ALGEBRAIC AND TOPOLOGICAL INDICES OF MOLECULAR PATHWAY NETWORKS IN HUMAN CANCERS

Peter Hinow*

Department of Mathematical Sciences
University of Wisconsin - Milwaukee
P.O. Box 413, Milwaukee, WI 53201, USA

Edward A. Rietman

Newman-Lakka Institute
Tufts University School of Medicine
Boston, MA 02111, USA

Sara Ibrahim Omar

Cross Cancer Institute
University of Alberta
Edmonton, T6G 2E1, Canada

Jack A. Tuszyński

Cross Cancer Institute and Department of Physics
University of Alberta
Edmonton, T6G 2E1, Canada

Abstract. Protein-protein interaction networks associated with diseases have gained prominence as an area of research. We investigate algebraic and topological indices for protein-protein interaction networks of 11 human cancers derived from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We find a strong correlation between relative automorphism group sizes and topological network complexities on the one hand and five year survival probabilities on the other hand. Moreover, we identify several protein families (e.g. PIK, ITG, AKT families) that are repeated motifs in many of the cancer pathways. Interestingly, these sources of symmetry are often central rather than peripheral. Our results can aide in identification of promising targets for anti-cancer drugs. Beyond that, we provide a unifying framework to study protein-protein interaction networks of families of related diseases (e.g. neurodegenerative diseases, viral diseases, substance abuse disorders).

1. **Introduction.** Biological networks have been an active area of research for some years, see e.g. [12, 17, 18] and the references therein. In earlier work [5] we reported that molecular signaling network complexity is correlated with cancer patient survival. In that work we reported the degree entropy, a statistical mechanics measure of network complexity, see also the related work of Takemoto and Kihara [23]. Here

we focus on the relative sizes of automorphism groups and the dimensions of the cycle spaces (cyclomatic numbers).

Complex real-world networks contain feedback loops to enable the network "communication" to continue in the face of node failure [1]. In the case of protein-protein interaction (PPI) networks this means that inhibition of a specific node may or may not have any effect. It is well known that targeting hub nodes in networks often causes the network to break up into multiple components and this could be lethal, because many protein hubs in PPIs for cancer are also important proteins in metabolic networks. As we argued in [5], targeting nodes with high-betweenness has higher potential for improved cancer treatment. Selective targeting of nodes in a PPI for cancer treatment is fraught with difficulties.

In this paper we apply two more algebraic and topological indices to study cancer PPI networks and show correlation with 5 year patient survival. In addition, we report the symmetry groups and orbit equivalence classes themselves. We identify several repeated motifs of proteins that are "interchangeable" in a sense to be specified below. In the long run we anticipate that the methods described here will aid identification of potential drug targets. Group theory has been extremely useful in physics and chemistry to study symmetries of crystals, molecules and elementary particles. As suggested earlier in [21], we believe that it also has a very valuable role to play in biology.

2. **Results.** A *network* is an undirected graph $G = (V, E)$ with vertex set V and edge set E. When applying this terminology to cancer networks, the vertices are proteins and two vertices are connected by an edge if there is a known interaction of the two partners, either by direct binding or by enzymatic catalysis. We emphasize that all graphs are undirected, which is, of course, a simplification. Beyond cancer pathways, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database also contains pathways related to immune diseases (e.g. asthma), neurodegenerative diseases (Alzheimer's disease, Parkinson's disease), substance dependence, cardiovascular diseases, viral diseases and many others [13, 14]. The KEGG networks are assembled from the literature by searches for experimental confirmation of the relevant interactions. Each interaction is always confirmed by two or more different experimental techniques such as pull-down mass spectrometry, yeast two-hybrid and various biochemical tests. Naturally, networks constructed from experimental results are likely to contain errors, which are however impossible to quantify. There is no correlation between low survival rate in a cancer and heightened research interest and increased knowledge about it [5], so we can rule out a systematic bias.

An *automorphism* is a permutation $\phi : V \to V$ that preserves the adjacency relation, that is,

$$(u, v) \in E \ \Leftrightarrow (\phi(u), \phi(v)) \in E.$$

With the operation of composition, the automorphisms form a group $Aut(G)$. The relation on the set of vertices

$$u \sim v \ \Leftrightarrow \ \text{there exists a } \phi \in Aut(G) \text{ such that } v = \phi(u)$$

is an equivalence relation and its equivalence classes are called the (group) orbits.

MacArthur *et al.* [19] list 20 examples of real world networks and their rich symmetry groups. This is in contrast to large random graphs, such as graphs from the Erdős-Rényi model $\mathcal{G}(n, p)$. Here n is the number of vertices. Edges are independently present with probability $0 < p < 1$. Such graphs have only the trivial

automorphism, with probability approaching one, in the limit $n \to \infty$ [4, Chapter IX]. The difference is not surprising if one realizes that real networks display a modular structure, with vertices organized in communities tightly connected internally and loosely connected to each other [8]. This results in the presence of symmetric subgraphs such as trees and complete cliques.

Figure 1 shows as an example the protein-protein interaction network of pancreatic cancer as retrieved from the KEGG database. We find that the automorphism group of this network is isomorphic to the direct product of symmetric groups (groups of permutations of n symbols)

$$Aut(G) \cong S_2^9 \times S_3^6 \times S_5^2 \times S_8,$$

see Table 1 for a complete list of automorphism groups. Tables S1-S11 contain the orbit equivalence classes of the respective groups (Supplementary Material). Remarkably, symmetries do not only arise due to tree subgraphs at the "ends" of the network, but also due to "central" nodes, that have high degree and whose neighbors do not have mostly degree one (highlighted in yellow in Figure 1). Thus any flow of information that passes through one node in such an orbit equivalence class may pass through any other node in the same equivalence class. The presence of such modular patterns indicates a high level of redundancy which confers robustness to the associated biological system (tumor cells). We suggest that to interrupt the flow through such a network most efficiently, the nodes adjacent to large central orbits are the best to be targeted for example by pharmacological agents that inhibit a specific protein-protein interaction pair. Similar suggestions have been made in [6, 7, 25]. The use of automorphism groups has, to the best of our knowledge, not yet been proposed.
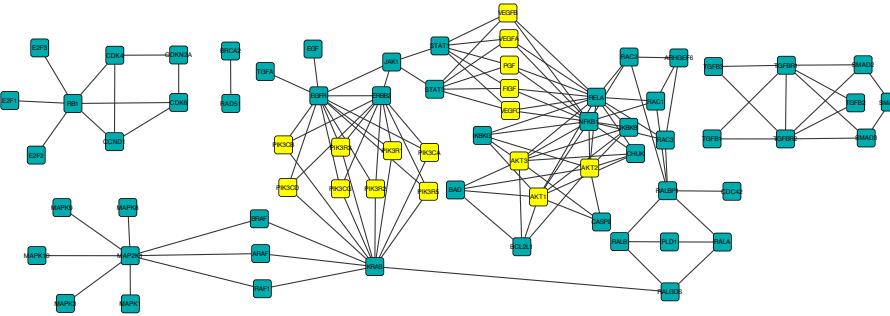

FIGURE 1. *The protein-protein interaction network of pancreatic cancer.* The network was retrieved from the KEGG database [13, 14] and contains 69 vertices and 137 edges. Its automorphism group was determined with `saucy` [15], namely $Aut(G) = S_2^9 \times S_3^6 \times S_5^2 \times S_8$. Highlighted in yellow are three central orbits of nodes of degrees 3, 4 and 8, respectively. Two of these are the PI3K (eight yellow nodes) and the AKT families (three yellow nodes), respectively.

Automorphism groups are often used to measure the complexity of a network [26]. In order to make automorphism group sizes of graphs with $n = |V|$ vertices comparable, we use the suggested normalization from [26] and compute the ratio

$$\beta_G = \left(\frac{|Aut(G)|}{n!} \right)^{\frac{1}{n}}.$$

This relates the size of $Aut(G)$ to the size of the automorphism group S_n of the complete graph on n vertices.

Cancer	n	$Aut(G)$	$|Aut(G)|$	c	β_G	μ_G
AML	62	$S_2^7 \times S_3^4 \times S_4 \times S_7^2$	$1.011 \cdot 10^{14}$	1	$7.03 \cdot 10^{-2}$	108
Bld	29	$S_2^5 \times S_3^3 \times S_4$	$1.658 \cdot 10^5$	3	$1.29 \cdot 10^{-1}$	20
CML	73	$S_2^8 \times S_3^7 \times S_4 \times S_5^2 \times S_8$	$9.98 \cdot 10^{17}$	3	$6.3 \cdot 10^{-2}$	115
Col	49	$S_2^5 \times S_3^4 \times S_4^2 \times S_8$	$9.63 \cdot 10^{11}$	3	$9.2 \cdot 10^{-2}$	58
End	44	$S_2^5 \times S_3^4 \times S_4 \times S_5 \times S_8$	$4.82 \cdot 10^{12}$	2	$1.1 \cdot 10^{-1}$	45
Gli	64	$S_2^8 \times S_3^6 \times S_4^4 \times S_6$	$2.85 \cdot 10^{15}$	2	$7.0 \cdot 10^{-2}$	128
NSCL	48	$S_2^7 \times S_3^5 \times S_4^3$	$1.37 \cdot 10^{10}$	1	$8.6 \cdot 10^{-2}$	75
Pc	69	$S_2^9 \times S_3^6 \times S_5^2 \times S_8$	$1.39 \cdot 10^{16}$	4	$6.4 \cdot 10^{-2}$	72
Ren	60	$S_2^8 \times S_3^2 \times S_6 \times S_8 \times S_{18}$	$1.7 \cdot 10^{27}$	3	$1.2 \cdot 10^{-1}$	52
SCL	77	$S_2^4 \times S_3^4 \times S_4 \times S_6 \times S_8 \times S_{12} \times S_{18}$	$4.4 \cdot 10^{37}$	2	$1.0 \cdot 10^{-1}$	150
Thyr	28	$S_2^4 \times S_3^2 \times S_4 \times S_7$	$7.0 \cdot 10^7$	3	$1.7 \cdot 10^{-1}$	24

TABLE 1. Automorphism groups of all cancers. Column n contains the number of vertices, column c contains the number of connected components of the protein-protein interaction network. The abbreviations are the same as in Figure 3.

A second graph invariant is of topological nature. A *cycle* (or *Eulerian subgraph*) is a sequence of adjacent edges that starts and ends at the same vertex and uses no edge twice. Note that a cycle need not be connected. The set of all cycles $\mathcal{C}(G)$ can be made a vector space over the field $\mathbb{Z}_2$ as follows. The sum of cycles C_1 and C_2 is the symmetric difference

$$C_1 \Delta C_2 = (C_1 \cup C_2) \setminus (C_1 \cap C_2)$$

(of the edge sets), the negation is $-C = C$, zero is the empty cycle and $1C = C$, $0C = \varnothing$. The dimension of this vector space is called the *cyclomatic number* μ_G, or the *circuit rank*. Loosely speaking, it is a count of the "independent" loops, see Figure 2. It is shown in [3, 16] that for a graph with n vertices, m edges and c connected components, μ_G is given by

$$\mu_G = m - n + c.$$

The quantities β_G and μ_G are listed in Table 1 for all cancers under consideration.

We plot these two indices against the five year survival probability p, obtained from the Surveillance, Epidemiology and End Results (SEER) database [20] for 11 types of cancer in Figure 3. Interaction networks with larger values of β_G or equivalently greater symmetry are associated with better chances of survival. A large value of μ_G indicates high topological complexity and correlates with decreased chance of survival. We find that both coefficients of determination of a linear fit are $R^2 = 0.52$ with corresponding P-value $P = 0.011$ (the equality is coincidental). These coefficients of determination are comparable to those reported by Breitkreutz *et al.* [5] and Takemoto and Kihara [23] for the modularity of the cancer signaling networks. A fit of the survival probabilities to uniformly distributed random numbers on the same range as β_G results in $R^2 = 0.006$, Supplementary Figure S1. There are widespread differences in detection stage, metastasis status, treatment and general health of the patient that are unfortunately not accessible from the SEER database. Nevertheless, given this large amount of natural uncertainty in the data, this indicates a strong correlation of averages. It would be invaluable for future research to classify database entries according to some of the parameters
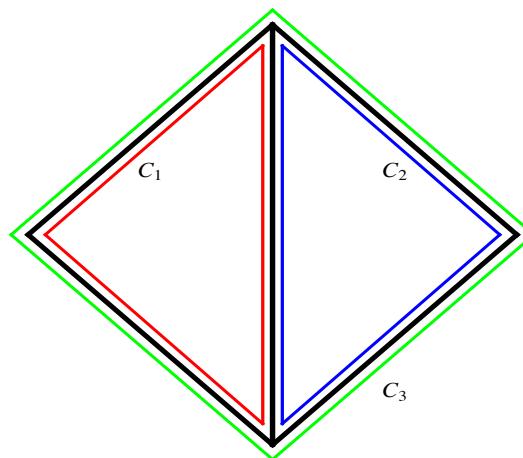
FIGURE 2. *The cycle space of a simple graph with four vertices.*
The graph G is shown in black and three cycles C_1, C_2 and C_3
are marked in red, blue and green, respectively. Any of the sets
$\{C_1, C_2\}$, $\{C_1, C_3\}$ and $\{C_2, C_3\}$ is a basis of the cycle space $\mathcal{C}(G) =$
$\{\varnothing, C_1, C_2, C_3\}$.

mentioned above. Since both the automorphism group size β_G and the cyclomatic number μ_G are correlated to the five year survival probability p, it is to be expected that these two quantities are correlated to each other, for the the protein-protein interaction networks of cancers that are the object of our study, see Figure 4, left panel. However, it is easy to construct examples of graphs that show no correlation between β_G and μ_G, see Figure 4, right panel.

3. **Discussion.** The nodes of the orbit equivalence class of the highest degree in each cancer were analyzed for the protein functions using the the Database for An-notation, Visualization and Integrated Discovery (DAVID) [10, 11]. This database provides bioinformatics tools that identify and characterize several aspects of genes and their encoding proteins including gene ontology (GO) terms, protein molecular functions and their interacting proteins. It also helps to visualize KEGG pathway maps of genes. We suggest that orbit-equivalent nodes of highest degree are the most significant to the respective type of cancer, since they provide the greatest possible redundancies in a signaling pathway and hence confer robustness on the cancer cell [8]. Below we discuss in detail each cancer analyzed and how the various member proteins of an equivalence class are related to one another.

KRAS, NRAS and HRAS form the orbits of the highest degree for thyroid cancer, AML and glioma with degrees 8, 12 and 20, respectively. These proteins have GTPase activity and they all play a role in integrin affinity modulation. AKT1, AKT2 and AKT3 are the equivalent nodes of highest degree for colorectal cancer, NSCL and CML with degrees 8, 12 and 14, respectively. These proteins belong to a class of serine/threonine protein kinases and are involved in many pathways including the MAPK signaling, ErbB signaling and apoptosis. All three proteins have the same molecular functions such as nucleotide and ATP binding. However, AKT1 has additional molecular functions as a nitric oxide synthase and can bind lipids and enzymes. In bladder cancer, MAPK1 and MAPK3 are equivalent nodes
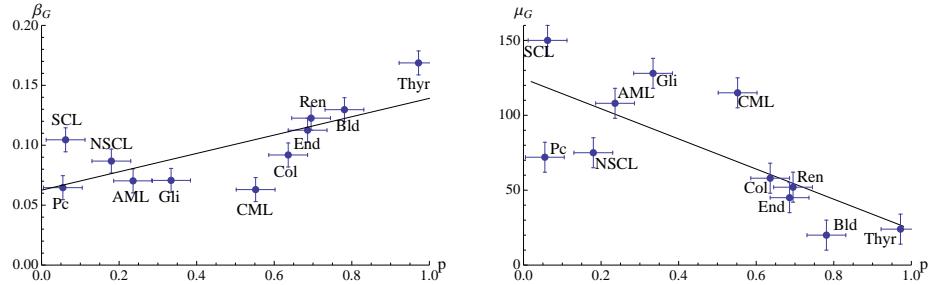
FIGURE 3. *Automorphism group sizes and cyclomatic numbers.* A plot of relative automorphism group size β_G (left) and cyclomatic number μ_G (right) against five year survival probability p for 11 types of cancer. In both cases we have the coefficients of determination of the linear fits $R^2 = 0.52$ and P-value $P = 0.011$. The cancer types are *AML*: acute myelogeneous leukemia, *Bld*: bladder cancer, *CML*: chronic myelogeneous leukemia, *Col*: colorectal cancer, *End*: endometrial cancer, *Gli*: Glioma, *NSCL*: non-small cell lung cancer, *Pc*: pancreatic cancer, *Ren*: renal cancer, *SCL*: small cell lung cancer, *Thyr*: thyroid cancer. The width of the horizontal error bars is 0.1, the width of the vertical error bars is 0.02 in the left panel and 20 in the right panel. These errors are estimated as the actual error is unknown.
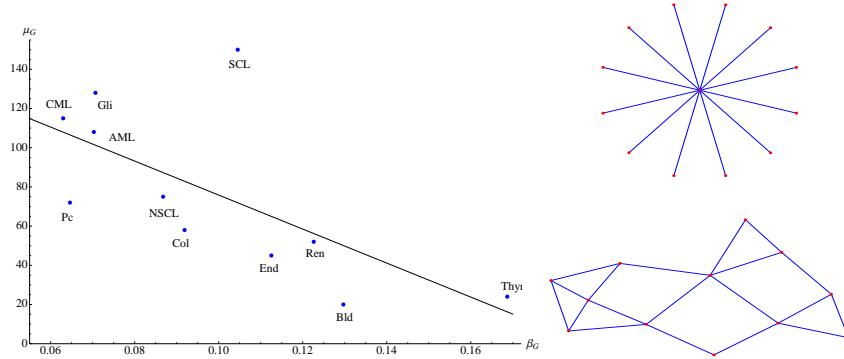
FIGURE 4. *Mutual relation between β_G and μ_G.* (Left) A plot of cyclomatic number μ_G against relative automorphism group size β_G for the 11 protein-protein interaction networks of cancers. (Right) In general, these quantities are not related. Shown are a highly symmetric graph with trivial cycle space (top, $\beta_G = \left(\frac{1}{13}\right)^{\frac{1}{13}} = 0.82$, $\mu_G = 0$) and an asymmetric graph with large cycle space (bottom, $\beta_G = 0$, $\mu_G = 19 - 12 + 1 = 8$).

of degree 6. Since both belong to the same family of kinase-related transforming proteins, they are involved in the same pathways including the MAPK, mTOR and chemokine pathways. MAPK1 has additional functions by being able to bind to phosphoproteins.

RELA and NFκB1 are the nodes of highest degree in pancreatic cancer and SCL (14 and 13, respectively). It is interesting to note that the proteins are involved in the same pathways despite the fact that they belong to two different families. They both share common molecular functions by being able to bind to DNA and have transcription regulation activities. RELA has additional functions such as ion and enzyme binding. On the other hand, NFκB1 can bind to promoters and has protein dimerization activity.

In endometrial cancers, the equivalent PDPK1 and ILK proteins have the highest degree 10. Although both proteins share common molecular functions, ILK can bind to integrin, SH3 domains and has protein tyrosine kinase activity. PDPK1, however, can bind to insulin receptors and enzymes, including kinases, and has 3-phosphoinositide-dependent protein kinase activity. Interestingly, they are both part of the peroxisome proliferator-activated receptors (PPAR) signaling pathway and they play a role in focal adhesion processes. Additionally, PDPK1 plays a role in other pathways such as mTOR signaling, insulin signaling and aldosterone-regulated sodium reabsorption. Finally, EPAS1 and HIF1A, have the highest degree 18 in the renal cancer network. These proteins can bind DNA, enzymes, transcription factors and histone acetyltransferases, while only HIF1A can bind to Hsp90. The two proteins also play roles such as RNA polymerase II transcription factor, in transcription regulation as well as protein dimerization.

Further study of the automorphism groups reveals repeated motifs in several interaction networks. The eight proteins from the PIK3C{A,B,D,G} and PIK3R{1,2,3,5} family form a single orbit equivalence class in seven of the networks (AML, CML, colorectal, endometrial, pancreatic, renal and SCL cancers) and are split in two orbit equivalence class in two more networks (glioma and NSCL). The three proteins AKT{1,2,3} are orbit equivalent in eight networks (CML, colorectal, endometrial, glioma, NSCL, pancreatic, renal and SCL cancers), that is, whenever they appear in the network to begin with. These players have been known for a long time to be of crucial importance to the initiation and progression of cancer, mainly due to the various biological and biochemical assays performed on cancer cells. However, our conclusions stem directly from a group-theoretic analysis of the PPI networks and they are network specific. Since its initial discovery as a proto-oncogene, the serine/threonine kinase AKT has become a major focus of attention because of its critical regulatory role in diverse cellular processes, including cancer progression and insulin metabolism. The AKT cascade is activated by receptor tyrosine kinases, integrins, B and T cell receptors, cytokine receptors, G-protein-coupled receptors and other stimuli that induce the production of phosphatidylinositol (3,4,5)-triphosphates (PtdIns(3,4,5)P3) by phosphoinositide 3-kinase (PI3K). These lipids serve as plasma membrane docking sites for proteins that harbor pleckstrin-homology (PH) domains, including AKT and its upstream activator PDK1. The tumor suppressor PTEN is recognized as a major inhibitor of AKT and is frequently lost in human tumors. There are three highly related isoforms of AKT (AKT1, AKT2, and AKT3), which represent the major signaling arm of PI3K. For example, germline mutations of AKT have been identified in pathological conditions of cancer and insulin metabolism. AKT regulates cell growth through its effects on the TSC1/TSC2 complex and mTOR pathways, as well as cell cycle and cell proliferation through its direct action on the CDK inhibitors p21 and p27, and its indirect effect on the levels of cyclin D1 and p53. AKT is a major mediator of cell survival through direct inhibition of pro-apoptotic signals such as

the pro-apoptotic regulator BAD and the FOXO and Myc family of transcription factors. AKT has been demonstrated to interact with Smad molecules to regulate TGF-β signaling. These findings make AKT an important therapeutic target for the treatment of cancer.

Interestingly, the network of small cell lung cancer contains an enormous orbit of 18 equivalent nodes of degree six. This orbit consists of laminines, collagens and a fibronectin that are major proteins in the basal lamina. All nodes are connected to six members of the integrin family of transmembrane receptors.

4. **Conclusion.** We have shown that the relative size of the automorphism groups and the cyclomatic numbers for cancer pathway networks from the KEGG database are both correlated with five-year survival of cancer patients. This is in line with similar earlier results by Breitkreutz *et al.* [5] and Takemoto and Kahira [23] on degree entropy and network modularity. Determination of the specific reasons for these great discrepancies in survival rates remains a topic for future research. Interestingly, cancers with more symmetric interaction networks are associated with better survival rates. This may be due to a greater robustness to failure, which, somewhat counterintuitive, is a positive feature in this context.

We suggest that selective removal of nodes from the network (clinically equivalent to protein inhibition) and reinterpolation on the linear curves helps to identify potential drug targets. This indicates that complexity of a biochemical network involved in a deregulated cell cycle as exemplified by cancer cells is of crucial importance to its robustness. This is manifested by various redundancies in the PPI network that make the search for a therapeutic "silver bullet" an impossible task. We suggest that selective removal of nodes from the network (clinically equivalent to protein inhibition) and re-interpolation on the linear curves helps to identify potential drug targets. We have shown that PI3K and AKT families of proteins appear to be the most suitable targets for pharmacological inhibition in the most number of cancer types studied. It is encouraging that there are several AKT pathway inhibitors in clinical development, e.g. perifosine (KRX-0401, Aeterna Zentaris/Keryx), MK-2206 (Merck), and GSK-2141795 (Glaxo-SmithKline) [2]. Similarly, Bayer, GlaxoSmithKline (GSK), Novartis, Merck & Co., Roche and Sanofi are just a few of the companies that have placed great importance on the development of a spectrum of agents targeting the PI3K pathway. Drug candidates including pan-PI3K inhibitors, PI3K isoform-specific inhibitors, AKT inhibitors and mTOR inhibitors are currently tested alone and in combinations in an array of cancer indications [9]. While the motivation for this focus has been stated as: "The pathway is almost invariably on in cancer", our methodology identifies this pathway as the most crucial using mathematical analysis of the network. Moreover, we are able to identify those types of cancer where the pathway should be the main target and those types where targeting it may not produce the expected clinical outcomes.

Methods. The cancer pathways were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [14] with the help of the open source software packages KEGGgraph [27] and cytoscape [22]. The automorphism groups of the networks were found with saucy [15] and gap [24] (see Supplementary Material, Tables S1-S11 for the complete group lists). Bases of the cycle spaces were found using python networkX. The DAVID bioinformatics database [10, 11] was used to analyze the equivalent nodes (proteins) of highest degree. The corresponding genes were obtained using their official gene symbol and were limited to *Homo sapiens* species.

REFERENCES

[1] R. Albert, H.Jeong and A. L. Barabási, Error and attack tolerance of complex networks, *Nature*, **406** (2000), 378–382.

[2] W. Alexander, Inhibiting the Akt pathway in cancer treatment, Three leading candidates, *Pharmacy and Therapeutics*, **36** (2011), 225–227.

[3] F. Berger, P. Gritzmann S. de Vries, Minimum cycle bases for network graphs, *Algorithmica*, **40** (2004), 51–62.

[4] B. Bollobás, *Random Graphs*, Cambridge University Press, Cambridge, 2001.

[5] D. Breitkreutz, L. Hlatky, E. Rietman, J. A. Tuszynski, Molecular signaling network complexity is correlated with cancer patient survivability, *Proc. Natl. Acad. Sci. USA*, **109** (2012), 9209–9212.

[6] N. Chandra and J. Padiadpu, Network approaches to drug discovery, *Expert Opin. Drug Discov.*, **8** (2013), 7–20.

[7] P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London, R. Nussinov, Structure and dynamics of molecular networks: A novel paradigm of drug discovery. A comprehensive review, *Pharmacol. Therapeut.*, **138** (2013), 333–408.

[8] D. Garlaschelli, F. Ruzzenenti and R. Basosi, Complex networks and symmetry I: A review, *Symmetry*, **2** (2010), 1683–1709, arXiv:1006.3923.

[9] D. Holmes, PI3K pathway inhibitors approach junction, *Nat. Rev. Drug Discov.*, **10** (2011), 563–564.

[10] D. W. Huang, B. T. Sherman and R. A. Lempicki, Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acid Res.*, **37** (2009), 1–13.

[11] D. W. Huang, B. T. Sherman and R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature Protocols*, **4** (2009), 44–57, http://david.abcc.ncifcrf.gov/.

[12] B. H. Junker, F. Schreiber (editors), *Analysis of Biological Networks*, John Wiley & Sons, Hoboken, NJ, 2008.

[13] M. Kanehisa and S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acid Res.*, **28** (2000), 23–30, http://www.genome.jp/kegg/.

[14] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno and M. Hattori, The KEGG resource for deciphering the genome, *Nucleic Acid Res.*, **32** (2004), D277–D280.

[15] H. Katebi, K. A. Sakallah and I. L. Markov, Graph symmetry detection and canonical labeling: Differences and synergies, In A Voronkov, editor, *The Alan Turing Centenary*, pages 181–195. EasyChair, 2012, http://vlsicad.eecs.umich.edu/BK/SAUCY/.

[16] T. Kavitha, C. Liebchen, K. Mehlhorn, D. Michail, R. Rizzie, T. Ueckerdt and K. A. Zweig, Cycle bases in graphs characterization, algorithms, complexity, and applications, *Computer Science Review*, **3** (2009), 199–243.

[17] E. V. Koonin, Y. I. Wolf and G. P. Karev (editors), *Power Laws, Scale-Free Networks and Genome Biology*, Landes Bioscience, Austin, TX, 2006.

[18] A. Ma'ayan and B. D. MacArthur (editors), *New Frontiers of Network Analysis in Systems Biology*, Springer Verlag, Dordrecht, Heidelberg, New York, London, 2012.

[19] B. D. MacArthur, R. J. Sánchez-García and J. W. Anderson, Symmetry in complex networks, *Discr. Appl. Math.*, **156** (2008), 3525–3531.

[20] National Cancer Institute, Surveillance, Epidemiology and End Results (SEER) Program, 2013, http://seer.cancer.gov/.

[21] E. A. Rietman, R. L. Karp and J. A. Tuszynski, Review and application of group theory to molecular systems biology, *Theor. Biol. Med. Model.*, **8** (2011), p21.

[22] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13** (2003), 2498–2504. http://cytoscape.org.

[23] K. Takemoto and K. Kihara, Modular organization of cancer signaling networks is associated with patient survivability, *BioSystems*, **113** (2013), 149–154.

[24] The GAP Group, *GAP – Groups, Algorithms, and Programming*, University of St Andrews, St Andrews, United Kingdom, 2013. http://www.gap-system.org.

[25] W. Winterbach, P. van Mieghem, M. Reinders, H. Wang and D. de Ridder, Topology of molecular interaction networks, *BMC Syst. Biol.*, **7** (2013), p90.

[26] Y. Xiao, B. D. MacArthur, H. Wang, M. Xiong and W. Wang, Network quotients: Structural skeletons of complex systems, *Phys. Rev. E*, **78** (2008), 046102, arXiv:0802.4318.

[27] J. D. Zhang and S. Wiemann, KEGGgraph: A graph approach to KEGG PATHWAY in R and Bioconductor, *Bioinformatics*, **25** (2009), 1470–1471, http://bioconductor.org.

E-mail address: hinow@uwm.edu

E-mail address: erietman@gmail.com

E-mail address: sara.ibrahim.omar@gmail.com

E-mail address: jack.tuszynski@gmail.com

Supplementary material.
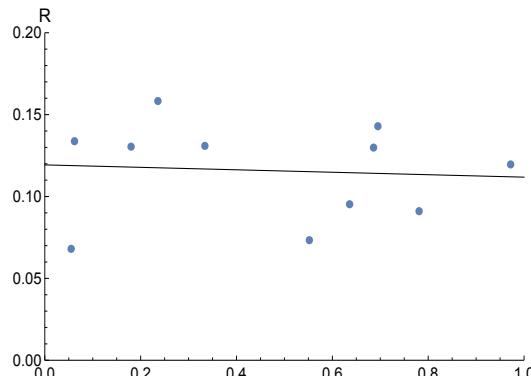
FIGURE S1. *Random data assigned to survival probabilities.* The 11 survival probabilities are assigned uniformly distributed random numbers in the same range as β_G. The coefficient of determination of the linear fit is $R^2 = 0.006$.

Nodes							Degree
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3R1	PIK3R2	PIK3R3	9
CCND1	MYC	PPARD	LEF1	TCF7	TCF7L2	TCF7L1	1
KRAS	NRAS	HRAS					12
RPS6KB1	RPS6KB2	EIF4EBP1					5
STAT3	STAT5A	STAT5B					5
RARA	PML	LOC652346	LOC652671				3
ARAF	RAF1	BRAF					1
KIT	LOC652799						12
IKBKB	IKBKG						5
MAP2K1	MAP2K2						5
SOS1	SOS2						4
PIM1	PIM2						3
NFKB1	RELA						3
MAPK1	MAPK3						2

TABLE S1. Orbit equivalence classes for the AML group.

Nodes								Degree
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	4
RHOA	RAC1	RAC2	RAC3					4
LEF1	TCF7	TCF7L1	TCF7L2					4
AKT1	AKT2	AKT3						11
MAPK8	MAPK9	MAPK10						4
ARAF	RAF1	BRAF						2
SMAD4	TGFBR1	TGFBR2						2
MAPK1	MAPK3							4
JUN	FOS							3
DCC	CASP3							3
SMAD2	SMAD3							3

TABLE S4. Orbit equivalence classes for the colorectal cancer group.

Nodes				Degree
DAPK1	DAPK3	DAPK2	RPS6KA5	2
KRAS	NRAS	HRAS		5
ARAF	RAF1	BRAF		5
E2F1	E2F2	E2F3		1
MAPK1	MAPK3			6
MAP2K1	MAP2K2			5
RASSF1	FGFR3			3
CCND1	CDK4			3
EGFR	ERBB2			1

TABLE S2. Orbit equivalence classes for the bladder cancer group.

Nodes								Degree
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	9
CRK	CRKL	CBLC	CBL	CBLB				9
SHC2	SHC4	SHC3	SHC1	MYC				2
HDAC1	HDAC2	CTBP1	CTBP2					2
AKT1	AKT2	AKT3						14
ARAF	RAF1	BRAF						5
KRAS	NRAS	HRAS						5
CHUK	IKBKB	IKBKG						4
CDK4	CDK6	CCND1						3
E2F1	E2F2	E2F3						1
ABL1	BCR							13
MAP2K1	MAP2K2							5
SOS1	SOS2							4
MECOM	RUNX1							4
STAT5A	STAT5B							3
MAPK1	MAPK3							2
TGFRB1	TGFRB2							2
NFKB1	RELA							1

TABLE S3. Orbit equivalence classes for the CML group.

Nodes								Degree
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	3
APC2	APC	AXIN1	AXIN2	GSK3B				1
LEF1	TCF7	TCF7L1	TCF7L2					3
AKT1	AKT2	AKT3						5
KRAS	NRAS	HRAS						5
ARAF	RAF1	BRAF						5
BAD	CASP9	FOXO3						3
PDPK1	ILK							10
MAP2K1	MAP2K2							5
MAPK1	MAPK3							3
SOS1	SOS2							4
MYC	CCND1							4
DCC	CASP3							3

TABLE S5. Orbit equivalence classes for the endometrial cancer group.

Nodes						Degree
CALML6	CALML5	CALM1	CALM2	CALM3	CALML3	3
PIK3CA	PIK3CB	PIK3CD	PIK3CG			9
CAMK2A	CAMK2B	CAMK2D	CAMK2G			8
PIK3R5	PIK3R1	PIK3R2	PIK3R3			7
SHC1	SHC2	SHC3	SHC4			5
KRAS	NRAS	HRAS				20
ARAF	RAF1	BRAF				5
PRKCA	PRKCB	PRKCG				3
CCND1	CDK4	CDK6				3
AKT1	AKT2	AKT3				1
E2F1	E2F2	E2F3				1
PDGFRA	PDGFRB					16
MAP2K1	MAP2K2					9
PLCG1	PLCG2					4
SOS1	SOS2					4
MDM2 ↔ TP53 (2)			CDKN2A ↔ CDKN1A (4)			-
MAPK1	MAPK3					2
PDGFA	PDGFB					2
EGF	TGFA					1

TABLE S6. Orbit equivalence classes for the glioma group. The
nodes MDM2 and TP53 can only be permuted simultaneously with
the nodes CDKN1A and CDKN2A.

Nodes				Degree
PIK3CA	PIK3CB	PIK3CD	PIK3CG	7
PIK3R5	PIK3R1	PIK3R2	PIK3R3	5
PLCG1	PLCG2	TGFA	EGF	2
AKT1	AKT2	AKT3		12
ARAF	RAF1	BRAF		8
PRKCA	PRKCB	PRKCG		3
CCND1	CDK4	CDK6		3
BAD	CASP9	FOXO3		3
E2F1	E2F2	E2F3		1
EGFR	ERBB2			9
NRAS	HRAS			5
MAP2K1	MAP2K2			4
SOS1	SOS2			4
MAPK1	MAPK3			3
STK4	RASSF5			2
CDK4	CDK6			2

TABLE S7. Orbit equivalence classes for the NSCL group.

Nodes								Degree
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	3
FIGF	PGF	VEGFA	VEGFB	VEGFC				4
MAPK1	MAPK3	MAPK8	MAPK9	MAPK10				1
AKT1	AKT2	AKT3						8
CHUK	IKBKB	IKBKG						5
RAC1	RAC2	RAC3						4
ARAF	RAF1	BRAF						2
TGFB1	TGFB2	TGFB3						2
E2F1	E2F2	E2F3						1
RELA	NFKB1							14
STAT1	STAT3							6
TGFBR1	TGFBR2							6
BAD	BCL2L1							4
RALA	RALB							3
CDK4	CDK6							3
SMAD2	SMAD3							3
BRCA2	RAD51							1
TGFA	EGF							1

TABLE S8. Orbit equivalence classes for the pancreatic cancer group.

Nodes								Degree
ARNT	ARNT2	CREBBP	EP300	EGLN2	EGLN3	EGLN1	SLC2A1	
FIGF	PGF	VEGFA	VEGFB	VEGFC	TGFB1	TGFB2	TGFB3	2
PDGFB	TGFA							
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	4
PAK1	PAK2	PAK3	PAK4	PAK6	PAK7			2
AKT1	AKT2	AKT3						8
ARAF	RAF1	BRAF						2
EPAS1	HIF1A							18
RAC1	CDC42							6
MAPK1	MAPK3							5
MAP2K1	MAP2K2							5
ETS1	JUN							2
CRK	CRKL							2
SOS1	SOS2							1
RAP1A	RAP1B							1

TABLE S9. Orbit equivalence classes for the renal group.

Nodes								Degree
LAMA1	LAMA2	LAMA3	LAMA4	LAMA5	LAMB1	LAMB2	LAMB3	
LAMB4	LAMC1	LAMC2	LAMC3	COL4A1	COL4A2	COL4A4	COL4A5	6
COL4A6	FN1							
TRAF1	TRAF2	TRAF3	TRAF4	TRAF5	TRAF6			
PTGS2	NOS2	BIRC2	BIRC3	XIAP	BCL2L1			2
PIK3R5	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	4
ITGA2	ITGA2B	ITGA3	ITGA6	ITGAV	ITGB1			19
PIAS1	PIAS2	PIAS3	PIAS4					3
AKT1	AKT2	AKT3						11
CHUK	IKBKB	IKBKG						4
CCNE1	CCNE2	CDK2						2
E2F1	E2F2	E2F3						1
RELA	NKFB1							13
MAX	MYC							8
CDK4	CDK6							4
APAF1	CYCS							1

TABLE S10. Orbit equivalence classes for the SCL group.

Nodes							Degree
RET	CCDC6	NCOA4	TFG	NTRK1	TPM3	TPR	3
LEF1	TCF7	TCF7L1	TCF7L2				3
KRAS	NRAS	HRAS					8
RXRA	RXRB	RXRG					2
MYC	CCND1						4
MAP2K1	MAP2K2						3
PPARG	PAX8						3
MAPK1	MAPK3						2

TABLE S11. Orbit equivalence classes for the thyroid cancer group.