Prognostic interaction patterns in diabetes mellitus II: A random-matrix-theory relation

Aparna Rai,[1] Amit Kumar Pawar,[1] and Sarika Jalan[1,2,*]

[1]*Centre for Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Indore 452017, India*

[2]*Complex Systems Lab, Discipline of Physics, Indian Institute of Technology Indore, Indore 452017, India*

(Received 7 March 2015; revised manuscript received 28 May 2015; published 11 August 2015)

We analyze protein-protein interactions in diabetes mellitus II and its normal counterpart under the combined framework of random matrix theory and network biology. This disease is the fifth-leading cause of death in high-income countries and an epidemic in developing countries, affecting around 8% of the total adult population in the world. Treatment at the advanced stage is difficult and challenging, making early detection a high priority in the cure of the disease. Our investigation reveals specific structural patterns important for the occurrence of the disease. In addition to the structural parameters, the spectral properties reveal the top contributing nodes from localized eigenvectors, which turn out to be significant for the occurrence of the disease. Our analysis is time-efficient and cost-effective, bringing a new horizon in the field of medicine by highlighting major pathways involved in the disease. The analysis provides a direction for the development of novel drugs and therapies in curing the disease by targeting specific interaction patterns instead of a single protein.

DOI: 10.1103/PhysRevE.92.022806 PACS number(s): 89.75.Hc, 82.39.Rt

I. INTRODUCTION

Diabetes mellitus II (DM-II) is a metabolic disorder which is known to be a complex and multifactorial disease [1–4]. The World Health Organization statistics reveal that DM-II had affected nearly 325 million people up to 2013, of which 46% of cases are undiagnosed, and the disease shows a rising trend worldwide [5–7], making it the fifth-leading cause of death in high-income countries and an epidemic in developing countries [8,9]. It is one of the most important noncommunicable diseases of the 21st century, in terms of both mortality and morbidity [10]. Indecent lifestyle, increased urbanization, unhealthy behaviors, and obesity are some of the major causes behind the disease [11,12]. According to a recent study, individuals with this complex disease spend on average more than $85,000 on treatment and its complications over their entire lifetimes, anda trillion dollars is spent on health and diseases worldwide [13,14]. But the research pertaining to the development of novel treatments and methods of cure is exhaustive [8]. Even with the enormous investment in pharmacological research and clinical trials in the past decades, there has not been a proportionate advancement in clinical results. All these factors demand a new perspective in disease research, which we provide here using the combined framework of random matrix theory (RMT) and network theory. This framework predicts important structural patterns crucial for the disease and is very time-efficient and cost-effective.

A few network studies done on DM-II reveal some important pathways associated with insulin signaling using microarray data sets [15–18]. These studies are based on a handful of disease-associated proteins, while our investigation for the first time performs an extensive analysis of all the pancreatic cell proteins. We investigate DM-II at the proteomic level using network biology along with the eigenvector localization under the RMT framework. This mathematical theory was formulated six decades ago in order to understand the spectra of compound nuclei [19]. Later on, the theory has shown remarkable success in clarifying various complex systems ranging from quantum chaos to galaxy, stock market, power grid, etc. [20–22]. In recent years, this theory has shown credibility in studies of various biological systems like gene coexpression networks, protein-protein interaction (PPI) networks, understanding of the genetic variance among species, etc. [23–25]. In the current work, we construct the protein-protein interaction network of DM-II by analyzing normal and disease states of pancreatic cells and investigating their structural properties, which are further compared with the properties of random networks. The analysis reveals specific structural patterns, as well as nodes contributing significantly to the most localized eigenvector, that are crucial for the occurrence of the disease.

II. MATERIALS AND METHODS

A. Data assimilation and network construction

In the PPI network of DM-II, nodes are the proteins and edges denote the interactions between these proteins. After diligent and enormous effort, we collected the protein interaction data from various literature and bioinformatic sources. To preserve the authenticity of the data, we take into account only proteins which are reviewed and cited. We consider two widely used bioinformatic databases, namely, GenBank from NCBI [26] and UniProtKB, consisting of data available from other resources like the European Bioinformatic Institute, the Swiss Institute of Bioinformatics, and the Protein Information Resource [27]. The protein data from the above resources are very few in number. To add more information, we take highly studied DM-II cell lines whose protein expression data are known. Here, we use the protein expression data of EndoC-βH1 cells in the disease data set [28]. Since there are very few human cell lines available, we take the mouse and rat cell lines, namely, MIN and INS-1, the proteins of which have already been proved to behave similarly in humans [29,30]. After collecting the proteins for both data sets, their interacting partners are downloaded from the STRING database [31]. We

*sarika@iiti.ac.in

take into account only the largest subnetworks from the normal as well as from the disease data and investigate their structural and spectral properties.

B. Structural measures

The adjacency matrix of the PPI networks can be represented as follows:

$$A_{ij} = \begin{cases} 1 & \text{if} \quad i \sim j, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The degree of a node (k_i) is defined as the number of neighbors the node has ($k_i = \sum_j A_{ij}$). The degree distribution $p(k)$, revealing the fraction of vertices of degree k, is known as the fingerprint of the network [32]. Another important parameter is the (CC) of the node (i), which is defined as the ratio of the number of connections a particular node has to the possible number of connections that it can have [33]. The average CC of the network characterizes the overall tendency of nodes to form clusters or groups. Further, the betweenness centrality of a node is defined as the fraction of shortest paths between node pairs that passes through the said node of interest [34]. Another important parameter is the diameter of the network, which measures the longest of the shortest paths between all the pairs of nodes [32].

C. Spectral techniques

Eigenvalues of the adjacency matrix [Eq. (1)], denoted $\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_N$, are referred to as network spectra. Further, we use the IPR to analyze localization properties of the eigenvectors. The IPR of the kth eigenvector E^k, with its lth component denoted E^k_l, can be defined as

$$I^k = \frac{\sum_{l=1}^{N} \left[E^k_l\right]^4}{\left(\sum_{l=1}^{N} \left[E^k_l\right]^2\right)^2} \tag{2}$$

which shows two limiting values: (i) a vector with identical components $E^k_l \equiv 1/\sqrt{N}$ has $I^k = 1/N$, whereas (ii) a vector with one component $E^k_1 = 1$, and the remainder being 0, has $I^k = 1$. Thus, the IPR quantifies the reciprocal of the number of eigenvector components that contribute significantly. An eigenvector whose components follow a Porter-Thomas distribution yields $I^k = 3/N$ [35], and those which deviate from this value provide system-dependent information [36]. We further calculate the average value of the IPR, in order to measure the overall localization of the network calculated as $\langle \text{IPR} \rangle = \sum_{k=1}^{N} I^k/N$.

III. RESULTS: STRUCTURAL PROPERTIES OF NORMAL AND DISEASE NETWORKS

The normal data set has 4613 nodes (proteins) with 26,035 connections (interactions) among them, while the disease data set comprises 1100 nodes and 5578 connections, demonstrating that the disease data consist of fewer proteins and connections than the normal data. One of the possible reasons behind this is the limited availability of data for the disease state as discussed in Sec. II A. Another possible reason is that in the disease state, many pathways are silenced or levels of protein expression are altered, leading to fewer proteins [37]. The normal and the disease data sets lead to many connected clusters or subnetworks, whose sizes are summarized in the second column in Table I. Here, we perform statistical analysis of the two largest clusters for both the normal and the disease data sets. The third column in Table I conveys that even with fewer nodes, the disease subnetworks possess almost the same $\langle k \rangle$ as the normal ones. Further, the disease subnetworks exhibit relatively smaller diameters compared to the normal ones. Since a small diameter facilitates rapid communication [33], up-regulation and down-regulation of pathways may be one of the reasons behind the faster signaling leading to the disease state. Further, we calculate the average CC of all the subnetworks along with the total number of nodes having CC = 1 ($N_{\text{clus}=1}$) as listed in Table I. The percentages of nodes having this property for the normal and all the disease subnetworks, except $D2$, are approximately same. For $D2$, the number of nodes having CC = 1 is slightly higher compared to that in other subnetworks, indicating the presence of more clique structures in this subnetwork. The importance of cliques and relevance of this structure for the disease become more clear in Sec. III A. In addition, there is a high degeneracy at -1 and 0, denoted λ_{-1} and λ_0 (Table I), for all the subnetworks. The occurrence of -1 eigenvalues indicates the existence of complete subgraphs or structures close to complete subgraphs [38]. The number of zero degeneracies has a direct relation to the number of exact and partial duplicates, as both of them contribute to lowering the rank of the corresponding matrix [39]. A large number of λ_0 in real networks indicates a high number of exact and partial duplicates. Note that an isolated node also contributes to lowering the rank of the matrix and hence adding to the zero degeneracy. This trivial situation does not arise here as we consider connected clusters. The values of λ_0 for all the real networks are much higher than those for their corresponding model networks as reported in Tables I and II. We return to this point in Sec. V A when discussing nodes which are localized as well as functionally important.

The degree distributions $p(k)$ of both the normal and the disease subnetworks follow power law [40], indicating that few nodes have a very high degree. Earlier structural analyses of DM-II networks also emphasize that proteins having a large number of interactions are functionally important [41].

TABLE I. Detailed parameters for normal and disease subnetworks. The columns represent the total number of proteins (nodes) in the network, N, collected using various databases (described in the Method section), number of connections in the subnetworks N_C, average degree $\langle k \rangle$, diameter (D), average clustering coefficient $\langle \text{CC} \rangle$, number of nodes having CC = 1 in the whole network ($N_{\text{clus}=1}$), average inverse participation ratio $\langle \text{IPR} \rangle$, and number of degenerate eigenvalues in the whole network at -1 (λ_{-1}) and $0\lambda_0$ for the subnetworks of both the normal and the disease state.

Network	N	N_C	$\langle k \rangle$	D	$\langle \text{CC} \rangle$	$N_{\text{clus}=1}$	$\langle \text{IPR} \rangle$	λ_{-1}	λ_0
$N1$	2083	11017	10	17	0.35	6.1%	0.010	1.8%	3.1%
$N2$	1705	9888	11	12	0.33	6.2%	0.008	1.8%	3.5%
$D1$	656	3628	11	10	0.36	6.0%	0.015	1.06%	3.6%
$D2$	384	1882	10	9	0.46	7.8%	0.032	4.9%	7.2%

TABLE II. The corresponding Erdös-Rényi (ER), small world (SM), and configuration (conf) models for all the normal and disease data sets. The parameters show all the results as an ensemble average of 10 realizations of the ER, SM, and conf networks with the same N and $\langle k\rangle$ as for real networks.

Network	$\langle CC\rangle^{ER}$	$N^{ER}_{clus=1}$	λ^{ER}_{-1}	λ^{ER}_{0}	D^{ER}	$\langle CC\rangle^{SW}$	$N^{SW}_{clus=1}$	λ^{SW}_{-1}	λ^{SW}_{0}	D^{SW}	$\langle CC\rangle^{conf}$	$N^{conf}_{clus=1}$	λ^{conf}_{-1}	λ^{conf}_{0}	D^{conf}
$N1$	0.005	0	0	0	6	0.36	0	0	0	7	0.023	7	0	1.5%	7
$N2$	0.007	0	0	0	6	0.36	0	0	0	7	0.027	4	0	1.7%	7
$D1$	0.017	0	0	0	7	0.32	0	0	0	6	0.083	12	0	4.5%	7
$D2$	0.027	0	0	0	7	0.44	0	0	0	6	0.123	7	0	2.1%	6

Since the present investigation uses a much larger number of proteins and interactions compared to any other work done on DM-II, our data sets reveal new proteins that turn out to be on top in the degree distributions. The bigger data set used in our investigation arises for the following two reasons: first, we made an exhaustive literature search for almost all the authenticated databases to construct our data set; and second, new proteins and interactions have been revealed since 2008 [41], contributing additional interactions to the proteins in our data set.

High-degree nodes are involved in insulin production and are known to be regulated in the disease state, leading to lower production of insulin. A detailed functional description of high-degree nodes is deferred to the Supplementary Material [40]. Further, since the power-law degree distribution is known to confer robustness to the underlying network against random external perturbations as well as to instill complexity in the corresponding system [32], the similar behaviors exhibited by the disease and normal networks bring them into the same universality class as other complex biological systems [42]. It follows that the overall structural properties, such as diameter, CC, average degree, and degree distribution, of disease and normal subnetworks are similar, indicating complex interactions and fast dissemination of information in the pancreatic cell, which is in line with other complex systems. Important differences between the disease and the normal states are revealed when we analyze nodes forming complete subgraphs as well as those appearing in spectral analysis under the RMT framework.

A. Preserved structures through clique formation

All the subnetworks have a very high value of $\langle CC\rangle$ as reported in Table I. This implicates the presence of a high number of triangles or cliques of order 3 [33]. Cliques indicate preserved interactions in the networks and are believed to be conserved during evolution [43]. Further, these structures are also considered to be the building blocks of a network, making the underlying system more robust [44] and stable [45]. Therefore, we hope to reveal important information by analyzing these patterns in detail. We focus on nodes that have $CC = 1$ and are common in both the networks. It turns out that there are 34 nodes appearing commonly in the normal and disease data sets. We perform an extensive functional analysis [40] of these nodes to study the differences between the states, as these proteins enjoy special structural features in the networks. Of 34 proteins, 28 are directly responsible for the occurrence of DM-II, for instance in regulation of the insulin level in the blood [46], insulin resistance and aberrant glucose metabolism [47], and lipid metabolism and obesity [48]. Overall, nodes having $CC = 1$ and common in both the disease and the normal data sets play a significant role in the occurrence of the disease [40]. We further compare these structural properties to various other model networks, namely, random, small world (SW), and configuration networks, to examine the deviation of the disease and normal networks from the random controls.

IV. COMPARISON WITH VARIOUS RANDOM CONTROL NETWORKS

A. Erdös-Rényi (ER) random model

To create corresponding ER networks [49], we connect nodes with a fixed probability p, calculated as $\langle k\rangle/N$. Properties of corresponding ER random networks thus constructed are summarized in Table II. Owing to the very nature of construction of the ER networks, we get a very small CC ($CC \sim k/N$). In addition, the ER random network depicts no degeneracy at zero eigenvalue.

B. Small world network

Since all the disease networks have a significantly high value of $\langle CC\rangle$, indicating hidden importance of network interactions contributing to this property, we attempt to model this using a corresponding SW network algorithm (Table II). The SW networks are generated using the Watts-Strogatz algorithm [33]. Starting from a regular lattice, connections are rewired with probability p. The rewiring probability is chosen in such a manner that it leads to the desired CC matching the corresponding real networks. It turns out that while the model network contribution to the CC comes from almost all the nodes, for real-world networks only 75%–80% of nodes contribute to the CC of the network and the rest of the nodes do not have any interacting neighbors, i.e., they form starlike structures. Additionally, real networks have a significant number of nodes with $CC = 1$, a property which, again, is not found in the corresponding model network. As elaborated in the previous section (Sec. III A), nodes with $CC = 1$ have a special role in the occurrence of the disease, making the deviation of $CC = 1$ properties of the nodes from the model networks more crucial. Next, SW networks also do not exhibit any degeneracy at 0, which is very much expected from the way this network is constructed. In SW networks almost all the nodes form a ring structure, and hence this rules out two (or more) nodes sharing the same neighbors.

C. Configuration model

The ER random and SW networks, being the simplest models for capturing various properties of real systems, fail to model one of the very important aspects of many real-world networks including those considered here. The degree distribution of ER random networks follows a Poisson distribution, whereas those investigated here follow power-law behavior, as the configuration model preserves the exact degree sequence of a network [50]. The configuration model generates a random network with a given degree sequence of an array of size $m = \frac{1}{2}\sum_{i=1}^{N} k_i$ which has random connections among different elements. We generate 10 such realizations for a given degree sequence.

The CC and IPR of all the configuration model networks are all the same as for the ER random networks (Table II), as expected from the configuration model algorithm. Further, we note that all these model networks, while acting under one or the other constraint, are basically built upon random interactions, i.e., the pairs of nodes that should interact are chosen randomly out of many possible configurations. There are features which real-world systems are supposed to lack that in turn are captured in various spectral properties. For instance, none of the random controls are able to reproduce nodes having CC = 1 and high degeneracy at −1. But there is a high value of the zero degeneracy. The occurrence of the zero degeneracy for the configuration model is not surprising owing to the power-law nature of the degree distribution which is known to contribute into zero degeneracy [39].

It follows that while networks corresponding to various random models manifest diameters different from those of the real networks, the SW networks model the high-CC property and the configuration model captures the high-zero-degeneracy property possessed by the real networks. In the next sections, we delve into understanding the localization properties of the four subnetworks.

V. LOCALIZATION BEHAVIOR OF THE NORMAL AND DISEASE NETWORKS

We calculate the eigenvector localization using the IPR [Eq. (2)], and based on the localization behavior, we can divide the individual spectrum into two components. The first component follows RMT predictions of the Porter-Thomas distribution of Gaussian Orthogonal Ensemble statistics [51], and the other one deviates from this universality and indicates localization (Fig. 1).

The very distinct interpretation of this behavior of universal and nonuniversal (deviation from universality) components is

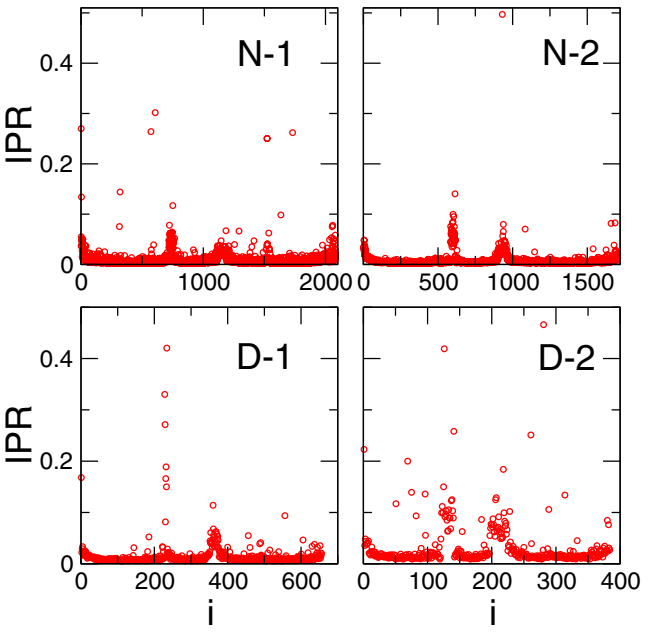

FIG. 1. (Color online) Eigenvector localization for both the normal and the disease subnetworks. The IPR of both the normal ($N1$ and $N2$) and the disease ($D1$ and $D2$) networks, clearly reflecting three regions: (i) a degenerate part in the middle, (ii) a large nondegenerate part which follows Gaussian Orthogonal Ensemble statistics of the RMT, and (iii) a nondegenerate part both at the end and near the zero eigenvalues which deviate from the RMT.

that the underlying system has random interactions leading to the universal part of the spectra, as well as nonrandom, pattern-specific interactions contributing to the part of the spectra deviating from universality. Further, the average IPR calculated for the normal and disease subnetworks reveals that the normal subnetworks are less localized than the subnetworks of the disease networks (Table I), indicating that the normal subnetworks are more random than the disease subnetworks. The fact that the disease state has a reduced number of connections compared to the normal state suggests that there are some pathways or proteins which are silenced, resulting in hampered interactions, perhaps due to mutation leading to the disease state. This interpretation, combined with the result that the ⟨IPR⟩ of the disease is higher than the normal, indicates that these hampered pathways may correspond to or should be treated as random pathways whose removal results in diminished randomness in the disease state. This result can

TABLE III. Top contributing nodes (TCNs) of the disease-1 ($D1$) subnetwork. The top localized eigenvectors (E^k) for the $D1$ data set representing the index of the localized eigenvector (column 1), followed by their TCNs (column 2) and network parameters, namely, degree, clustering coefficient (CC), and betweenness centrality.

E^k	TCN	k	CC
234	PSMD1, PSMD11, PSMB8	4, 4, 4	1, 1, 1
228	PFKP, PFKL, PSMB8, H3.3B	5, 5, 4, 8	0.9, 0.9, 1, 0.75
229	COL1A2, COL3A1, PSMB8, H3.F3B	6, 6, 4, 8	0.5, 0.53, 1, 0.75
232	PSMB8, PSME1, PSMD11, H3F3B, H3.3B, PSMD1	4, 4, 4, 8, 8, 4	1, 1, 1, 0.75, 0.75, 1

TABLE IV. Top contributing nodes (TCNs) of the disease-2 ($D2$) subnetwork. The top most localized eigenvectors (E^k) for the $D2$ data set (column 1), their top contributing proteins (column 2), and network parameters, namely, degree, clustering coefficient (CC), and betweenness centrality.

E^k	TCN	k	CC
281	AVPR2, MT-ND5, MT-ND4	2, 13, 13	1, 0.846, 0.846
126	MT-ND1, MT-CO3, MT-ND6	12, 1, 12	0.985, 0.985, 0.984
141	SEL1L, SELS, C1S, HLA-DQA1	1, 2, 9, 6	0, 0, 0.5, 0.46
261	YWHAE, CALM1, CALM3, CALM2	1, 6, 5, 5	0, 0.53, 0.80, 0.80

be considered crucial, as randomness has been emphasized as an essential ingredient for the robustness of a system [52], and lack of *sufficient randomness* might lead to the disease state.

As discussed, the eigenvector localization technique offers a platform to distinguish random and nonrandom parts of the spectra. The eigenvector component deviating from the RMT can be explored further to gain insight into the important interactions revealed through the top contributing nodes (TCNs). In order to do that, we investigate the properties of TCNs of both subnetworks, $D1$ and $D2$, as reported in Tables III and IV, respectively. There turn out to be 31 TCNs in the disease subnetworks, of which 24 are unique. We analyze the structural properties, namely, degree, CC, and betweenness centrality, of these TCNs in both subnetworks. The top contributing proteins lie in the low-degree regime, demonstrating that they do not take part in many pathways. Additionally, the betweenness centrality of all the TCNs is nearly 0, indicating very poor connectivity of these nodes with the rest of the nodes in their individual subnetworks. Together, these rule out a trivial importance of the nodes in terms of their interactions or connectivity, and hence we shift our focus to the biological significance of these TCNs and interaction patterns they form in the network.

A. Functional importance and interaction pattern of top contributing nodes

These TCNs, when examined for their functional properties, are found to be significantly important in DM-II, as they are involved in major functions related to the occurrence of DM-II. Detailed functional information on all these proteins is discussed in the Supplementary Material [40].

Next, we analyze the structural patterns of the TCNs in both disease subnetworks. The $D1$ subnetwork reveals the formation of cliques (Fig. 2). Also, they are involved in the same kind of pathways, which is confirmed by studying the functional properties of these proteins briefly [40]. Further, the TCNs of the $D2$ subnetwork also comprise complete subgraphs (Fig. 2). The functional properties discussed above demonstrate that these TCNs form complexes, as they perform the same kinds of functions, the details of which are explained in the Supplementary Material [40]. Thus, the interaction patterns of these TCNs in both networks of disease data sets, which are revealed through the IPR, pertain to nodes that form cliques indicating a robust and stable system as discussed in Sec. III A.

Thereafter, we compare the interaction patterns of TCNs in the disease state with their interaction patterns in the normal state. Of 24 TCNs in the disease state, 10 are also found in the normal data set. The interaction patterns of these TCNs possess cliquelike structures in the normal subnetworks. These TCNs in the normal data set have a comparatively higher degree than in the disease state, indicating more interactions. This may be due to silencing of various pathways (e.g., insulin signaling and glucose metabolism) from the normal to the disease state due to change in the levels of protein expression as discussed earlier. The structural analysis indicates that even after reduced interactions in the disease state, cliques of order 3 are not destroyed, reflecting that the minimum criteria for maintaining the functionality of the networks are preserved. This may be due to the addition of new proteins in the diseased state. Additionally, the interacting partners of these TCNs illustrate the phenomenon of gene duplication, i.e., a pair of nodes having exactly the same interacting partners [53]. Gene duplication is already known to be important for evolution [54,55]. Degeneracy at the zero eigenvalue, which has a direct relation with the complete and the partial duplication [39], reveals that there is an increase in the exact and partial duplications from the normal to the disease network, as indicated by the increased number of λ_0 degeneracies. It is interesting that although overall there is an increase in the duplicate nodes, the duplicate nodes carrying other special structural properties proceed differently, and of the 10 TCNs common in both states, two pairs of TCNs possess the phenomenon of gene duplication in the normal data set as indicated in Fig. 3. Interestingly, this duplication phenomenon is lost in the disease state (Fig. 3). The gene duplication has structural and functional significanc, as fluctuations in or destruction of the duplicated nodes can either cause an imbalance of

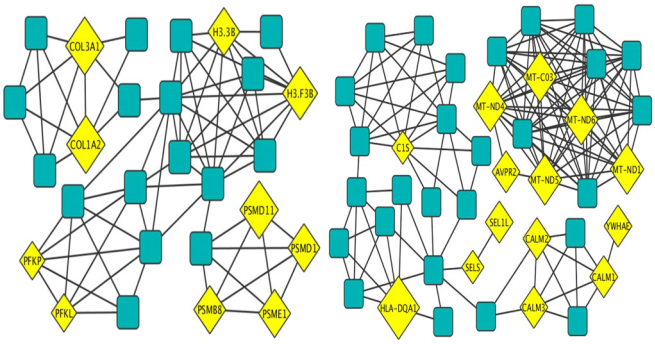

FIG. 2. (Color online) Interaction patterns of top contributing nodes (TCNs) in disease data sets. Left: Local structure of all TCNs in the $D1$ subnetwork. Right: Interaction pattern of TCNs in the $D2$ sub-network. Diamonds (yellow) represent the (TCNs); boxes (green), the first interacting partners of these TCNs.

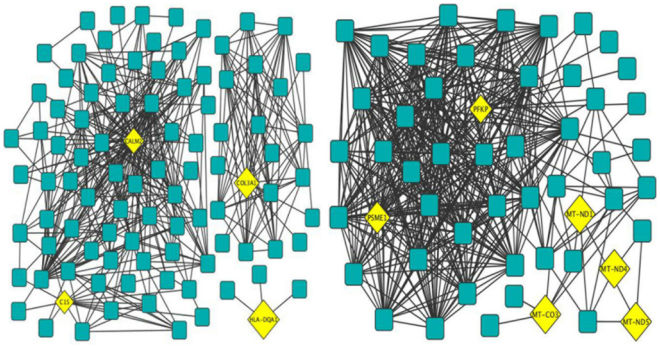

FIG. 3. (Color online) Interaction patterns of disease top contributing nodes (TCNs) in normal data sets. Local structural patterns for the disease TCNs in the normal network. Left: $N1$ subnetwork. Right: $N2$ subnetwork. Diamonds (yellow) represent the TCNs; boxes, the first interacting partners of these TCNs. The same TCNs in normal subnetworks possess more interactions than the disease subnetworks.

genetic material or lead to the generation of new gene products resulting in diseases [56]. The protein corresponding to the TCN possessing gene duplication behavior affects the major pathways, namely, insulin resistance and pathways promoting obesity, in turn resulting in DM-II [40]. Thus, proteins revealed through the localization properties of spectra have interactions which may be important targeting sites for drug development in DM-II.

VI. CONCLUSION

To conclude, we provide a platform for detecting important proteins revealed through structural patterns in DM-II using the network approach and RMT. The analysis suggests that, instead of targeting individual proteins, a group of proteins forming particular interaction patterns should be taken into consideration for drug development. This approach, though based on sophisticated mathematical techniques, is time-efficient and cost-effective and paves the way to looking at diseases from a different perspective by taking the whole system into consideration. For the first time the localization of eigenvectors under RMT combined with the network theory framework is employed to analyze DM-II. Being the first step, the approach holds the potential to provide a new dimension to disease research. RMT is a very well-developed branch of physics and continues to witness emerging techniques and concepts [57,58]. Appreciating the applicability of this technique to uncovering crucial information about DM-II, other well-developed tools of RMT can be used to gain more insight into the complexity of DM-II and to decipher important entities and their structural patterns significant in other diseases. Moreover, our work, on one hand, presents a new tool for identification of the proteins responsible for the occurrence of the disease and predicts interaction patterns for drug targets, while, on the other hand, providing insight into the complexity of the disease at the rudimentary level. The analysis, exemplifying the structural and functional significance of these proteins, indicates that, rather than individual units, the whole interaction pattern should be treated as a target for drug development, providing a direction toward treatment of the disease. This approach can also be extended to the composition of novel drugs, conceptualization of single-drug therapy for multiple diseases [59,60], and derivation of personalized therapy [61], which would be especially beneficial for low-income countries.

ACKNOWLEDGMENTS

A.R. is thankful to Complex Systems Lab members Camellia Sarkar and Sanjiv K Dwivedi for assistance. It is a pleasure to acknowledge Professors Seyed E. Husnain (School of Biological Sciences, IIT Delhi, India) and Stefano Boccaletti (CNR-Istituto dei Sistemi Complessi, Italy) for encouraging discussions. S.J. is grateful to Department of Science and Technology, Government of India, and Council of Scientific and Industrial Research, India, project grant nos. EMR/2014/000368 and 25(0205)/12/EMR-II respectively for financial support.

[1] W. Y. Fujimoto, Am. J. Med. **108**, 9 (2000).
[2] E. A. McIntyre and M. Walker, Clin. Endocrinol. **57**, 303 (2002).
[3] L. M. T. Tusie, Arch. Med. Res. **36**, 210 (2005).
[4] M. Kasuga, J. Clin. Invest. **116**, 1756 (2006).
[5] H. King, R. E. Aubert, and W. H. Herman, Diabetes Care **21**, 1414 (1998).
[6] G. Danaei, M. M. Finucane, Y. Lu, G. M. Singh, M. J. Cowan, C. J. Paciorek, J. K. Lin, F. Farzadfar, Y.-H. Khang, G. A. Stevens, M. Rao, M. K. Ali, L. M. Riley, C. A. Robinson, and M. Ezzati, Lancet **378**, 31 (2011).
[7] World Health Organization, Diabetes fact sheet; http://www.who.int/mediacentre/factsheets/fs312/en/ (Accessed 30 June 2014).
[8] International Diabetes Federation (IDF) (2013) *Diabetes Atlas*, 6th ed. (IDF, Brussels, Belgium); http://www.idf.org/diabetesatlas (Accessed 25 June 2014).
[9] P. Zimmet, K. G. Alberti, and J. Shaw, Nature **414**, 782 (2001).
[10] J. Sepúlveda and C. Murray, Science **345**, 1275 (2014).
[11] E. J. Lin *et al.*, Diabetes Care **27**, 2154 (2004).
[12] L. Dean and J. McEntyre, *The Genetic Landscape of Diabetes* (National Library of Medicine, Washington, DC, 2004).
[13] American Diabetes Association, Diabetes Care **21**, 296 (1998).
[14] International Diabetes Federation, *World Health Organization. The Economics of Diabetes and Diabetes Care* (IDF, Brussels, Belgium, 1996). (Accessed 12 July 2014).
[15] M. Liu *et al.*, PLoS Genet. **3**, e96 (2007).
[16] A. Rasche, H. Al -Hasani, and R. Herwig, BMC Genom. **9**, 310 (2008).
[17] B. F. Voight *et al.*, Nat. Genet. **42**, 579 (2010).
[18] V. I. Grolmusz, R. Soc. Open Sci. **2**, 140252 (2015).
[19] E. P. Wigner, Ann. Math. **62**, 548 (1955).
[20] T. Papenbrock and H. A. Weidenmuller, Rev. Mod. Phys. **79**, 997 (2007).
[21] J. Kwapień and S. Drożdż, Phys. Rep. **515**, 115 (2012).
[22] S. Jalan, C. Sarkar, A. Madhusudanan, and S. K. Dwivedi, PLoS ONE **9**, e88249 (2014).

[23] S. M. Gibson *et al.*, PLoS ONE **8**, e55871 (2013).
[24] A. Agrawal, C. Sarkar, S. K. Dwivedi, N. Dhasmana, and A. Jalan, Physica A **404**, 359 (2014).
[25] M. W. Blows and K. McGuigan, Mol. Ecol. **24**, 2056 (2014).
[26] D. A. Benson *et al.*, Nucleic Acids Res. **41**, D36 (2013).
[27] A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, and B. Boeckmann, Nucleic Acids Res. **33**, D154 (2005).
[28] G. C. Weir and S. B. Weir, J. Clin. Invest. **121**, 3395 (2011).
[29] M. Skelin, M. Rupnik, and A. Cencić, Altex **27**, 105 (2010).
[30] L. F. Waanders *et al.*, Proc. Natl. Acad. Sci. USA **106**, 18902 (2009).
[31] D. Szklarczyk *et al.*, Nucleic Acids Res. **39**, D561 (2011).
[32] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[33] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[34] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[35] Y. V. Fyodorov and A. D. Mirlin, Phys. Rev. Lett. **67**, 2049 (1991).
[36] S. Jalan *et al.*, Europhys. Lett. **99**, 48004 (2012).
[37] A.-L. Barabasi, N. Gulbahce, and J. Loscalzo, Nat. Rev. Genet. **12**, 56 (2011).
[38] P. V. Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, UK, 2011).
[39] A. Yadav and S. Jalan, Chaos Interdis. J. Nonlin. Sci. **25**, 043110 (2015).
[40] For the degree distribution and functional properties of the nodes having $CC = 1$ and TCNs, see Fig. S3 in the Supplementary Material http://link.aps.org/supplemental/10.1103/PhysRevE.92.022806.
[41] H. Lu, Y. Yang, E. M. Allister, N. Wijesekara, and M. B. Wheeler, Mol. Cell. Proteom. **7**, 1434 (2008).
[42] A.-L. Barabasi and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).
[43] E. Y. Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, and R. Milo, Proc. Natl. Acad. Sci. USA **101**, 5934 (2004).
[44] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, London, 2006).
[45] S. K. Dwivedi and S. Jalan, Phys. Rev. E **90**, 032803 (2014).
[46] B. C. Martin *et al.*, Lancet **340**, 925 (1992).
[47] W. T. Garvey, T. P. Hueckstead, S. Matthaei, and J. M. Olefsky, J. Clin. Invest. **81**, 1528 (1988).
[48] C. A. Fagot *et al.*, Int. J. Epidemiol. **27**, 808 (1998).
[49] P. Erdös and A. Renyí, Acta Math. Hung. **10**, 21 (1959).
[50] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[51] K. Zyczkowski, in *Quantum Chaos*, edited by H. A. Cerderia, R. Ramaswami, M. C. Gutzwiller, and G. Casati (World Scientific, Singapore, 1991), p. 153.
[52] S. Jalan and J. N. Bandyopadhyay, Europhys. Lett. **87**, 48010 (2009).
[53] U. Bergthorsson, D. I. Andersson, and J. R. Roth, Proc. Natl. Acad. Sci. USA **104**, 17004 (2007).
[54] J. Zhang, Trends Ecol. Evol. **18**, 292 (2003).
[55] M. Nowak, M. Boerlijst, J. Cooke, and J. Smith, Nature **388**, 167 (1997).
[56] J. R. Lupski, Trends Genet. **14**, 417 (1998).
[57] H. A. Weidenmuller and G. E. Mitchell, Rev. Mod. Phys. **81**, 539 (2009).
[58] T. Guhr, A. Muller, and H. A. Weidenmuller, Phys. Rep. **299**, 189 (1998).
[59] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, and M. Vidal, Proc. Natl. Acad. Sci. USA **104**, 8685 (2007).
[60] M. A. Yildirim, K. I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, Nat. Biotechnol. **25**, 1119 (2007).
[61] J. Woodcock, Clin. Pharm. Ther. **81**, 164 (2007).