Spectral Analysis of Protein-Protein Interactions in *Drosophila melanogaster*

Christel Kamp*
Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2AZ, United Kingdom

Kim Christensen†
Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2AZ, United Kingdom and
Physics of Geological Processes, University of Oslo, PO Box 1048, Blindern, N-0316 Oslo, Norway

Within a case study on the protein-protein interaction network (PIN) of *Drosophila melanogaster* we investigate the relation between the network's spectral properties and its structural features such as the prevalence of specific subgraphs or duplicate nodes as a result of its evolutionary history. The discrete part of the spectral density shows fingerprints of the PIN's topological features including a preference for loop structures. Duplicate nodes are another prominent feature of PINs and we discuss their representation in the PIN's spectrum as well as their biological implications.

PACS numbers: 89.75.-k, 89.20.-a, 89.75.Hc, 89.75.Fb, 87.16.Yc, 87.16.-b, 87.10.+e, 02.50.Fz

I. INTRODUCTION

Network structures can be observed in most diverse domains ranging from biological and technological systems to social or economical systems [1]. Genetic regulatory networks, protein-protein interaction networks and metabolic networks support the functions of life in any living organism. Technological networks such as the internet or the World Wide Web have a huge impact on our lives and societies. Networks of acquaintances and the exchange of information within these networks shape social and economical systems. Considering the omnipresence of networks, their investigation has a long tradition in graph theory [2, 3]. However, during the last few years high quality data on real-world networks has revealed that they cannot be adequately described by standard models from random graph theory and the topic has attracted growing interest. Still, much attention has been devoted to the derivation of rather specific quantities like degree distributions or clustering coefficients [4] that do not allow for a classification and understanding of network topologies within a broader and self-consistent framework.

Making an attempt towards a more comprehensive description spectral graph theory [5, 6, 7, 8] can be considered as one promising ansatz. A network of N nodes can be described by its adjacency matrix $\mathbf{A} = (a_{ij})$ with entries

$$a_{ij} = \begin{cases} 1 & \text{if there is a link between node } i \text{ and } j \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The adjacency matrix is a symmetric, non-negative matrix in the case of undirected networks and accordingly has real eigenvalues $\{\lambda_j\}$, $j = 1, ..., N$, being solutions of $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. The relation between features of a network and properties of its spectral density

$$\rho(\lambda) \;=\; \frac{1}{N}\sum_{j=1}^{N}\delta(\lambda - \lambda_j) \quad (2)$$

with respect to its adjacency matrix is a topic of current research. While dense classical random networks exhibit a semi-circular spectral density of the adjacency matrix [9], networks with a broad or scale free degree distribution give rise to a broader spectrum [10, 11, 12, 13, 14, 15]. A striking feature of sparse random networks' spectral density is the emergence of peaks at eigenvalues of finite trees [9], similar to those found in large random trees [16], due to the strong prevalence of these subgraphs [32]. Here we address whether these findings are applicable more generally, that is whether peaks in the spectral density of sparse random networks can be associated to a strong prevalence of specific subgraphs. The search for subgraphs that are statistically overrepresented relative to a null-model, so-called motifs, recently gained much attention [17, 18]. As a case study on the relation between these two approaches, we investigate the spectral properties of the protein-protein interaction network (PIN) of the fruit fly *Drosophila melanogaster* [19]. While no simple correspondence between network motifs and a network's spectral proprieties can be derived, on a more abstract level, we infer from the PIN's spectrum a prevalence of loop structures. Furthermore, some properties specific to a network that has evolved by duplication of nodes are studied and discussed within the context of spectral analysis.

II. THE SPECTRUM OF THE PIN OF *DROSOPHILA MELANOGASTER*

For our study we used the PIN of *Drosophila melanogaster* as given in [19] and available via the Database of Interacting Proteins [20]. The protein-

*Electronic address: mail@christelkamp.de, c.kamp@imperial.ac.uk
†Electronic address: k.christensen@imperial.ac.uk

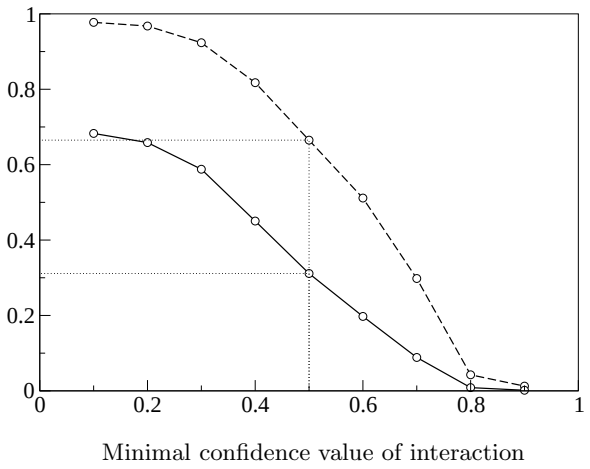

FIG. 1: The number of nodes divided by 10000 (solid line) and the fraction of nodes (dashed line) in the largest connected component in the PIN as a function of the minimal confidence value of protein-protein interaction. We focus on the PIN defined by a minimal confidence value of 0.5, see dashed line.

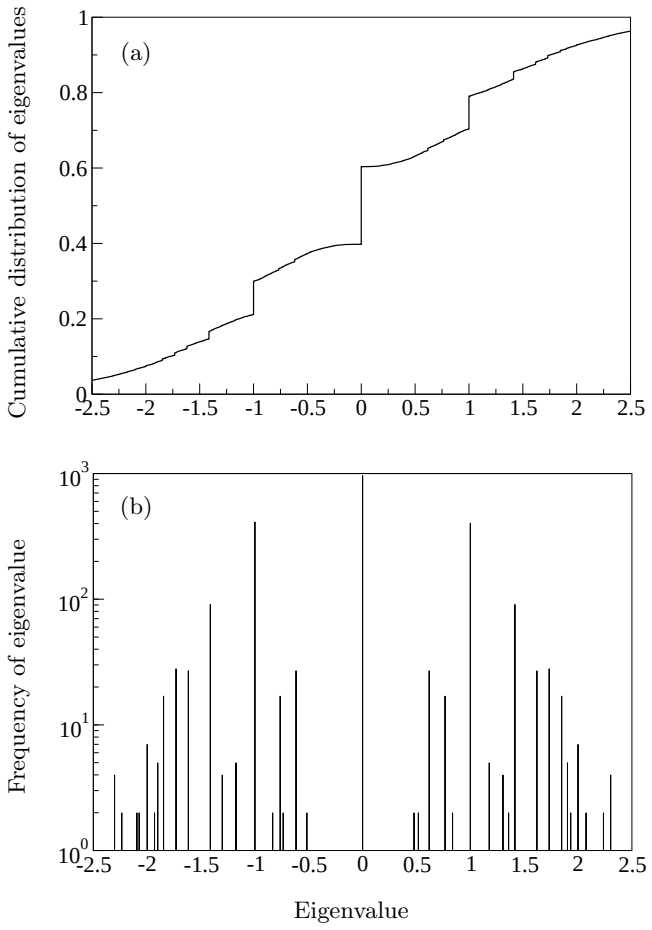

FIG. 2: Spectral analysis of the protein-protein interaction network of *Drosophila melanogaster*. (a) The cumulative spectral density. (b) The discrete frequency spectrum containing 49% of all eigenvalues.

protein interactions have been derived using the two-hybrid method which, however, is known to generate many false positives. Therefore each interaction in the network is classified by a confidence value between zero and one defining a hierarchy of networks with increasing minimal confidence value for the protein-protein interactions. In Fig. 1 the size of the largest connected component in a network with a given minimal confidence value of interactions is shown.

For our further analysis we choose a network with a minimal confidence value of 0.5 which contains 4681 proteins and 4794 interactions corresponding to an average degree $\langle k \rangle = 2.05$. The network is enriched with biologically meaningful interactions while it still shows a strong largest connected component (i.e. a giant component) containing about 2/3 of its nodes.

We determined the eigenvalues of the adjacency matrix corresponding to this PIN. The cumulative spectral density in Fig. 2 (a) exhibits jumps at various eigenvalues which are represented by the discrete spectrum in Fig. 2 (b) [33]. Since about 2/3 of the network's nodes belong to its giant component and 49% of the eigenvalues in the network's spectrum are in the discrete spectrum, the emergence of spectral peaks cannot be explained by small isolated clusters alone.

A. The discrete spectrum and network motifs

To get a better understanding of the emergence of spectral peaks we compare the discrete spectrum with the corresponding spectra of two reference networks. First, we look at a network of the same size and degree sequence but randomized links following the procedure of [21] (a randomized PIN). Second, we consider a classical random network of the same size and average degree $\langle k \rangle = 2.05$ (a random network), that is a network with a probability $p = 0.000438$ for a link between any two nodes. In Fig. 3 the discrete spectrum of the adjacency matrix of the protein-protein interaction network of *Drosophila melanogaster* as well as of the two reference networks are shown, the latter being averages over 10 reference networks.

To get more reliable results, we concentrate our further analysis only on eigenvalues that can be found more than twice in the spectrum of the original network. Qualitatively, we see that while the classical random network shows only a few peaks of that size corresponding to the eigenvalues of simple tree-graphs (2.1, 3.1, 4.1, 4.2. in Fig. 4), additional eigenvalues appear in the discrete spectrum of the randomized PIN with the same degree sequence as the original network and eventually the original network (see Fig. 3). This change in the spectral properties indicates some differences in the structural organization of the underlying networks. In the following paragraphs, we will discuss how the observed hierarchy of spectral peaks reflects the networks' topologies and relates to other concepts like the search for motifs.

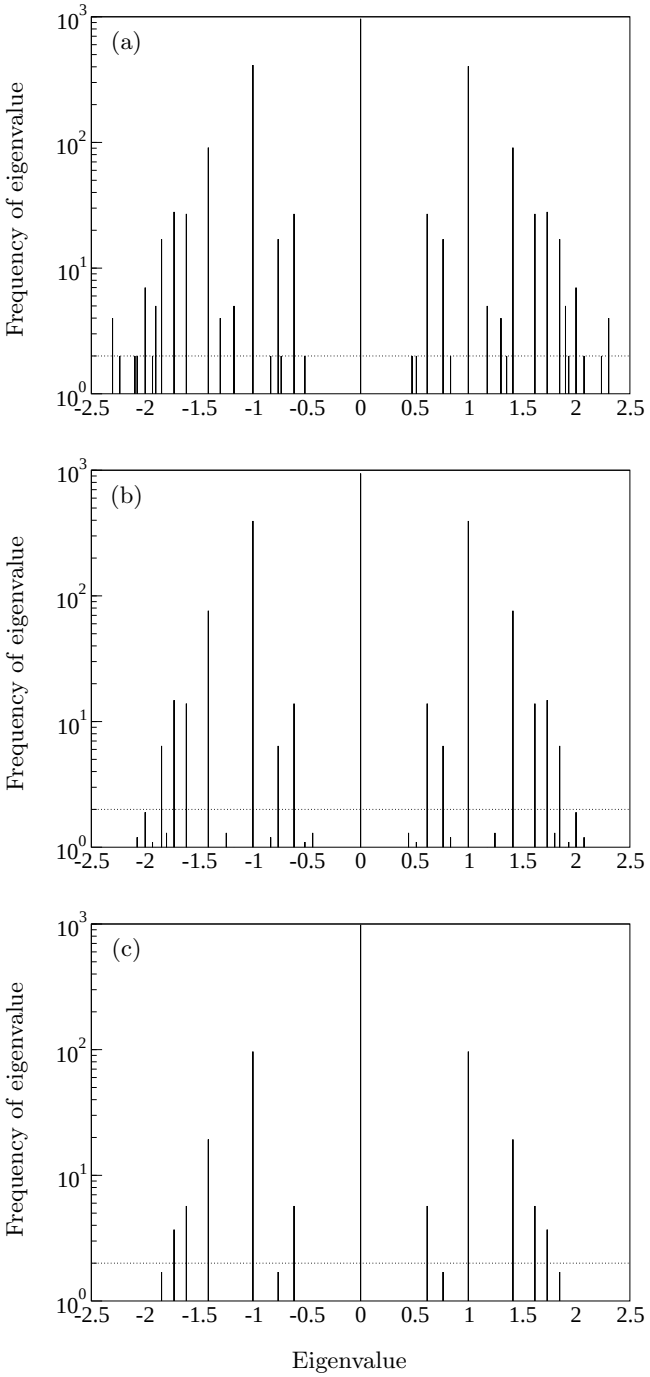

FIG. 3: The discrete frequency spectrum of (a) the PIN of *Drosophila melanogaster* containing 49% of all eigenvalues, (b) a randomized PIN with identical degrees at each node containing 43% of all eigenvalues, and (c) a classical random network of identical size and average degree $\langle k \rangle = 2.05$ containing 27% of all eigenvalues.

Following the arguments of [16], we suggest that the prevalence of specific peaks in the discrete spectrum of a network corresponds to a strong representation of certain subgraphs. It has recently been shown that networks from different contexts show characteristic overrepresentation of specific subnetworks which are usually referred to as motifs [17, 18]. Although motifs can be expected to leave marks in a network's spectrum, there is seemingly no simple correspondence between the eigenvalues of small subgraphs and spectral peaks. First, subgraphs are not generally represented by their eigenvalues in the spectrum of the whole network. Second, isospectral graphs are not necessarily isomorphic [23]. Nevertheless, a thorough comparative study of the discrete spectrum can provide some insight into the networks' structure. In Figs. 4 and 5 we show the connected subgraphs up to size 5 with the full set of eigenvalues present in the discrete spectrum of the whole network. It shows that the spectrum of the PIN is more consistent with loop-structures (cf. graphs 3.2, 4.3, 5.4, 5.5, 5.6, 5.8) than any randomized version which might hint to regulatory functionality supplied by this network. The eigenvalues behind these structures might correspond to eigenvalues of trees, e.g., the eigenvalues of a triangle (graph 3.2) or the box (graph 4.3, often also referred to as bi-fan structure) might well be explained by graphs 2.1 and 5.3. However, to represent the eigenvalues of graphs 5.5 and 5.6 one has to consider trees of minimum size 8 and 7, respectively. The eigenvalues of graph 5.8 cannot be found among trees of size up to 10. Considering that the

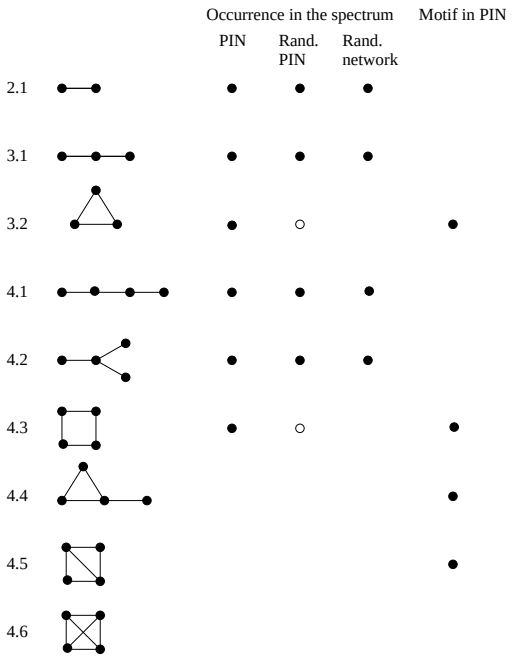

FIG. 4: Connected subgraphs with up to 4 nodes: A bullet (•) in the three middle columns denotes that the eigenvalues of this graph can be found in the spectrum of the original network (PIN), the randomized network (Rand. PIN) or the random network (Rand. network), respectively. The rightmost column shows whether the subgraph is a motif according to the `mfinder` software (default settings, [17, 22]). White bullets (○) correspond to single eigenvalue occurrences.

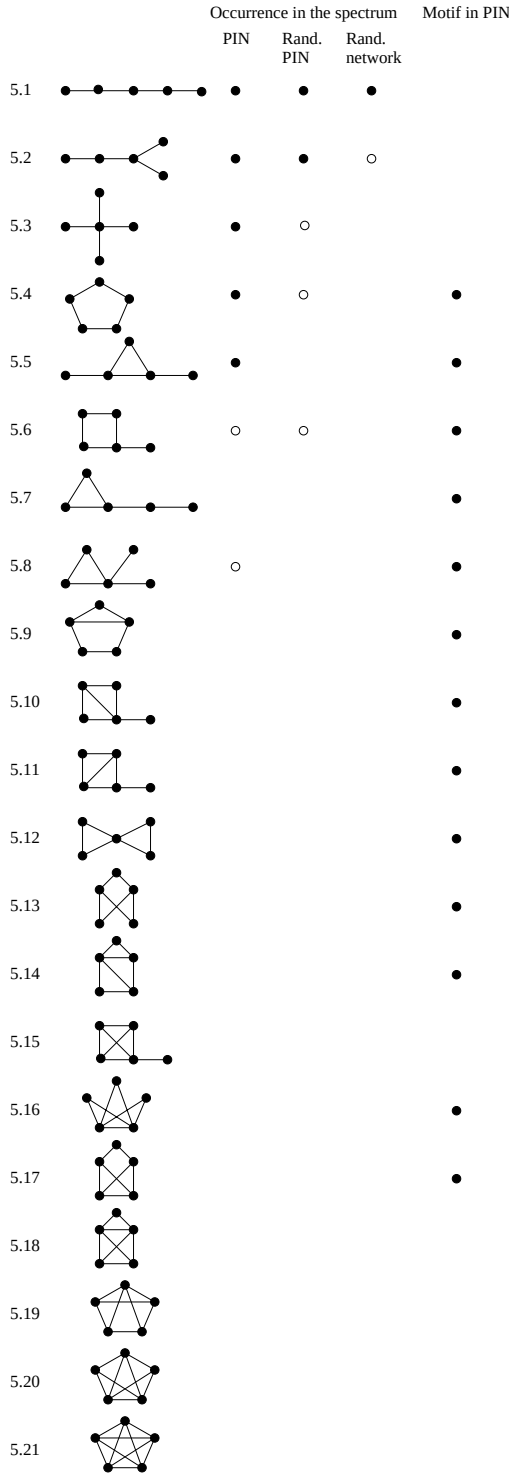

FIG. 5: Connected subgraphs with 5 nodes: A bullet (•) in the three middle columns denotes that the eigenvalues of this graph can be found in the spectrum of the original network (PIN), the randomized network (Rand. PIN) or the random network (Rand. network), respectively. The rightmost column shows whether the subgraph is a motif according to the `mfinder` software (default settings, [17, 22]). White bullets (○) correspond to single eigenvalue occurrences.

frequency of a given tree of size n in a sparse network decreases exponentially with n [9] and relating the findings in the PIN to those in the randomized reference networks we hypothesize that the spectral peculiarities reflect the loop structure in the original network.

To quantify the correspondence between the number of specific subgraphs in the PIN and the PIN's discrete spectrum we tried to decompose the spectrum into the contributions of connected subgraphs up to size 5. This, however, was not feasible indicating that higher order contributions, though being individually small, cannot be neglected as a whole.

Although we have to ascertain that there is no simple correspondence between subgraphs of a network and the prevalence of their eigenvalues in the discrete spectrum of the whole network we want to discuss the relation of spectral properties to the notion of motifs. According to the definition introduced in Ref. [17], a motif is a subnetwork that shows strong prevalence within the network relative to a randomized network. For our analysis we refer to the default requirements implemented in the `mfinder` software [17, 22], that is a motif is a subgraph that occurs at least by two standard deviations more than in 100 randomized networks with the same degree sequence. In Figs. 4 and 5 the rightmost columns show which connected subgraphs up to size 5 are motifs in the PIN according to these criteria. There exist a lot of highly connected motifs while the spectrum reflects more the tree-structures in the network. However, the fingerprints of trees in the spectrum of both the original and the randomized PIN are consistent with the fact that they do not show up as motifs according to the above definition. One might further speculate whether some motifs are hidden for spectral analysis because they are in fact building blocks of larger units. For example graph 5.5 as well as its subgraph 4.4 is a motif according to [17, 22]. But only the eigenvalues of 5.5 can be found in the spectrum of the PIN. Moreover highly connected motifs do not occur in high (absolute) numbers and might accordingly be drowned in spectral analysis.

B. The circuitry of the PIN

In section II A we have shown that the discrete spectrum of the PIN of *Drosophila melanogaster* favors the eigenvalues of loopy subgraphs. This observation derived from the investigation of distinct local structures and their eigenvalue representations can be confirmed by an assessment of the whole set of eigenvalues. Evaluating the trace of the matrix $\mathbf{A}^k$

$$Tr(\mathbf{A}^k) = \sum_{i=1}^{N} \lambda_i^k \tag{3}$$

yields the number of directed loops of length k in the underlying network [6, 10] as shown in Fig. 6, though neglecting details of the graphs underlying the loops. Note,

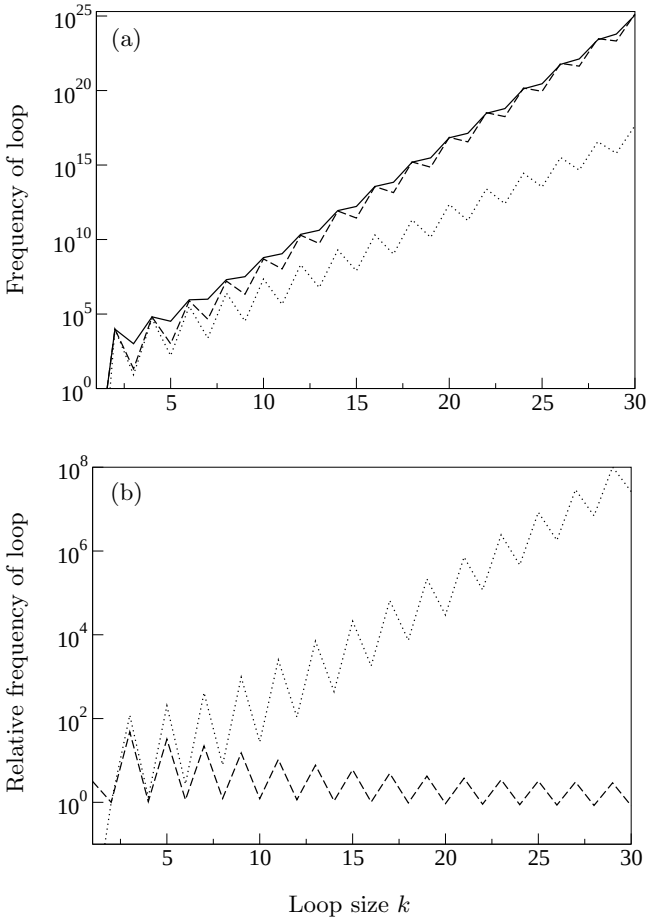

FIG. 6: (a) The frequency of loops of size k in the PIN of *Drosophila melanogaster* (solid line), a randomized PIN with identical degrees at each node (dashed line) and a classical random network of identical size and average degree $\langle k \rangle = 2.05$ (dotted line). Odd cycles represents non-trivial loops, that is, deviations from a tree-like structure in the networks. (b) The relative frequency of loops of size k in the PIN of *Drosophila melanogaster* with respect to the randomized PIN (dashed line) and the classical random network (dotted line).

that even loops might be trivial going back and forth in a tree while odd loops are non-trivial. The difference in growth rates of the numbers of loops of growing size between the original network and the classical random network is likely due to the strong fragmentation of the latter one (many isolated nodes). However, the strong relative prevalence of loops of odd length in the original network is more remarkable with respect to the networks' topologies. This becomes more obvious from Fig. 6 (b) showing the number of loops of a given size in the original PIN normalized to the numbers in the two reference networks. While tree graphs only have trivial loops of even length, loops of odd length indicate non-trivial loops which confirms the results derived from the evaluation of the discrete spectrum on the basis of eigenvalue representations of small subgraphs.

The analysis of a network's discrete spectrum could reveal some structural information about the network as a whole. This information is less specific than an analysis in terms of motifs, that is only conclusions about more general properties like the prevalence of loops are possible instead of exact motif counts. However, it should be emphasized that spectral analysis is not hampered by an a priori *bias* towards predefined quantities like motifs of a given size. It is a challenging question of future research to investigate the relationship between a network's spectrum and its topological features, e.g. in terms of motifs, in more detail to get a more rigorous and *unbiased* characterization of a network's topological features.

III. FINGERPRINTS OF DUPLICATION

The evolution of many biological networks and specifically PINs is assumed to be strongly driven by duplication (and diversification) of nodes in the network [24, 25]. The genomes underlying the PIN of many organisms have undergone a few whole genome duplications complemented by many single-gene duplications [26]. After duplication, one of the duplicates usually diverges from its original appearance, possibly providing new functionality. The concept of duplication has similarly been recognized to be important for functional roles in a network motif [27]. The search for fingerprints of the evolutionary history of a PIN naturally has to include an assessment of duplicate nodes, that is those that share the same interaction partners. Each set of duplicate nodes represents an equivalence class also referred to as an orbit. The reduced network is a network in which all nodes of an orbit are reduced to one node.

	Duplicate nodes	Duplicate links
PIN *D. melanogaster*	686	728
Randomized PIN	626.0 ± 22.0	629.1 ± 22.0
Random network	151.0 ± 14.2	151.0 ± 14.2

TABLE I: The table shows the number of nodes that are duplicates (duplicate nodes) and the number of neighbors associated to these nodes (duplicate links) in the original network and the two reference networks, that is the difference in the number of nodes and links between the original network and the reduced network. Isolated nodes have been neglected.

Tab. I shows that the PIN has more duplicate nodes (with associated links) than the reference networks. Fig. 7 shows the frequency of orbits of a given size in the original as well as in the reference networks. The distribution of orbit sizes in the original network is very close to the one found in the randomized PIN with the same degree sequence, but much broader than that of the classical random network.

Again, spectral analysis offers a complementary approach to the topic. Using the results in Appendix A, we determined the eigenvalues of those graphs that arise from duplication of the two simplest reduced graphs: a line

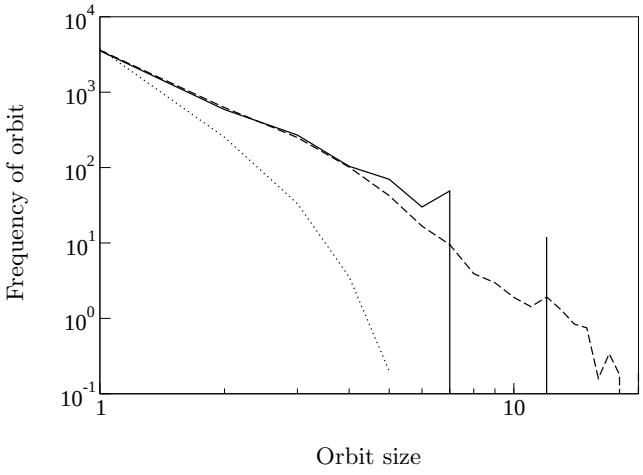

FIG. 7: The frequency of orbit size in the PIN of *Drosophila melanogaster* (solid line), a randomized PIN with identical degrees at each node (dashed line) and a classical random network of identical size and average degree $\langle k \rangle = 2.05$ (dotted line). Isolated nodes have been neglected.

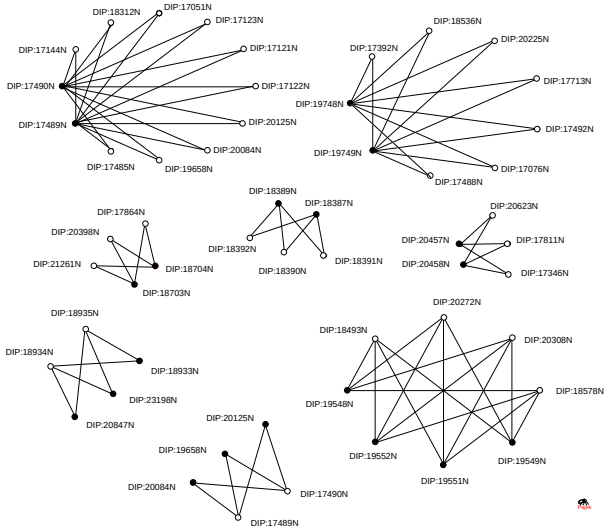

FIG. 8: Subgraphs of nodes that form orbits of size ≥ 2 and that have degree ≥ 2 (black), orbits of size 2 of nodes with degree 2 have been omitted. The white nodes are the neighbors of duplicate nodes that may have more neighbors than shown. The nodes' labels are their identifiers from the Database of Interacting Proteins [20], cf. Appendix B for more details.

and a triangle (graphs 2.1 and 3.2 in Fig. 4). We allowed for up to ten duplications of each node of the reduced network and searched for the eigenvalues of the resulting subgraphs. However, spectral analysis is only consistent with the emergence of star graphs and the original triangle as well as the box or bi-fan structure (graph 4.3 in Fig. 4).

Considering the representation of star graphs in the spectrum of the PIN one might guess that the high frequency of large orbits mainly reflects nodes with many leaves. A look at the joint distribution of the size of an orbit and the degree of its nodes in the original and the reference networks supports this hypothesis. In 100 reference networks (of both kind) the nodes in an orbit larger than one have degree one, that is only nodes with degree one have duplicates. Only in extremely rare cases do nodes with degree two have a single duplicate.

This matches the global situation in the original network, however, there are some remarkable exceptions with nodes of high degree in large orbits shown in Fig. 8 that cannot be found in the reference networks. This is also well in accordance with the values found in Tab. I. Different from the original network, in both reference networks the number of duplicate links is practically the same as the number of duplicate nodes.

From the Database of Interacting Proteins [20] and FlyBase [28] we derived names and descriptions (if available) for the proteins in Fig. 8 as shown in Appendix B. We find that duplicate proteins are likely to have similar functionality in accordance with results in the yeast PIN [29].

IV. SUMMARY AND CONCLUSIONS

Recent developments in the research on complex networks have brought up a better understanding of a network's topology and its connection to functionality. However, a comprehensive theory of networks incorporating classical graph theory as well as recent findings into a self-consistent framework has still to be worked out. Considering spectral graph theory to be a promising ansatz for this attempt, we have done a case study on the PIN of *Drosophila melanogaster*. The eigenvalues of a network's adjacency matrix (and of related matrices) provide information about a network's structural properties like the number of connected components, its diameter or characteristics of its degree distribution. Here, we have put special emphasis on the investigation of the discrete spectrum of a sparse network relating it to prevalent substructures. Although it will probably not be possible to derive the densities of specific subgraphs from the spectrum of a network we could show that structural prevalences on a more abstract level are reflected in the (discrete) spectrum of the PIN under investigation. While we here focused on the appearance of loops in subgraphs as well as the whole PIN future analysis might reveal further topological features.

Considering the evolutionary history of PINs we also discussed the appearance of proteins that share their neighbors together with the fingerprints of these structures that can be found in the network's spectrum. Studying structures of duplicate proteins in more details we find that they often have close functional relationships in accordance with earlier findings in yeast.

The requirement applied here for the members of an orbit to show exactly the same neighborhood is very restrictive, though required to allow for transitivity. This might be generalized by the definition of a similarity measure that quantifies the overlap of the neighborhoods of two nodes. This similarity measure can be defined as a distance measure between nodes and the application of a clustering algorithm in the associated metric space might give further insight into local structures.
This case study shows that a more systematic assessment of the relation between a network's spectral and topological properties has to be a topic of future research. It is a challenging task, however, it can bring important insight into a network's structure in a less biased and more systematic way than currently available.

Acknowledgments: C. Kamp would like to thank A. Bunten for providing computing facilities and software and for many helpful discussions on technical problems. I am as grateful to N. Farid, S.A. Teichmann and J. Leal for the discussions and many helpful comments on the manuscript. Also, we would like to acknowledge the hospitality of the department of physics of the university of Oslo during the period of finalizing this manuscript. This work was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

Appendix A

Let $\mathbf{A}$ be a $N \times N$ matrix representing an undirected graph, i.e. a symmetric matrix with entries $a_{ij} \in \{0,1\}$ and $a_{ii} = 0$. Let $\mathbf{D}$ be the matrix that is obtained after m perfect duplications of nodes or in other words by m duplications of rows and columns, respectively. Let i_1,...,i_l, $l \in \{1,...,N\}$ the number (identifier) of (mutually different) nodes that have been duplicated and m_{i_1},...,m_{i_l} be the corresponding number of duplications per node with $m = \sum_{j=1}^{l} m_{i_j}$. Let $\mathbf{A}_{i_j}$ the matrix $\mathbf{A}$ but with the element $a_{i_j i_j}$ replaced by $a_{i_j i_j} + \lambda$. Analogously, the matrix $\mathbf{A}_{i_1 \ldots i_l}$ corresponds to the matrix $\mathbf{A}$ but with $a_{i_1 i_1}$,...,$a_{i_l i_l}$ replaced by $a_{i_1 i_1} + \lambda$,... $a_{i_l i_l} + \lambda$. Let furthermore $\mathbf{I}$ be the identity matrix. Then the following equation holds

$$\begin{aligned}
&\det(\mathbf{D} - \lambda\mathbf{I}) = \qquad (4)\\
&(-\lambda)^m \det(\mathbf{A} - \lambda\mathbf{I})\\
&+ (-\lambda)^m \sum_{r \le l} m_{i_r} \det(\mathbf{A}_{i_r} - \lambda\mathbf{I})\\
&+ (-\lambda)^m \sum_{r \le l-1} \sum_{r < j \le l} m_{i_r} m_{i_j} \det(\mathbf{A}_{i_r i_j} - \lambda\mathbf{I})\\
&+ (-\lambda)^m \sum_{r \le l-2} \sum_{r < j \le l-1} \sum_{j < s \le l} m_{i_r} m_{i_j} m_{i_s} \det(\mathbf{A}_{i_r i_j i_s} - \lambda\mathbf{I})\\
&\vdots\\
&+ (-\lambda)^m \sum_{r \le 1} \sum_{r < j \le 2} \ldots \sum_{x < y \le l} m_{i_r} \ldots m_{i_y} \det(\mathbf{A}_{i_r \ldots i_y} - \lambda\mathbf{I}).
\end{aligned}$$

Note that the last term is equivalent to $m_{i_1} \ldots m_{i_l} \det(\mathbf{A}_{i_1 \ldots i_l} - \lambda\mathbf{I})$. It gets obvious from this formula that perfect duplication of nodes only adds zeros to the spectrum of the graph.
Equation (4) can be proven by induction. Considering a graph with adjacency matrix $\mathbf{A}$ in which an arbitrary node i is duplicated m times leading a duplication matrix $\mathbf{D}$ one can show that

$$\det(\mathbf{D} - \lambda\mathbf{I}) = (-\lambda)^m [\det(\mathbf{A} - \lambda\mathbf{I}) + m \det(\mathbf{A}_i - \lambda\mathbf{I})]. \quad (5)$$

After validating the case of $l = 0$, $m = 0$ of equation (4) we do the induction by evaluation of the adjacency matrix $\tilde{\mathbf{D}}$ of a graph generated from the duplication graph represented by $\mathbf{D}$ by duplicating (a non-duplicate) node i_{l+1} $m_{i_{l+1}}$ times. Therefore, we apply (5)

$$\begin{aligned}
&\det(\tilde{\mathbf{D}} - \lambda\mathbf{I})\\
&= (-\lambda)^m_{i_{l+1}} [\det(\mathbf{D} - \lambda\mathbf{I}) + m_{i_{l+1}} \det(\mathbf{D}_{i_{l+1}} - \lambda\mathbf{I})]
\end{aligned}$$

and derive $\det(\mathbf{D} - \lambda\mathbf{I})$ and $\det(\mathbf{D}_{i_{l+1}} - \lambda\mathbf{I})$ using the assumption (4) yielding the formula (4) for $m + m_{i_{l+1}}$ duplications of $l + 1$ mutually different nodes.
As an example, this formula is applied to the 2×2-matrix $\mathbf{A}$ corresponding to two connected nodes. Then, $i_1 = 1$, $i_2 = 2$ and one gets for the matrix D after $m = m_1 + m_2$ duplications:

$$\begin{aligned}
\det(\mathbf{D} - \lambda\mathbf{I}) &= (-\lambda)^m (\lambda^2 - 1 - m_1 - m_2 - m_1 m_2)\\
\lambda_{1;2} &= \pm\sqrt{1 + m + m_1 m_2}.
\end{aligned}$$

Translating this into the number of nodes per orbit $n_i = m_i + 1$ leads to the eigenvalues

$$\lambda_{1;2} = \pm\sqrt{n_1 n_2}.$$

Appendix B

The following tables contain the information on the proteins shown in Fig. 8 extracted from the Database of Interacting Proteins [20] and FlyBase [28].

DIP ID	Protein name/description
DIP:17489N	CG11719-PA open reading frame, Mst98Ca, (Male-specific RNA 98Ca)
DIP:17490N	CG18396-PA open reading frame, Mst98Cb, (Male-specific RNA 98Cb)
DIP:17144N	CG4015-PA open reading frame, Fcp3C, (Follicle cell protein 3C)
DIP:18312N	CG17777-PA open reading frame
DIP:17051N	CG17666-PA open reading frame
DIP:17123N	CG15781-PA open reading frame
DIP:17121N	CG15032-PA open reading frame
DIP:17122N	CG15489-PA open reading frame
DIP:20125N	CG1981-PA open reading frame, Thd1, G/T-mismatch-specific-thymine-DNA-glycosylase, double-stranded DNA-binding, mismatch repair
DIP:20084N	CG13363-PA open reading frame
DIP:19658N	CG12212-PA open reading frame, peb, (pebbled), transcription factor activity

DIP:17485N	CG10154-PA open reading frame, structural constituent of peritrophic membrane, (sensu Insecta)

TABLE II: Proteins found in the 2-orbit with nodes of degree 10, both duplicates (bold) are male specific RNA (with corresponding polypeptides).

DIP ID	Protein name/description
DIP:18704N	CG2789-PA open reading frame, bonzodiazepine receptor activity, transporter activity, metabolism and transport
DIP:18703N	CG1341-PA open reading frame, Rpt1, endopeptidase activity, ATPase activity, proteolysis and peptidolysis
DIP:21261N	CG3173-PA open reading frame
DIP:20398N	CG12096-PA open reading frame
DIP:17864N	CG10694-PA open reading frame damaged DNA-binding, base excision repair
DIP:20457N	CG12405-PA open reading frame, Prx2540-1, (Peroxiredoxin 2540), peroxidase, antioxidant activity, defense response, oxygen species metabolism
DIP:20458N	CG12896-PA open reading frame, peroxidase activity defense response, oxygen species metabolism
DIP:20623N	CG9624-PA open reading frame
DIP:17811N	CG5576-PA open reading frame, imd, (immune deficiency), antimicrobial humoral response, (sensu Invertebrata)
DIP:17346N	CG12470-PA open reading frame
DIP:18389N	CG18779-PA open reading frame
DIP:18387N	open reading frame CG-10530/4-PA, Lcp65Ag1/Lcp65Ag2 protein, (larval cuticle protein), structural constituent of larval cuticle, (sensu Insecta), larval cuticle biosynthesis, (sensu Insecta)
DIP:18392N	CG2082-PA open reading frame, signal transduction
DIP:18391N	CG16978-PA open reading frame
DIP:18390N	CG12907-PA open reading frame

TABLE III: Proteins found in a 2-orbit with nodes of degree 3, lines separate different orbits, bold proteins are duplicates.

DIP ID	Protein name/description
DIP:20847N	CG8284-PA open reading frame, UbcD4, (Ubiquitin conjugating enzyme 4), ubiquitin conjugating enzyme activity, ligase activity, protein metabolism, ubiquitin cycle
DIP:23198N	CG30344-PA open reading frame
DIP:18933N	CG10862-PA open reading frame, ubiquitin conjugating enzyme activity, ligase activity, protein metabolism
DIP:18935N	CG8974-PA open reading frame, transcription regulatory activity, "nucleo-metabolism", transcription,
DIP:18934N	CG32581-PA open reading frame, transcription regulatory activity, "nucleo-metabolism", transcription
DIP:20125N	CG1981-PA open reading frame, Thd1, G/T-mismatch-specific-thymine-DNA-glycosylase, double-stranded DNA-binding, mismatch repair
DIP:19658N	CG12212-PA open reading frame, peb, (pebbled), transcription factor activity
DIP:20084N	CG13363-PA open reading frame
DIP:17489N	CG11719-PA open reading frame, Mst98Ca, (Male-specific RNA 98Ca)
DIP:17490N	CG18396-PA open reading frame, Mst98Cb, (Male-specific RNA 98Cb)

TABLE IV: Proteins found in a 3-orbit with nodes of degree 2, lines separate different orbits, bold proteins are duplicates. Note, that Mst98Ca and Mst98Cb form a 2-orbit of degree 10, too (cf. Tab. II).

DIP ID	Protein name/description
DIP:19548N	CG31366/18743-PA open reading frame, Hsp70A, (Heat shock protein 70A), heat, defense response, protein complex assembly and folding
DIP:19549N	CG31449/31359/6489-PA open reading frame, Hsp70B, (heat shock protein 70B), heat, defense response, protein complex assembly and folding
DIP:19551N	open reading frame CG31449-PA, Hsp70Ba, (heat shock protein 70Ba), heat, defense response, protein complex assembly and folding
DIP:19552N	CG5834-PA open reading frame, Hsp70Bbb, (heat shock protein 70Bbb)
DIP:18493N	CG7945-PA open reading frame, chaperone activity
DIP:20272N	CG5203-PA open reading frame, CHIP, chaperone activity, protein folding and metabolism
DIP:20308N	CG32130-PA open reading frame
DIP:18578N	CG13165-PA open reading frame

TABLE V: Proteins found in the 4-orbit with nodes of degree 4, all duplicates (bold) are heat shock proteins (Hsp), released after heat shock or other stress.

DIP ID	Protein name/description
DIP:19748N	CG1252-PA open reading frame, Ccp84Ab, (cuticle cluster 7), structural constituent of larval cuticle, (sensu Insecta)
DIP:19749N	CG2360-PA open reading frame, Ccp84Aa, (cuticle cluster 8), structural constituent of larval cuticle, (sensu Insecta)

DIP:17392N	CG9949-PA open reading frame, sina, (seven in absentia), sensory organ development
DIP:18536N	CG6615-PA open reading frame, scaf6, RNA binding, nuclear mRNA splicing via spliceosome, spliceosome complex
DIP:20225N	CG2341-PA open reading frame, Ccp84Ad, (cuticle cluster 5), structural constituent of larval cuticle, (sensu Insecta)
DIP:17713N	CG15422-PA open reading frame
DIP:17492N	CG12723-PA open reading frame
DIP:17076N	CG6945-PA open reading frame
DIP:17488N	CG11505-PB open reading frame

TABLE VI: Proteins found in the 2-orbit with nodes of degree 7, duplicates (bold) are constituents of the larval cuticle. Note, that peb, Thd1, and CG13363-PA also form a 3-orbit with respect to Mst98Ca and Mst98Cb (cf. Tab. IV).

[1] S. H. Strogatz, Nature **410**, 268 (2001).
[2] B. Bollobás, *Modern Graph Theory* (Springer, New York, 1998).
[3] B. Bollobás, *Random Graphs* (Cambridge University Press, Cambridge, 2001).
[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[5] F. Chung, *Spectral Graph Theory*, vol. 92 of *Regional Conference Series in Mathematics* (American Mathematical Society Providence, Rhode Island, 1994).
[6] D. Cvetković, M. Doob, and H. Sachs, *Spectra of graphs*, Pure and applied mathematics (Academic Press, 1980).
[7] D. Cvetković, M. Doob, I. Gutman, and A. Torgašev, *Recent results in the theory of graph spectra*, no. 36 in Annals of discrete mathematics (North Holland, 1988).
[8] N. Biggs, *Algebraic graph theory* (Cambridge University Press, Cambridge, 1996).
[9] M. Bauer and O. Golinelli, J. Stat. Phys. **103**, 301 (2001).
[10] I. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek, Phys. Rev. E **64**, 026704 (2001).
[11] S. Dorogovtsev, A. Goltsev, J. Mendes, and A. Samukhin, Phys. Rev. E **68**, 046109 (2003).
[12] F. Chung, L. Lu, and V. Vu, Annals of Combinatorics **7**, 21 (2003).
[13] F. Chung, L. Lu, and V. Vu, Proc. Natl. Acad. Sci. USA **100**, 6313 (2003).
[14] M. Mihail and C. Papadimitriou, in *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques* (2002), pp. 254–262.
[15] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. E **64**, 051903 (2001).
[16] O. Golinelli, http://arxiv.org/abs/cond-mat/0301437v1 (2003).
[17] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[18] S. Milo, R. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004).
[19] L. Giot, J. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. Hao, C. Ooi, B. Godwin, E. Vitols, et al., Science **302**, 1727 (2003).
[20] *Database of Interacting Proteins (DIP)*, URL `http://dip.doe-mbi.ucla.edu/`.
[21] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[22] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, `Mfinder` *Software*, URL `http://www.weizmann.ac.il/mcb/UriAlon/groupDownloadableData`
[23] D. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs*, Pure and Applied Mathematics (Academic Press, 1979).
[24] S. Ohno, *Evolution by gene duplication* (Springer, Berlin, 1970).
[25] S. Teichmann and M. Babu, Nature Genetics **36**, 492 (2004).
[26] G. Amoutzias, D. Robertson, S. Oliver, and E. Bornberg-Bauer, EMBO **5**, 1 (2004).
[27] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, Bioinformatics (2004), in press.
[28] *FlyBase, A Database of the Drosophila Genome*, URL `http://flybase.bio.indiana.edu/`.
[29] M. Samanta and S. Liang, Proc. Natl. Acad. Sci. USA **100**, 12579 (2003).
[30] S. Kirkpatrick and T. Eggarter, Phys. Rev. B **6**, 3598 (1972).
[31] S. Evangelou, Phys. Rev. B **27**, 1397 (1983).
[32] A similar phenomenon has earlier been discussed in the context of a model of quantum percolation [30, 31].
[33] An eigenvalue is considered to belong to the discrete spectrum if there exists at least one other eigenvalue that does not deviate more than 10^{-12}.