RESEARCH

Statistical properties of the MetaCore network of protein-protein interactions

Ekaterina Kotelnikova[1], Klaus M. Frahm[2], José Lages[3*] and Dima L Shepelyansky[2]

*Correspondence:
jose.lages@utinam.cnrs.fr
[3]Institut UTINAM, CNRS, Université Bourgogne Franche-Comté, Besançon, France
Full list of author information is available at the end of the article

Abstract

The MetaCore commercial database describes interactions of proteins and other chemical molecules and clusters in the form of directed network between these elements, viewed as nodes. The number of nodes goes beyond 40 thousands with almost 300 thousands links between them. The links have essentially bi-functional nature describing either activation or inhibition actions between proteins. We present here the analysis of statistical properties of this complex network applying the methods of the Google matrix, PageRank and CheiRank algorithms broadly used in the frame of the World Wide Web, Wikipedia, the world trade and other directed networks. We specifically describe the Ising PageRank approach which allows to treat the bi-functional type of protein-protein interactions. We also show that the developed reduced Google matrix algorithm allows to obtain an effective network of interactions inside a specific group of selected proteins. This method takes into account not only direct protein-protein interactions but also recover their indirect nontrivial couplings appearing due to summation over all the pathways passing via the global bi-functional network. The developed analysis allows to espablish an average action of each protein being more oriented to activation or inhibition. We argue that the described Google matrix analysis represents an efficient tool for investigation of influence of specific groups of proteins related to specific diseases.

Keywords: Complex networks; MetaCore; Google matrix; PageRank; protein-protein interactions network

1 Introduction

The MetaCore database [1] provides a large size network of Protein-Protein Interactions (PPI). It has been shown to be useful for analysis of specific biological problems (see e.g. [2, 3]) and finds various medical applications. At present, the network has $N = 40079$ nodes with $N_\ell = 292904$ links and an average of $n_l = N_l/N \approx 7.3$ links per node. The nodes are composed mainly by proteins but in addition there are also certain molecules and molecular clusters catalyzing the interactions with proteins. This PPI network is directed and non-weighted. Its interesting feature is the bi-functional nature of the links leading to either the activation or the inhibition of one protein by another one. In some cases, the link action is neutral or unknown.

In the present work, we describe the statistical properties and the Google matrix analysis (GMA) of the MetaCore network. The GMA and the related PageRank algorithm has been at the foundation of the Google search engine with important applications to the World Wide Web (WWW) analysis [4, 5]. A variety of GMA applications to directed networks are presented in [6]. The first application of the

GMA to PPI network was reported for the SIGNOR PPI network [7] in [8]. However, the size of the SIGNOR network is by a factor ten smaller than the MetaCore one and thus the GMA of SIGNOR network can be considered only as a test bed for more detailed studies of PPI.

An important feature of the PPI networks is the bi-functional character of the directed links representing activation or inhibition actions. Usually, the directed networks have been considered without functionality of links (see e.g. [4, 5, 6]). The Ising-Google matrix analysis (IGMA) [9] extends the GMA for bi-functional links. A test application to the SIGNOR PPI network [7] can be found in [10]. The Ising-Google matrix analysis (IGMA) represents each node by two states $\uparrow$ and $\downarrow$, like Ising spins up and down. A link is then represented by a 2×2-matrix describing the actions of activation or inhibition [9, 10]. By contrast with the case of links without functionality, this description leads to a doubling of the number of nodes $N_I = 2N$. In the present work, we apply the IGMA to the MetaCore network which provide bi-functional interactions between multiple proteins.

In addition, we also use the reduced Google matrix analysis (RGMA), developed in [11, 12], to describe the effective interactions between a subset of $N_{\rm r} \ll N$ selected nodes taking account of all the indirect pathways connecting each couple of these $N_{\rm r}$ nodes throughout the global PPI network. The efficiency of the RGMA has been demonstrated for large variety of directed networks including Wikipedia and the world trade network (see e.g. [13, 14]). The RGMA adapted to the IGMA for bi-functional links is called hereafter the RIGMA.

The paper is composed as follows: the data sets and the methods are described in the Section 2, the results are presented in the Section 3 and the discussion and the conclusion are given in the Section 4.

2 Data sets and methods

2.1 Google matrix construction of the MetaCore network

At the first step, we start the construction of the Google matrix G of the MetaCore network neglecting the bi-functional character of the links and considering unweighted links. Considering the adjacency matrix A, the elements A_{ij} of which are equal to 1 if node j points to node i and equal to 0 otherwise, the stochastic matrix S of the node-to-node Markov transitions is obtained by normalizing to unity each column of the adjacency matrix A. For dangling nodes, the corresponding column is filled with elements with value $1/N$. The stochastic matrix S describes a Markov chain process on the network: a random surfer hops from node to node in accordance with the network structure and hops anywhere on the network if it reaches a dangling node. The elements of the Google matrix G takes then the standard form

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

where $0.5 \leq \alpha < 1$ is the damping factor. The random surfer obeying to the stochastic process encoded in G explores, with a probability α, the network in accordance to the stochastic matrix S and hops, with a complementary probability $(1 - \alpha)$, to any node of the network. The damping factor allows the random surfer to escape from possible isolated communities. Here, we use the standard value

$\alpha = 0.85$ [5, 6]. The PageRank vector P is the right eigenvector of the Google matrix G corresponding to the leading eigenvalue, here $\lambda = 1$. The corresponding eigenproblem equation is then $GP = P$. According to the Perron-Frobenius theorem, the PageRank vector P has positive elements. The PageRank vector element $P(j)$ gives the probability to find the random surfer on the node j once the Markov process has reached the stationary regime. Consequently, all the nodes can be ranked by decreasing PageRank probability. We define the PageRank index $K(j)$ giving the rank of the node j. The node j with the highest (lowest) PageRank probability $P(j)$ corresponds to $K(j) = 1$ ($K(j) = N$). On average, the PageRank probability $P(j)$ is proportional to the number of ingoing links pointing to node j.

It is also useful to consider a network obtained by the inversion of all the directions of the links. For this inverted network, the corresponding Google matrix is denoted G^* and the corresponding PageRank vector is called the CheiRank vector P^* and is defined such as $G^*P^* = P^*$. The importance and the detailed statistical analysis of the CheiRank vector have been reported in [15, 16] (see also [6, 14]). 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. We define also a CheiRank index $K^*(j)$ giving the rank of the node j according to its CheiRank probability $P^*(j)$.

2.2 Reduced Google matrix

The concept of the reduced Google matrix analysis (RGMA) was introduced in [11] and applied with details to Wikipedia networks in [12]. The RGMA determines effective interactions between a selected subset of N_r nodes embedded in a global network of size $N \gg N_r$. These effective interactions are determined taking into account that there are many indirect links between the N_r nodes via all the other $N_s = N - N_r$ nodes of the network. As an example, we may have two nodes A and C which belongs to the selected subset of N_r nodes and which are not coupled by any direct link. However, it may exist a chain of links from A to B_1, then from B_1 to B_2,..., and then from B_m to C where $B_1, \ldots, B_m$ are nodes not belonging to the subset of N_r nodes. Although A and C are not directly connected, the is a chain of $m + 1$ directed links indirectly connecting A and C. The RGMA allows to infer an effective weighted link between any couple of two nodes of the N_r subset of interest taking account of the possible direct link existing between these two nodes and taking account of all the possible chains of links connecting them throughout the remaining global network of size $N_s = N - N_r \gg N_r$. It is important to stress that rather often the network analysis is done taking only into account the direct links between the N_r nodes and, as a consequence, completely omitting their indirect interactions via the global network. It is known that such a simplified approach produces erroneous results as it happened for the network of historical figures extracted from Wikipedia when only direct links between historical figures were taking into account and all other links had been omitted [17] (see discussion at [18]).

It is convenient to write the Google matrix G associated to the global network as

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

where the label "r" refers to the nodes of the reduced network, ie 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 associated to the subset of the N_r nodes is a $N_r \times N_r$ matrix defined as

$$G_R P_r = P_r \tag{3}$$

where P_r is a N_r size vector the components of which are the normalized PageRank probabilities of the N_r nodes of interest, $P_r(j) = P(j)/\sum_{i=1}^{N_r} P(i)$. The RGMA consists in finding an effective Google matrix for the subset of N_r nodes keeping the relative ranking between these nodes. To ensure the relation (3), the reduced Google matrix G_R has the form [11, 12]

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

As shown in [11, 12], the reduced Google matrix G_R can be represented as the sum of three components

$$G_R = G_{rr} + G_{pr} + G_{qr}. \tag{5}$$

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 with all the columns being approximately equal to the reduced PageRank vector P_r; the third component, G_{qr}, describes all the indirect pathways passing through the global network. Thus, the component G_{qr} represents the most nontrivial information related to indirect hidden transitions. We also define G_{qrnd} matrix which is the G_{qr} matrix deprived of its diagonal elements. The contribution of each component is characterized by their weights $W_R, W_{pr}, W_{rr}, W_{qr}$ (W_{qrnd}) respectively for $G_R, G_{pr}, G_{rr}, G_{qr}$ (G_{qrnd}). The weight of a matrix is given by the sum of all the matrix elements divided by its size, here N_r (by definition $W_R = 1$). Examples of reduced Google matrices associated to various directed networks are given in [8, 12, 14, 10].

2.3 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 [9, 10] 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. To construct the doubled "Ising" network of proteins, each elements of the initial adjacency matrix is replaced by one of the following 2×2 matrices

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

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 (see Section 2.1). The initial simple MetaCore network has $N = 40079$ nodes and $N_\ell = 292904$ links; the ratio of the number of activation/inhibition links is $N_{\ell+}/N_{\ell-} = 65379/49384 \simeq 1.3$ and the number of neutral links is $N_{\ell n} = N_\ell - N_{\ell+} - N_{\ell-} = 178141$. The doubled Ising MetaCore network corresponds to $N_I = 80158$ nodes and $N_{I,\ell} = 942090$ links (according to 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. We have numerically verified that the identity $P(j) = P_+(j) + P_-(j)$ holds up to the numerical precision $\sim 10^{-13}$.

As in [9], we characterize each node by its PageRank "magnetization" given by

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

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.

2.4 Sensitivity

The reduced Google matrix $G_{\rm R}$ of bi-functional (or Ising) MetaCore network describes effective interactions between $N_{\rm r}$ nodes taking into account the activation or inhibition nature of the interactions.

Following [13], it is useful to determine the sensitivity of the PageRank probabilities in respect to small variation of the matrix elements of $G_{\rm R}$. The PageRank sensitivity of the node j with respect to a small variation of the $b \to a$ link is defined as

$$D_{(b\to a)}(j) = \frac{1}{P_{\rm r}(j)} \left. \frac{dP_{{\rm r}\varepsilon}(j)}{d\varepsilon} \right|_{\varepsilon=0} = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon P_{\rm r}(j)} [P_{{\rm r}\varepsilon}(j) - P_{\rm r}(j)] \tag{8}$$

where $P_{{\rm r}\varepsilon}(j)$ is the PageRank vector computed from a perturbed matrix $G_{{\rm R}\varepsilon}$ the elements of which are defined by $G_{{\rm R}\varepsilon}(a,b) = G_{\rm R}(a,b)(1+\varepsilon)/[1+\varepsilon G_{\rm R}(a,b)]$ for the element (a,b), $G_{{\rm R}\varepsilon}(c,b) = G_{\rm R}(c,b)/[1+\varepsilon G_{\rm R}(a,b)]$ for the other elements (c,b) in the same column b, and $G_{{\rm R}\varepsilon}(c,d) = G_{\rm R}(c,d)$ for the elements (c,d) in the other columns. The factor $1/[1+\varepsilon G_{\rm R}(a,b)]$ ensures the correct sum normalization of the modified column b.

We use here an efficient algorithm described in [19] to evaluate the derivative in (8) exactly without usage of finite differences.

As proposed in [13], we define the symmetric matrix (see Eq.15 of [13])

$$D_{(a\leftrightarrow b)}(j) = D_{(b\to a)}(j) + D_{(a\to b)}(j)\ . \tag{9}$$

and furthermore the two symmetric and anti-symmetric sensitivity matrices

$$F_+(a,b) = D_{(a\leftrightarrow b)}(a) + D_{(a\leftrightarrow b)}(b) \quad , \quad F_-(a,b) = D_{(a\leftrightarrow b)}(a) - D_{(a\leftrightarrow b)}(b) \; . \tag{10}$$

These two sensitivity matrices characterize a variation of PageRank with a small variation of coupling matrix element between b and a nodes.

3 Results

Below, we describe various statistical properties of the MetaCore network obtained by the methods described above. More detailed data are available at [20].

3.1 CheiRank and PageRank of the MetaCore network

Let us sort the PageRank probabilities from the highest value to which we associate the $K = 1$ rank to the smallest value to which we associate to the $K = N$ rank.

The dependence $P(K)$ of the PageRank probabilities on the PageRank index K and the dependence $P^*(K^*)$ of the CheiRank probabilities on the CheiRank index K^* are shown in Fig. 1 for the simple MetaCore network and the Ising (doubled) MetaCore network. The decay of the probabilities is approximately proportional to an inverse index in a power $\beta \approx 2/3$, ie $P(K) \propto 1/K^{2/3}$. This exponent β is approximately the same for the PageRank and the CheiRank probabilities, and for both network types. The situation is different from the networks of WWW, Wikipedia, and Linux for which one usually have $\beta \approx 0.9$ for the PageRank probabilities and $\beta \approx 0.6$ for CheiRank probabilities [5, 6, 15]. We assume that in PPI networks both ingoing and outgoing links are of equal importance while, in the other above cited networks, ingoing links are more robust and stable than outgoing links which have a more random character.

The top 40 PageRank and CheiRank nodes of the MetaCore network are given in Tables 1 and 2 respectively. The top 3 PageRank positions are occupied by specific molecules actively participating in various reactions with proteins. The top 3 CheiRank positions are occupied by the transcription factor c-Myc, the generic enzyme eIF2C2 (Argonaute-2), and the generic binding protein IGF2BP3. In a certain sense, we can say that top PageRank nodes are like workers in a company, who receive many orders, while top CheiRank nodes are like company administrators who submit many orders to their workers (such a situation was discussed for a company management network [21]).

The density distribution of nodes of the MetaCore network on the PageRank-CheiRank (K, K^*)-plane is shown in Fig. 2. Comparing to the case of Wikipedia networks [6, 16] the distribution is globally more symmetric in respect to the diagonal $K = K^*$. This reflects the fact that the decay of the PageRank and the CheiRank probabilities in Fig. 1 is approximately the same. However, the top nodes are rather different for the PageRank and CheiRank rankings that is also visible from Tables 1 and 2. As an example, the top 40 PageRank and the top 40 CheiRank share only 7 nodes in common (Beta-catenin, p53, ESR1, STAT3, Androgen receptor, c-Myc, RelA) which are transcription factors with the exception of Beta-catenin which is a generic binding protein. As a consequence, depending on the considered biological process, these biological elements trigger the multiple cascades of

interactions or are at the very end of these cascades. In contrast, there are biological elements with low K and high K^* and *vice versa*. For example, the phosphate compound PO_4^{3-}, with PageRank-CheiRank indexes ($K = 16, K^* = 14888$), is mainly a residue of biological processes and the passage of the Potassium ion K^+ from the cytosol to the extracellular region, with PageRank-CheiRank indexes ($K = 19$, $K^* = 26346$), can be considered as the final step of some biological process.

Among the top 40 PageRank nodes, we select a subset of 12 nodes which are more directly related to proteins. These 12 nodes are represented by white stars in the Fig. 2. The list of these nodes is given in Table 1. Below, we present the RIGMA analysis of these 12 nodes taking into account of the bi-functionality of the links (activation - inhibition).

3.2 Magnetization of nodes of the Ising MetaCore network

From the PageRank probabilities of the Ising MetaCore network, we determine the magnetization $M(K)$ of each node given by (7). The dependence of the magnetization $M(K)$ on the PageRank index K is shown in Fig. 3. For $K \leq 10$, only few nodes have a significant positive magnetization. In the range $10 < K \leq 10^3$, some nodes have almost the maximal positive or negative values of the magnetization with M being close to 1 or -1. Such nodes perform mainly activation or inhibition actions, respectively. For the range $K > 10^3$, we see an envelope restricting the maximal or the minimal values M. At present, we have no analytical description of this envelope. We suppose that nodes with high K values have a majority of outgoing links which are more fluctuating in this range thus giving a decrease of the maximal/minimal values of M.

Focusing on the top 40 PageRank in Fig. 3, we mainly observe that the nodes are either *non-magnetized* $M \approx 0$, or positively magnetized $M \gtrsim 1$. These two situations correspond to biological elements which are equally activated/inhibited ($M \approx 0$) and mainly activated ($M \gtrsim 0$), respectively. Among the top 40 PageRank nodes, the non-magnetized elements are mainly inorganic ions, such as H^+, Na^+, K^+, Ca^{2+}, and Cl^-, which are involved in many elementary interactions. As non-magnetized nodes, we observe also very important biological molecules such as DNA and the ADP compound which should occupy a central place in the protein interaction network. Among positively magnetized nodes, we observe reactions ($M \gtrsim 0.75$), RNA ($M \simeq 0.7$), protein kinase ($M \simeq 0.25 - 0.4$) and phosphatase ($M \simeq 0.55$), which respectively are known to *turn on* and *turn off* proteins. Let us remark that, as DNA, RNA occupies a very central role in the protein interaction network ($K = 6$) but has a relatively high magnetization $M \simeq 0.7$ which indicates that RNA is mainly activated at the end of major biological processes. The other positively magnetized nodes correspond to some transcription factors, such as PPAR-gamma and STAT3, generic binding proteins, such as PI3K and GRB2, members of RAS superfamily, such as Rac1, and generic proteins, such as ITGB1. We nevertheless note that among the top 40 PageRank nodes, there are some mainly inhibited proteins ($M \simeq -0.2$) such as the generic binding protein Bcl-2, the generic enzyme MDM2, and the lipid phosphatase PTEN.

We return to the magnetization properties of the selected subset of 12 nodes and the top 40 PageRank nodes in the next Section.

3.3 RIGMA analysis of the Ising MetaCore network

We illustrate the RIGMA analysis of the Ising MetaCore network by applying it to the subset of the 12 nodes given in Table 1. They are selected from the top 40 PageRank list of Table 1 by excluding simple molecules and keeping best ranked proteins according to PageRank probabilities. Each of the 12 nodes of the subset are doubled into a (+) component and a (−) component. We order these 24 nodes by ascending PageRank index K and alternating the (+) and the (−) components. This ordering is used to represent, in Fig. 4, the reduced Ising Google matrix G_R and its three matrix components G_{rr}, G_{pr}, and G_{qr}. The weights of these components are respectively $W_{rr} = 0.015$, $W_{pr} = 0.952$, and $W_{qr} = 0.033$. As in the case of Wikipedia networks [12], the component G_{pr} has the highest weight, but as discussed, it is rather close to a matrix with identical columns, each one similar to the PageRank column vector. Thus, the G_{pr} matrix component does not provides more information than the standard PageRank/GMA analysis. We also see that the weight W_{qr} of the indirect links generated by long indirect pathways passing through the global Ising MetaCore network has approximately twice higher weight than the weight W_{rr} of direct links. Consequently, the contribution of indirect links are very important.

In the G_{rr} matrix component, each element i of the jth column corresponds to the direct action of the protein j on the protein i. The action is either an activation (+) or an inhibition (-). As a consequence, the G_{rr} matrix component simply mimics the Ising MetaCore network matrix adjacency (the elements of G_{rr} with a value equal to (greater than) $(1-\alpha)/2N \approx 0$ correspond to values 0 (1 or 1/2) in the adjacency matrix of the Ising MetaCore network). It is interesting to compare the G_{qr} matrix elements with those of the G_{rr} matrix. Each one of the G_{qr} matrix elements either modifies, generally enhances, the weight of an existing link, for which a non zero matrix element exists in the G_{rr} matrix, or, interestingly, quantifies the strength of a hidden effective interaction between two proteins. As an example of the enhancement of an existing direct link, we observe, in Fig. 4, that the known activation of FAK1 by ITGB1 is enhanced by indirect links, ie, by pathways passing by the elements outside the set of the twelve chosen proteins. Also, we clearly observe also an enhancement of the self-activation of FAK1 and the appearance of its indirect self-inhibition.

Let us focus on the possible hidden interactions between the chosen set of twelve proteins. For that purpose, we show, in Fig. 5 (left panel), the matrix sum $G_{rr} + G_{qr}^{(nd-block)}$ which summarizes both the information concerning the direct and hidden interactions between the set of twelve proteins. Here, we use the $G_{qr}^{(nd-block)}$ matrix which is the G_{qr} matrix from which the diagonal elements (self-interaction terms) have been removed. In Fig. 5 (right panel), the $G_{rr} + G_{qr}^{(nd-block)}$ matrix elements associated to direct links have been masked to highlight only hidden interactions. Hence, although the ARX protein (aristaless related homeobox) is not directly connected to the other eleven proteins, ie, there is no direct action of the ARX protein onto the other eleven proteins and *vice versa*, it indirectly strongly inhibits the tumor suppressor protein p53. Secondarily, the ARX protein indirectly acts on different other proteins as it is indicated by blue shades on the ARX column in Fig. 5: hence, the ARX protein indirectly activates the EGFR and ESR1 proteins

(epidermal growth factor receptor and estrogen receptor, respectively) and inhibits the c-Myc protein (proto-oncogene protein). Similarly, according to the G_{rr} matrix component (see Fig. 4), the c-Myc protein does not act on the chosen twelve proteins. But, the blues shades of the c-Myc column on the right panel of Fig. 5 gives us information on which proteins it indirectly contributes to activate or deactivate. Among strong weights of the $G_{rr} + G_{qr}^{(nd-block)}$ matrix sum, we observe also the SHP-2 phosphatase protein indirectly strongly interacts with with the ARX protein and the estrogen receptor protein ESR2. In return, the ESR2 protein, which directly inhibits ESR1 and c-Myc proteins, also indirectly activates the SHP-2 protein.

In contrast to the adjacency matrix and the Google matrix, the matrix sum $G_{rr}+G_{qr}$ allows to discriminate the directed links outgoing from a given protein by assigning different weights to them. This discrimination is possible as the RGMA and the RIGMA takes account of not only the direct linkage of the twelve chosen proteins but all the knowledge encoded in the MetaCore complex network. Moreover, possible hidden links between proteins, which are non directly connected in the MetaCore network, can be inferred from non negligible weights in G_{qr}. We propose to construct a reduced network highlighting the most important, direct and hidden, interactions between the twelve chosen proteins. Hence, for each protein *source* of the chosen subset, we retain, in the corresponding column of the $G_{rr} + G_{qr}$ matrix, the two most important weights revealing the most important protein *target* of the protein *source*. Here, we do not consider self-inhibition and self-activation matrix elements in the matrix sum $G_{rr} + G_{qr}$. The constructed reduced network associated to the twelve chosen proteins is presented in Fig. 6. We observe that it captures the above mentioned direct and hidden activation/inhibition actions between the considered proteins.

3.4 Sensitivity of the chosen subset

The PageRank sensitivity of the chosen subset of 12 proteins is obtained from the RIGMA and presented in Fig. 7 following the definitions given by (8) and (9). We remind that $F_+(a,b)$ gives the symmetric PageRank sensitivity of the nodes a and b to a variation of the link weight between them (in both directions from a to b and from b to a). The asymmetric PageRank sensitivity $F_-(a,b)$ determines what node is more sensitive to such weight variation. Thus, for $F_-(a,b) > 0$ we obtain that node a is more influenced by node b and for $F_-(a,b) < 0$ that node b is more influenced by node a.

In Fig. 7, the symmetric PageRank sensitivity (left panel) shows that the activation or the inhibition of the p53 protein affect or are affected by all the other chosen proteins. Indeed, the p53 protein with $K = 4$ occupies a very central role in the protein interactions network as it contributes to the stability of the genome preventing damage biological information to be spread [22, 23, 24]. The reddish horizontal and vertical lines on the symmetric PageRank sensitivity panel (Fig. 7 left) indicate that the activation of the EGFR, STAT3, FAK1, SHP-2 and the GRB2 proteins are affected or affect all the other proteins of the chosen set. The right panel of the Fig. 7 shows the asymmetric PageRank sensitivity. We clearly observe that in fact it is the p53 protein which influences the activation/inhibition of the other proteins, and in a stronger manner the inhibition of the GRB2, SHP-2, ITGB1, and

FAK1 proteins. In general, the inhibition of these four cited proteins is influenced by most of the other proteins (see greenish horizontal lines) and in return their respective activation influences also the other proteins (see greenish vertical lines).

3.5 Examples of magnetization of nodes

In Fig. 8, left panel, we show, in the PageRank-CheiRank (k, k^*)-plane (see Tables 1 and 2 for the relative PageRank and CheiRank indexes k and k^*), the PageRank magnetization M of the chosen 12 proteins. These nodes have global PageRank indexes $K \leq 26$ (see Table 1). In agreement with data presented in Fig. 3, for such K values, the magnetization is indeed mainly positive. So, these proteins are primarily activated. More precisely, as they belong to the top PageRank of the proteome ($K/N \leq 0.6‰$), these proteins are activated as the result of most important cascade of interactions along the causality pathways. The magnetization M is also presented in Fig. 7, right panel, but for every nodes with $K \leq 40$ (see Table 1). Here, for these top PageRank indexes, we have both positive and negative magnetization values, but the majority of the nodes have a magnetization close to zero, as discussed in Fig. 3.

For the top 40 PageRank ($K \leq 40$), the top 3 most activated nodes are the K^+ Potassium ion in cytosol ($K = 19$, $M(K) = 0.962815$), the $CO_2+H_2O \rightarrow H^+ + HCO_3^-$ reaction ($K = 29$, $M(K) = 0.757892$), and the intracellular mRNA ($K = 6$, $M(K) = 0.708154$), and the top 3 most inhibited nodes are the generic binding protein Bcl-2 ($K = 34$, $M(K) = -0.220751$), the generic enzyme MDM2 ($K = 36$, $M(K) = -0.208082$), and the generic binding protein E-cadherin ($K = 28$, $M(K) = -0.114311$).

4 Discussion

In this work, we have presented a detailed description of the statistical properties of the protein-protein interactions MetaCore network obtained with extensive Google matrix analysis. In this way, we find the proteins and molecules which are at the top PageRank and CheiRank positions playing thus an important role in the influence flow through the whole network structure. With a simple example of a subset of selected proteins (subset of selected nodes), we show that the reduced Google matrix analysis (RGMA) allows to determine the effective interactions between these proteins taking into account all the indirect pathways between these proteins through the global MetaCore network, in addition to direct interactions between selected proteins. We stress that the approach with the reduced Ising Google matrix algorithm, based on Ising spin description, allows to take into account the bi-functional nature of the protein-protein interactions (activation or inhibition) and to determine the average action type (or magnetization) of each protein .

Here, we have presented mainly the statistical properties of the MetaCore network without entering into detailed analysis of related biological effects. We plan to address, in further studies, the biological effects obtained from the reduced Google matrix analysis of the MetaCore network.

Abbreviation

PPI: protein-protein interactions
GMA: Google matrix analysis
IGMA: Ising Google matrix analysis
RGMA: reduced Google matrix analysis
RIGMA: reduced Ising Google matrix analysis
WWW: World Wide Web

Author details

[1]Clarivate Analytics, Barcelona 08025, Spain. [2]Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France. [3]Institut UTINAM, CNRS, Université Bourgogne Franche-Comté, Besançon, France.

References

1. MetaCore. https://clarivate.com/cortellis/solutions/early-research-intelligence-solutions/
2. Ekins, S., Bugrim, A., Brovold, L., Kirillov, E., Nikolsky, Y., Rakhmatulin, E., Sorokina, S., Ryabov, A., Serebryiskaya, T., Melnikov, A., Metz, J., Nikolskaya, T.: Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. Xenobiotica **36**(10-11), 877–901 (2006). doi:10.1080/00498250600861660. https://doi.org/10.1080/00498250600861660
3. Bessarabova, M., Ishkin, A., JeBailey, L., Nikolskaya, T., Nikolsky, Y.: Knowledge-based analysis of proteomics data. BMC bioinformatics **13 Suppl 16**, 13 (2012). doi:10.1186/1471-2105-13-S16-S13
4. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Networks **30**(1-7), 107–117 (1998). doi:10.1016/S0169-7552(98)00110-X
5. Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond - the Science of Search Engine Rankings. Princeton University Press, Princeton (2006)
6. Ermann, L., Frahm, K.M., Shepelyansky, D.L.: Google matrix analysis of directed networks. Rev. Mod. Phys. **87**, 1261–1310 (2015). doi:10.1103/RevModPhys.87.1261
7. Perfetto, L., Briganti, L., Calderone, A., Perpetuini, A.C., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., Pavlidou, T., Peluso, D., Petrilli, L.L., Pirrò, S., Posca, D., Santonico, E., Silvestri, A., Spada, F., Castagnoli, L., Cesareni, G.: SIGNOR: a database of causal relationships between biological entities. Nucleic Acids Res. **44**(Database-Issue), 548–554 (2016). doi:10.1093/nar/gkv1048
8. Lages, J., Shepelyansky, D.L., Zinovyev, A.: Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks. PLOS ONE **13**(1), 1–28 (2018). doi:10.1371/journal.pone.0190812
9. Frahm, K.M., Shepelyansky, D.L.: Ising-PageRank model of opinion formation on social networks. Physica A: Statistical Mechanics and its Applications **526**, 121069 (2019). doi:10.1016/j.physa.2019.121069
10. Frahm, K.M., Shepelyansky, D.L.: Google matrix analysis of bi-functional SIGNOR network of protein–protein interactions. Physica A: Statistical Mechanics and its Applications **559**, 125019 (2020). doi:10.1016/j.physa.2020.125019
11. Frahm, K.M., Shepelyansky, D.L.: Reduced Google matrix. 1602.02394
12. Frahm, K.M., Jaffrès-Runser, K., Shepelyansky, D.L.: Wikipedia mining of hidden links between political leaders. Eur. Phys. J. B **89**(269) (2016). doi:10.1140/epjb/e2016-70526-3
13. El Zant, S., Jaffrès-Runser, K., Shepelyansky, D.L.: Capturing the influence of geopolitical ties from Wikipedia with reduced Google matrix. PLOS ONE **13**(8), 1–31 (2018). doi:10.1371/journal.pone.0201397
14. Coquidé, C., Ermann, L., Lages, J., Shepelyansky, D.L.: Influence of petroleum and gas trade on eu economies from the reduced Google matrix analysis of UN COMTRADE data. The European Physical Journal B **92**(171) (2019). doi:10.1140/epjb/e2019-100132-6
15. Chepelianskii, A.D.: Towards physical laws for software architecture (2010). 1003.5455
16. Zhirov, A.O., Zhirov, O.V., Shepelyansky, D.L.: Two-dimensional ranking of Wikipedia articles. Eur. Phys. J. B **77**, 523–531 (2010). doi:10.1140/epjb/e2010-10500-7. 1006.4270
17. Aragon, P., Laniado, D., Kaltenbrunner, A., Volkovich, Y.: Biographical social networks on Wikipedia: A cross-cultural study of links that made history. In: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration. WikiSym '12. Association for Computing Machinery, New York, NY, USA (2012). doi:10.1145/2462932.2462958. https://doi.org/10.1145/2462932.2462958
18. Eom, Y.-H., Aragón, P., Laniado, D., Kaltenbrunner, A., Vigna, S., Shepelyansky, D.L.: Interactions of cultures and top people of Wikipedia from ranking of 24 language editions. PLOS ONE **10**(3), 1–27 (2015). doi:10.1371/journal.pone.0114825
19. Frahm, K.M., Shepelyansky, D.L.: Linear response theory for Google matrix. 1908.08924
20. MetaCoreNet. http://quantware.ups-tlse.fr/QWLIB/metacorenet/
21. Abel, M., Shepelyansky, D.L.: Google matrix of business process management. Eur. Phys. J. B **84** (2011). doi:10.1140/epjb/e2010-10710-y. 1009.2631
22. Prives, C., Hall, P.A.: The p53 pathway. The Journal of Pathology **187**(1), 112–126 (1999). doi:10.1002/(SICI)1096-9896(199901)187:1¡112::AID-PATH250¿3.0.CO;2-3
23. Joerger, A.C., Fersht, A.R.: The p53 pathway: Origins, inactivation in cancer, and emerging therapeutic approaches. Annual Review of Biochemistry **85**(1), 375–404 (2016). doi:10.1146/annurev-biochem-060815-014710. PMID: 27145840. https://doi.org/10.1146/annurev-biochem-060815-014710
24. Toufektchan, E., Toledo, F.: The guardian of the genome revisited: p53 downregulates genes required for telomere maintenance, dna repair, and centromere structure. Cancers **10**(5) (2018). doi:10.3390/cancers10050135

Competing interests

The authors declare that they have no competing interests.

Author's contributions

The authors contributed equally to this work. All authors read and approved the final manuscript.

Acknowledgements

We thank Andrei Zinovyev (Institut Curie) for useful discussions.

Funding

This research has been partially supported through the grant NANOX N° ANR-17-EURE-0009 (project MTDINA) in the frame of the *Programme des Investissements d'Avenir, France*. This research has been also supported by the Programme Investissements d'Avenir ANR-15-IDEX-0003, ISITE-BFC (GNETWORKS project) and the council of Bourgogne Franche-Comté region (APEX project and REpTILs project).

Tables

Table 1 Top 40 PageRank nodes of the simple MetaCore network. These nodes are sorted by descending PageRank probabilities $P(K)$ and consequently by ascending PageRank index K. The corresponding name, class and bio-localization of the node is given. The values $M(K)$ of the PageRank magnetization (7) are also given. The node highlighted in yellow corresponds to the twelve proteins selected for the RGMA and RIGMA analysis. These twelve proteins are ordered by the relative PageRank index k. Here, NA means not applicable.

K	$P(K)$ (10^{-2})	k	$M(K)$	Name	Class	Localization
1	0.2506		0	H^+ cytosol	Inorganic ion	Cytosol
2	0.2376		0	Na^+ cytosol	Inorganic ion	Cytosol
3	0.1741		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
4	0.1701	1	-0.028308	p53	Transcription factor	Nucleus
5	0.1469		0.256018	c-Src	Protein kinase	Cytoplasm
6	0.1435		0.708154	mRNA intracellular	RNA	Intracellular
7	0.1352		0	H^+ extracellular region	Inorganic ion	Extracellular region
8	0.1189	2	0.105603	EGFR	Receptor with enzyme activity	Plasma membrane
9	0.1180		-0.014278	DNA	DNA	Nucleus
10	0.1125	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.1125		0	K^+ extracellular region	Inorganic ion	Extracellular region
12	0.1056		0	ADP cytoplasm	Compound	Cytoplasm
13	0.1023	4	0.250910	STAT3	Transcription factor	Nucleus
14	0.0997		0.062046	Androgen receptor	Transcription factor	Nucleus
15	0.0947		0.287801	Rac1	RAS superfamily	Cytoplasm
16	0.0946		0	PO_4^{3-} cytoplasm	Compound	Cytoplasm
17	0.0940	5	0.006332	c-Myc	Transcription factor	Nucleus
18	0.0919	6	0.360271	FAK1	Protein kinase	Cytoplasm
19	0.0899		0.962815	cytosol $K^+ \rightarrow$ extracellular region K^+	Reaction	NA
20	0.0889	7	0.003377	ESR2 (nuclear)	Transcription factor	Nucleus
21	0.0884		0	K^+ cytosol	Inorganic ion	Cytosol
22	0.0849	8	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
23	0.0834	9	0.004567	ARX	Transcription factor	Cytoplasm
24	0.0828	10	0.208984	ITGB1	Generic receptor	Plasma membrane
25	0.0787	11	0.548888	SHP-2	Protein phosphatase	Cytoplasm
26	0.0776	12	0.364614	GRB2	Generic binding protein	Cytoplasm
27	0.0760		0.479956	PI3K reg class IA (p85)	Generic binding protein	Cytoplasm
28	0.0759		-0.114311	E-cadherin	Generic binding protein	Plasma membrane
29	0.0754		0.757892	$CO_2 + H_2O \rightarrow H^+ + HCO_3^-$	Reaction	NA
30	0.0753		-0.098664	p21	Generic binding protein	Nucleus
31	0.0752		0.148707	Caveolin-1	Generic binding protein	Cytoplasm
32	0.0749		0.007470	Ca^{2+} cytosol	Inorganic ion	Cytosol
33	0.0744		0.381345	PI3K reg class IA (p85-alpha)	Generic binding protein	Cytoplasm
34	0.0727		-0.220751	Bcl-2	Generic binding protein	Mitochondrion
35	0.0720		0	Cl^- intracellular	Inorganic ion	Intracellular
36	0.0712		-0.208082	MDM2	Generic enzyme	Nucleus
37	0.0707		-0.169004	PTEN	Lipid phosphatase	Cytoplasm
38	0.0702		0.391984	PPAR-gamma	Transcription factor	Nucleus
39	0.0698		0.031543	ACTB	Generic binding protein	Cytoplasm
40	0.0679		0	Acetyl-CoA intracellular	Compound	Intracellular

Figures

Table 2 Top 40 CheiRank nodes of the simple MetaCore network. These nodes are sorted by descending CheiRank probabilities $P^*(K^*)$ and consequently by ascending PageRank index K^*. The corresponding name, class and bio-localization of the node is given. The values $M(K^*)$ of the PageRank magnetization (7) are also given. The node highlighted in green corresponds to proteins from the subset of the twelve proteins chosen in Table 1 with $K^* \leq 40$. These proteins are ordered by the relative PageRank index k^*.

K^*	$P^*(K^*)$ (10^{-2})	k^*	$M(K^*)$	Name	Class	Localization
1	1.1464	1	0.006332	c-Myc	Transcription factor	Nucleus
2	0.8172		0.035667	eIF2C2 (Argonaute-2)	Generic enzyme	Cytoplasm
3	0.6722		-0.174071	IGF2BP3	Generic binding protein	Cytoplasm
4	0.4890		0.680968	Ubiquitin	Generic binding protein	Cytoplasm
5	0.3719		0.110759	SOX9	Transcription factor	Nucleus
6	0.3529	2	-0.028308	p53	Transcription factor	Nucleus
7	0.3373		0.228978	c-Fos	Transcription factor	Nucleus
8	0.3276		0	CUX1 (p110)	Transcription factor	Nucleus
9	0.2989		-0.057557	SP1	Transcription factor	Nucleus
10	0.2770	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.2769	4	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
12	0.2534		-0.010911	eIF2C1 (Argonaute-1)	Generic binding protein	Cytoplasm
13	0.2354		0.062046	Androgen receptor	Transcription factor	Nucleus
14	0.2350		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
15	0.2330		-0.075622	BRD4	Generic binding protein	Nucleus
16	0.2308		0.153950	Oct-3/4	Transcription factor	Nucleus
17	0.2259		-0.001577	PUM2	Generic binding protein	Cytoplasm
18	0.2239		0.188479	EZH2	Generic enzyme	Nucleus
19	0.2193		0.208146	p300	Generic enzyme	Nucleus
20	0.2072		-0.407833	TUG1	RNA	Cytoplasm
21	0.2072		-0.118501	E2F1	Transcription factor	Nucleus
22	0.2062		0	ASCC2	Generic binding protein	Nucleus
23	0.2005		0	LIMR	Generic receptor	Plasma membrane
24	0.1903		0.148471	BRG1	Generic enzyme	Nucleus
25	0.1871	5	0.250910	STAT3	Transcription factor	Nucleus
26	0.1811		0.381258	RBM24	Generic binding protein	Cytoplasm
27	0.1789		0.746981	SUMO-1	Generic binding protein	Nucleus
28	0.1728		0.140357	c-IAP2	Generic binding protein	Cytoplasm
29	0.1699		0.038221	HIF1A	Transcription factor	Nucleus
30	0.1677		0	Zn^{2+} cytosol	Inorganic ion	Cytosol
31	0.1623		-0.013644	CDK9	Protein kinase	Cytoplasm
32	0.1587		-0.223816	MeCP2	Generic binding protein	Nucleus
33	0.1533		-0.053592	ELAVL1 (HuR)	Generic binding protein	Nucleus
34	0.1497		0.120649	HDAC1	Generic enzyme	Nucleus
35	0.1473		-0.034082	BRD7	Generic binding protein	Nucleus
36	0.1452		0.131956	CREB1	Transcription factor	Nucleus
37	0.1449		0	Zn^{2+} nucleus	Inorganic ion	Nucleus
38	0.1423		0.096830	SUMO-2	Generic binding protein	Cytoplasm
39	0.1400		-0.051730	BRD2	Protein kinase	Cytoplasm
40	0.1343		0.228824	C/EBPbeta	Transcription factor	Nucleus

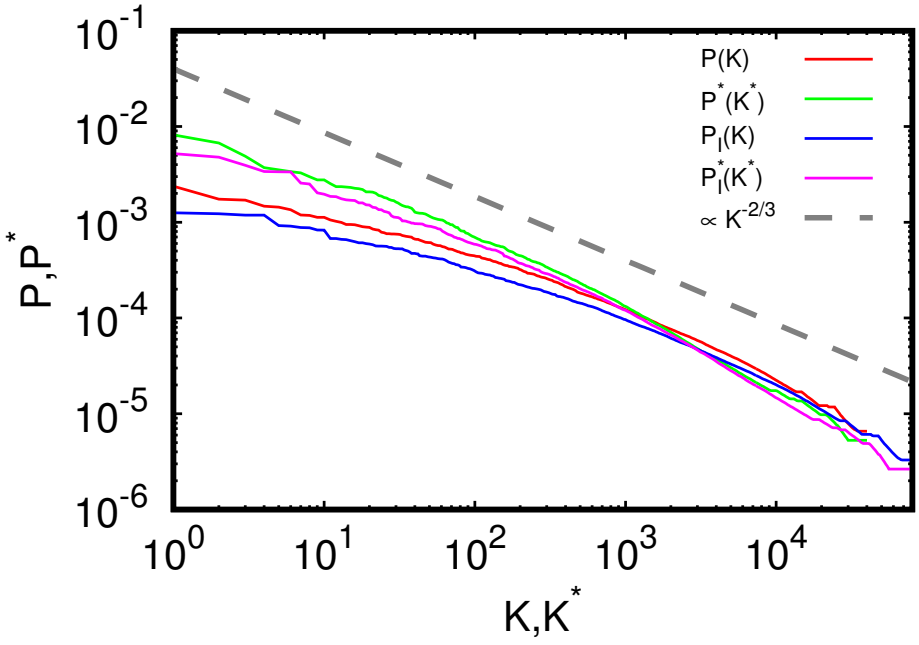

Figure 1 PageRank probability $P(K)$ ($P_I(K)$) and CheiRank probability $P^*(K^*)$ ($P^*_I(K)$) are shown as a function of the corresponding rank indexes K and K^* for the simple (Ising) MetaCore network. For comparison, the dashed gray line corresponds to the power decay $P \propto K^{-2/3}$.

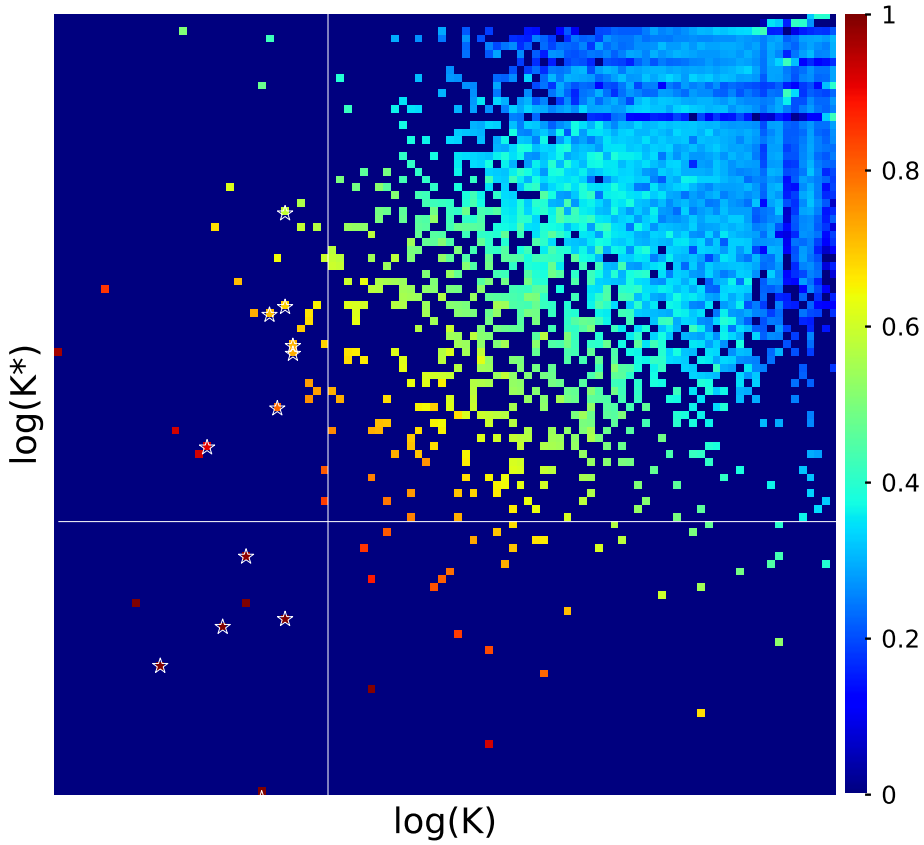

Figure 2 Density of nodes of the MetaCore network on the PageRank-CheiRank (K, K^*)-plane. The numbers of the color bar are a linear function of the logarithm of the density (with maximum values corresponding to 1 (red); minimum non-zero and zero values of the density corresponding to 0 (blue); the distribution is computed for 100×100 cells equidistant in logarithmic scale). The white stars indicate the positions of the 12 selected nodes presented in Table 1. The white vertical and horizontal lines represent nodes with $K \leq 40$ and $K^* \leq 40$, respectively.

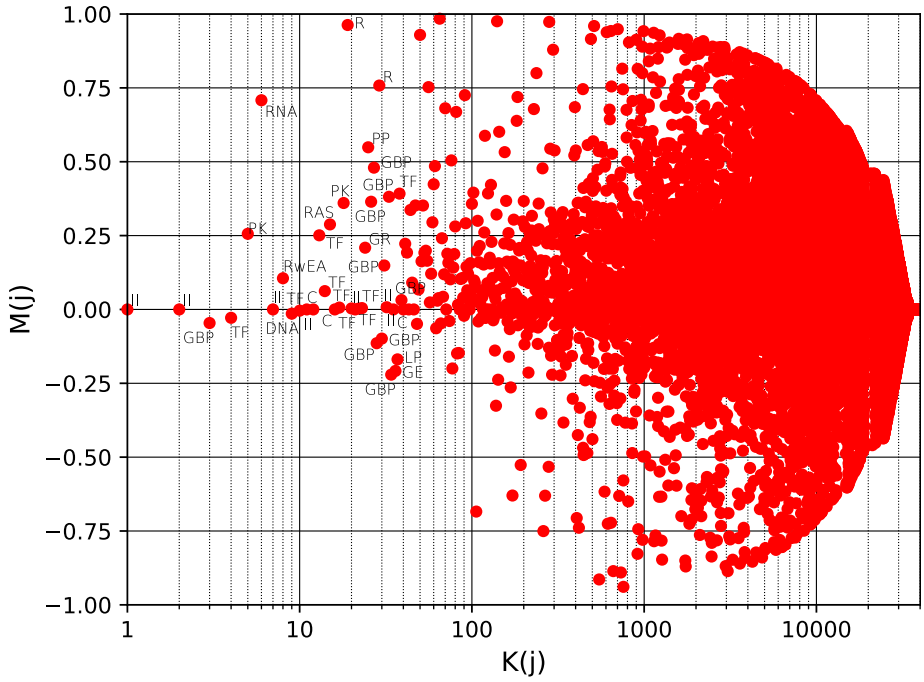

Figure 3 PageRank magnetization $M(j) = (P_+(j) - P_-(j))/(P_+(j) + P_-(j))$ for the Ising MetaCore network. Here, j is the node index and $K(j)$ is the PageRank index of the node j in the simple Metacore network (without node doubling). The biological class is reported for the top 40 PageRank nodes ($K \leq 40$, see Table 1): Inorganic ions (II), Generic binding protein (GBP), Transcription factor (TF), Protein kinase (PK), RNA, Receptor with enzyme activity (RwEA), DNA, Compound (C), RAS superfamily (RAS), Reaction (R), Generic receptor (GR), Protein phosphatase (PP), Generic enzyme (GE), Lipid phosphatase (LP).

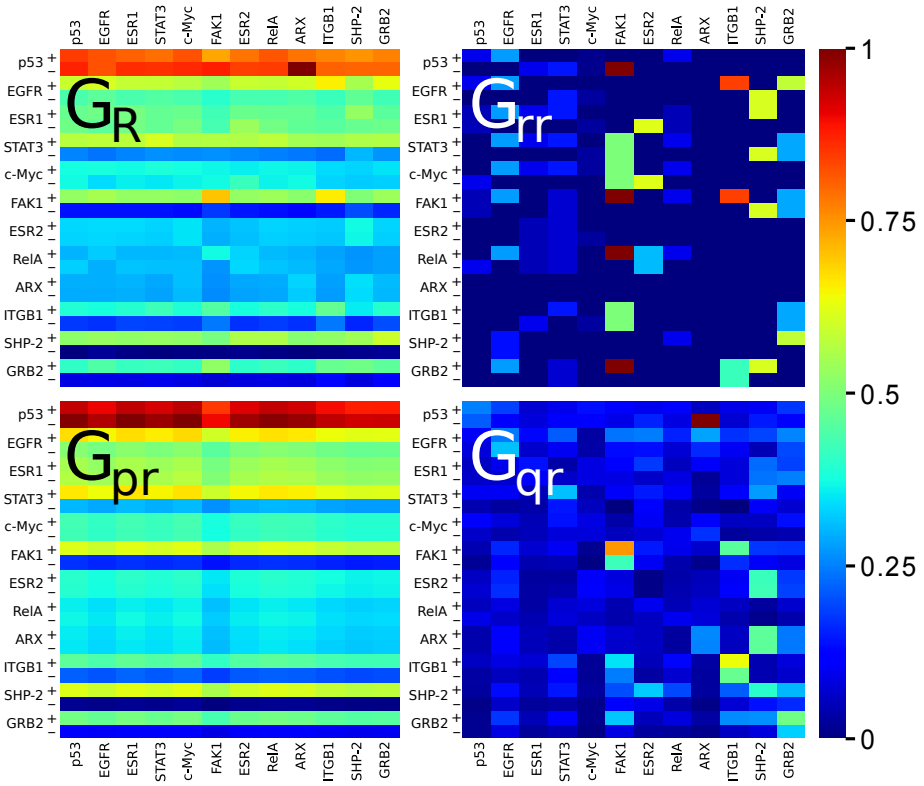

Figure 4 Reduced Google matrix G_{R} and its three matrix components G_{pr}, G_{rr} and G_{qr} associated to the subset of nodes presented in Table 1 and belonging to the Ising MetaCore network. The weights of the matrix components are $W_{\mathrm{pr}} = 0.952$, $W_{\mathrm{rr}} = 0.015$, and $W_{\mathrm{qr}} = 0.033$. Each colored cell corresponds to a $G_{\mathrm{X}ij}$ elements with X standing for R, rr, pr, or qr. A $G_{\mathrm{X}ij}$ matrix element is associated to the $j \to i$ link where the j index corresponds to proteins read on the bottom or the top axis of the panels and the i index corresponds to proteins read on the left axis of the panels. The + and − signs correspond to the activated and inhibited state of the node, respectively. The values of the color bar correspond to the ratio of the matrix element over its maximum value. Note that the elements of G_{qr} may be possibly negative. There are only few and very small negative values (between -2.6×10^{-5} and -5.3×10^{-5}) which are not distinguishable from zero and have the same blue color code. Therefore, the color bar is only shown for positive values.

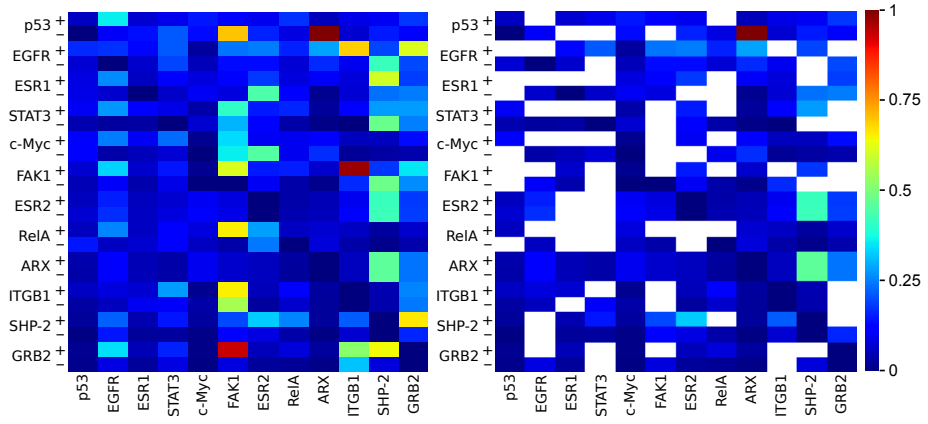

Figure 5 Sum of the two matrix components $G_{\mathbf{rr}} + G_{\mathbf{qr}}{}^{(\text{nd-block})}$. The matrix components are the same as in Fig. 4 with the exception of $G_{\mathrm{qr}}{}^{(\text{nd-block})}$ which is obtained from G_{qr} by excluding 2×2 diagonal blocks, each one of these blocks corresponding to a protein self-loop. The right panel is the same as the left panel with the exception of the white cells which hide the direct links $j \to i$ between the 12×2 chosen nodes in the Ising MetaCore network. The values of the color bar correspond to the ratio of the matrix element over its maximum value.

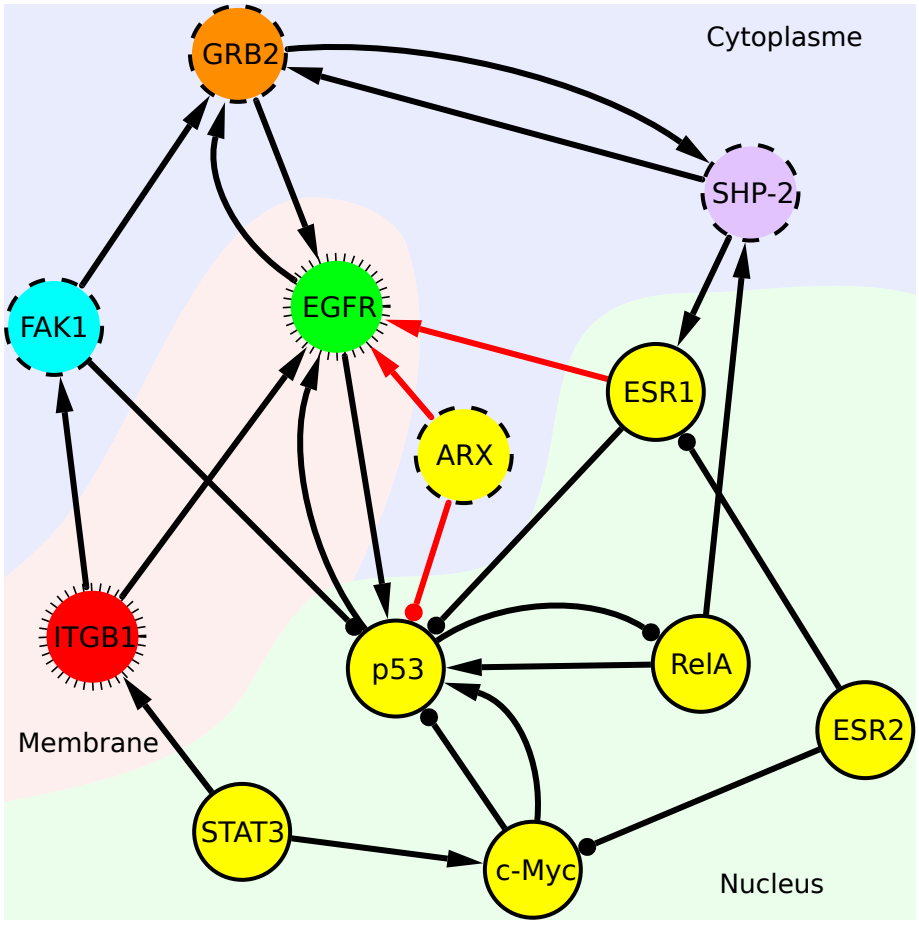

Figure 6 Reduced network of the chosen twelve proteins (see Table 1). The construction procedure of this network is given in the main text. Arrow edges (→) represent activation links and dot edges (⊸) represent inhibition links. The arrow tips and the dots are on the side of the target nodes. The black edges represent direct links. The red edges represent hidden links. The color of the nodes correspond to the type of proteins: transcription factors (yellow), protein kinase (cyan), generic receptor (red), receptor with enzyme activity (green), general binding protein (orange), and protein phosphatase (violet). The border style of the node correspond to the location of the proteins: nucleus (solid line), cytoplasm (dashed line), and plasma membrane (hairy line).

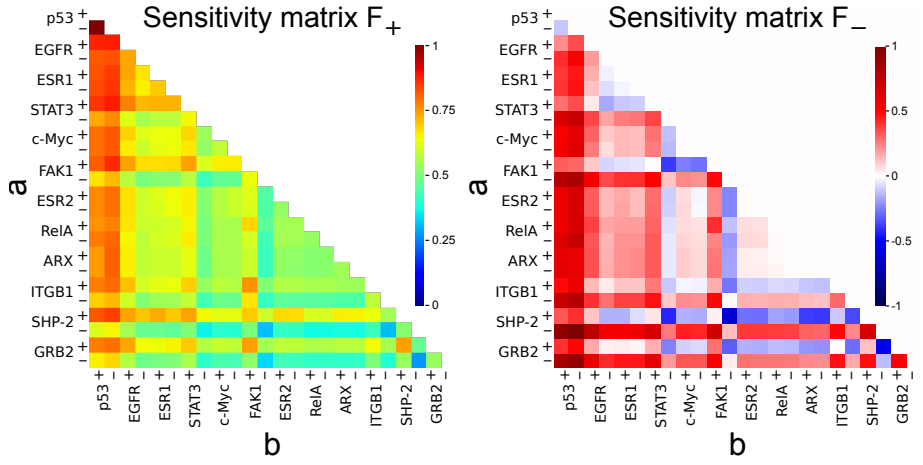

Figure 7 PageRank sensitivity matrices $F_+(a,b)$ (left panel) and $F_-(a,b)$ (right panel) associated to the subset of nodes presented in Table 1 and belonging to the Ising MetaCore network. The values of the color bar correspond to $F_+(a,b)/\max_{a,b}(F_+(a,b))$ (left panel) and to $F_-(a,b)/\max_{a,b}|F_-(a,b)|$ (right panel).

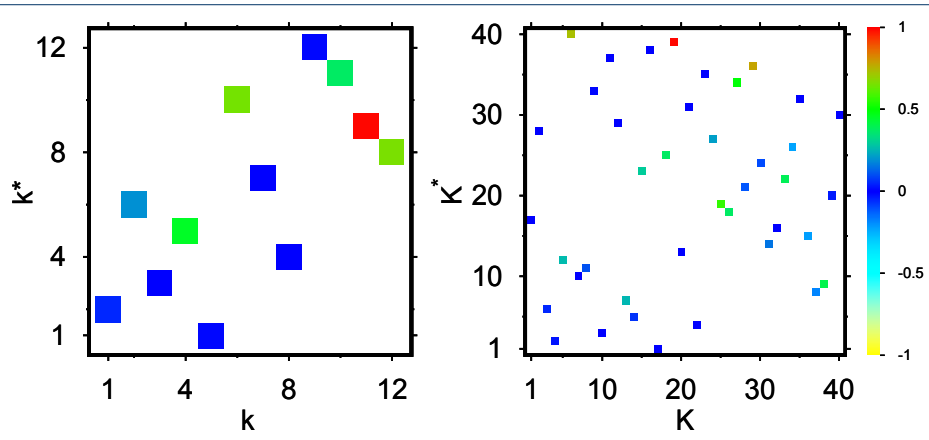

Figure 8 PageRank magnetization $M(K) = (P_+(K) - P_-(K))/(P_+(K) + P_-(K))$ presented in the PageRank-CheiRank (K, K^*)-plane. Left panel: PageRank magnetization $M(k)$ for the chosen twelve proteins presented in the relative indexes (k, k^*)-plane (see k adn k^* indexes in Table 1). Right panel: PageRank magnetization $M(K)$ for nodes with $K \leq 40$. Here, $P_\pm(K)$ is the PageRank probability of the $(\pm)$ component of the Ising MetaCore network node associated with the K PageRank (see text). The values of the color bar correspond to $M/\max|M|$ with $\max_{k\leq 12}|M(k)| = 0.549$ (left panel) and $\max_{K\leq 40}|M(K)| = 0.963$ (right panel). On the right panel, the K^* index is here the relative CheiRank index inside the set of the first $K \leq 40$ nodes.