Check for updates

Stability of Imbalanced Triangles in Gene Regulatory Networks of Cancerous and Normal Cells

Abbas Karimi Rizi[1], Mina Zamani[1], Amirhossein Shirazi[1], G. Reza Jafari[1,2]* and János Kertész[1]

[1] Physics Department, Shahid Beheshti University, Tehran, Iran, [2] Department of Network and Data Science, Central European University, Budapest, Hungary

Genes communicate with each other through different regulatory effects, which lead to the emergence of complex network structures in cells, and such structures are expected to be different for normal and cancerous cells. To study these differences, we have investigated the Gene Regulatory Network (GRN) of cells as inferred from RNA-sequencing data. The GRN is a signed weighted network corresponding to the inductive or inhibitory interactions. Here we focus on a particular of motifs in the GRN, the triangles, which are imbalanced if the number of negative interactions is odd. By studying the stability of imbalanced triangles in the GRN, we show that the network of cancerous cells has fewer imbalanced triangles compared to normal cells. Moreover, in the normal cells, imbalanced triangles are isolated from the main part of the network, while such motifs are part of the network's giant component in cancerous cells. Our result demonstrates that due to genes' collective behavior the structure of the complex networks is different in cancerous cells from those in normal ones.

Keywords: gene regulatory networks, cancerous cells, stability, sign network, balance theory, max entropy, inverse problem

INTRODUCTION

Cancers are a large family of diseases that involve abnormal cell growth with the potential to invade or spread to other parts of the body (Pezzella et al., 2019). From the reductionist perspective, cancer is known as a disease of the genes. From this perspective, related studies focus on finding particular genes for each type of cancer and, consequently, diagnosing or curing cancer face formidable challenges. On the other hand, from the complexity theory perspective, collective behaviors emerged from the interactions of systems with many interacting units, are not describable solely by knowing the behavior of the system's building blocks (genes), and we cannot understand what happens at a higher level of organization by just studying how each element works at a lower scale. In other words, we need a holistic point of view to study the collective behavior of the genes (Zhou et al., 2014). The human body contains more than 10 trillion (10^{13}) cells, originating from a single one. Cells differ from each other, depending on which genes are turned on (Bianconi et al., 2013). The process by which information from a gene is used to synthesize functional gene products (often proteins) is called gene expression. Today, there are several projects globally, compiling genomic information related to cancers, and recent advances with sequencing technology reveal the high importance of these projects. Despite all the advances in technology and analysis in genome sequences, it seems that cancer remains indomitable to a large extent. While we know

some genes play an essential role in specific cancers, we are often far from controlling, let alone curing them (Goh et al., 2007; Jeyashree Krishnan et al., 2020).

Gene expressions are not independent (Demicheli and Coradini, 2011). They communicate with each other through regulatory effects, in a sense that some genes can up-regulate or down-regulate the expression level of other genes. These complex interactions between the genes can lead to collective behavior and result in changing the state of the cell. Complex systems consist of heterogeneous agents mutually influenced via interactions of different intensities over multiple spatio-temporal scales. This heterogeneity encompassed in both the participating components and their varying interactions makes complex systems difficult to decipher. To understand and control these complex systems, the network theory provides an effective mathematical modeling framework that enables the encoding of the entities (nodes) of a complex system and their heterogeneous interactions (links) of different strength (weights) into a topological network configuration implicitly embedded in metric spaces, where the distance among nodes is decided both by the structural configuration of the system (topology) and the intrinsic nature of the inter-node couplings (e.g., social affinity, chemical bonds, traffic intensity, or neural connectivity strength). In some cases, the properties of the inter-couplings among system components and the corresponding spatial embeddings even play a far more dominant role in regulating the overall system behaviors and dynamics. For instance, the atomic and molecular interactions among a chain of amino acids definitively dictate not only the dynamical spatial conformation of the corresponding protein but also its biological functionality. The disturbance of normal protein interactions can lead to irreversible pathological consequences known as proteopathies like Alzheimer's, Parkinson's, and Huntington's disease. Therefore, the study of structural organization, formation and dynamics of the complex systems can benefit from studying their geometrical properties and discovering new relationships between geometrical characteristics and network problems (e.g., community structure identification; Xue and Bogdan, 2017). In this scenario, there is a network of interactions, in which each gene is represented as a node, and its regulatory effect on other genes is considered the links connecting it to other nodes. These links can have zero (no effect), positive (up-regulation), or negative (down-regulation) weight, forming a weighted signed network. Such networks are called Gene Regulatory Networks (GRN) (Barabasi and Oltvai, 2004; Hempel et al., 2011; Walhout, 2011; Peter and Davidson, 2015; Costanzo et al., 2016; Liesecke et al., 2018; Huynh-Thu and Sanguinetti, 2019; Tieri et al., 2019). Different methods exist to build a GRN by computing a similarity, correlation or information-theory-based measure between the vectors associated to genes (Hempel et al., 2011).

Since the advent of high-throughput measurement technologies in biology in the late 90s, reconstructing gene regulatory networks' structure has been a central computational problem in systems biology (Huynh-Thu and Sanguinetti, 2019). Despite the efforts, the exact causal relationships between each pair of genes are unknown. Previous studies report the individual gene expression dynamics as well as the cross-dependency between them in the context of gene regulatory network the dynamics between genes are fractal and long-range cross-correlated (Ghorbani et al., 2018). Advanced analytical tools to analyse the multiscale patterns that occur in natural and synthetic biological systems, such as the methods reported in previous study (Xue and Bogdan, 2017), will be needed to develop a more complete and predictive understanding of the mechanisms and consequences of collective behavior in biological networks. Furthermore, discussion of gene expression and interactions is highly complex, which is why higher-order interactions are expected. One of the simplest interactions of a higher than two orders is a third-order called Balance theory (Marvel et al., 2011). We use Balance theory as the simplest model that does not consider interactions independent of each other and regards them as triadic interactions (Fritz, 1958; Antal et al., 2005; Moradimanesh et al., 2020). Thus, we use the simplification of considering the network as undirected and independent of time.

In this step, Even though we know there are time lags in our case, as we use Balance theory to discuss the characteristics of the weighted gene networks, we need to consider the interaction of genes statically, and this may be the next step in how to incorporate the effect of time lag into the theory of balance. This action requires improving and modifying the theory of balance. One of the other limitations that we confronted was our computational calculations limitation, which forced us to reduce our network size. We tried to use methods that reduce the quality of the deleted information as much as possible and achieve significant results in the end. To assess the pairwise interaction network structure, we use a maximum-entropy (Abellán and Castellano, 2017) probability model to explore the properties of the GRN. Such maximum entropy models have been widely used in statistical physics, e.g., for Ising type interacting models (Belaza et al., 2017; Nguyen et al., 2017). Physical systems in thermal equilibrium are described by the Boltzmann distribution, which has the maximum possible entropy given the mean energy of the system (Jaynes, 1957; Hedayatifar et al., 2017).

METHOD

From Real Data to Gene Interaction Network

The mRNA data (expression level) of 20,532 genes in the case of Breast Cancer (BRCA: Breast invasive carcinoma) has been downloaded from The Cancer Genome Atlas (TCGA) project (NIH, 2006-2014; Weinstein et al., 2013). The data contain 114 normal and 764 cancerous samples, and the measurement of the expression levels has been done with the technique of RNA sequencing (RNA-Seq). We have used the Reads Per Kilobase transcript per Million reads (RPKM) normalized data. RPKM puts together the ideas of normalizing by sample and by the gene. When we calculate RPKM, we are normalizing for both the library size (the sum of each column) and the gene length. In the following we had to reduce the number of genes because it was difficult to handle a 20,532∗20,532 matrix computationally. For each gene, we have calculated the variance

of its expression level over its samples, and finally we have stored the first 483 genes with the highest variance, which is due to more different activity patterns these genes show (CCNSD, 2019). Note that there are so-called housekeeping genes that typically get transcribed continually. These genes are required to maintain basic cellular function and are expressed in all cells of an organism under normal and pathophysiological conditions (Butte et al., 2001; Eisenberg and Levanon, 2003; Zhu et al., 2008). Some housekeeping genes are expressed at relatively constant rates in most non-pathological situations.

Measuring interactions is difficult within a living cell, but measuring abundances of components (mRNA levels) is considerably easier. Therefore, from the experimental data we wanted to reconstruct the gene-gene interactions computationally based on a model, following the practice that collective behaviors in such systems are described quantitatively by models that capture the observed pairwise correlations but assume no higher-order interactions (Schneidman et al., 2006). By assuming a maximum entropy pairwise model, we were looking for the interaction matrix J, whose every element J_{ij} is the strength of the net interaction between gene i and gene j. In other words, the strength and sign of the interaction represent the mutual influence on each other of a pair of genes' expression levels. From the maximum entropy probability distribution, we have constructed the energy function, which in this case is an Ising-like model with long-range Ferro- as well as antiferromagnetic couplings, which may lead to frustrated triangles. The energy function for our problem can be written as:

$$H = - \sum_{i<j} J_{ij} S_i S_j, \tag{1}$$

where the expression level of gene i as a continuous real-valued variable (a Gaussian field) is represented by S_i. Using the energy function above, we can write down the Boltzmann equilibrium distribution as:

$$P(\{S_i\}) = \frac{1}{Z} \exp\left(- \sum_{i<j} J_{ij} S_i S_j\right). \tag{2}$$

Z is the partition function, and we have subsumed temperature into the couplings J_{ij} without loss of generality. The interaction matrix, J, is not known, and we wanted to learn/infer it (Nguyen et al., 2017) from the experimental data. We want to infer all the J_{ij} as the parameters of our model. To this end, we have restricted ourselves to a probabilistic model with terms up to second order, which we have derived for continuous, real-valued variables. In other words, our model is constrained to generate the first and the second moments which are exactly the same as what we find from the experimental data (Stein et al., 2015). Thus, P must maximize the Gibbs-Shannon entropy to infer the parameters of the model.

$$S[P] = - \sum_{i<j} P(\{S_i\}) \ln P(\{S_i\}). \tag{3}$$

Using Lagrange multipliers, it can be shown (Stein et al., 2015) that the desired model is a multivariate Gaussian distribution, twice of its covariance is minus the inverse of the interaction matrix.

$$P(S; \langle S \rangle; C) = \frac{e^{-\frac{1}{2}(S-\langle S\rangle)^T C^{-1}(S-\langle S\rangle)}}{(2\pi)^{\frac{L}{2}} \det(C)^{\frac{1}{2}}}. \tag{4}$$

So, within this approximation, we can write $J_{ij} = -C_{ij}^{-1}$. L is the number of genes based on which we have built the distribution. The elements of the matrix J are, by definition, the effective pairwise gene interactions that reproduce the gene profile covariances (Lee et al., 2014) exactly while maximizing the entropy of the system. The inverse of the covariance matrix, C^{-1}, which is commonly referred to as the precision matrix, displays information about the partial correlations of variables. In practice, the precision matrix can be estimated by simply inverting the sample covariance matrix, if a sufficiently large number of samples are available. In our study, due to the lack of enough samples, the inverse of the covariance matrix has been obtained by means of the Graphical Lasso (GLasso) algorithm (Friedman et al., 2008). GLasso is an algorithm to estimate the inverse of the covariance matrix from the observations from a multivariate Gaussian distribution. In statistics and machine learning, lasso [least absolute shrinkage (Peterson and Ford, 2012) and selection operator] is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. G-Lasso sparse the network in such a way that it does not disrupt the overall properties of the network. In sparsing a matrix, One of the problems is that the threshold method in the network is severe. In this way, in networks the threshold may eliminate weak links in favor of solid links. But we know that some links are fragile, and their share in the network is very high. For example, it connects part of the network to another part, but it can be a strong link between the network and the node that does not matter to us. The threshold method eliminates the important weak link that connects the two network parts—in contrast, keeping a strong link connected to the trivial part of the network. We know that removing a strong link that is only connected to an insignificant node does not destroy the network properties while removing a weak link that affects the network properties, G-Lasso is wary of such issues.

Following are step by step calculations in brief:

- Import Row data from TCGA Database, The mRNA data(expression level) of 20,532 genes.
- Dimension reduction, keep genes with the highest variance (483 genes).
- Calculate the covariance matrix of genes (483*483).
- Calculate J, inverse of the covariance matrix by G-Lasso (Mazumder and Hastie, 2012) approach to make it sparse, with penalty = 0.09.
- Calculate Energy-Energy matrix.

All of the calculations have been done in Python and MatLab. All codes and results are available upon request[1].

---

[1]https://ccnsd.ir/research/cancer-project/

Frustration in Interaction Network

The positive (negative) value of the interactions implies that increasing (or decreasing) a gene's expression results in up-regulating (down-regulating) of the other gene(s)'s expression(s), respectively. J is the generalized adjacency matrix (Newman, 2018), representing the presence and weight of a link. J_{ij} is the strength of the interaction between gene i and gene j or in network terms, the weight of the link $i - j$.

Let us now consider the local triangles; Groups with three interacting genes forming a triangle of interactions in the network. The triangle $\Delta(i, j, k)$ is defined as balanced if the sign of the product of its links is positive; $J_{ij}J_{jk}J_{ki} > 0$, otherwise, the triangle is imbalanced or frustrated; $J_{ij}J_{jk}J_{ki} < 0$. We define a triangle to be of type Δ_k if it contains k negative links. Thus, Δ_{+++} and Δ_{-+-} are balanced, while Δ_{+-+} and Δ_{---} are imbalanced (Heider, 1946). The statistics of the analogues of these imbalanced triangles have been shown to be relevant in systems with signed interactions like random magnets (Fischer and Hertz, 1991) and social networks (Antal et al., 2005).

The notion of balance allows us to define an "energy landscape" for such networks (Marvel et al., 2009; Górski et al., 2017). For a triangle this is:

$$\Delta_{ijk} = E_{ijk} = -J_{ij}J_{jk}J_{ki} \tag{5}$$

and by summing over all the E_{ijk} the energy of the whole network can be obtained (Krawczyk et al., 2019).

$$E_{total} = \frac{1}{N} \sum_{i,j,k=1}^{N} \Delta_{ijk} \tag{6}$$

Note that this energy is different from that of (1) and serves to characterize the triangles, while H was used to calculate the interactions from the measured expression strengths. Energy counts the number of triangles and does not indicate where the triangles are. The correlation between triangles shows which triangle with energy E_k has a common link with which triangle with energy E_l.

$$C_{kl} = \frac{1}{N} \sum_{i,j,k=1}^{N} \Delta_{kij}\Delta_{lij} \tag{7}$$

This equation answers the question that a triangle with preferred energy is adjacent to which triangle. C_{kl} can be positive, negative, or zero. A positive C_{kl} means that a balanced triangle links to another balance triangle. A negative C_{kl} indicates that a balanced triangle links to another imbalance triangle. Finally, C_{kl} zero means there is no preference and link between two triangles.
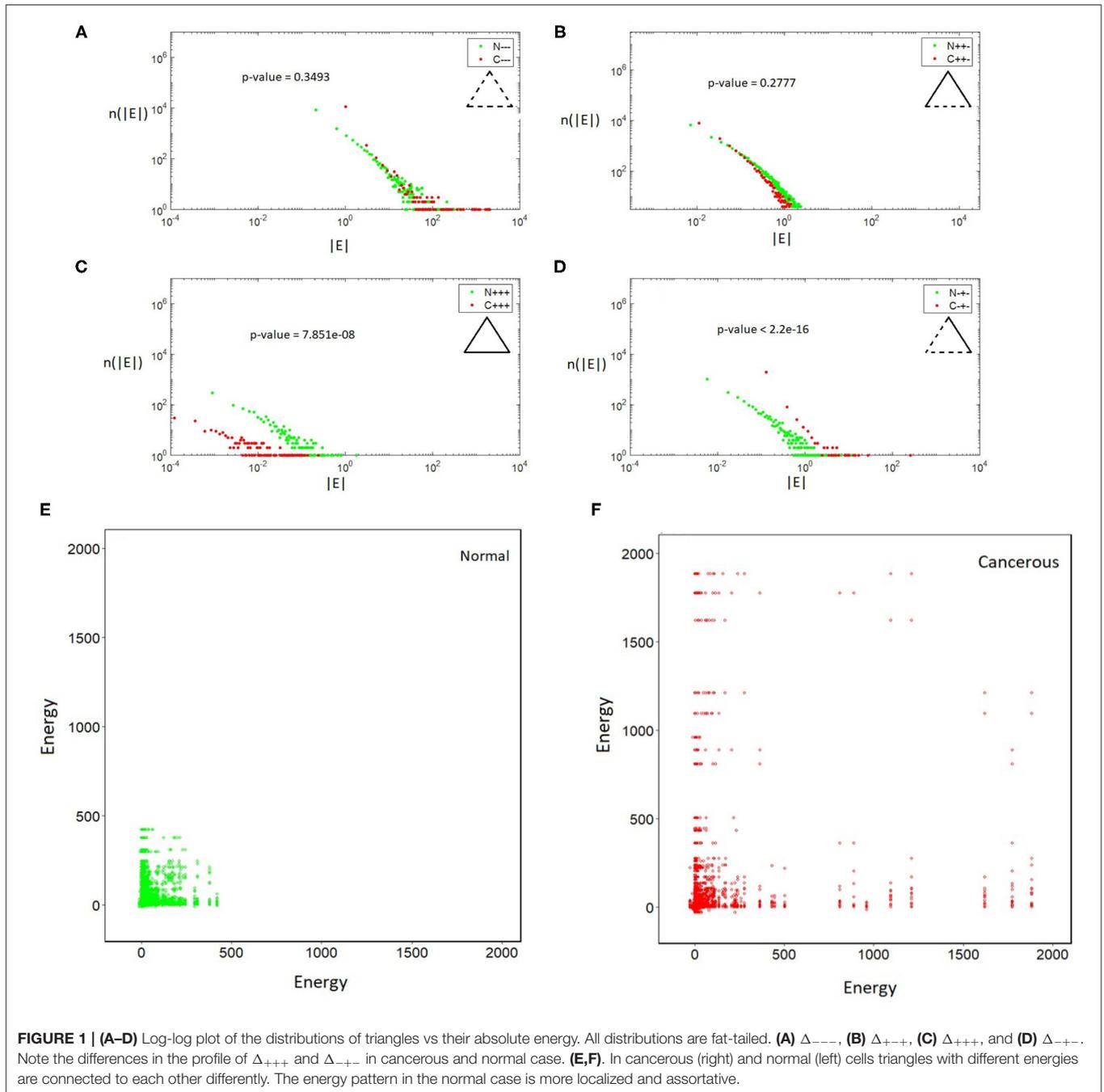
RESULTS

We have calculated the distributions of the energies of different types of triangles in both cancerous and normal data-sets and observed the following results (**Figure 1**). (i) In all the cases, the energy distributions of all types of triangles are fat-tailed. (ii) The distributions of imbalanced (frustrated) triangles, Δ_{+-+}

(**Figure 1B**) and Δ_{---} (**Figure 1A**), do not show noticeable differences between cancerous and normal data. (iii) In the cancerous network Δ_{+++} (**Figure 1C**) triangles and normal network Δ_{-+-}-types (**Figure 1D**) are less. The total energy of the cancerous network is 27,239 units and total energy of the normal network is 35,984 units. So the total energy of the cancerous network is lower than that of normal network.

In order to see if the effect comes from structural correlations specific to the differences between the normal and cancerous data, we have shuffled the links in the networks. This was carried out by swapping endpoints of randomly selected pairs of links many times, which is a standard procedure to produce degree preserving random reference networks. The energy difference between the shuffled networks is 280 units which is one order of magnitude less than in the original case. Moreover, the distribution profiles change dramatically for the shuffled network.

The next question we have studied was about the distribution of triangles with different energies in the networks and their relationships. For this purpose, we coarse grain the network such that balanced and imbalanced triangles are represented as green and red nodes, respectively. Two coarse-grained nodes are connected if their corresponding triangles have one edge in common. We calculate the energy-energy mixing pattern (Newman, 2002) between the triangles. The plots in **Figures 1E,F** shows how many triangles with different energies are connected. Notice that this matrix is rather sparse reflecting that only low number of the triangles have links in common. In the normal network, frustrated triangles are packed together and they form a kind of module while in the cancerous network they have a more heterogeneous pattern of connections and they are mixed with balanced triangles. Moreover, triangles with higher absolute valued energies are connected to ones with lower absolute valued energies. In both cases, we see triangles with lower energies are more connected to each other. Triangles in the cancerous network do not tend to distribute evenly in a particular region of energy-energy space. In fact the energy pattern in the normal case is more localized and assortative. Another result is that in both of the networks so many triangles do not have a link in common.
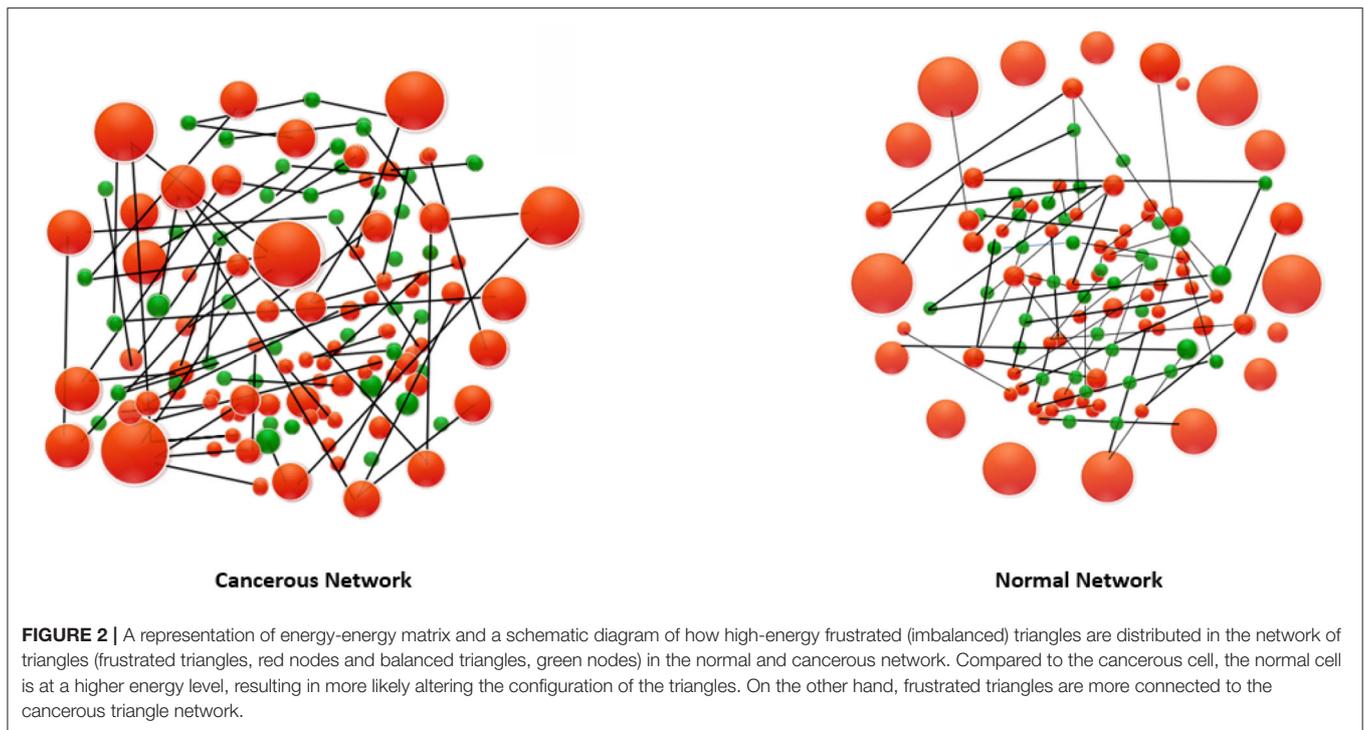
Having more energy for a cell, in this context, means that there is more tendency toward changing the states of the triangles. In the case of cancerous network, we have seen that triangles exhibit a lower chance of being changed. On the other hand, we see frustrated triangles are somehow uniformly distributed in the cancerous coarse-grained network while they are more localized in the normal coarse-gained case. These facts are mimicked in the **Figure 2**. Inspired by the concept of Balance theory in social science (Sheykhali et al., 2020), we saw that the interaction network of the normal case has more imbalanced (frustrated) triangles and more energy as a consequence. This energy has been defined in a social context giving a good clue to look at the system of genes as a social system. Not only genes cannot live independent of each other, but they also must pay the cost of living together! Note that changing the expression of a gene can have drastic consequences (Witthaut and Timme, 2013). Our analysis reveals the fact that to get a true picture of biology at the cell level, it is

FIGURE 1 | (A–D) Log-log plot of the distributions of triangles vs their absolute energy. All distributions are fat-tailed. **(A)** Δ_{---}, **(B)** Δ_{+-+}, **(C)** Δ_{+++}, and **(D)** Δ_{-+-}. Note the differences in the profile of Δ_{+++} and Δ_{-+-} in cancerous and normal case. **(E,F)**. In cancerous (right) and normal (left) cells triangles with different energies are connected to each other differently. The energy pattern in the normal case is more localized and assortative.

essential to know the connections and their type between the genes.

Applying maximum entropy and Ising models for identifying the interactions between genes and the use of balance theory is a new perspective discussed in this article, which also has its limitations. One of the limitations is that the maximum entropy assumes the gene expression as an equilibrium process which lacks time-varying properties. Several studies have mentioned the existence and implications of multi-fractal dynamics in gene expression, proteomics, and physiological processes. However,

there are various valuable studies on GRN as well that, despite this limitation, have considered gene expression in equilibrium. In this step, even though we know there are time lags in our case, as we use balance theory to discuss the characteristics of the weighted gene networks, we need to consider the interaction of genes statically. Incorporating the effect of time lag into the theory of balance may be the next step, which requires the extension and modification of the balance theory. The other limitation we have confronted was the computational calculations limitation, which forced us to reduce the network

FIGURE 2 | A representation of energy-energy matrix and a schematic diagram of how high-energy frustrated (imbalanced) triangles are distributed in the network of triangles (frustrated triangles, red nodes and balanced triangles, green nodes) in the normal and cancerous network. Compared to the cancerous cell, the normal cell is at a higher energy level, resulting in more likely altering the configuration of the triangles. On the other hand, frustrated triangles are more connected to the cancerous triangle network.

size. We have conducted methods that reduce the quality of the deleted information as much as possible and achieve significant results in the end.

CONCLUSION

Cancer has been commonly known as a group of diseases of the genes and there has been a huge effort to find the effective genes responsible for different cancers. Thanks to such reductionist approaches, we now know some specific genes for some cancers. Genes, however, are not independently functioning in the cell and their expressions are strongly correlated with each other. Recently, it has been recognized that the regulatory effects between the genes can be represented by a gene-gene interaction network and the structure of this network is essential in understanding the collective phenomena, which play a role in developing cancer-related studies. Our results contribute to this line of research (Rabbani et al., 2019). We have presented a formalism, by which we arrived from the data about gene expressions to an interacting network model, where the interactions were inferred using the maximum entropy principle. The resulting signed weighted network (Saeedian et al., 2017) was analyzed from the balanced and imbalanced triangles perspective. We have found significant differences between normal and cancerous cell GRN-s: There are more imbalanced triangles in

normal GRN-s than in cancerous ones and the correlations between such triangles are also different in these two networks. Further investigations are indeed valuable to study when the observed differences develop and whether our observations can be used for diagnostic purposes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AR, MZ, and AS conceived the model. AR and MZ did the computations and prepared the manuscript. GJ supervised the work. AR, MZ, AS, JK, and GJ analyzed the results. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

The authors would like to appreciate Professor Tamás Vicsek and Professor David Saakian for reading the manuscript and their constructive comments.

REFERENCES
Abellán, J., and Castellano, J. G. (2017). Improving the naive bayes classifier via a quick variable selection method using maximum of entropy. *Entropy* 19:247. doi: 10.3390/e19060247

Antal, T., Krapivsky, P. L., and Redner, S. (2005). Dynamics of social balance on networks. *Phys. Rev. E* 72:036121. doi: 10.1103/PhysRevE.72.036121

Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Belaza A. M., Hoefman, K., Ryckebusch, J., Bramson, A., van den Heuvel, M., Schoors, K. (2017). Statistical physics of balance theory. *PLoS ONE* 12:e0183696. doi: 10.1371/journal.pone.0183696

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., et al. (2013). An estimation of the number of cells in the human body. *Ann. Hum. Biol.* 40, 463–471. doi: 10.3109/03014460.2013.807878

Butte, A. J., Dzau, V. J., and Glueck, S. B. (2001). Further defining housekeeping, or "maintenance," genes focus on "a compendium of gene expression in normal human tissues". *Physiol. Genomics* 7, 95–96. doi: 10.1152/physiolgenomics.2001.7.2.95

CCNSD. *Cancer Project* (2019). Available online at: https://ccnsd.ir/research/cancer-project/.

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353:aaf1420. doi: 10.1126/science.aaf1420

Demicheli, R., and Coradini, D. (2011). Gene regulatory networks: a new conceptual framework to analyse breast cancer behaviour. *Ann. Oncol.* 22, 1259–1265. doi: 10.1093/annonc/mdq546

Eisenberg, E., and Levanon, E. Y. (2003). Human housekeeping genes are compact. *Trends Genet.* 19, 362–365. doi: 10.1016/S0168-9525(03)00140-9

Fischer, K. H., and Hertz, J. A. (1991). *Spin Glasses (Cambridge Studies in Magnetism)*. Cambridge University Press.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441. doi: 10.1093/biostatistics/kxm045

Fritz, H. (1958). *The Psychology of Interpersonal Relations*. New York, NY: John Wiley and Sons.

Ghorbani, M., Jonckheere, E. A., and Bogdan, P. (2018). Gene expression is not random: scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. *Front. Physiol.* 9:1446. doi: 10.3389/fphys.2018.01446

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., Barabási, A. L., et al. (2007). The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8685–8690. doi: 10.1073/pnas.0701361104

Górski, P. J., Kułakowski, K., Gawroński, P., and Hołyst, J. A. (2017). Destructive influence of interlayer coupling on heider balance in bilayer networks. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-15960-y

Hedayatifar, L., Hassanibesheli, F., Shirazi, A., Farahani, S. V., and Jafari, G. (2017). Pseudo paths towards minimum energy states in network dynamics. *Phys. A* 483, 109–116. doi: 10.1016/j.physa.2017.04.132

Heider, F. (1946). Attitudes and cognitive organization. *J. Psychol.* 21, 107–112.

Hempel, S., Koseska, A., Nikoloski, Z. and Kurths, J. (2011). Unraveling gene regulatory networks from time-resolved gene expression data–a measures comparison study. *BMC Bioinformatics* 12:292. doi: 10.1186/1471-2105-12-292

Huynh-Thu, V. A., and Sanguinetti, G. (2019). "Gene regulatory network inference: an introductory survey," in *Gene Regulatory Networks*, ed G. Sanguinetti (Springer), 1–23. doi: 10.1007/978-1-4939-8882-2_1

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.* 106, 620–630.

Jeyashree Krishnan, A. S. E. D. N., Reza Torabi (2020). A modified ising model of barabási-Albert network with gene-type spins. *J. Math. Biol.* 81, 769–798. doi: 10.1007/s00285-020-01518-6

Krawczyk, M. J., Wołoszyn, M., Gronek, P., Kułakowski, K., and Mucha, J. (2019). The heider balance and the looking-glass self: Modelling dynamics of social relations. *Sci. Rep.* 9, 1–8. doi: 10.1038/s41598-019-47697-1

Lee, J. A., Dobbin, K. K., and Ahn, J. (2014). Covariance adjustment for batch effect in gene expression data. *Stat. Med.* 33, 2681–2695. doi: 10.1002/sim.6157

Liesecke, F., Daudu, D., Dugé de Bernonville, R., Besseau, S., Clastre, M., Courdavault, V., et al. (2018). Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks. *Sci. Rep.* 8, 1–16. doi: 10.1038/s41598-018-29077-3

Marvel, S. A., Kleinberg, J., Kleinberg, R. D., and Strogatz, S. H. (2011). Continuous-time model of structural balance. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1771–1776. doi: 10.1073/pnas.1013213108

Marvel, S. A., Strogatz, S. H., and Kleinberg, J. M. (2009). Energy landscape of social balance. *Phys. Rev. Lett.* 103:198701. doi: 10.1103/PhysRevLett.103.198701

Mazumder, R., and Hastie, T. (2012). The graphical lasso: New insights and alternatives. *Electron. J. Stat.* 6, 2125–2149. doi: 10.1214/12-EJS740

Moradimanesh, Z., Khosrowabadi, R., Gordji, M. E., and Jafari, G. (2020). Altered structural balance of resting-state networks in autism. *arXiv preprint arXiv:2010.09441*. Available online at: https://arxiv.org/abs/2010.09441

Newman, M. (2018). *Networks, 2nd Edn.* Oxford University Press.

Newman, M. E. (2002). Assortative mixing in networks. *Phys. Rev. Lett.* 89:208701. doi: 10.1103/PhysRevLett.89.208701

Nguyen, H. C., Zecchina, R., and Berg, J. (2017). Inverse statistical problems: from the inverse ising problem to data science. *Adv. Phys.* 66, 197–261. doi: 10.1080/00018732.2017.1341604

NIH (2006-2014). *The Cancer Genome Atlas*. Available online at: https://www.cancer.gov/tcga

Peter, I. S., and Davidson, E. H. (2015). *Genomic Control Process Development and Evolution*. Elsevier Science.

Peterson, L. E., and Ford, C. E. (2012). "Random matrix theory and covariance matrix filtering for cancer gene expression," in *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics, CIBB* (Berlin; Heidelberg: Springer). doi: 10.1007/978-3-642-38342-7_15

Pezzella, F., Tavassoli, M., and Kerr, D. (eds.). (2019). *Oxford Textbook of Cancer Biology*. Oxford University Press.

Rabbani, F., Shirazi, A. H., and Jafari, G. (2019). Mean-field solution of structural balance dynamics in nonzero temperature. *Phys. Rev. E* 99:062302. doi: 10.1103/PhysRevE.99.062302

Saeedian, M., Azimi-Tafreshi, N., Jafari, G., and Kertesz, J. (2017). Epidemic spreading on evolving signed networks. *Phys. Rev. E* 95:022314 doi: 10.1103/PhysRevE.95.022314

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Sheykhali, S., Darooneh, A. H., and Jafari, G. R. (2020). Partial balance in social networks with stubborn links. *Phys. A Stat. Mech. Appl.* 548:123882. doi: 10.1016/j.physa.2019.123882

Stein, R. R., Marks, D. S., and Sander, C. (2015). Inferring pairwise interactions from biological data using maximum-entropy probability models. *PLoS Comput. Biol.* 11:e1004182. doi: 10.1371/journal.pcbi.1004182

The Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw K. R.M., Ozenberger, B. A., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764

Tieri, P., Farina, L., Petti, M., Astolfi, L., Paci, P., and Castiglione, F. (2019). "Network inference and reconstruction in bioinformatics," in *Encyclopedia of Bioinformatics and Computational Biology*, eds S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach (Academic Press), 805–813. doi: 10.1016/B978-0-12-809633-8.20290-2

Walhout, A. J. (2011). "Gene-centered regulatory network mapping," in *Methods in Cell Biology, Vol. 106*, eds J. H. Rothman and A. Singson (Worcester, MA: Elsevier), 271–288.

Witthaut, D., and Timme, M. (2013). Nonlocal failures in complex supply networks by single link additions. *Eur. Phys. J. B* 86:377. doi: 10.1140/epjb/e2013-40469-4

Xue, Y., and Bogdan, P. (2017). Reliable multi-fractal characterization of weighted complex networks: algorithms and implications. *Sci. Rep.* 7, 1–22. doi: 10.1038/s41598-017-07209-5

Zhou, X., Menche, J., Barabási, A.-L., and Sharma, A. (2014). Human symptoms–disease network. *Nat. Commun.* 5, 1–10. doi: 10.1038/ncomms5212

Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. *Trends Genet.* 24, 481–484. doi: 10.1016/j.tig.2008.08.004