background”, by Figures 5 and 6 for D_{ab} and Appendix Figure A3 for $D_{ab}^{(166)}$. However, there are typically some exceptional peaks at a few values of the (a, b) index pair where strong deviations from this simple expression are possible and which are due to the components of G_{rr} and G_{qr} in G_R .

Essentially $D_{ab} \sim P_r(b)$ does not (strongly) depend on a explaining that the values of the partial sum $D_s^{(\text{TGF}-\beta)}(a) = \sum_{b=1}^4 D_{ab}$ show only modest size variations. Furthermore, Table 2, containing the largest D_{ab} values (with b being either an X or a TGF- β protein and a being an up- or down protein), is dominated by X-proteins which have mostly larger $P_r(b)$ values than the TGF- β proteins.

We also determine the global influence on the whole group of fibrosis up- and down-proteins by computing the sum $D_s^{(u/d)}(b) = \sum_{a=5}^{44} D_{ab}$ (i.e. the a -sum is over up- and down-proteins) for each X or TGF- β protein b . The resulting values of this quantity are provided in Table 3. According to the simple expression for D_{ab} we have a linear dependence of $D_s^{(u/d)}(b)$ on $P_r(b)$ and due to the a -sum the effect of exceptional peaks is strongly reduced. This linear dependence is clearly visible in Table 3 and Appendix Figure A5. A simple linear fit $D_s^{(u/d)}(b) = \eta P_r(b)$ provides the value $\eta = 39.5 \pm 1.4$ for the coefficient and a more general power law fit $D_s^{(u/d)}(b) = \tilde{\eta}[P_r(b)]^\kappa$ results in a similar coefficient $\tilde{\eta} = 41.9 \pm 4.3$ and an exponent $\kappa = 1.017 \pm 0.028$ close to unity.

However, Table 3 also shows that at the ranking positions 9 ($K_x = 9$ for MMP-14) and 10 ($K_t = 2$ for TGF- β 1) there is one ranking inversion between $D_s^{(u/d)}(b)$ and $P_r(b)$. The value of $D_s^{(u/d)}(K_x = 9)$ is roughly 30% larger than $D_s^{(u/d)}(K_t = 2)$ while the PageRank value of the former is very slightly (0.15%) smaller than the value of the latter (both PageRank values are nearly identical). In Appendix Figure A5 both of these proteins correspond to two data points with a certain visible (vertical) difference for $D_s^{(u/d)}(b)$ but with no visible (horizontal) difference for $P_r(b)$.

Table 3. Values of the sum $D_s^{(u/d)}(b) = \sum_{a=5}^{44} D_{ab}$ (i.e. the a -sum is over up- and down-proteins) for b belonging to the TGF- β or the X-proteins subgroups. The list is ordered with respect to decreasing $D_s^{(u/d)}(b)$ values with the first column giving the corresponding ranking index; the 2nd and 3rd columns giving the $K_g, K_{t,x}$ indexes; the 4th and 5th columns containing the local PageRank index K and the name of the protein b and the 6th and 7th columns giving the values of $D_s^{(u/d)}(b)$ and the local PageRank probability $P_r(b)$. Both K and $P_r(b)$ correspond to the group of 54 fibrosis proteins of Table 1.

Rank	$K_g(b)$	$K_{t,x}(b)$	K	Protein(b)	$D_s^{(u/d)}(b)$	$P_r(b)$
1	45	$K_x = 1$	1	β -catenin	6.809993	0.175768
2	46	$K_x = 2$	2	p53	6.789229	0.171249
3	47	$K_x = 3$	3	ESR1	4.513399	0.113285
4	48	$K_x = 4$	4	STAT3	4.109638	0.104088
5	49	$K_x = 5$	5	RelA	3.343309	0.085443
6	50	$K_x = 6$	6	PPAR- γ	3.086237	0.070668
7	51	$K_x = 7$	7	IKK- β	1.696330	0.043249
8	52	$K_x = 8$	8	SNAIL1	1.477019	0.034269
9	53	$K_x = 9$	10	MMP-14	1.368302	0.029121
10	2	$K_t = 2$	9	TGF- β 1	1.081828	0.029166
11	54	$K_x = 10$	12	Flotillin-1	0.787569	0.016863
12	3	$K_t = 3$	13	TGF- β 2	0.333981	0.012451
13	4	$K_t = 4$	20	TGF- β 3	0.159633	0.004157
14	1	$K_t = 1$	30	TGF- β 0	0.081261	0.002090

We argue that the obtained high sensitivity values shown in Figures 5, 6 and Table 2 can be tested in experiments similar to those reported in [5]. Also the global influence $D_s^{(u/d)}$ from Table 3 gives us a prediction of the globally stronger influence of the X-proteins than the TGF- β proteins. These results open new perspectives for external proteins influence on fibrosis.

3.5. Bi-functional of fibrosis network

Here we present in short certain results for the bi-functional MetaCore network. The doubled Ising MetaCore network has $N_I = 80158$ nodes and $N_{I,\ell} = 939808$ links. We compute the reduced Google matrix G_R for the doubled number of nodes $2 \times 54 = 108$ (by attributing (+) and (-) labels to each node) for the fibrosis proteins of Table 1. Here we present only some selected characteristics, all data for the Ising Google matrix are available at [22].

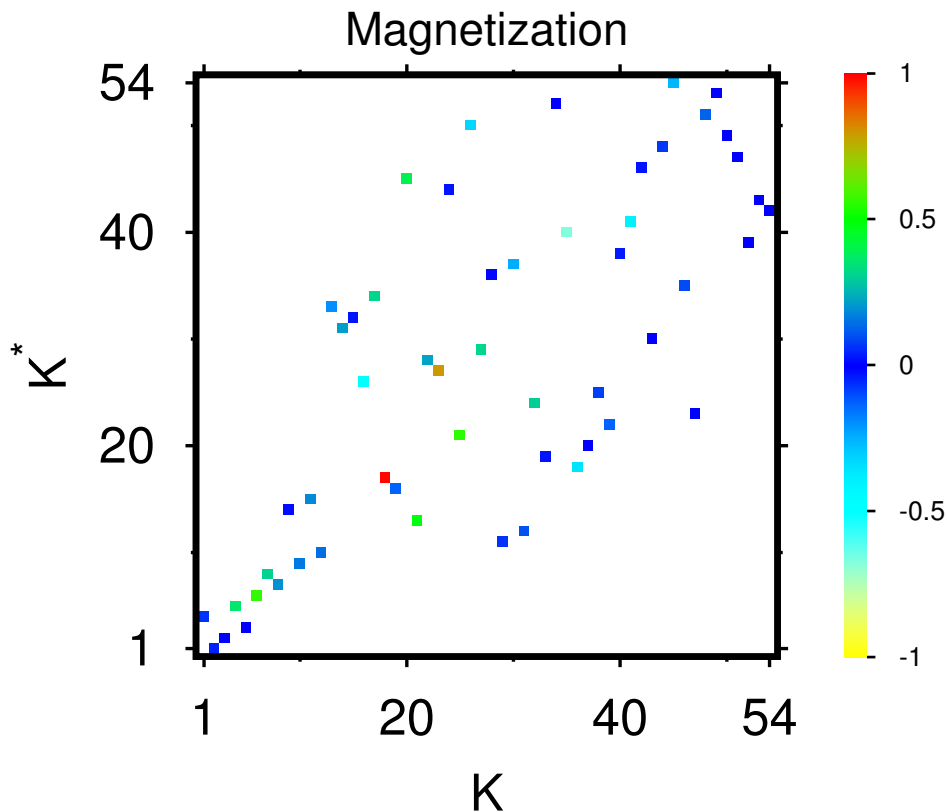


Figure 7. PageRank “magnetization” $M(j) = (P_+(j) - P_-(j)) / (P_+(j) + P_-(j))$ of proteins of Table 1 shown on the PageRank-CheRank plane (K, K^*) of local indices; here j represents a protein node in the initial single protein network and $P_{\pm}(j)$ are the PageRank components of the bi-functional Ising MetaCore network (see text). The values of the color bar correspond to $M / \max |M|$ with $\max |M| = 0.690937$ being the maximal value of $|M(j)|$ for the shown group of proteins. Note that the positions in PageRank-CheRank plane are identical to the positions of Appendix Figure A1 and the corresponding K, K^* values are given in the 3rd and 4th column of Table 1.

In Figure 7 we show the magnetization $M(j) = (P_+(j) - P_-(j)) / (P_+(j) + P_-(j))$ of proteins of Table 1 with their location on PageRank-CheRank plain (K, K^*) . We remind that $P_{\pm}(j)$ is the PageRank value of the node j with label (\pm) and that the sum satisfies $P(j) = P_+(j) + P_-(j)$ where $P(j)$ is the PageRank value of the node j of the simple network. The magnetization is positive for nodes which are more likely to be activated or in other words which have on average more incoming activation links (and/or coming from other nodes with larger PageRank values) than inhibition links while negative values correspond to nodes being more likely to be inhibited by other nodes.

According to Figure 7 the majority of proteins have values of M being close to zero (neutral action on average coming from other nodes) but also there are some nodes with with significant positive values such as RAPGEF4 (at $K = 18, K^* = 17, K_g = 20, K_u = 16, M = 0.690937$) corresponding to the only red box (maximum value of 1 in units of the color bar) and HMGCS2 (at $K = 23, K^* = 27, K_g = 35, K_d = 11, M = 0.550286$) with an orange-brown box (value of 0.8 in units of the color bar). There are about further 9 proteins with various degrees of green color (M values between 0.2 and 0.4 corresponding to 0.3 to 0.6 in units of the color bar). The two proteins with strongest negative values of M are CLEC3B (at $K = 35, K^* = 40, K_g = 25, K_d = 1, M = -0.463912$) with a light cyan box (value of -0.7 in units of the color bar) and ACAN (at $K = 16, K^* = 26, K_g = 8, K_u = 4, M = -0.342585$) with a cyan box (value of -0.5 in units of the color bar). There are about 5 further proteins with various degrees of cyan color (M values between -0.28 and -0.17 corresponding to -0.4 to -0.25 in units of the color bar). We note that CELC3B is also selected in both network diagrams of Figure 4 and Appendix Figure A2 as one of the two down-top-nodes either because it is the first protein in the list of

down-proteins or because it appears at the top position of Table 2 for the strongest sensitivity value D_{ab} (with a being CELC3B and b being the X-protein MMP-14). One may also note that Appendix Figure A1 shows the same (K, K^*) positions as Figure 7 and allows to identify which of the boxes belong to the subgroups of TGF- β proteins, up- or down-proteins, or X-proteins, The complete table of magnetization values used for Figure 7 including the values of K, K^*, K_g etc. is available in one of the data files provided in [22].

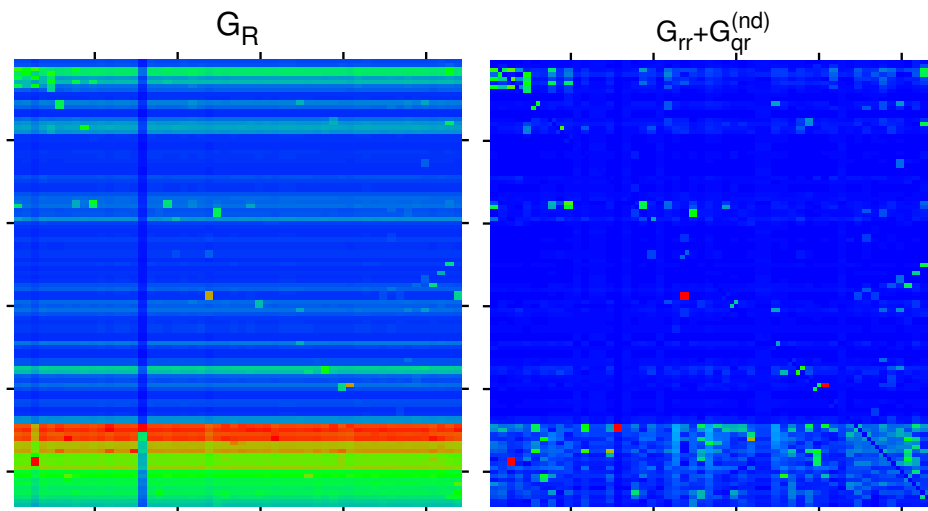


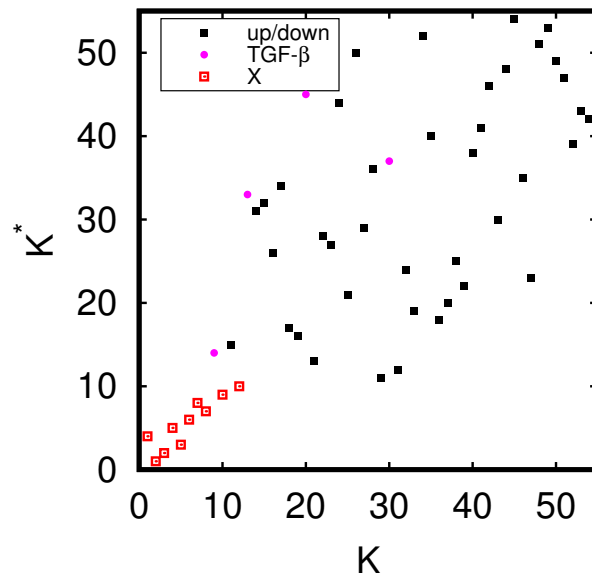
Figure 8. Color density plots of G_R and $G_{rr} + G_{qr}^{(nd)}$ for the bi-functional Ising MetaCore network and the extended group of 108 nodes by attribution of labels (+) and (-) to each node of Table 1. The matrix plot style is similar as in Figure 2 with outside tics indicating multiples of 20 of the index values. The color bar is as in Figure 2 with the same translation of colors to matrix values. The saturation value is for both panels the 6th largest value for each matrix and larger values are reduced to this value. The strongest cell values are reduced from 0.437575 (0.424939) to 0.101874 (0.060717) for G_R ($G_{rr} + G_{qr}^{(nd)}$).

In Figure 8 we show the matrices components G_R and $G_{rr} + G_{qr}^{(nd)}$ for the group of selected 108 nodes corresponding to the Ising MetaCore network. Their structure is quite similar to the corresponding components for the group of 54 nodes for the simple network shown in Figures 2 and 3, i.e. G_R is dominated by the uniform background due to the component G_{pr} with some exceptional peak values and large values if the first (vertical) matrix index correspond to an X-protein with large PageRank probability. For $G_{rr} + G_{qr}^{(nd)}$ the structure is more sparse showing the most significant direct and relevant indirect transitions. We note that for the Ising case, the matrix values are identical for the two labels of given node in horizontal position (except for the diagonal elements of $G_{rr} + G_{qr}^{(nd)}$ which have been artificially put to zero) which is a mathematical property of these matrices. However, in vertical direction there are significant differences between the two Ising labels especially for $G_{rr} + G_{qr}^{(nd)}$.

Further detailed analyses of the Ising MetaCore network with applications on fibrosis interactions are kept for future studies. However, an interested reader can find additional numerical results at [22]. In particular figures for the Ising network diagrams obtained from the Ising versions of G_R and $G_{rr} + G_{qr}^{(nd)}$ in the same way as in subsection 3.3 are available there.

4. Conclusion

In this work we presented the Google matrix analysis of protein-protein interactions of fibrosis. The group of 54 proteins actively participating in the fibrosis process is determined on the basis of INSERM experimental results presented in [5] which give 44 proteins and in addition we add 10 external proteins with strongest sensitivity (on certain 10 proteins of the 44-group) which is computed in the context of the REGOMAX approach applied to the MetaCore network [6]. Our results allow to identify the most important interactions between this fibrosis group of 54 proteins. We hope that these



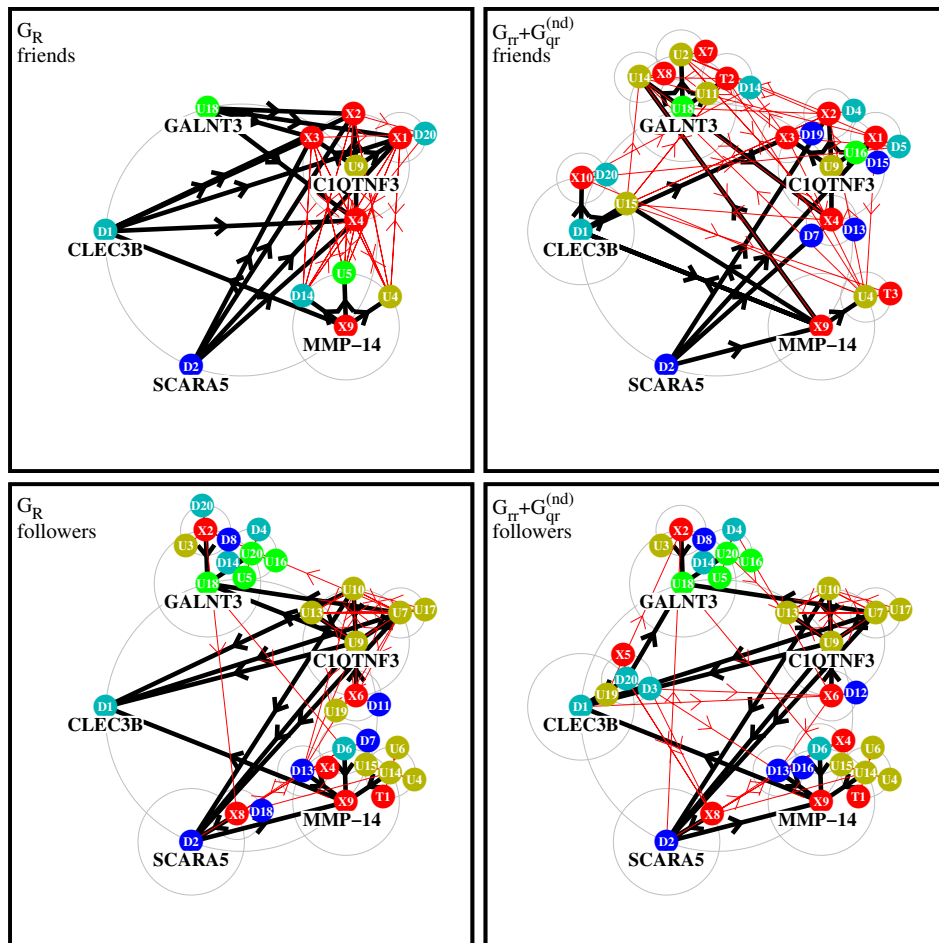
Appendix Figure A1. Positions of the 54 proteins of Table 1 in the local PageRank-CheiRank. Note that these positions are identical to the positions of Figure 7 and the corresponding K, K^* values are given in the 3rd and 4th column of Table 1. Pink full circles correspond to the subgroup of TGF- β nodes, full black boxes correspond to the subgroups of up- and down-proteins and red squares correspond to the subgroup of X-proteins.

algorithms allow to predict those proteins which will produce a significant influence on the fibrosis process. We expect that future experiment will allow to check these predictions obtained from the Google matrix analysis.

Appendix

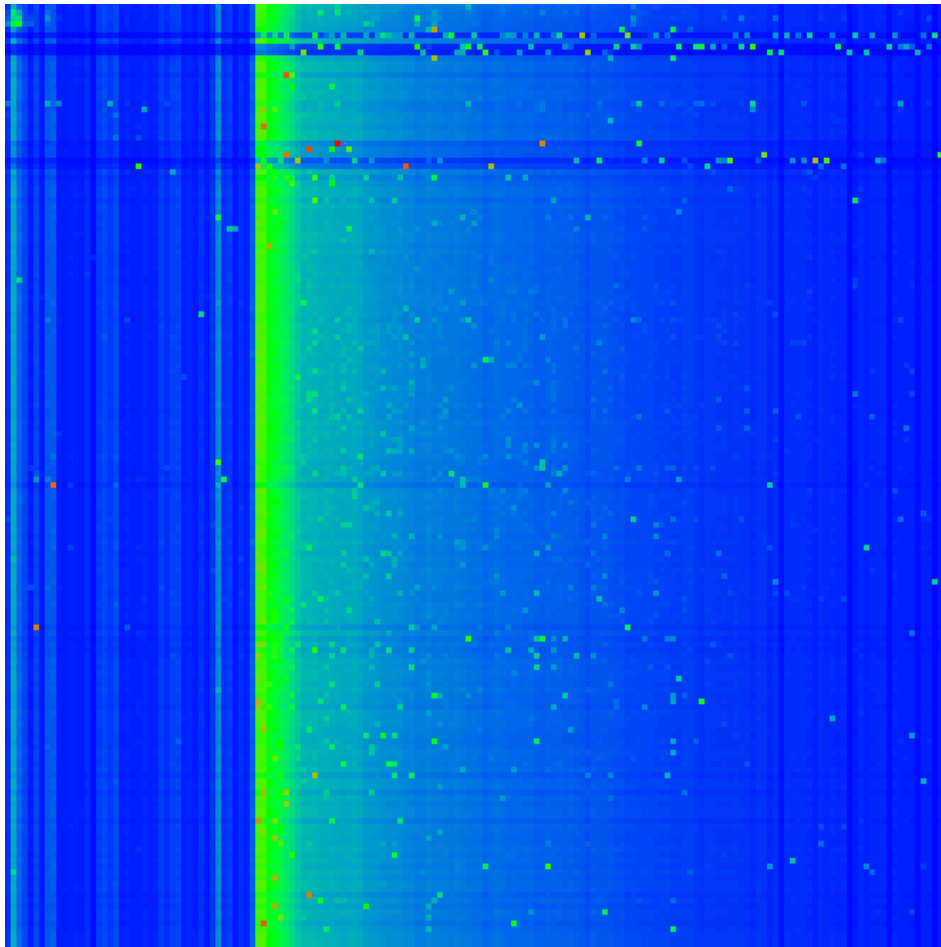
A1. Additional figures for REGOMAX results

Here we present additional Appendix Figures A1, . . . , A5 for the main part of this article. Appendix Figure A1 provides complementary information to Figure 1.



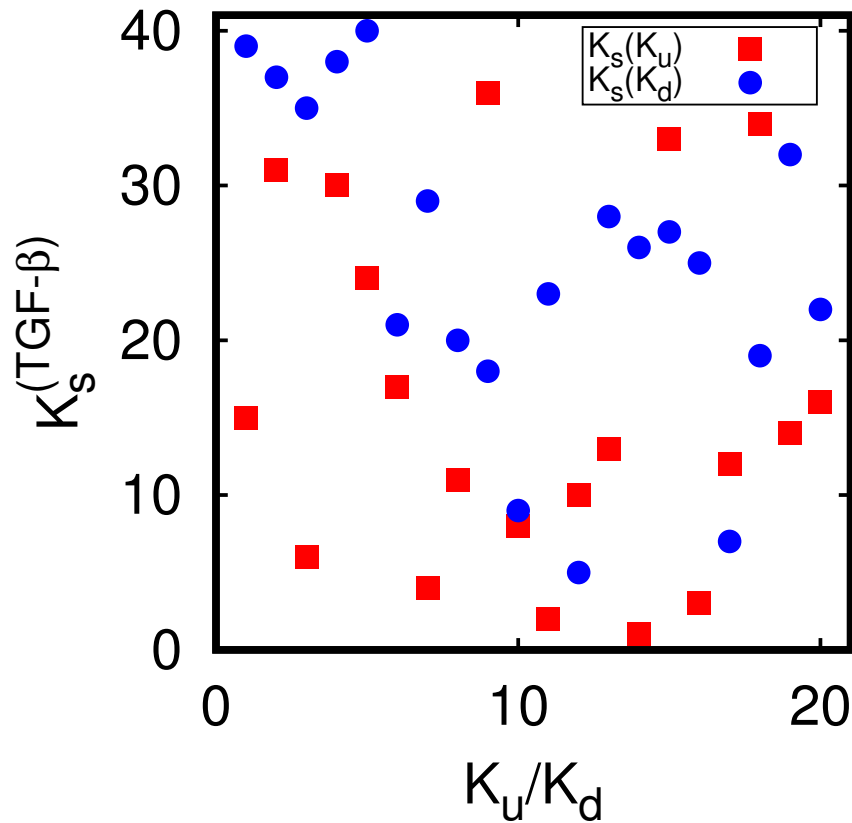
Appendix Figure A2. Effective network diagram for the same cases as in Figure 4 but using different 5 top nodes being the first X-node, the first two up-nodes and the first two down-nodes according to Table 2.

Appendix Figure A2 provides the network diagrams similar as in Figure 4 but with a different choice of 5 top nodes based on the criterion of top positions in Table 2.



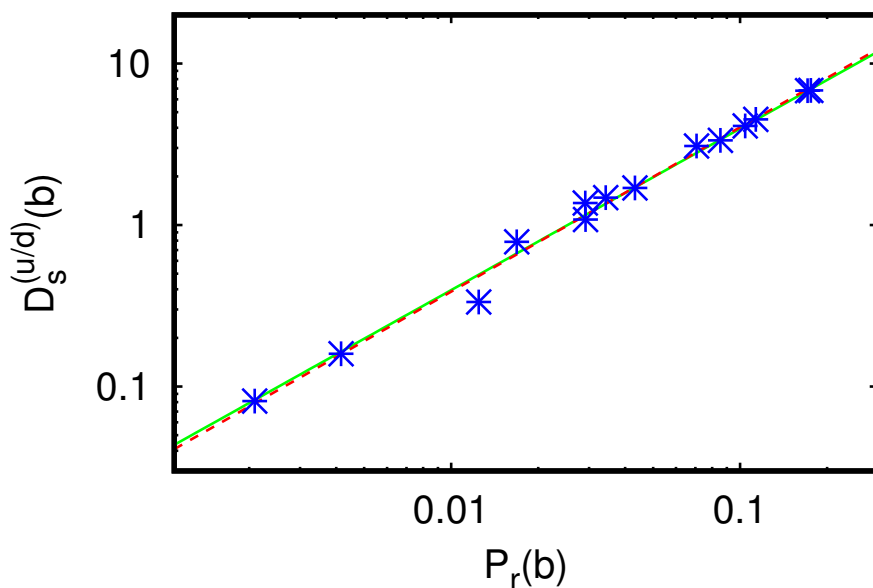
Appendix Figure A3. Color density plot of the sensitivity matrix $D_{ab}^{(166)}$ for the intermediary group of 166 proteins being the first 44 proteins of Table 1 (TGF- β , up- and down-subgroups) and 122 further proteins (in PageRank order) determined by having a direct link to one of the top 5 up-nodes ($K_u \leq 5$) or top 5 down-nodes ($K_d \leq 5$; see also text).

Appendix Figure A3 shows the sensitivity matrix $D_{ab}^{(166)}$ for the intermediary group of 166 proteins which was used to determine the additional 10 X-proteins as explained in subsection 2.6.



Appendix Figure A4. Effective ranking $K_s^{(TGF-\beta)}$ index of the TGF- β sensitivity versus K_u/K_d of up- (red boxes) and down-proteins (blue full circles). The ranking index $K_s^{(TGF-\beta)}$ is determined by ordering the sum $D_s^{(TGF-\beta)}(a) = \sum_{b=1}^4 D_{ab}$ in decreasing order for $a = 5, \dots, 44$ (i.e. a belongs to one of the sets of up- or down-proteins) and where D_{ab} is the sensitivity matrix for the 54 nodes of Table 1 (see also Figure 5).

Appendix Figure A4 provides an additional analysis of the overall influence of the TGF- β proteins on the up- and down-proteins which is discussed in subsection 3.4.



Appendix Figure A5. Dependence of the sum of sensitivities $D_s^{(u/d)}(b)$ from Table 3 on the (local) PageRank probability $P_r(b)$; the straight green line shows the fit dependence $D_s^{(u/d)}(b) = \eta P_r(b)$ with the obtained numerical value $\eta = 39.5 \pm 1.4$; the dashed red line corresponds to the power law fit $D_s^{(u/d)}(b) = \tilde{\eta} [P_r(b)]^\kappa$ with $\tilde{\eta} = 41.9 \pm 4.3$ and $\kappa = 1.017 \pm 0.028$.

Appendix Figure A5 provides the graphical and fit verification of the linear behavior between the two quantities $D_s^{(u/d)}(b)$ and $P_r(b)$ appearing the last two columns of Table 3.

A2. Simple estimate for the sensitivity matrix

In the second part of this appendix we remind some details (see [29]) about the numerical computation of the sensitivity (8) and provide an analytic approximation based on a simplified model. Let (a, b) be an arbitrary index pair and G_ε be the perturbed Google matrix obtained from a general unperturbed Google matrix G_0 by multiplying its element $G_0(a, b)$ at position (a, b) by $(1 + \varepsilon)$ and then sum-renormalizing the column b to unity. The elements in the other columns are not modified. In a more explicit formula we have:

$$\forall_{c,d} \quad G_\varepsilon(c, d) = \frac{(1 + \varepsilon \delta_{ca} \delta_{db}) G_0(c, d)}{1 + \varepsilon \delta_{db} G_0(a, b)} \quad (\text{A1})$$

where $\delta_{ca} = 1$ (or 0) if $c = a$ (or $c \neq a$). Note that the denominator is either 1 if $d \neq b$ or the modified column sum $1 + \varepsilon G_0(a, b)$ of column b if $d = b$. Expanding (A1) up to first order in ε we obtain $G_\varepsilon = G_0 + \varepsilon \Delta G + \dots$ with ΔG having the elements:

$$\forall_{c,d} \quad \Delta G(c, d) = \delta_{ca} \delta_{db} G_0(c, d) - \delta_{db} G_0(a, b) G_0(c, d). \quad (\text{A2})$$

Let P_ε be the sum-normalized PageRank vector of G_ε determined by the conditions $G_\varepsilon P_\varepsilon = P_\varepsilon$ and the normalization $E^T P_\varepsilon = 1$ where $E^T = (1, \dots, 1)$ is a (row) vector with unit entries. Note that the column sum condition of G_ε can be written as $E^T G_\varepsilon = E^T$ and of course for $\varepsilon = 0$ we also have $G_0 P_0 = P_0$, $E^T P_0 = 1$ and $E^T G_0 = E^T$. Furthermore we write the perturbed PageRank vector in the form $P_\varepsilon = P_0 + \varepsilon \Delta P + \dots$ where the ΔP must satisfy the condition $E^T \Delta P = 0$. Then the sensitivity (8) is directly related to ΔP by :

$$D_{(b \rightarrow a)}(j) = \frac{\Delta P(j)}{P_0(j)}. \quad (\text{A3})$$

Expanding the PageRank equation $G_\varepsilon P_\varepsilon = (G_0 + \varepsilon \Delta G + \dots)(P_0 + \varepsilon \Delta P + \dots) = P_\varepsilon = P_0 + \varepsilon \Delta P + \dots$ to order one we first obtain the unperturbed PageRank equation $G_0 P_0 = P_0$ and a further inhomogeneous equation :

$$\Delta P = G_0 \Delta P + \Delta G P_0 \quad (\text{A4})$$

which can be efficiently numerically solved by iteration (choosing initially $\Delta P = 0$ on the right hand side) once P_0 has been computed (see [29] for details on this point). This provides a numerical precise scheme to compute the sensitivity in the limit $\varepsilon \rightarrow 0$ without the need to take finite ε -differences.

Now, we consider a particular very simple model where G_0 has identical columns being the PageRank P_0 , i.e. $G_0 = P_0 E^T$ or more explicitly $G_0(c, d) = P_0(c)$ for all values of c, d . Then we obtain from (A2)

$$\forall_{c,d} \quad \Delta G(c, d) = \delta_{ca} \delta_{db} P_0(c) - \delta_{db} P_0(a) P_0(c) \quad (\text{A5})$$

and from (A4)

$$\Delta P = (P_0 E^T) \Delta P + \Delta G P_0 = \Delta G P_0 \quad (\text{A6})$$

since $E^T \Delta P = 0$. Inserting (A5) in (A6) we obtain (replacing $c = j$ and performing the d -sum for the matrix vector product)

$$\forall_j \quad \Delta P(j) = [\delta_{ja} - P_0(a)] P_0(j) P_0(b) \quad (\text{A7})$$

and from (A3)

$$D_{(b \rightarrow a)}(j) = [\delta_{ja} - P_0(a)] P_0(b) . \quad (\text{A8})$$

Choosing $j = a$ this gives the sensitivity matrix

$$D_{ab} = D_{(b \rightarrow a)}(a) = [1 - P_0(a)] P_0(b) \approx P_0(b) \quad (\text{A9})$$

where the last approximation holds if typically $P_0(a) \ll 1$.

This result is of course only valid for the simplified model of identical columns (being the PageRank vector) in G_0 . However, when G_0 represents a typical reduced Google matrix, with $N_r \ll N$, the component G_{pr} which has the strongest numerical weight (typically $\sim 95\%$) is of the form $G_{pr} = \tilde{P}_0 \tilde{E}^T$ where $\tilde{P}_0 \approx P_0$ and $\tilde{E}^T \approx E^T$ except for a few number of components j where strong deviations between $\tilde{P}_0(j)$ and $P_0(j)$ (and similarly between $\tilde{E}(j)$ and $E(j)$) are possible.

Our examples of D_{ab} visible in Figure 5 and of $D_{ab}^{(166)}$ of Appendix Figure A3 confirm the typical behavior $D_{ab} \sim P_0(b)$ for a “uniform background” but there are some exceptional peak values which arise from the deviations from G_{pr} to the simplified model and also from the contributions of G_{rr} and G_{qr} . This also explains our numerical finding that all matrix elements of D_{ab} are positive. Actually, according to (A8) we expect that $D_{(b \rightarrow a)}(j)$ is typically positive if $j = a$ and negative if $j \neq a$.

Furthermore, when taking the partial a -sum over up- and down-nodes of D_{ab} the effect of exceptional peaks is strongly reduced thus explaining the linear behavior $D_s^{(u/d)}(b) \approx \eta P_r(b)$ visible in Table 3 and Figure A5.

Author Contributions: All authors equally contributed to all stages of this work.

Funding: This research was supported in part through the grant NANOX N° ANR-17-EURE-0009, (project MTDINA) in the frame of the Programme des Investissements d’Avenir, France; it was granted access to the HPC resources of CALMIP (Toulouse) under the allocation 2021-P0110.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Murtha L.A., Schuliga M.J., Mabotuwana N.S, Hardy S.A., Waters D.W., Burgess J.K., Knight D.A. and Boyle A.J., *The processes and mechanisms of cardiac and pulmonary fibrosis*, Front Physiol. **128** (2017) 777.
2. Liu T., Song D., Dong J., Zhu P., Liu J., Liu W., Ma X., Zhao L. and Ling S., *Current understanding of the pathophysiology of myocardial fibrosis and its quantitative assessment in heart failure*, Front. Physiol. **8** (2017) 238.

3. Meng Xm., Nikolic-Paterson D. and Lan H. *TGF- β : the master regulator of fibrosis*, Nat. Rev. Nephrol. **12** (2016) 325.
4. Wynn T.A., *Cellular and molecular mechanisms of fibrosis*, Jour. Pathology **214**(2) (2007) 199.
5. Pintus S.S., Sharipov R.N., Kel A., Timotin A., Keita S., Martinelli I., Boal F., Tronchere H., Kolpakov F. and Kunduzova O., *Drug repositioning for cardiac fibrosis through molecular signature of aberrant fibroblast activation*, INSERM preprint, to be published (2021).
6. MetaCore, Available: <https://clarivate.com/cortellis/solutions/early-research-intelligence-solutions/>. Accessed October 2021.
7. Ekins S., Bugrim A., Brovold L., Kirillov E., Nikolsky Y., Rakhmatulin E., Sorokina S., Ryabov A., Serebryiskaya T., Melnikov A., Metz J. and Nikolskaya T., *Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms*, Xenobiotica **36**(10-11) (2006) 877; doi:10.1080/00498250600861660.
8. Bessarabova M., Ishkin A., JeBailey L., Nikolskaya T., and Nikolsky Y., Knowledge-based analysis of proteomics data, BMC bioinformatics **13 Suppl 16** (2012) 13; doi:10.1186/1471-2105-13-S16-S13.
9. Kotelnokova E., Frahm K.M., Lages J. and Shepelyansky D.L., *Statistical properties of the MetaCore network of protein-protein interactions*, <https://doi.org/10.1101/2021.04.02.438245> (2021).
10. Brin S. and Page L., *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems **30** (1998) 107.
11. Langville A.M. and Meyer C.D., *Google's PageRank and beyond: the science of search engine rankings*, Princeton University Press, Princeton (2006).
12. Ermann L., Frahm K.M. and Shepelyansky D.L., *Google matrix analysis of directed networks*, Rev. Mod. Phys. **87** (2015) 1261.
13. Markov A.A., *Rasprostranenie zakona bol'shikh chisel na velichiny, zavisyaschie drug ot druga*, Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya (in Russian) **15** (1906) 135; English translation: *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*, reprinted in Appendix B of: Howard R., *Dynamic Probabilistic Systems 1: Markov chains*, John Wiley and Sons, New Jersey USA (1971).
14. Frahm K.M. and Shepelyansky D.L., *Reduced Google matrix*, <https://arxiv.org/abs/1602.02394> (2016).
15. Frahm K.M., Jaffres-Runser K. and Shepelyansky D.L., *Wikipedia mining of hidden links between political leaders*, Eur. Phys. J. B **89** (2015) 269.
16. Lages J., Shepelyansky D.L. and Zinovyev A., *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE **13**(1) (2018) e0190812.
17. Frahm K.M. and Shepelyansky D.L., *Google matrix analysis of bi-functional SIGNOR network of protein-protein interactions*, Physica A **559** (2020) 125019.
18. Rollin G., Lages J. and Shepelyansky D.L., *World influence of infectious diseases from Wikipedia network analysis*, IEEE Access **7** (2019) 26073.
19. Rollin G., Lages J. and Shepelyansky D.L., *Wikipedia network analysis of cancer interactions and world influence*, PLoS ONE **14**(9) (2019) e0222508.
20. Coquide C., Ermann L., Lages J. and Shepelyansky D.L., *Influence of petroleum and gas trade on EU economies from the reduced Google matrix analysis of UN COMTRADE data*, Eur. Phys. J. B **92** (2019) 71.
21. Coquide C., Lages J. and Shepelyansky D.L., *Interdependence of sectors of economic activities for world countries from the reduced Google matrix analysis of WTO data*, Entropy **22** (2020) 1407.
22. <https://www.quantware.ups-tlse.fr/QWLIB/fibrosisPPInetwork/>, Accessed October (2021).
23. Chepelianskii A.D., *Towards physical laws for software architecture*, <https://arxiv.org/abs/1003.5455> (2010).
24. Zhironov A.O., Zhironov O.V. and Shepelyansky D.L., *Two-dimensional ranking of Wikipedia articles*, Eur. Phys. J. B **77** (2010) 523.
25. *The Schur complement and its applications*, Ed. Fushen Zhang, Springer, Berlin (2005).
26. Beenakker C.W.J., *Random-matrix theory of quantum transport*, Rev. Mod. Phys. **69** (1997) 731.
27. Gaspard P., *Quantum chaotic scattering*, Scholarpedia **9**(6) (2014) 9806.
28. Meyer C.D., *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Review **31**(2) (1989) 240.
29. Frahm K.M. and Shepelyansky D.L., *Linear response theory for Google matrix*, <https://arxiv.org/abs/1908.08924> (2019).