Google matrix analysis of bi-functional SIGNOR network of protein-protein interactions

Klaus M. Frahm and Dima L. Shepelyansky

1Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France.

∗To whom correspondence should be addressed.

Received on August 28, 2019; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Directed protein networks with only a few thousand of nodes are rather complex and do not allow to extract easily the effective influence of one protein to another taking into account all indirect pathways via the global network. Furthermore, the different types of activation and inhibition actions between proteins provide a considerable challenge in the frame work of network analysis. At the same time these protein interactions are of crucial importance and at the heart of cellular functioning.

Results: We develop the Google matrix analysis of the protein-protein network from the open public database SIGNOR. The developed approach takes into account the bi-functional activation or inhibition nature of interactions between each pair of proteins describing it in the frame work of Ising-spin matrix transitions. We also apply a recently developed linear response theory for the Google matrix which highlights a pathway of proteins whose PageRank probabilities are most sensitive with respect to two proteins selected for the analysis. This group of proteins is analyzed by the reduced Google matrix algorithm which allows to determine the effective interactions between them due to direct and indirect pathways in the global network. We show that the dominating activation or inhibition function of each protein can be characterized by its magnetization. The results of this Google matrix analysis are presented for three examples of selected pairs of proteins. The developed methods work rapidly and efficiently even for networks with several million of nodes and can be applied to various biological networks.

Availability: The Google matrix data and executive code of described algorithms are available at http://www.quantware.ups-tlse.fr/QWLIB/google4signornet/

1 Introduction

Protein-protein interactions (PPI) are at the heart of information processing and signaling in cellular functions. It is natural to present and analyze these PPI by presenting them as a directed network of actions between proteins (or network nodes). The development of related academic databases of PPS networks with an open public access is a challenging task with various groups working in this direction (see e.g. Liberzon et al., 2011; Perfetto, L. et al., 2016; Kanehisa et al., 2017; Fabregat et al., 2018; Splender et al., 2018). A typical example is the SIGNOR directed network of PPI links for about 4000 proteins of mammals and 12000 bi-functional directed links as reported by Perfetto, L. et al., 2016.

On the scale of the past twenty years, modern society has created a variety of complex communication and social networks including the World Wide Web (WWW), Facebook, Twitter, Wikipedia. The size of these networks varies from a several millions for Wikipedia to billions and more for Facebook and WWW. The description of generic features of these complex networks can be found e.g. in Dorogotsev, 2010.

An important tool for the analysis of directed networks is the construction of the Google matrix of Markov transitions and related PageRank algorithm invented by Brin and Page in 1998 for ranking of all WWW sites (see Brin and Page, 1998; Langville and Meyer, 2006). This approach has been at the foundations of the Google search engine used world wide. A variety of applications of Google matrix analysis to various directed networks is described by Ermann et al., 2015.

Here we apply recently developed extensions of Google matrix analysis, which include the REDuced GOogle MAtrix (REGOMAX)
algorithm (Frahm et al., 2016) and the Linear Response algorithm for GQogle MaxNetX (LRIGOMAX) (Frahm and Shepelyansky, 2019a), to the SIGNOR PPI network. The efficiency of these algorithms has been demonstrated for Wikipedia networks of politicians (Frahm et al., 2016) and world universities (Frahm and Shepelyansky, 2019a; Coquide et al., 2019a), and multi product world trade of UN COMTRADE database (Coquide et al., 2018b). Thus it is rather natural to apply these algorithms to PPI networks which have a typical size being significantly smaller than Wikipedia and WWW.

From a physical view-point the LRIGOMAX approach corresponds to a small probability pumping at a certain network node (or group of nodes) and absorbing probability at another specific node (or group of nodes). This algorithm allows first to determine the most sensitive group of nodes involved in this pumping-absorption process tracing a pathway connecting two selected proteins. In a second stage one can then apply the REGOMAX algorithm and obtain an efficient reduced Google matrix, and in particular effective interactions, for the found subset of most sensitive nodes. These interactions are due to either direct or indirect pathways in the global huge network in which is embedded the selected relatively small subset of nodes.

The REGOMAX and LRIGOMAX algorithms originate from the scattering theory of nuclear and mesoscopic physics, field of quantum chaos and linear response theory of electron transport (Frahm et al., 2016; Frahm and Shepelyansky, 2019a).

We point out that the analysis of the SIGNOR PPI network already found biological applications reported by Sacco et al., 2016, Lun X-K et al., 2017; Kanhaiya et al., 2017; Dimitrakopoulos et al., 2018. The detailed review of various applications of the PPI signaling networks is given by Invergo and Beltrao, 2018. However, the Google matrix analysis has not been used in these studies.

The challenging feature of PPI networks is the bi-functionality of directed links which produce activation or inhibition actions. While in our previous analysis of SIGNOR network by Lages et al., 2018 this feature was ignored, here we apply the Ising-PageRank approach developed in Frahm and Shepelyansky, 2019b for opinion formation modeling. In this Ising-type approach the number of nodes in the PPI network is doubled, with a (+) or (−) attribute for each protein, and the links between doubled nodes are described by 2 × 2 matrices corresponding to activation or inhibition actions.

In this work we apply the LRIGOMAX and REGOMAX algorithm to the bi-functional PPI network of SIGNOR. We show that this approach allows to determine the effective sensitivity with direct and indirect interactions between a selected pair of proteins. As particular examples we choose three protein pairs implicating the Epidermal growth factor receptor (EGFR) which is considered to play an important role in the context of lung cancer (see e.g. Bethune et al., 2010; Zamay et al., 2017).

The paper is constructed as follows: In section 2 we describe the construction of Google matrix from links between proteins and related LRIGOMAX and REGOMAX algorithms, in section 3 we characterize data sets and the Ising-PPI network for bi-functional interactions between proteins, results are presented in section 4 and the conclusion is given in section 5. Supplementary Material provides additional matrix data and executable code for the described algorithms for the SIGNOR Ising-PPI network.

2 Methods of Google matrix analysis

2.1 Google matrix construction

The Google matrix G of N nodes (proteins or proteins with (+)/ (−) attribute) is constructed from the adjacency matrix Aij, with element 1 if node j points to node i and zero otherwise. The matrix G has the standard form $G_{ij} = \alpha S_{ij} + (1 - \alpha)/N$ (see Brin and Page, 1998; Langville and Meyer, 2006; Ermann et al., 2015), where S is the matrix of Markov transitions with elements $S_{ij} = a_{ij}/\sum k a_{kj}(j)$ and $a_{ij}(j) = \sum_{i=1}^{N} A_{ij} \neq 0$ being the out-degree of node j (number of outgoing links); $S_{ij} = 1/N$ if j has no outgoing links (dangling node). The parameter $0 < \alpha < 1$ is known as the damping factor with the usual value $\alpha = 0.85$ (Langville and Meyer, 2006) which we use here. For the range 0.5 $\leq \alpha \leq 0.95$ the results are not sensitive to $\alpha$ (Langville and Meyer, 2006; Ermann et al., 2015). A useful view on this G matrix is given by the concept of a random surfer, moving with probability $\alpha$ from one node to another via one of the available directed links or with a jump probability $(1 - \alpha)$ to any node.

The right PageRank eigenvector of G is the solution of the equation $GP = \lambda P$ for the leading unit eigenvalue $\lambda = 1$ (Langville and Meyer, 2006). The PageRank $P(j)$ values represent positive probabilities to find a random surfer on a node $j \sum_i P(i) = 1$. All nodes can be ordered by decreasing probability $P$ numbered by PageRank index $K = 1, \ldots, N$ with a maximal probability at $K = 1$ and minimal at $K = N$. The numerical computation of $P(j)$ is done efficiently with the PageRank iteration algorithm described by Langville and Meyer, 2006. The idea of this algorithm is simply to start with some initial, sum normalized, vector $P^{(0)}$ of positive entries, e.g. being 1/N for simplicity, and then to iterate $P(n+1) = G P^{(n)}$ which typically converges after $n = 150 - 200$ iterations (for $\alpha = 0.85$).

It is also useful to consider the original network with inverted direction of links. After inversion the Google matrix G* is constructed via the same procedure with $G^* P^* = P^*$. The matrix $G^*$ has its own PageRank vector $P^*(j)$ called CheiRank (Chepelianski, 2010; Ermann et al., 2015). Its values give probabilities to find a random surfer of a given node and they can be again ordered in a decreasing order with CheiRank index $K^*$ with highest $P^*$ at $K^* = 1$ and smallest at $K^* = N$. On average, the high values of $P^*(P)$ correspond to nodes with manyingo (outgoing) links (Langville and Meyer, 2006; Ermann et al., 2015).

2.2 Reduced Google matrix (REGOMAX) algorithm

The REGOMAX algorithm is described in detail by Frahm et al., 2016; Lages et al., 2018; Coquide et al., 2019a. It allows to compute efficiently a "reduced Google matrix" $G_{K}$ of size $N_r \times N_r$ that captures the full contributions of direct and indirect pathways appearing in the full Google matrix G between $N_r$ nodes of interest selected from a huge global network with $N \gg N_r$ nodes. For these $N_r$ nodes their PageRank probabilities are the same as for the global network with $N$ nodes, up to a constant multiplicative factor taking into account that the sum of PageRank probabilities over $N_r$ nodes is unity. The computation of $G_{K}$ determines a decomposition of $G_{K}$ into matrix components that clearly distinguish direct from indirect interactions: $G_{K} = G_{rr} + G_{qr} + G_{sr}$ (Frahm et al., 2016). Here $G_{rr}$ is given by the direct links between the selected $N_r$ nodes in the global G matrix with $N$ nodes. We note that $G_{rr}$ is rather close to the matrix in which each column is approximately proportional to the PageRank vector $P_r$, satisfying the condition that the PageRank probabilities of $G_{rr}$ are the same as for $G$ (up to a constant multiplicative due to normalization). Hence, in contrast to $G_{rr}$, $G_{qr}$ doesn’t give much new information about direct and indirect links between selected nodes.

The most interesting role is played by $G_{sr}$, which takes into account all indirect links between selected nodes happening due to multiple pathways via the global network of nodes $N$ (see Frahm et al., 2016). The matrix $G_{qr} = G_{Qdr} + G_{Qsr}$ has diagonal ($G_{Qdr}$) and non-diagonal ($G_{Qsr}$) parts with $G_{Qdr}$ describing indirect interactions between selected nodes. The exact formulas for all three components of $G_{sr}$ are given in Frahm et al., 2016. It is also useful to compute the weights $W_{rr} , W_{qr} , W_{sr} , W_{rr}$ of $G_{K}$ and its 3 matrix components $G_{rr}, G_{qr}, G_{sr}$ given by the sum of
all its elements divided by the matrix size $N_{\ell}$. Due to the column sum normalization of $G_{\ell}$ we obviously have $W_{\ell} = W_{\ell 1} + W_{\ell 2} + W_{\ell 3} = 1$.

We note that the matrix elements of $G_{\ell}$ may have negative values (only the full reduced matrix $G_{\ell}$ should have positive elements; $G_{\ell}$ also has only positive matrix elements) but these negative values are found to be small for the Ising-PPI-networks and do not play a significant role. A similar situation for Wikipedia networks is discussed by Frahm et al., 2016; Frahm and Shepelyansky, 2019a.

2.3 LIRGOMAX algorithm
The detainted description of the LIRGOMAX algorithm is given by Frahm and Shepelyansky, 2019a. It performs an infinitely weak $\epsilon$-probability injection (pumping) at one node (a protein or a protein with $(+)/(−)$ attribute) and absorption at another node of interest. This process is described by the modified PageRank iteration $P_{\ell}^{(k+1)} = G F(\epsilon, P_{\ell})$ where the vector valued function $F(\epsilon, P)$ has the components $F(\epsilon, P)$ for $1$ being the index of the injection/pumping node, $P(j) - \epsilon$ for $j$ being the index of the absorption node and simply $P(k)$ for all other nodes $k$. In this way the vector $F(\epsilon, P)$ is also sum normalized if $P$ is sum normalized and obviously $F(0, P) = P$ is the identity operation. In Frahm and Shepelyansky, 2019a a more general version of $F(\epsilon, P)$ was considered with potentially different prefactors for the $\epsilon$ contributions, injection/absorption at possibly more than two nodes and an additional renormalization factor to restore the sum normalization (which is automatic in the simple version). However, for the applications in this work the above given simple version of $F(\epsilon, P)$ is sufficient.

In principle one can solve iteratively the above modified PageRank iteration formula which converges at the same rate as the usual PageRank iteration algorithm and provides a modified $\epsilon$-depending PageRank $P(\epsilon)$. Then one can compute the linear response vector $P_1 = dP(\epsilon)/d\epsilon|_{\epsilon=0} = \lim_{\epsilon \to 0} [P(\epsilon) - P(0)]/\epsilon$ where $P_0$ is the PageRank obtained for $\epsilon = 0$. However the naive direct evaluation of this limit is numerically not stable in the limit $\epsilon \to 0$. Fortunately as shown by Frahm and Shepelyansky, 2019a it is possible to compute $P_1$ directly in an accurate and efficient way by solving the inhomogeneous PageRank equation

$$P_0 = G P_1 + V_0 , \quad V_0 = G W_0$$

where the vector $W_0$ has only two non-zero components for the two particular injection or absorption nodes $W_0(i) = 1$ or $W_0(j) = 1$ respectively. Therefore a more explicit expression for the vector $V_0$ appearing in (1) is $V_0(k) = G_{\ell k} - G_{\ell j}$ (for all nodes $k$). We mention that the three vectors $P_0$, $V_0$ and $W_0$ are orthogonal to the vector $k^T = (1, \ldots, 1)$ composed of unit entries, i.e. $\sum_j P_0(k) = \sum_j V_0(k) = \sum_j W_0(k) = 0$. Furthermore, all of these vectors, especially $P_0$ have real positive or negative entries (note that in general eigenvectors of a non-symmetric real matrix may be complex).

A formal solution of the inhomogeneous PageRank equation is: $P_1 = \sum_{m=0}^{\infty} G_0^m V_0 = (1 - G)^{-1} V_0$ which is well defined since $V_0$, when expanded in the basis of (generalized) eigenvectors of $G$, does NOT have a contribution of $P_0$ (the only eigenvector of $G$ with eigenvalue $1$) such that the singularity of the matrix inverse does not constitute a problem. Of course numerically, we compute $P_0$ in a different way, as described by Frahm and Shepelyansky, 2019a one can iterate the equation $P_1^{(n+1)} = G F^{(n)} P_0 + V_0$ with $P_0^{(0)} = 0$ which converges with the same rate as the usual PageRank iteration.

In a similar as the PageRank $P_0$ is characterized by the index $K$ we introduce the index $K_{\ell}$ by ordering $|P_1|$ such that $K_{\ell 1} = 1$ corresponds to the node with largest value of $|P_1|$ and $K_{\ell \ell} = N$ to the node with smallest value of $|P_1|$. Once $P_0$ is computed for the pair of chosen injection/absorption nodes we determine the 20 top nodes with strongest negative values of $P_0$ and further 20 top nodes with strongest positive values of $P_1$ which constitute a subset of 40 nodes which are the most significant nodes participating in the pathway between the pumping node $i$ and absorbing node $j$. We also require that these two particular nodes $i$ and $j$ belong to this subset. If this is not automatically the case we replace the node at total position 20 (position 20 for strongest negative values of $P_0$) with the absorption node $j$ and/or the node at total position 40 (position 20 for strongest positive values of $P_1$) with the injection node $i$. This situation happens once for the absorption node of the third example below which has a very low ranking position $K_{\ell} \approx 2000$ for $|P_1|$.

In general from a physical/biological point of view we indeed expect that the two particular injection/absorption nodes belong automatically to the selected subset of most sensitive nodes. However, there is no simple or general mathematical argument for this.

Using this subset of top nodes in the $K_{\ell}$ ranking we then apply the REGOMAX algorithm to compute the reduced Google matrix and its components and in particular we determine the effective direct and indirect interactions of this reduced network. The advantage of the application of the LIRGOMAX at the initial stage is that it provides an automatic and more rigorous procedure to determine an interesting subset of protein nodes related to the pumping between nodes $i$ and $j$ instead of using an arbitrary heuristic choice for such a subset.

3 Data sets and Ising-PPI-network construction
We use the open public SIGNOR PPI network Perfetto, L. et al., 2016 (April 2019 release for human, mouse and rat). This network contains $N = 4341$ nodes (proteins) and $N_{\ell} = 12547$ directed hyperlinks between nodes. Each protein (node) is described by its name and identifier.

A new interesting feature of this PPI directed network is that its hyperlinks have activation and inhibition actions. For some links the functionality is unclear and they are considered to be neutral. This feature raises an interesting mathematical challenge for the Google matrix description of such bi-functional networks. To meet this challenge we use the Ising-PageRank approach developed by Frahm and Shepelyansky, 2019b for a model of opinion formation on social networks. In this approach each node is doubled getting two components marked by $(+)$ and $(-)$. The activation links point to the $(+)$ components and inhibition links point to the $(-)$ components. Such transitions between doubled nodes are described by $2 \times 2$ block matrices $\sigma_{\ell}(+) \sigma_{\ell}(-)$ matrices with entries $1(0)$ in the first row and $0(1)$ in the second row as for Ising spins $1/2$ (see details described in Supplementary Material). A neutral transition is described by $2 \times 2$ block matrices $\sigma_{\ell}$ with all elements being $1/2$. Thus for this Ising-network (doubled-size network) we have doubled number of node $N = 8682$ and the total number of hyperlinks being $N_{\ell} = 27266$; among them there are $N_{act} = 14941$ activation links, $N_{inh} = 7975$ inhibition links and $N_{neutral} = 4344$ neutral links ($N_{act} + N_{inh} + N_{neutral}$). From this weighted Ising-PPI-network with $N_{\ell} = 27266$ nodes we construct the Google matrix following the standard rules described by Langville and Meyer, 2006; Ermann et al., 2015 and also given above.

Below we apply the Google matrix analysis taking into account the bi-functionality PPI and illustrate the efficiency of the LIRGOMAX and REGOMAX algorithms for the SIGNOR Ising-PPI network.

The details of Ising-PPI-network construction, its main statistical properties and an executable code for the described algorithms are provided in the Supplementary Material and in Frahm and Shepelyansky, 2019c. Below we discuss the results obtained with the LIRGOMAX and REGOMAX algorithms for three examples of specific pathways between two specific proteins.
4 Results

Here we present results obtained with LIRGOMAX and REGOMAX algorithms for pathways between several pairs of selected proteins.

4.1 Case of pathway EGFR - JAK2 proteins

As a first example we choose the node \textit{EGFR P00533 (+)} for injection (pumping) and \textit{JAK2 O60674 (-)} for absorption. It is known that mutations affecting the protein EGFR expression or activity could result in lung cancer (see e.g. Bethune et al. 2010, Zayam et al., 2017). This protein interacts with the protein JAK2 whose mutations have been implicated in various types of cancer. We argue that the injection (pumping) at \textit{EGFR P00533 (+)} and absorption at \textit{JAK2 O60674 (-)} should involve certain variations of the PageRank probability, represented by \(P_1\), showing interactions between various proteins actively participating in the pathway from \textit{EGFR P00533 (+)} to \textit{JAK2 O60674 (-)}. The pumping process can be viewed as a result of disease development and absorption as a certain mutation of this disease into another one.

![Linear response vector of PageRank](image)

The global PageRank indices of these two nodes are \(K = 90\) (PageRank probability \(P(90) = 0.00009633\) for \textit{EGFR P00533 (+)} and \(K = 470\) (PageRank probability \(P(470) = 0.0003444\) for \textit{JAK2 O60674 (-)}. As described above in the LIRGOMAX computations we choose the vector in \(V_1\) which appears in the inhomogeneous PageRank equation (1) as \(V_0 = G W_0\) with \(W_0(K = 90) = +1\), \(W_0(K = 470) = -1\) and \(W_0(K) = 0\) for all other values of the index \(K\). We remind that both \(W_0\) and \(V_0\) are orthogonal to the left leading eigenvector \(E^T = (1, \ldots, 1)\) of \(G\) according to the general description of the LIRGOMAX algorithm given above and in Frahm and Shepelyansky, 2019c.

For comparison we let us note that the top 4 PageRank nodes are \(K = 1\) \((P(1) = 0.00003041)\) for \textit{CASPI3 P42574 (+)} and \(K = 2(P(2) = 0.0002821)\)
for NOTCH1 P46531 (+), K = 3 (P(3) = 0.002433) for PIK3CD 006329 (-), K = 4 (P(4) = 0.002413) for CTNNB1 P35222 (-) (other values/data are available at Frahm and Shepelyansky, 2019a).

Similar to the two Wikipedia examples analyzed by Frahm and Shepelyansky, 2019a the LIRGOMAX algorithm selects the proteins mostly affected by injection/absorption process with 20 most positive (EGFR block) and 20 most negative (JAK2 block) values of $P_i$ shown in Table 1. Here the pumped protein EGFR P00333 (+) is on the third position in its block of positive $P_i$ values ($i = 23$) and with $K_L = 11$ (where $K_L$ is the ranking index obtained by ordering the components of $|P_i|$) while the protein with absorption JAK2 060674 (-) has the second position in its block of negative $P_i$ values ($i = 2$) with $K_L = 4$. Thus these two nodes are not at the first positions in their respective blocks but still are placed at very high positions.

The dependence of $|P_i|$ of the index $K_L$ is shown in the top panel of Figure 1. The decay of $|P_i|$ is relatively slow for $K_L \leq 40$ followed by a more rapid drop for $K_L > 40$. The bottom panel shows the dependence of positive (blue) and negative (red) values of $P_i$ in $K_L$. We note that the top absolute values $|P_i|$ for blue and red components have comparable values being of the order of $|P_i| \sim 0.1$ for approximately $K_L \leq 40$. However, in this range the number of positive (blue) values of $P_i$ is significantly smaller compared to the number of negative (red) values of $P_i$. This point can also be seen from the column of $K_L$ values in Table 1. Another feature visible from Table 1 is that the number of proteins with negative component (−) is significantly smaller than those with a positive component (+) for 1 ≤ i ≤ 20 and for 21 ≤ i ≤ 40. We return to the properties of positive and negative components a bit later.

After the selection of most significant 40 nodes of the pathway between the two injection/absorption proteins (see Table 1) we apply the REGOMAX algorithm which determines all matrix elements of Markov transitions between these 40 nodes including all direct and indirect pathways via the large global Ising PPI-network with 3862 nodes.

The reduced Google matrix $G_G$ and its three components $G_{rr}, G_{pr}$, $G_{qr}$ are shown in Figure 2 for proteins of Table 1 (1 ≤ i ≤ 40). The weight of the component $G_{rr}$ is $W_{rr} = 0.761$ being not so far from unity but this value is below $W_{rr} \approx 0.95$ appearing usually in Wikipedia networks (Frahm et al., 2016; Frahm and Shepelyansky, 2019a). We attribute this to a significantly smaller number of links per node being $\ell = N_i/N \approx 3.1$ for the Ising PPI-network while for the English Wikipedia network of 2017 we have $\ell \approx 22.5$ (Frahm and Shepelyansky, 2019a). Indeed, the weight $W_{rr} = 0.220$ of direct transitions of $G_{rr}$ is significantly larger than the corresponding values for the Wikipedia case with $W_{rr} \approx 0.04$. However, the weights $W_{pr} = 0.019$ are comparable for both reduced networks.

The matrix structure of direct transitions $G_G$ has a clear two block structure with dominant transitions inside each block associated to EGFR and JAK2 with only 4 significant matrix elements from the EGFR to the JAK2 block. These matrix elements correspond to links from EGFR (+) to JAK2 (+) and STAT1 (+) and have the same value $g = 0.0167$ while all other matrix elements (of this EGFR to JAK2 block) are very small with the value $g \approx 1.73 \times 10^{-10}$ corresponding to the minimal value $(1 - \alpha)/N$ in $G$ related to the damping factor $\alpha = 0.85$.

The matrix $G_{qr}$ (which is exactly of rank 1) has a very simple structure with all columns being (approximately) proportional to the (local) PageRank of $G_{rr}$ (which is itself proportional to the global PageRank projected onto the subset of 40 nodes) and one clearly sees that the strong horizontal red lines correspond to index positions $i$ of Table 1 where the corresponding index $K$ is quite low below $\sim 100$ corresponding to a relatively high PageRank position. The full reduced matrix $G_{rr}$ is numerically dominated by $G_{rr}$ (but less clearly as for typical Wikipedia cases) and has at first sight a similar structure as $G_{rr}$ but with somewhat smaller values. However, some of the strongest direct links (from $G_{rr}$) are also visible. Similarly to the Wikipedia network of politicians as discussed in Frahm et al., 2016 both matrix components $G_{rr}$ and $G_{qr}$ are not very useful to identify the indirect links.

The indirect links are visible in the matrix $G_{qr}$. As explained and shown mathematically in Frahm et al., 2016 they correspond to pathways where a given node $i$ of the small subset points to a certain node outside the subset (in the big surrounding PPI network) which itself points eventually to another node outside the subset and comes after a finite number of
iterations finally back to a different node $i_2$ inside the subset. This provides an indirect link from $i_2$ to $i_2$ and the weight or strength of this indirect link is characterized by the value of the matrix element $G_{i_2i_2}$, sometimes with negative values (note that the matrix elements of $G_{i_2i_2}$ may be negative).

According to Figure 2 there are now also significant interactions between the two blocks of EGFR and JAK2 for the matrix $G_{ij}$, sometimes with negative values (note that the matrix elements of $G_{ij}$ may be negative).

Figure 3 shows the sum of the two components $G_{ij} + G_{ij}^{quadr}$ ($G_{ij}^{quadr}$ corresponds to $G_{ij}$ without its diagonal elements) which confirms this observation. Actually, we consider that the elements of $G_{ij} + G_{ij}^{quadr}$ describe best the combined direct and indirect links for the given subset.

Due to the contribution of indirect transitions there are additional transitions between these two blocks where the four strongest additional elements of $G_{ij}$ have values $g = 0.0106$ (GRB2 P62993 (+) to JAK2 O60674 (+)); $q = 0.0099$ (GRB2 P62993 (+) to STAT1 P42224 (+)); $g = 0.0059$ (GAB1 Q13480 (+) to GTF2I P78347 (+)); $g = 0.0039$ (PIK3R1 P27986 (+) to GTF2I P78347 (+)). There are also 11 additional transitions with $g > 0.1$. Thus even if the weight of $G_{ij}$ is not high it provides important new indirect interactions between proteins from the EGFR block to the JAK2 block.

The situation is even more striking when we consider the transitions from the JAK2 block to the EGFR block. There are no direct links between these blocks in this direction from the global network but due to the construction of the Google matrix described above there are still numerically very small values $q$ for the matrix elements of $G_{ij}$ due to dangling nodes (nodes with no outgoing links) with $q = 1/N \approx 1.15 \times 10^{-4}$ (in certain columns) or due to the damping factor term $(1 - \alpha)/N \approx 1.73 \times 10^{-5}$ (for the other columns). On the other side concerning the indirect links described by $G_{ij}$ we find rather significant transitions from the JAK2 block to the EGFR block with the four largest values: $g = 0.0122$ (CCCR2 P41597 (+) to EGFR P00533 (-)) and to ESR1 P03372 (+); $g = 0.006$ (CSF2RAC/CSF2RB SIGNOR-C121 (+) to ESR1 P03372 (+) and to PIK3R1 P27986 (+)). There are also 9 additional transitions with $g > 0.001$. Complete data files for the matrix elements of matrix components (for all examples) are available at Frahm and Shepelyansky, 2019c.

It is convenient to present the interactions between proteins, generated by the matrix elements of the sum of two components $G_{ij} + G_{ij}^{quadr}$ from Figure 3, in the form of a network shown in Figure 4. To construct the network of effective friends, we select first five initial nodes which are placed on a (large) circle: the two nodes with injection and absorption (EGFR (+)/injection node, blue) and JAK2 (-)/absorption node, olive) and three other nodes with a rather top position in the $K_l$ ranking: JAK2 (+) (related to JAK2 (-) with $K_L = 1$, $i = 1$, red), GRB2 (+) (with $K_L = 2$, $i = 21$, green) and FES (+) (with $K_L = 3$, $i = 22$, cyan). For each of these five initial nodes we determine four friends by the criterion of largest matrix elements (in modulus) in the same column, i.e. corresponding to the four strongest links from the initial node to the potential friends. The friend nodes found in this way are added to the network and drawn on circles of medium size around their initial node (if they do not already belong to the initial set of 5 top nodes). The links from the initial nodes to their friends are drawn as thick black arrows. For each of the newly added nodes (level 1 friends) we continue to determine the four strongest friends (level 2 friends) which are drawn on small circles and added to the network (if there are not already present from a previous level). The corresponding links from level 1 friends to level 2 friends are drawn as thin red arrows.

Each node is marked by the index $i$ from the first column of Table 1. The colors of the nodes are essentially red for nodes with strong negative values of $P_i$ (corresponding to the index $i = 1, \ldots , 20$) and blue for nodes with strong positive values of $P_i$ for $i = 21, \ldots , 40$. Only for three of the initial nodes we choose different colors which are olive for JAK (-), green for GRB2 (+) and cyan for FES (+). This procedure generates the directed friendship network shown in Figure 4. The obtained network of Figure 4 has a rather clear separation between the two blocks related to EGFR (mainly blue nodes) and JAK2 (mainly red nodes). There is only one link of first level (black arrow) from the EGFR block (GRB2 (+)) to the JAK2 block (JAK2 (+)). Of course, there are other strong direct transitions from the EGFR block to the JAK2 block as described above, but these links are weaker than the 4 closest friends and therefore they do not appear in the network structure of Figure 4. However, we see that there are many links between the two blocks on the secondary level of red arrows.

The block of JAK2 (red nodes) is rather compact with only 6 nodes (one red node at $i = 20$) is more linked to the EGFR block. In contrast the EGFR block contains 15 (blue) nodes showing that this group of proteins is characterized by broader and more extensive interconnections. We think that such a network presentation provides a useful qualitative image of the effective interactions between the two groups of proteins.

Network figures, for this example and the other two examples discussed below, constructed in the same way using the other matrix components $G_{ij}$, $G_{ij}^{quadr}$ instead of $G_{ij}$ generate follower networks that are available at Frahm and Shepelyansky, 2019c.

**4.2 Magnetization of proteins of EGFR - JAK2 pathway**

In the Ising-PPI-network each protein is described by two components which can be considered as spin up or down state. The PageRank probability of a protein is given by the sum of probabilities of its two components with $P(j) = P_u(j) + P_d(j)$. It can be shown that due to the structure of the matrix transitions given by the matrices $\sigma_+ \sigma_- \sigma_0$ the sum of probabilities $P(j)$ for a given protein $j$ is the same as for the directed PPI network without doubling (see Supplementary Material). Thus the activation or inhibition links in the Ising-PPI network of doubled size only redistribute PageRank probability for a given protein between up and down components. The physical meaning of these up and down

---

*“signor” — 2019/8/29 — page 6 — #6*
component probabilities \( P_i \) and \( P_j \) is qualitatively related to the fact that on average the PageRank probability \( P \) of a node is proportional to the number of ingoing activation links and \( P_i \) is proportional to the number of ingoing inhibition links. Thus we can characterize each node by its normalized ingoing activation links and the number of ingoing links. Thus we can also study the magnetization of CheiRank probability of proteins.

We take all different 38 proteins present in Table 1 and consider their PageRank-CheiRank plane shown in Figure 5 and the colors of the square boxes presents the values of \( M(j) \) (see caption of Figure 5). The three proteins with the strongest positive magnetizations are PLECG1 P19174 (\( M = 0.8959 \)), GRB2 P06290 (\( M = 0.8719 \)) and with the strongest negative values are BCR P11274 (\( M = -0.7799 \)), PIK3CD O00329 (\( M = -0.7328 \)). In total there are only 5 proteins of Table 2 with negative magnetization values. We attribute this to the fact that the number of inhibition links is smaller than the number of activation ones. We think that the magnetization of proteins can provide new interesting information about the functionality of proteins.

![Mag., PPI, EGFR/JAK2](image)

**Fig. 5.** PageRank “magnetization” \( M(j) = (P^+(j) - P^-(j)) / (P^+(j) + P^-(j)) \) of proteins of Table 2 shown on the PageRank-CheiRank plane \((K, K^*)\) of local indices; here \( j \) represents a protein node in the initial single protein network and \( P^+(j), P^-(j) \) are the PageRank components of the Ising-PPI-network (see text). The values of the color bar correspond to \( M_i / \max |M| \) with \( \max |M| = 0.8967 \) being the maximal value of \(|M(j)|\) for the given group of proteins.

### 4.3 Examples of other protein pathways

We also consider two other protein pathways for injection (pumping) and absorption which we analyzed in the same way. Again we compute the vector \( Y_i = GW_i \) where \( W_i \) has only two non-zero components being 1 at the pumping node and -1 at the absorption node, we solve the inhomogeneous PageRank equation (1) to obtain the linear response vector \( P_i \) from which we determine a set of 40 nodes composed with 20 strongest negative and 20 strongest positive values. In order to ensure that the two initial injection and absorption nodes also belong to this subset we eventually replace the node at position 20 for strongest positive and/or negative values with the injection and/or absorption node respectively.

**Table 2.** Group of 38 nodes of the single protein network obtained from the group of Table 1 by removing the (+) and (−) attributes. \( K (K^*) \) represent the local rank indices obtained from the PageRank (CheiRank) ordering of the single protein network. The index \( i \) is the same as in Table 1 where the two values \( i = 2 \) and \( i = 26 \) do not appear here since they correspond to the two nodes where both components (+) and (−) are present in Table 1.
Here we only show and discuss the list of the obtained subsets and the effective network schemes corresponding to Table 1 and Figure 4 for these two examples while Tables and Figures analogous to Table 2, Figures 1, 2, 3, 5 are given in the Supplementary Material.

First we discuss the case of injection at MAP2K1 Q02750 (+) and absorption at EGF(P) P00533 (-). The protein MAP2K1 is a member of the dual-specificity protein kinase family that acts as an integration point for multiple biochemical signals. There is no direct link between MAP2K1 and EGFR. The global PageRank indices of these two nodes are \( K = 84 \) (PageRank probability \( P(84) = 0.0009794 \) ) for MAP2K1 Q02750 (+) and \( K = 136 \) (PageRank probability \( P(136) = 0.0007817 \) ) for EGF(P) P00533 (-). The subset of most sensitive proteins obtained from the LIBROMAX algorithm for this protein pair is given in Table 3. These proteins are different from those of Table 1. We note now that the injection and absorption proteins have lower positions in the rank indices \( K_L \) and \( i \) of Table 3. We attribute this somehow unexpected result of the \( P_I \) ranking to rather nontrivial vortex flows on the Ising-PPI-network.

The friendship network for this case is shown in Figure 6 (the construction method is the same as Figure 4). The 5 proteins of the initial large circle are EGF (+) (olive), FES (+) (red), MAP2K1 (+) (cyan), MAPK14 (-) (green), CEBPA (+) (blue). In this network we find a number of strong indirect links from the block of MAP2K1 Q02750 (+) (blue nodes) to EGF(P) P00533 (-) (red nodes) for which there is no direct link (e.g. from \( i = 21 \) to \( i = 14 \) proteins of Table 3). In the opposite direction from red to blue nodes there are only two strong direct matrix elements of \( G_{ij} \) being from PI3K SIGNOR-C156 (+) \( i = 13 \) to IRS1 P15560 (-) \( i = 21 \) with \( g = 0.08501 \) and from STAT3 P40763 (+) \( i = 18 \) to CASP3 P42574 (+) \( i = 32 \) with \( g = 0.03543 \) with all other elements being below \( 1.8 \times 10^{-5} \). However, in this direction there are 9 new indirect links with elements \( g > 0.01 \) and 20 with \( g > 0.005 \). This results in a rather dense network with many links shown in Figure 6. From the network structure we see that the proteins \( i = 25, 40 \) of the blue block are more closely related with proteins of the red block and inversely the proteins \( i = 10, 18, 20 \) of the red block are more closely related with proteins of the blue block.

Fig. 6. Same as Figure 4 but for the pathway of Table 3.

| \( i \) | \( K_L \) | \( K \) | Node name |
|---|---|---|---|
| 1 | 15 | 172 | FES P07332 (+) |
| 2 | 16 | 29 | GRB2 P62993 (+) |
| 3 | 17 | 90 | EGF(P) P00533 (+) |
| 4 | 18 | 3 | PIK3CD P00329 (+) |
| 5 | 19 | 126 | CBL P28681 (+) |
| 6 | 20 | 648 | EGR P15311 (+) |
| 7 | 21 | 136 | EGF(P) P00533 (+) |
| 8 | 22 | 38 | PTK2 P05397 (+) |
| 9 | 23 | 30 | JAK2 Q60674 (+) |
| 10 | 24 | 57 | STAT1 P22242 (+) |
| 11 | 25 | 46 | GAB1 P31480 (+) |
| 12 | 26 | 424 | BCR P12174 (+) |
| 13 | 27 | 9 | PI3K SIGNOR-C156 (+) |
| 14 | 28 | 326 | P40763 (+) |
| 15 | 29 | 746 | SHC1 P92529 (+) |
| 16 | 30 | 2398 | VAV2 P25735 (+) |
| 17 | 31 | 299 | ERK2 P04626 (+) |
| 18 | 32 | 15 | STAT3 P40763 (+) |
| 19 | 33 | 388 | ERK1/2 SIGNOR-PF1 (+) |
| 20 | 34 | 40 | JAK1 P35458 (+) |
| 21 | 35 | 125 | CEBPA P49715 (+) |
| 22 | 36 | 144 | MAPK14 Q16539 (-) |
| 23 | 37 | 5 | GSK3B P49841 (+) |
| 24 | 38 | 543 | TAL1 P17542 (+) |
| 25 | 39 | 74 | CASP9 P55211 (+) |
| 26 | 40 | 16 | PPIK P27371 (+) |
| 27 | 41 | 1491 | ARRB2 P31211 (+) |
| 28 | 42 | 156 | MAP2K1 Q02750 (+) |
| 29 | 43 | 84 | MAP2K1 Q02750 (+) |
| 30 | 44 | 246 | MAPK1 P28452 (+) |
| 31 | 45 | 80 | IRS1 P35568 (-) |
| 32 | 46 | 1 | CASP3 P42574 (-) |
| 33 | 47 | 523 | KIF3A Q99496 (+) |
| 34 | 48 | 7 | ERK1/2 SIGNOR-PF1 (+) |
| 35 | 49 | 258 | ERG P11308 (+) |
| 36 | 50 | 106 | MEK2 P06413 (+) |
| 37 | 51 | 826 | ANGPT2 Q15123 (+) |
| 38 | 52 | 290 | TEK Q02673 (+) |
| 39 | 53 | 181 | CPT1B Q02673 (+) |
| 40 | 54 | 86 | JUN P05412 (+) |

As a further example we also briefly discuss the pathway generated by injection at EGF(P) P00533 (+) and absorption at PI3KCA P42336 (-). These two proteins are conventional bio markers of lung cancer (see e.g. Zamay et al., 2017). The global PageRank indices of these two nodes are \( K = 90 \) (PageRank probability \( P(90) = 0.0009653 \) ) for EGF(P) P00533 (+) and \( K = 1604 \) (PageRank probability \( P(1604) = 0.0001366 \) ) for PI3KCA P42336 (-).

The most sensitive proteins obtained by the LIBROMAX algorithm, are shown in Table 4. However, now the absorption node PI3KCA P42336 (-) has a very low value (in modulus) of \( L_I \) \( P_L = -4.59 \times 10^{-5} \), \( K_L = 2860 \) and does initially not belong to the group of nodes with 20 top strongest negative values. Therefore we replace the node AKT3 Q99243 (+) \( K_L = 70 \) which was initially selected for \( i = 20 \) by the
absorption node PIK3CA P42336 (+). Furthermore, also its (+) component PIK3CA P42336 (+) ($P_i = -0.004546$ and $K_{iK} = 1.38$) does not appear in Table 4 showing that the influence of EGFR P00533 (+) on the protein PIK3CA P42336 is rather low.

The friendship network structure of shown in Figure 7 shows a clear separation between the two blocks of positive (blue) and negative (red) $P_i$ values. However, some proteins of one block happen to be closer to proteins of the other block (e.g. proteins $i = 10, 14$ from the red block are closer to the blue block and blue block protein $i = 29$ is closer to the proteins of the red block). We also note that concerning the links from the blue to the red block there are 9 significant direct transitions (matrix elements of $G_{rr}$ larger than $0.01$) and 35 significant indirect and direct transitions (matrix elements of $G_{rr} + G_{qr}^{(nd)}$ larger than $0.01$). For the opposite direction of transitions from the red to the blue block the increase is less significant but still there are new transitions due to indirect pathways (2 significant transitions for $G_{rr}$ and 3 for $G_{rr} + G_{qr}^{(nd)}$). The significance of indirect transitions is also well visible in the friendship network of Figure 7 with many red arrows between the two blocks.

The same results for the original list, where the node AKT3 P91243 (+) at position $i = 20$ has not been replaced by PIK3CA P42336 (+), are available at Frahm and Shepelyansky, 2019c.

### 5 Conclusion

In this work we describe the properties of Google matrix analysis of the bi-functional SIGNOR PPI network from Perfetto, L. et al., 2016. The main elements of this approach are: the activation and inhibition actions of proteins on each other are described by Ising spin matrix transitions between the protein components in the doubled size Ising-PPI network; the recently developed LIRGOMAX (Frahm and Shepelyansky, 2019a) algorithm determines the most sensitive proteins on the pathway between two selected proteins with probability injection (pumping) at one protein and absorption at another protein, the set of most sensitive proteins are analyzed by the REGOMAX algorithm which treats efficiently all direct and indirect interactions in this subset taking into account all their effective interactions through the global PPI network. We illustrated the efficiency of this approach on several examples of two selected proteins. The obtained results show the efficiency of the LIRGOMAX and REGOMAX algorithms. We also show that the bi-functionality of protein-protein interactions leads to a certain effective magnetization of proteins which characterizes their dominant action on the global PPI network.

The executive codes and reduced Google matrix data are open and publicly available at Frahm and Shepelyansky, 2019c and interested researchers can easily study any example of a pathway between any pair of proteins from the SIGNOR network.

The described LIRGOMAX and REGOMAX algorithms can be applied also to other type of biological networks (e.g. metabolic networks discussed by Frainay et al., 2019).
We mention that the described Google matrix algorithms have been tested for networks with 5 million nodes and thus they can operate efficiently on other PPI networks of significantly larger size (e.g. MetaCore network which has several tens of thousands of nodes and about 2 million links). Thus we expect that the Google matrix approach, or in short Geoglomics, will find broad applications for the analysis of protein-protein interactions.

Funding
This work was supported in part by the Programme Investissements d’Avenir ANR-11-IDEX-0002-02, reference ANR-10-LABX-0037-NEXT (project THETRACOM); it was granted access to the HPC resources of CALMIP (Toulouse) under the allocation 2019-P0110.

References
Bethune, G. et al. (2010) Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. J Thorac Dis. 2, 48.

Brin, S. and Page, L. (1998) The anatomy of a large-scale hyperertextual Web search engine. Computer Networks and ISDN Systems 30, 107-117.

Chepelianskii, A.D. (2010) Towards physical laws for software architecture arXiv:1003.5455 [cs.SE]

Cogado, C. et al. (2019) World influence and interactions of universities from Wikipedia networks. Eur. Phys. J. B 92, 3.

Cogado, C. et al. (2019) Influence of petroleum and gas trade on EU economies from the reduced Google matrix analysis of UN COMTRADE data. Eur. Phys. J. B 92, 171.

Dimitrakopoulos, C. et al. (2018) Network-based integration of multi-omics data for prioritizing cancer genes. Bioinformatics 34(14), 2441-2448.

Dorogovtsev, S. (2010) Lectures on complex networks. Oxford University Press, Oxford.

Ermann, L. et al. (2015) Google matrix analysis of directed networks. Rev. Mod. Phys. 87, 1261-1310.

Fabregat, A. et al. (2018) The Reactome pathway knowledgebase. Nucleic Acids Res. 46 Database issue, D649-D655.

Frahm, K.M. et al. (2016) Wikipedia mining of hidden links between political leaders. Eur. Phys. J. B 89, 269.

Frahm, K.M. and Shepelyansky, D.L. (2019)a. Linear response theory for Google matrix. arXiv:1908.06924 [cs.SI]

Frahm, K.M. and Shepelyansky, D.L. (2019)b. Ising-Pagerank model of opinion formation on social networks. Physica A 526, 123069.

Frahm, K.M. and Shepelyansky, D.L. (2019)c. Google matrix analysis for SIGNOR network. Available at http://www.quantware.ups-tlse.fr/QWLIB/google4signornet/ Accessed August (2019).

Frainay, C. et al. (2019) MetaBisRank: network-based recommendation system to interpret and enrich metabolomics results. Bioinformatics 35(2), 274-283.

Invergo, B.M. and Beltrao, P. (2018) Reconstructing phosphorylation signalling networks from quantitative phosphoproteomic data. Essays Biochemistry 62, 525-534.

Kanhaiya, K. et al. (2017) KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45 Database issue, D353-D361.

Kostasuringa, K. et al. (2017) Controlling directed protein interaction networks in cancer. Sci. Reports 7, 10027.

Lages, J. et al. (2018) Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks. PLoS ONE 13(1), e0190812.

Langville, A.M. and Meyer, C.D. (2006) Google’s PageRank and beyond: the science of search engine rankings, Princeton University Press, Princeton.

Liberon, A. et al. (2011) Molecular signatures database (MSigDB) - 3.9. Bioinformatics 27, 1739-1740.

Lin, X.-K. et al. (2017) Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry. Nature Biotechnol. 35, 164-172.

MetaCore (Clarivate Analytics). Available at https://portal.genego.com/. Accessed August (2019).

Perfetto, L. et al. (2016) SIGNOR: a database of causal relationships between biological entities. Nucleic Acids Res. 44 Database issue, D548-D554.

Sacco, F. et al. (2016) Deep proteomics of breast cancer cells reveals that metformin rewires signaling networks away from a pro-growth state. Cell Systems 2, 159-171.

Splender, D.N. et al. (2018) The Reactome pathway knowledgebase. Nucleic Acids Res. 46 Database issue, D661-D667.

Zamay, T.N. et al. (2017) Current and prospective protein biomarkers of lung cancer. Cancers MDPI 9, 155.
SUPPLEMENTARY MATERIAL
for “Google matrix analysis of bi-functional SIGNOR network of protein-protein interactions”
by K.M. Frahm and D.L. Shepelyansky

S.1 Statistical properties of the SIGNOR protein-protein interactions network PPI

Using the Signor database a network of \( N = 4341 \) proteins with \( N_I = 12547 \) interactions was created. In a first version, called the “single protein network”, the links do not contain the information if the interaction corresponds to activation, inhibition or is neutral/unknown. As usual we first construct an adjacency matrix with entries \( A_{ij} = 1 \) if there is a link from node \( j \rightarrow i \) and \( A_{ij} = 0 \) if there is no such link. However, in certain rare cases there are multiple types of links between two proteins (e.g. activation and inhibition) in which case we choose \( A_{ij} \) being a multiplicity factor of 2 or 3 (instead of the usual entry 1). Once the adjacency matrix is fixed the Google matrix of this (single) protein network is constructed in the usual way: column sum normalization, taking into account the effect of dangling nodes (nodes with no outgoing link) by replacing each zero column by a uniform column with entries 1/\( N \) and with the application of the standard damping factor \( \alpha = 0.85 \).

In Fig. S.1 we show the PageRank \( P \) (CheiRank \( P' \)) for this single network versus the corresponding rank index \( K (K^*) \) showing a typical decay (roughly) comparable to a power law \( P \sim 1/K^\beta \) \( (P' \sim 1/(K^*)^\beta) \) with \( \beta \approx 0.7 \) \((0.8) \) for \( K \gtrsim 100 \) \((K^* \gtrsim 10) \). Fig. S.2 shows the density of nodes in the PageRank-CheiRank plane \((K, K^*)\) and the positions of the subgroup of nodes corresponding to Table 2 for this network.

To take into account the information about the nature of the links we use the approach of the Ising-PageRank to construct a larger network where each node is doubled with two labels (+) and (−). To construct the doubled “Ising” network of proteins each unit entry of the initial adjacency matrix is replaced by \( 2 \times 2 \) matrices which are:

\[
\sigma_+ = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad \sigma_0 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}
\]  

(S.1)

where \( \sigma_+ \) applies to “activation”, \( \sigma_- \) to “inhibition” and \( \sigma_0 \) to “neutral” or “unknown”. For the rare cases with multiple types of links between two proteins we use the sum of the corresponding \( \sigma \) matrices which increases the weight of the adjacency matrix elements. After this the corresponding Google matrix is constructed in the usual way. The doubled Ising protein network corresponds to \( N_I = 8682 \) nodes and \( N_{I,2} = 27266 \) links (according to the non-zero entries of the used \( \sigma \) matrices).

Now the PageRank vector (of this doubles Ising network) has components \( P_+(j) \) and \( P_-(j) \). Due to the particular structure of the \( \sigma \) matrices (S.1) one can show analytically the exact identity \( P(j) = P_+(j) + P_-(j) \) where \( P(j) \) is the PageRank of the initial single protein network. For this we have to replace in Eq. (4) of (Frahm and Shepelyansky (2019)b) the value \( n_i \) by \( n_{+i} \) with \( n_{+ij} = 1, 0, 1/2 \) for the matrix \( \sigma_+, \sigma_- \) or \( \sigma_0 \) respectively. The additional dependence of \( n_{+ij} \) on \( j \) takes into account that the choice of the \( \sigma \) matrix may be different for each link (and is not identical inside each row as it was the case for the model used in (Frahm and Shepelyansky (2019)b). Then the analytical argument of this work also applies in exactly the same way to the case of the doubled Ising protein network. We have also numerically verified that the identity \( P(j) = P_+(j) + P_-(j) \) holds up to numerical precision (\( \sim 10^{-13} \)).

As in (Frahm and Shepelyansky (2019)b) we introduce the PageRank “magnetization” by:

\[
M(j) = \frac{P_+(j) - P_-(j)}{P_+(j) + P_-(j)}
\]  

(S.2)

for a node \( j \). The dependence of \( M(j) \) on nodes is shown in Fig. S.3 for the whole network and in Fig. 5 for the subgroup of nodes of Table 2.
Fig. S.1. PageRank $P(K)$ and CheiRank $P^*(K^*)$ for the single protein network.

Fig. S.2. Density of nodes $W(K, K^*)$ of the single protein network on PageRank-CheiRank plane $(K, K^*)$ averaged over $100 \times 100$ logarithmically equidistant grids for $0 \leq \ln K \leq \ln N$. The density averaged over all nodes inside each cell of the grid, the normalization condition is $\sum_{K, K^*} W(K, K^*) = 1$. The color bar of Fig. 2 applies (for positive values) and its values correspond to $(W/\max W)^{1/4}$. In order to increase the visibility large density values have been reduced to (saturated at) $1/16$ of the actual maximum density. The x-axis corresponds to $\ln K$ and the y-axis to $\ln K^*$. The white crosses show the positions of the 38 nodes of Table 2 and in Fig. 5.

Fig. S.3. PageRank "magnetization" $M(j) = (P_+(j) - P_-(j))/(P_+(j) + P_-(j))$ in the Ising-PPI-network; here $j$ is the node index and $K(j)$ is the PageRank index of the initial SIGNOR network (without node doubling).
S.2 Pathway from MAP2K1 Q02750 (+) to EGFR P00533 (-)

Here we present additional figures and table for this pathway discussed in subsection 4.3. Table S.1 gives the proteins (extracted from Table 3) for which the magnetization $M$ is presented in Figure S.7.

Table S.1. Same as in Table 2 but for injection (pumping) at MAP2K1 Q02750 (+) and absorption at EGFR P00533 (-). The index $i$ is the same as in Table 3 where two values do not appear here since they correspond to the two nodes where both components (+) and (−) are present in Table 3.

| $K$ | $K^*$ | $i$ | Node name               |
|-----|-------|-----|-------------------------|
| 1   | 23    | 26  | PPARG P37231             |
| 2   | 9     | 32  | CASP3 P42574             |
| 3   | 20    | 25  | CASP9 P55211             |
| 4   | 37    | 4   | PIK3CD O00329            |
| 5   | 20    | 13  | PI3K SIGNOR-C156         |
| 6   | 8     | 18  | STAT3 P40763             |
| 7   | 5     | 34  | ERK1/2 SIGNOR-PF1         |
| 8   | 11    | 40  | JUN P05412               |
| 9   | 4     | 22  | MAPK14 Q16539            |
| 10  | 17    | 36  | MEF2C Q06413             |
| 11  | 9     | 9   | JAK2 Q04674              |
| 12  | 3     | 23  | GSK3B P49941             |
| 13  | 6     | 3   | EGFR P00533              |
| 14  | 18    | 21  | CEBPA P49715             |
| 15  | 7     | 14  | PLCG1 P19174             |
| 16  | 19    | 2   | GRB2 P62993              |
| 17  | 21    | 8   | PTK2 Q06397              |
| 18  | 22    | 20  | JAK1 P23486              |
| 19  | 2     | 28  | MAPK1 P28482             |
| 20  | 31    | 31  | IRS1 P15568              |
| 21  | 14    | 10  | STAT1 P42224              |
| 22  | 1     | 30  | MAPK1 P28482             |
| 23  | 12    | 29  | MAP2K1 Q02750            |
| 24  | 15    | 5   | CBL P22681               |
| 25  | 17    | 17  | ERBB2 P04626             |
| 26  | 27    | 1   | FES P07332               |
| 27  | 39    | 39  | CPT1B Q02523             |
| 28  | 26    | 38  | TEK Q12765               |
| 29  | 25    | 24  | TAL1 P17542              |
| 30  | 33    | 11  | GAB1 Q13480              |
| 31  | 18    | 40  | SORL1 P12174             |
| 32  | 6     | 3   | EZR P13311               |
| 33  | 38    | 33  | KIF3A Q09496             |
| 34  | 16    | 35  | ERG P11388               |
| 35  | 24    | 19  | ERBB3 P21060             |
| 36  | 15    | 35  | SHC1 Q02529              |
| 37  | 38    | 37  | ANO2 P12312              |
| 38  | 34    | 27  | ARRB2 P32121             |
| 39  | 16    | 32  | VAV2 P52735              |

Fig. S.7. Same as in Fig. 5 but for the pathway from MAP2K1 Q02750 (+) to EGFR P00533 (-) with proteins from Table S.1; the maximal magnetization used in the color bar normalization is $M_{\text{max}} = 0.963$. 
Here we present additional figures and table for this pathway discussed in subsection 4.3. Table S.2 gives the proteins (extracted from Table 4) for which the magnetization $M$ is presented in Figure S.11.

**Table S.2.** Same as in Table 2 but for injection (pumping) at EGFR P00533 (+) and absorption at PIK3CA P42336 (-). The index $\iota$ is the same as in Table 4 where two values do not appear here since they correspond to the two nodes where both components (+) and (−) are present in Table 4.

| $K^*$ | $K$ | $\iota$ | Node name |
|-------|-----|---------|-----------|
| 1     | 2   | 2       | AKT SIGNOR-PF24 |
| 2     | 1   | 3       | AKT1 P31749 |
| 3     | 35  | 24      | PIK3CD O00329 |
| 4     | 16  | 9       | PI3K SIGNOR-C156 |
| 5     | 3   | 38      | ERK1/2 SIGNOR-PF1 |
| 6     | 8   | 6       | Pdum(3,4,5)(P) CID:2475492 |
| 7     | 14  | 17      | mTORC1 SIGNOR-C3 |
| 8     | 6   | 27      | JAK2 P06674 |
| 9     | 13  | 8       | RAC1 P65000 |
| 10    | 4   | 23      | EGFR P00533 |
| 11    | 7   | 5       | MTOR P42336 |
| 12    | 15  | 22      | GRB2 P62903 |
| 13    | 28  | 19      | BAD P09294 |
| 14    | 25  | 14      | PIK3CB P42338 |
| 15    | 17  | 20      | PIK3CA P42336 |
| 16    | 18  | 35      | JAK1 P23458 |
| 17    | 29  | 7       | IRS1 P35586 |
| 18    | 10  | 28      | STAT1 P42224 |
| 19    | 12  | 32      | SRC1 P29533 |
| 20    | 5   | 4       | AKT2 P31751 |
| 21    | 9   | 26      | CBL P23861 |
| 22    | 11  | 36      | ERBB2 P04606 |
| 23    | 23  | 21      | FES P07332 |
| 24    | 20  | 1       | BTK Q06187 |
| 25    | 38  | 40      | STAT1/STAT3 SIGNOR-C118 |
| 26    | 31  | 30      | GAB1 Q13480 |
| 27    | 24  | 31      | BCR P1274 |
| 28    | 27  | 29      | EGR1 P1331 |
| 29    | 34  | 39      | JAK1/STAT3/STAT5 SIGNOR-C120 |
| 30    | 22  | 37      | ERBB3 P21860 |
| 31    | 33  | 33      | SRC3 Q52520 |
| 32    | 32  | 12      | BMX P51613 |
| 33    | 26  | 11      | TEC P42680 |
| 34    | 37  | 16      | PLCG2 P16885 |
| 35    | 21  | 18      | GATA P37347 |
| 36    | 19  | 13      | ITK Q08919 |
| 37    | 36  | 15      | DAPPI Q049919 |
| 38    | 30  | 34      | VAV2 P92735 |

**Fig. S.8.** Same as in Fig. 1 but for the pathway from EGFR P00533 (+) to PIK3CA P42336 (-).

**Fig. S.9.** Same as in Fig. 2 but for the pathway from EGFR P00533 (+) to PIK3CA P42336 (-).

**Fig. S.10.** Same as in Fig. 3 but for the pathway from EGFR P00533 (+) to PIK3CA P42336 (-).

**Fig. S.11.** Same as in Fig. 5 but for the pathway from EGFR P00533 (+) to PIK3CA P42336 (-) with proteins from Table S.2; the maximal magnetization used in the color bar normalization is $M_{\text{max}} = 0.961$. 