SCIENTIFIC REP⚙RTS

Graph Curvature for Differentiating Cancer Networks

Romeil Sandhu[1,*], Tryphon Georgiou[2], Ed Reznik[3], Liangjia Zhu[1], Ivan Kolesov[1], Yasin Senbabaoglu[3] & Allen Tannenbaum[1,*]

Cellular interactions can be modeled as complex dynamical systems represented by weighted graphs. The functionality of such networks, including measures of robustness, reliability, performance, and efficiency, are intrinsically tied to the topology and geometry of the underlying graph. Utilizing recently proposed geometric notions of curvature on weighted graphs, we investigate the features of gene co-expression networks derived from large-scale genomic studies of cancer. We find that the curvature of these networks reliably distinguishes between cancer and normal samples, with cancer networks exhibiting higher curvature than their normal counterparts. We establish a quantitative relationship between our findings and prior investigations of network entropy. Furthermore, we demonstrate how our approach yields additional, non-trivial pair-wise (i.e. gene-gene) interactions which may be disrupted in cancer samples. The mathematical formulation of our approach yields an exact solution to calculating pair-wise changes in curvature which was computationally infeasible using prior methods. As such, our findings lay the foundation for an analytical approach to studying complex biological networks.

The onset and proliferation of cancer stems from dynamic changes that result from a series of changes in cellular interactions governing a complex network[1–3]. Inspired by the previous work of Teschendorff[4], examining properties of such networks at various states may aid in the understanding of certain cellular processes leading to tumorigenesis. One key property is the notion of robustness, or the ability of a system to adapt to dynamic changes and perturbations while still maintaining functionality. From this perspective, a fundamental hurdle to cancer therapy is acquired tumor robustness[5]. On the other hand, quantification of robustness and in particular, that of cancer networks, has remained elusive. Understanding and exploiting such network properties from a biological perspective provides an alternative framework to viewing underlying mechanisms. In turn, this may guide and uncover new drug targets.

In this work, we demonstrate the role of curvature as system-level characteristic of certain cancer networks and its relationship to network functionality in terms of a notion of robustness[6,7], specifically at the local interaction level. Curvature, in the broad sense, is a measure by which a geometrical object deviates from being flat and is defined in varying manners given the context[8]. Our reference to curvature will be restricted to **Ricci curvature** and its contraction, **scalar curvature**. Similarly, "robustness" can be formally defined in terms of the rate function from the theory of large deviations by appealing to the Fluctuation Theorem[6,7]. Roughly speaking, robustness relates to the rate at which a given dynamical system returns to its original (normal) state following a perturbation or external disturbance. The key ingredient that intimately links curvature and robustness is the concept of entropy. Indeed, through a suitable characterization on the lower bound of Ricci curvature[9], one can show that entropy and curvature are positively correlated, a fact that we express as $\Delta S \times \Delta Ric \geq 0$ and where ΔS and ΔRic are the changes in entropy and Ricci curvature, respectively (see Methods)[10]. Now, if we consider random perturbations to the network, the Fluctuation Theorem asserts that $\Delta S \times \Delta R \geq 0$ where ΔR is a relative

[1]Departments of Computer Science and Applied Mathematics/Statistics, Stony Brook University, Stony Brook, NY 11794. [2]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455. [3]Computational Biology, Memorial Sloan Kettering, New York, NY 10065. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.T. (email: allen.tannenbaum@stonybrook.edu)
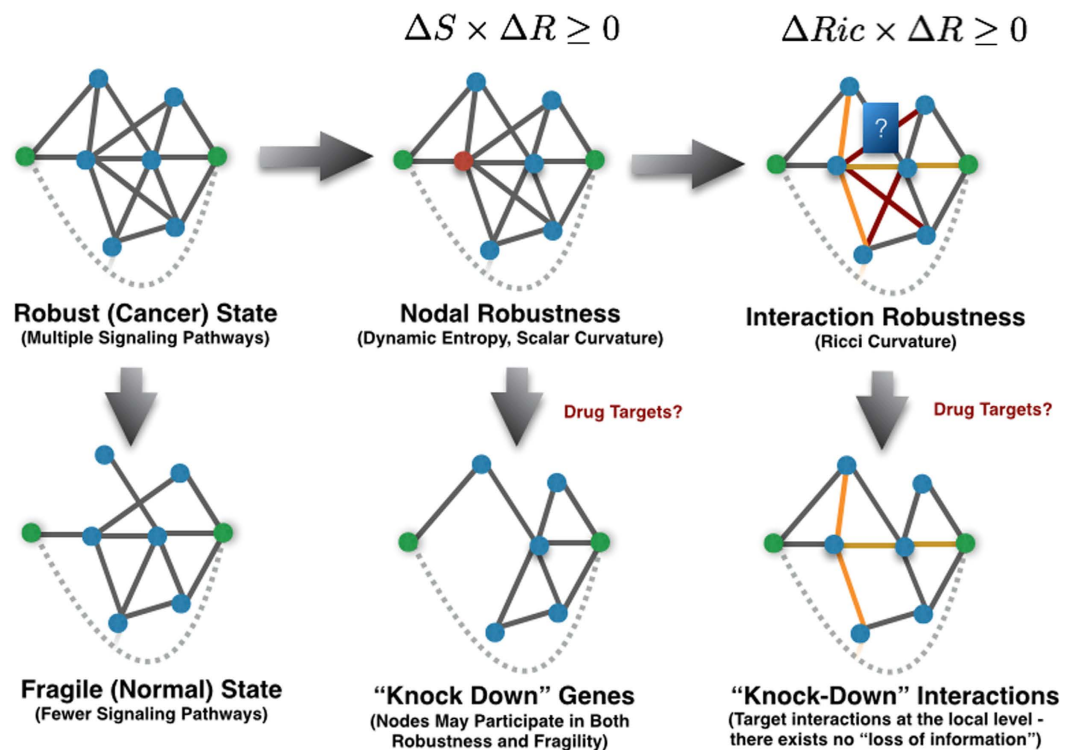
$$\Delta S \times \Delta R \geq 0 \qquad \Delta Ric \times \Delta R \geq 0$$

Figure 1. This work focuses on analyzing robustness with respect to pairwise interactions. Systems equipped with multiple signaling pathways can be framed in the context of robustness. Whereas previous work has shown dynamic entropy as a cancer "hallmark" through a nodal characterization, we expound upon this by providing a framework to analyzing gene-to-gene interaction robustness. In doing so, we will show that the method herein presents no "loss of information" and may be aptly suited to uncover particular pathways contributing to the robustness of cancer systems.

change in robustness and hence, the relationship $\Delta Ric \times \Delta R \geq 0$ holds. As we will argue in this work, this tacit relationship to robustness will allow curvature to serve as an alternative, yet powerful proxy (Fig. 1). This seems especially true for cancer networks.

Our work here differs from previous approaches of characterizing network robustness[4,6,7,11] in several important aspects. To the best our knowledge, it is the first to express general network functional robustness through curvature and to point out that this may provide an intrinsic cancer characteristic. With regards to entropy, our utilization of curvature holds the following key advantages: (I) Ricci curvature provides *pairwise information* over all possible pathways as opposed to network entropy, which is defined as a nodal measure[7]. This is particularly significant due to the fact that we are interested in specific gene-to-gene interactions contributing to the resilience of cancer (including those "hidden" interactions not necessarily defined by the underlying topology). In short, previous work of network entropy exhibits a "loss of information" with regards to the robustness of the interactions themselves[4,7]. (II) Ricci curvature can be formulated as a simple linear program and is well-behaved as compared to network entropy[12]. (III) Scalar curvature, in a similar manner to network entropy, is defined as a nodal measure in which interactions are not explicitly described.

In the present work, we compare gene co-expression networks from cancer and adjacent-normal tissue samples using network curvature. Motivated by previous entropic studies[4], we fix the underlying topology of the networks using prior data on known physical interactions between gene products allowing only the weights to evolve between normal and tumor networks. Then, by treating each network as a random walk, we attempt to exploit the underlying dynamics of specific gene-to-gene interactions.

Finally, we should note that the methods explicated in the present work are applicable not only to cancer networks, but may also assist in unifying several phenomena in molecular biology for which notions of robustness (via entropy and curvature) seem to be increasingly important[5,13–15]. For example, recent work has demonstrated that local signaling entropy may serve as a novel indicator of drug sensitivity[13] while at the same time, may operate as a proxy for the height or elevation in Waddington's differentiation landscape[14]. Furthermore, it has been argued that feedback loops are essential to the function of biological mechanisms and systems that arise from deliberate Darwinian-like principles[5,15]. In what follows, one can view Ricci curvature as a new feedback measure, i.e., the number of triangles in a network
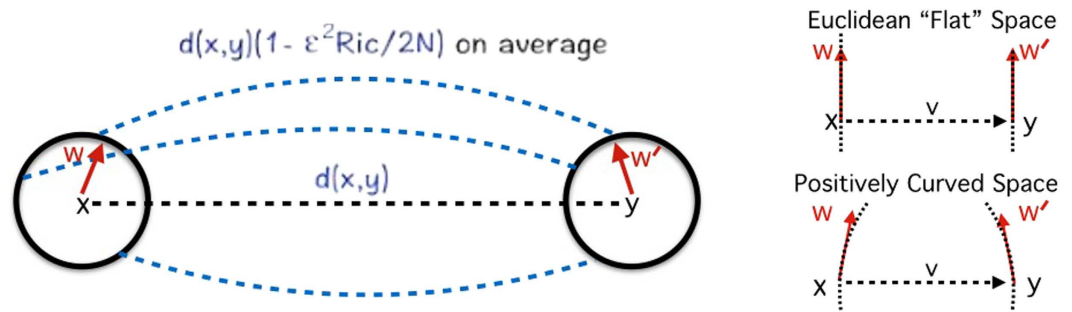
Figure 2. Positive Ricci curvature is reflected by the characteristic that for two very close points *x* and *y* with a tangent vector *v* connecting *xy* as well as tangent vectors *w* (at *x*) and *w′* (at *y*), in which *w′* is obtained by parallel transport of *w*, that the two corresponding geodesics will get closer. This can be compared to the traditional flat geometry of a Euclidean space where such distances are unaffected during the parallel transport. Equivalently, this may be formulated by the fact that the transportation distance between two small (geodesic balls) is less than the distance of their centers. Ricci curvature along the direction *xy* quantifies this, averaged on all directions *w* at *x*.

(redundant pathways) can be characterized by a lower bound of Ricci curvature[5,16,17]. This fascinating interplay between feedback, robustness, entropy, and now Ricci curvature is at the core of this work.

The remainder of this paper is outlined as follows. We first provide results to demonstrate that Ricci curvature, more precisely Ollivier-Ricci curvature[18,19], is a proxy for robustness as well as an apparent cancer characteristic. In particular, we discuss the importance (and previously unresolved) ability of quantifying robustness at the interaction level. We then show that several analogous nodal curvature measures, defined through varying contractions of Ricci curvature, achieve similar results to that of network entropy, which by construction, is a nodal attribute. We conclude with a discussion of the results with a primary focus on information loss from previous entropic methods, examination of robustness for specific gene-to-gene interactions in context of cancer biology, and analytic advantages of employing Ricci curvature as opposed to entropy. From this, we then offer a possible path forward that relates the well-known Ricci flow to the effect (and design thereof) of specific drug targets that can possibly mitigate the robust nature of cancer.

Results

We focus our investigation primarily on transcription networks composed of metabolic and cancer specific genes[20,21]. For each data set, gene co-expression networks were generated by calculating the non-parametric (Spearman) correlation between all pairs of genes. That is, for a given gene pair, correlation was computed across all samples within a given phenotype (normal or cancerous tissue). The metabolic data set consists of approximately 1600 metabolic genes (derived from the Recon2 human metabolic reconstruction[22]) of six different tumor types: breast cancer (BRCA^M), head and neck squamous cell carcinoma (HNSC^M), kidney papillary carcinoma (KIRP^M), liver cancer (LIHC^M), lung adenocarcinoma (LUAD^M), and thyroid cancer (THCA^M). We further supplemented the above study with corresponding networks that contain approximately 500 *cancer-related* genes derived from the Cosmic Cancer Gene Census[21] (denoted by ^{T}, e.g., BRCA^T). With regards to the topology, the networks analyzing metabolic genes possess a total of 33843 edges, average degree of 43, and a median degree of 34. For the networks composed of known cancer-related genes, the total number of edges, average degree, and median degree are 8162, 37, and 22 respectively (see Methods).

Gene-to-Gene Robustness: Ollivier-Ricci Curvature.

We employ a neat notion of a Ricci curvature[19] inspired through coarse geometry (Fig. 2). In particular, if we let (X, d) be a metric space equipped with a family of probability measures $\{\mu_x : x \in X\}$, we define the *Ollivier-Ricci curvature* $\kappa(x, y)$ along the geodesic connecting nodes x and y via

$$W_1(\mu_x, \mu_y) = (1 - \kappa(x, y))d(x, y),\qquad(1)$$

where W_1 denotes the Earth Mover's Distance (Wasserstein 1-metric)[23,24], and d is the geodesic distance on the graph. For the case of weighted graphs, we set
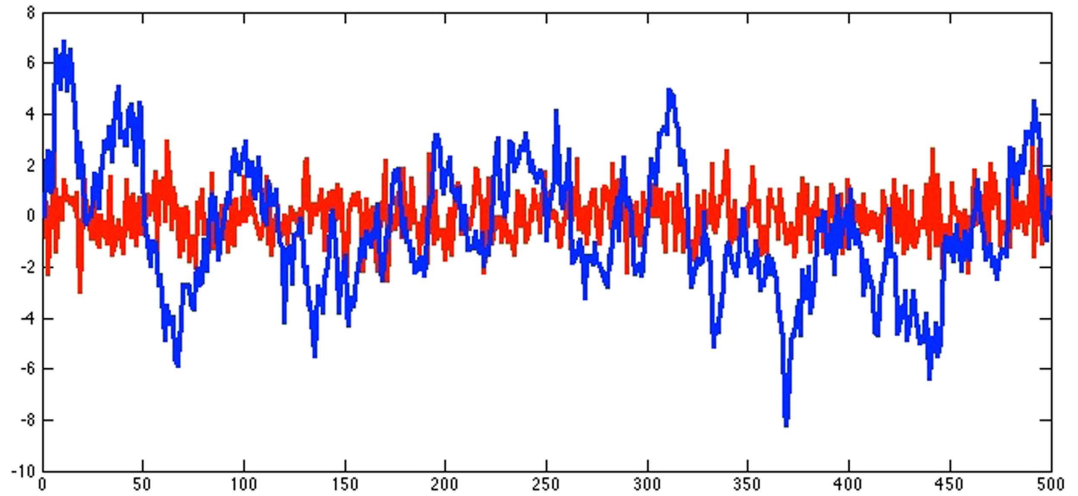
$$d_x = \sum_y w_{xy}\qquad(2)$$

Figure 3. We generated two Ornstein-Uhlenbeck processes with the same parameter set except for different α and in turn, exhibits different Ollivier-Ricci curvatures: $\kappa(x, y) = 0.6321$ (red) with $\alpha = 1.0$ and $\kappa(x, y) = 0.0952$ with $\alpha = 0.1$ For both signals, Ornstein-Uhlenbeck process parameters were initialized with $x(0) = 1$ with $\sigma = 1$. One can see that $\Delta\kappa \times \Delta\alpha \geq 0$ for this broad set of stochastic processes.
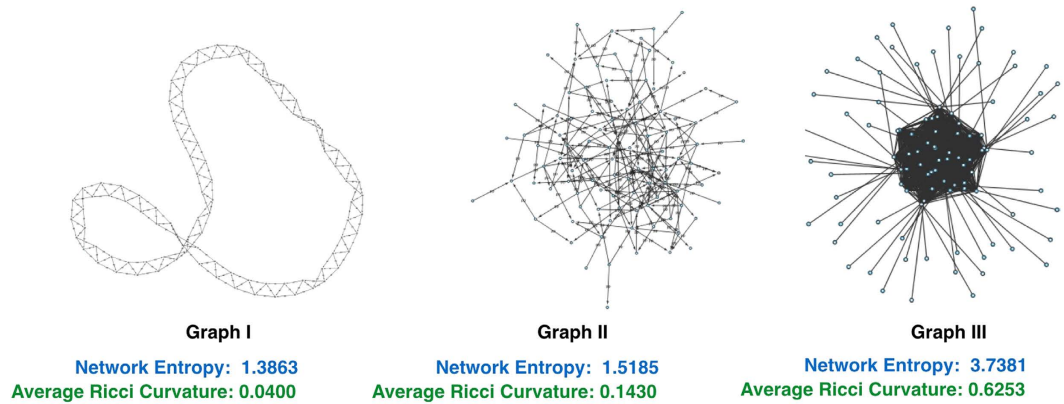

Graph I	Graph II	Graph III
Network Entropy: 1.3863	Network Entropy: 1.5185	Network Entropy: 3.7381
Average Ricci Curvature: 0.0400	Average Ricci Curvature: 0.1430	Average Ricci Curvature: 0.6253

Figure 4. We computed the average Ollivier-Ricci curvature for three different networks shown above as well as network entropy. To ensure a fair comparison, each of the networks is composed of 200 nodes with 400 (unweighted) edges - the only difference in the underlying structure. Although Ricci curvature is a local property, it nevertheless shows that, on average, Ricci curvature is higher for networks that exhibit higher entropy.

$$\mu_x(y) := \frac{w_{xy}}{d_x},$$

(3)

where d_x is the sum taken over all neighbors of node x and where w_{xy} denotes the weight of an edge connecting node x and node y ($w_{xy} = 0$ if $d(x, y) \geq 2$). The measure μ_x may be regarded as the distribution of a one-step random walk starting from x, with the weight w_{xy} quantifying the strength of interaction between nodal components or the diffusivity across the corresponding link (edge). To motivate this definition and highlight the role of curvature as a proxy for robustness, we compute the Ollivier-Ricci curvature for two Ornstein-Uhlenbeck[19] processes generated in an identical manner except with two different "mean-reversion" rates (see Methods). An Ornstein-Uhlenbeck process describes velocity of a Brownian particle (with mass) under the influence of friction, and is regarded as more realistic than simple Brownian motion. In particular, this illustrative example (Fig. 3) shows that the signal with higher curvature (red) is more capable of returning towards zero (equilibrium) in the face of the same noise (perturbations), illuminating its robustness and as argued previously via the Fluctuation Theorem. One may also consider, for motivational purposes, Ollivier-Ricci curvature on several networks with differing geometries and topologies, and their functionality with respect to robustness (Fig. 4, Fig. S1). Nevertheless, the positive correlation between the rate of return to equilibrium in the Ornstein-Uhlenbeck sense and

Ricci Curvature	BRCA^M	HNSC^M	KIRP^M	LIHC^M	LUAD^M	THCA^M
Δ Average	0.0195	0.0186	0.0229	0.0035	0.0075	0.0117
p-Value	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25
Δ 5% Left Tail	0.0059	0.0077	0.0077	0.0016	0.0022	0.0013
Δ 5% Right Tail	0.0059	0.0057	0.0059	0.0005	0.0017	0.0010

Table 1. A distribution analysis for changes in Ollivier-Ricci average curvature between cancer and normal tissue for all metabolic case studies. These statistics show that there exists a positive shift in the distribution signifying robustness. We also include the one tailed pair Wilcoxon signed rank test p-values to support the above statistics.

Gene Ranking	Δ Ricci Curvature (Cancer-Normal)	Differential Co-Expression	Gene X (Symbol)	Gene Y (Symbol)
1	0.3504	2.7316	RNF43	RSPO3
2	0.3444	2.2061	RNF43	RSPO2
3	0.3012	11.6989	ERG	ETV1
4	0.3001	1.1143	GPC3	PTCH1
5	0.2901	6.6544	GPC3	SDC4
6	0.2796	5.3742	POT1	SBDS
7	0.2538	5.1751	FGFR2	KDR
8	0.2509	−1.1166	ETV1	FEV
9	0.2460	5.2252	ERG	FOXA1
10	0.2410	3.0769	EXT1	SDC4

Table 2. Top 10 pairs with respect to changes in Ollivier-Ricci curvature in BRCA^T.

Ollivier-Ricci curvature, holds for higher dimensions and provides a simple yet informative example linking curvature to robustness.

We then compute the Ollivier-Ricci curvature on tumor and normal tissue networks for all the studied cancer types. We begin with a characterization of the distributions for all networks composed of metabolic genes (≈ 1.25 M possible pairs) as well as our supplemental corresponding networks for which we examine only known cancer related genes (≈ 100 K possible pairs). In particular, we provide an analysis in terms of average curvature, the difference in expected value on upper/lower 5% tails of distribution along with the p-value result of a paired one-tailed Wilcoxon signed-rank test[25] (Table 1, Table S1). This analysis is done, in part, to characterize the shift of distribution with respect to (cancer-normal) changes in Ollivier-Ricci curvature. As such, one can see that the difference between cancer and normal tissue distributions is "positive" with a low p-value signifying robustness. Further, one can consider the left tail of the distribution (at a given 0.1% 0.5%, 1%, 3%) as the lower bound of Ollivier-Ricci curvature as opposed to simply taking minimum value which is sensitive to topological errors. Then, it can be seen that this increase in lower bound points precisely to an increase in entropy[9] (Table S2, Table S3). In all cases of examining the left tail (12 cases at 5 given lengths), the lower bound for a particular cancer network was larger than its normal counterpart. The trend also became more apparent as we decreased the tail length. The largest tail length of 5% was chosen as this was in line with the Wilcoxon signed-rank test. We also note that while we do not restrict our computation to node degree or path length, i.e., curvature is assigned to every gene pair, the average statistic was taken over those interactions with $d(x, y) = 1$. Revisiting equation (1), one can see curvature (and changes in curvature) for interactions "far" from the underlying topology will decay due to the term $d(x, y)$. We should note that since a graph is a 1-geodesic space, if $\kappa(x, y) \geq k$ for $d(x, y) = 1$, then $\kappa(x, y) \geq k \ \forall \ x, y$[18]. Thus, computing statistics (i.e., averages) for adjacent vertices will suffice and results are still valid (in the sense of robustness) for $d(x, y) \geq 2$, e.g., non-adjacent pairs in general will contribute negligibly and can be treated as scaling such statistics.

We note that the primary advantage of employing Ollivier-Ricci curvature is its ability to characterize robustness at the interaction level (as opposed to genes where entropic measures are just defined at the nodal level). In particular, we first report the top and bottom ten interactions with respect to changes in Ollivier-Ricci curvature for the case of BRCA^T (Table 2, Table 3). The investigation of this network is particularly compelling as we sought to find a subset of *interactions* that contribute to the network resilience

Gene Ranking	Δ Ricci Curvature (Cancer-Normal)	Differential Co-Expression	Gene X (Symbol)	Gene Y (Symbol)
99226	−0.2203	−6.2393	POT1	RNF43
99227	−0.2214	1.8210	ELF4	FEV
99228	−0.2304	−7.0964	MUC1	RNF43
99229	−0.2489	−3.5280	DH1	SBDS
99230	−0.2496	−5.7021	ERCC2	RNF43
99231	−0.2839	−3.6199	PRDM1	RNF43
99232	−0.2851	−2.3147	CCND3	ETV1
99233	−0.3632	−6.8679	RNF43	SFPQ
99234	−0.3636	−8.0784	NONO	RNF43
99235	−0.3651	−8.6880	RNF43	TP53

Table 3. Bottom 10 pairs with respect to changes in Ollivier-Ricci curvature in BRCA^T.

(and/or fragility) amongst a set of known cancer related genes. We observe the gene RNF43 exhibits several robust and fragile pathways: RNF43-RSOP3, RNF43-RSOP2, RNF43-TP53, RNF43-NONO, RNF43-POT1. This is a surprising result given that RNF43 physically interacts with very few gene products and in general, is associated as a tumor suppressor in ovarian cancer[26]. On the other hand, RNF43 dominates the largest changes with respect to interaction robustness with several "hidden" non-adjacent pairs. We also computed the differential co-expression (see Methods) for case of breast cancer (both BRCA^M and BRCA^T) and refer the readers to previous work for computational details[20]. In particular, we observed the ranking of interactions of differential co-expression to that of differential Ollivier-Ricci curvature vastly differ (Fig. S2), i.e., we are uncovering hidden information of the underlying system. Similar observations were gleaned from the remaining sets; however, we focus on breast cancer for the sake of brevity.

To this end, we also applied our method to analyze metabolic genes for the case of breast cancer, i.e., BRCA^M (Table S4). While the data did not include various associated cancer genes (i.e., TP53, KRAS, BRAF), we were able to uncover several lesser known targets. In particular, we observed at the top of our list, the gene LPO which has been known to contribute to the initiation of breast cancer[27], SOD3 has been considered an important gene in the defense against oxidative stress and prevention of estrogen-mediated breast cancer[28], GOT2 has been noted to significantly affect cell growth[29], and over-expression of LRAT has lead to a poor prognoses in colorectal cancer[30]. While a complete analysis in the context of cancer biology will be a subject of future work, the above results should be placed in the context "lost information" due to the resolution limitations of network entropy (see Discussion). In short, we now have a proxy for robustness at the local interaction level.

Gene Robustness: Scalar Curvature. Until now, we have considered Ricci curvature (in the Ollivier sense), which is defined between *any* two vertices on a graph. This is the main focus of the present work. However, in order to compare the curvature based approach with that of network entropy[4], we now define several nodal measures based on the notion of "scalar curvature."

In standard geometry, scalar curvature represents the amount by which the volume of a geodesic ball in a curved Riemannian manifold deviates from that of the standard ball in Euclidean space[8]. On a weighted graph, it may be defined in an analogous manner as:

$$S(x) := \sum_y \kappa(x, y) \mu_x(y),$$

(4)

where we contract Ollivier-Ricci curvature with respect to measure $\mu_x(y)$. Analyzing this contraction, we note that the measure $\mu_x(y)$ serves as a normalization factor that attempts to remove biasing with regards to topology (i.e., node degree). We can analogously define the unnormalized scalar curvature by contracting with respect to the hop metric, i.e.,

$$\hat{S}(x) := \sum_y \kappa(x, y)$$

(5)

where the summation runs over all y such that $d(x, y) = 1$.

One may also consider measures where nodal curvature at x in its adjacent neighborhood can be defined as its minimum (maximum) Ollivier-Ricci curvature. Given that lower bounds of Ricci curvature are connected to entropy[9], attaching this bound as a measure yields yet another characterization of nodal

Measure	BRCA^M	HNSC^M	KIRP^M	LIHC^M	LUAD^M	THCA^M
Δ S	0.0119	0.8311	0.0139	0.0036	0.0036	0.0026
p-Value	<1e-25	<1e-25	<1e-25	<7e-14	<4e-21	2e-7
Δ Ŝ	0.8311	0.7925	0.9737	0.1483	0.3205	0.1365
p-Value	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25
Δ S$_e$	0.0116	0.0113	0.0144	0.0025	0.0047	0.0027
p-Value	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25
Δ Ŝ$_e$	0.0373	0.0353	0.0474	0.0070	0.0147	0.0077
p-Value	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25	<1e-25

Table 4. Comparison of different nodal measures for curvature and entropy on all networks composed of metabolic genes.

Gene Ranking	Δ Scalar Curvature (Cancer-Normal)	Differential Expression	Gene (Symbol)
457	−0.0761	−8.5248	SDHD
458	−0.0792	1.8463	SBDS
459	−0.0797	−18.6966	RNF43
460	−0.0841	1.4997	FLI1
461	−0.0884	−3.5989	MYCN
462	−0.1213	6.2235	ETV1
463	−0.1268	2.7942	SDC4
464	−0.1313	4.2820	TFEB
465	−0.1478	1.0500	ELN
466	−0.1870	10.0541	FEV

Table 5. Bottom 10 genes in BRCA^T ranked with respect to scalar curvature.

robustness. We should note that contracting with respect to the measure $\mu_x(y)$ is in the spirit of local normalized entropy defined in previous cancer studies to be $S_e(x) := -\frac{1}{\log \hat{d}_x} \sum_y \mu_x(y) \log(\mu_x(y))$[4]. Similarly, contracting with respect to hop metric above is very much in line with the unnormalized entropy, i.e., $\hat{S}_e(x) := -\sum_y \mu_x(y) \log(\mu_x(y))$[7].

After evaluating the above measures on all cancer networks for which we had data, we found that the results are consistent and comparable in the sense of cancer network differentiation and present an average nodal measure for each cancer study along with the p-value of a one-tailed paired Wilcoxon signed-rank test (Table 4, Table S5, Table S6). We see that there exists a positive shift in the distribution for both entropy and curvature with the exception of only one case where the p-value for ΔS in HNSC^T was insufficient (Table S6). Given that our primary focus of this work is on the interaction level, we present the top and bottom ten pairs in the BRCA^T network with respect to normalized scalar curvature (Table 5, Table S7). This is done, in part, to illustrate the unavoidable "information loss" of any nodal measure chosen. For example, we observe that although the some genes (i.e., RNF43, ETV1) possess the strongest robust interactions, they are listed in the bottom list with respect to scalar curvature. As we will argue in the next section, emphasis should be placed on interactions when analyzing network robustness.

Discussion

In this work, we have presented a framework to quantify interaction gene-to-gene robustness through the notion of Ollivier-Ricci curvature with an application to cancer networks. This was motivated through the intrinsic connection between entropy and Ricci curvature, and in turn, robustness via the Fluctuation Theorem. From this, we demonstrated that cancer tissue exhibits a higher curvature at the interaction and gene level on all the networks tested. While these two measures may provide important biological information[4], it is important to first discuss the differences in the context of our findings and in general, cancer biology. As the eventual goal is to uncover "knock-down" targets (and the effect thereof), we must also explore how one can alter network properties with respect to robustness including changes to the network geometry and topology.

We begin by revisiting the top and bottom gene-to-gene interactions of breast cancer in the studied network composed of known cancer-related genes. At the *interaction* level, changes in robustness need not be restricted to simply a negative/positive change–genes will tend to interact in a wide ranging manner and may contain seemingly important interactions not explicitly defined by the underlying geometry and topology. From equation (4), we clearly see that through the contraction, we "lose" information through a (weighted) average in two distinct manners. Firstly, RNF43 possesses two of the strongest and weakest pairs–averaging these together will cancel out their relative significance. Secondly, nodal measures take an average over an adjacent neighborhood thereby ignoring those interactions that are non-adjacent. As we can see, several important interacting pathways (e.g., RNF43-POT1) should not be ignored as these gene-to-gene interactions exhibit larger changes than many interactions that are adjacent. The same arguments hold for network entropy. Further, previous work on network entropy discusses the significance and careful attention one must have with respect to the topology biasing. Hence, normalization factors are often adopted to provide insight into the nodal robustness[4,7]. No such normalization is required when employing Ollivier-Ricci curvature.

Further, the development of a systematic approach to altering network properties to uncover potential drug targets is key. In particular, certain targets may not be directly "druggable" thereby requiring one to alter a set of genes/interactions that provide similar impact. That is, simply choosing a "knock-down" gene on nodal robustness may prove to be insufficient. To this end, one may consider the corresponding Ricci flow:

$$\frac{d}{dt}d\left(x,y\right) = -\kappa\left(x,y\right)d\left(x,y\right).$$

(6)

Not much is known about this flow, but the idea would be, while keeping the same topology, one would change the graph weights, or the network of links among the nodes, in such a way as to uniformize the curvature κ. In the engineering literature[31], this has been offered as an approach, in the case of certain wireless networks, to have the effect of removing some of the overloaded queues, and thus should have important implications for cancer networks. Understanding discrete analogues of Ricci flow in this connection will be considered as a future research topic in this connection.

Next, we would like to mention some very interesting work[32] that describes a metric geometry on the space of trees in connection with phylogenetics. It turns out that their space is a moduli space (universal parameter space) and has non-positive curvature. From previous results[33], this allows one to do statistics on this space since between any two points there is a unique geodesic. This has had a number of intriguing applications in cancer research[34]. It would be very interesting to generalize this to more general network structures, and instead of just looking at the geometric (curvature) property of an individual network to devise quantitative statistical methods based on the metric geometry comparing families of networks.

Finally, the work of Rabadan[34] has been largely motivated by the problem of cancer cell heterogeneity. Indeed, cancer progression is believed to follow Darwinian evolutionary pattern: fitter subtypes replace other less fit cells, which leads to disease. In combination with high-throughput genomics one can construct trees to study this process. This is an example of a deep relationship between the concepts of Darwinian evolution and Boltzmann thermodynamics[6]. The idea is that macroscopic entropy increases under microscopic molecular collisions, while macroscopic evolution can be (partially) explained via the concept of the increase of entropy. This reasoning is very much in line with the overall thrust of the present paper in which we are trying to use curvature (positively correlated with robustness) to quantify network robustness. The macroscopic theory is very much in line with Boltzmann thermodynamics. Evolutionary changes and network adaptability are key topics to be considered in future research.

Methods

Data.   All TCGA expression data were accessed using the Broad Institute Firehose on November 4, 2014.

Two distinct approaches were used to determine adjacency matrices for the networks under study. For our study of networks of cancer-related genes from Cosmic[21], we used the simple interaction data provided by Pathway Commons project (v6 - accessed in February, 2015 from http://www.pathwaycommons. org/pc2/downloads). To do this, we first downloaded the binary relationships between pairs of genes in Simple Interaction Format. We then filtered the data only for interaction type "neighborhood-of" that represents any type of pathway-based interaction between a pair of genes. We next filtered out all interactions in which either of the interacting genes was not in our cancer gene set, and therefore not of interest to us.

To identify adjacent edges in the metabolic gene data set, we used the Recon2 human metabolic reconstruction[22] to identify pairs of genes whose enzymatic products shared a common substrate or product. To do so, we pruned the stoichiometric matrix (S) for cofactors (ATP, ADP, NADH, NAD, NADPH, NADP, etc.) and other highly-connected metabolites which might adversely affect the adjacency calculations (*e.g.* water, hydrogen ions, metal cofactors). We then used this pruned stoichiometric matrix S_P, and the reaction-to-gene matrix (R) to generate a matrix encoding which metabolites and genes participated in common reactions ($MG = S \times R$). Finally, to generate an adjacency matrix (A)

indicating which genes participated in reactions sharing a common metabolite, we multiplied the transpose of MG by itself: $A = MG^T \times MG$. The matrix A is square, with the length of each dimension equal to the number of genes in the model. The curvature analysis was also repeated after removing highly connected reactions (*i.e.* with greater than 4 distinct metabolite substrates/products, after pruning for highly connected metabolites) from S with qualitatively similar results.

For each data set, gene co-expression networks were generated by calculating the non-parametric (Spearman) correlation between all pairs of genes. Note that since we are working with correlation data for which values can be less than zero, the analysis was conducted with respect to the transformed correlation coefficient: $\bar{c}_{xy} = 0.5 \cdot \left(1 + c_{xy}\right)$ in order to construct the random walk over the network[4]. We should also note, that one could examine and compute weights through an interesting mass action approach[13,14] as opposed to a more general computation of correlation values given above. The advantage of the mass action method is that from an statistical standpoint, it allows for the analysis to be carried out in a more sample specific manner. Moreover, given that biological networks involve both negative and positive weights representing specific activating and inhibiting interactions, a subject of future research will entail directly extending the current approach to more general directed graph case following related work[35].

The Wasserstein Distance. We begin by recording the basic definition of the L^p-Wasserstein distance from optimal transport theory that we will need below. Roughly speaking, on a metric measure space, one gets a natural distance on "small" balls around points or the "fuzzified" points. For full details about the Monge-Kantorovich (optimal mass transport) problem and the associated Wasserstein distance, we refer the reader to several works on this topic[23,24,36–38].

More precisely, let X be a metric measure space, equipped with distance d. Let μ_i, $i = 1, 2$, be two measures with the same total mass and finite p-th moment. A *coupling* between μ_1 and μ_2 is a measure μ on $X \times X$ such that

$$\int_y d\mu(x, y) = d\mu_1(x), \quad \int_x d\mu(x, y) = d\mu_2(y) \tag{7}$$

In other words, the marginals of μ are μ_1 and μ_2. Let $\Pi(\mu_1, \mu_2)$ be the set of couplings between μ_1 and μ_2. We then define the L^p *Wasserstein distance* as

$$W_p(\mu_1, \mu_2) := \left[\inf_{\mu \in \Pi(\mu_1, \mu_2)} \int \int d(x, y)^p d\mu(x, y) \right]^{1/p} \tag{8}$$

In this paper, we only consider the cases $p = 1, 2$. For $p = 1$, the Wasserstein distance is sometimes called the "Kantorovich-Rubinstein distance" or the Earth Mover's distance (EMD) and can be formulated as linear program[12]. In particular, let X denote a discrete metric measure space with n points denoted $\{x_1,\ldots,x_n\}$. Let μ_1 and μ_2 be two distributions, and let $d(x, y)$ denote the distance between $x, y \in X$ (for the case of graphs, this is simply taken to be the hop distance). We assume that μ_1 and μ_2 have the same total mass. Then, $W_1(\mu_1, \mu_2)$ may be defined as follows:

$$W_1(\mu_1, \mu_2) = \min_{\mu} \sum_{i,j=1}^{n} d(x_i, x_j) \mu(x_i, x_j) \tag{9}$$

where $\mu(x, y)$ is a coupling (or flow) subject to the following constraints:

$$\mu(x, y) \geq 0 \quad \forall \, x, y \in X \tag{10}$$

$$\sum_{i=1}^{n} \mu(x, x_i) = \mu_1(x) \quad \forall \, x \in X \tag{11}$$

$$\sum_{i=1}^{n} \mu(x_i, y) = \mu_2(y) \qquad \forall \, y \in X. \tag{12}$$

The cost above finds the optimal coupling of moving a set of mass from distributions μ_1 to μ_2 with minimal "work".

Curvature and Robustness. There have been a number of approaches[19,39–41] to extending the notion of Ricci curvature to more general metric measure spaces. At this point, the exact relationship of one approach as compared to another is unclear. Roughly, the techniques fall into two categories: the first generalizing the weak k-convexity of the entropy functional on the Wasserstein space of probability measures as in[9,39,42], and the second directly working with Markov chains to define the generalization[19,40,41] on networks. There is also a notion of "hyperbolicity" due to Gromov[43] based on the "thinness" or

"fatness" of triangles compared to the Euclidean case, and more generally a certain four-point criterion. Depending upon the application, each approach seems to be useful. In particular, we follow[19,39], because of connections to notions of metric entropy.

We first define the precise notion of "robustness" to which the Fluctuation Theorem[6,44], is applicable. One considers random fluctuations (perturbations) of a given network that result in deviations of some observable. Let $P_\varepsilon(t)$ denote the probability that the mean deviates by more than ε from the original (unperturbed) value at time t. Since $P_\varepsilon(t) \to 0$, we want to measure its relative rate, that is, we set

$$R := \lim_{t \to \infty} \left(-\frac{1}{t} \log P_\varepsilon(t) \right). \tag{13}$$

Therefore, large R means not much deviation and small R large deviations. In thermodynamics, it is well-known that entropy and rate functions from large deviations are very closely related.

Next we describe the relationship of curvature and entropy given in Lott and Villani[9]. Let (X, d, m) denote a geodesic space, and set

$$P(X, d, m) := \left\{ \mu \geq 0 : \int_X \mu \, dm = 1 \right\}, \tag{14}$$

$$P^*(X, d, m) := \left\{ \mu \in P(X, d, m) : \lim_{\varepsilon \searrow 0} \int_{\mu \geq \varepsilon} \mu \, \log \mu \, dm < \infty \right\}. \tag{15}$$

We define

$$H(\mu) := \lim_{\varepsilon \searrow 0} \int_{\mu \geq \varepsilon} \mu \, \log \mu \, dm, \quad \text{for } \mu \in P^*(X, d, m), \tag{16}$$

which is the negative of the *Boltzmann entropy* $S_e(\mu) := -H(\mu)$; note that the concavity of S_e is equivalent to the convexity of H. Then we say that X has *Ricci curvature bounded from below by k* if for every μ_0, $\mu_1 \in P(X)$ there exists a constant speed geodesic μ_t with respect to the Wasserstein 2-metric connecting μ_0 and μ_1 such that

$$S_e(\mu_t) \geq t S_e(\mu_0) + (1 - t) S_e(\mu_1) + \frac{kt(1 - t)}{2} W(\mu_0, \mu_1)^2, \quad 0 \leq t \leq 1. \tag{17}$$

This indicates that entropy and curvature are *positively correlated* that we will express as

$$\Delta S_e \times \Delta Ric \geq 0. \tag{18}$$

We note here that changes in *robustness*, i.e., the ability of a system to functionally adapt to changes in the environment (denoted as ΔR) is also positively correlated with entropy via the Fluctuation Theorem[6,44], and thus with network curvature:

$$\Delta R \times \Delta Ric \geq 0. \tag{19}$$

Since the curvature is very easy to compute for a network, this may be used as an alternative way of expressing functional robustness. This being said, we adopt the recent notion of Ollivier-Ricci curvature motivated from coarse geometry[18,19].

Ornstein-Uhlenbeck Process. It is very informative to consider the relationship of the the Ollivier-Ricci curvature and robustness via a simple example[18,19]. We consider the Ornstein-Uhlenbeck process. The latter is a modification of the Wiener process (random walk), in which there is a tendency to converge to a central location.

More precisely, consider the stochastic differential equation

$$dX_t = -\alpha X_t dt + \sigma dW_t, \quad X(0) = x_0, \tag{20}$$

where W is Brownian motion (Wiener process), and we take x_0 to be deterministic. We treat the 1-dimensional case for simplicity. Everything goes through in higher dimensions as well. The corresponding Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = \alpha \frac{\partial x p}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2}, \tag{21}$$

where $p = p(x, t | x_0, 0)$ is the transition probability of the underlying Markov process. One may show that $p(x, t | x_0, 0)$ is a Gaussian process with mean and variance given by[45]:

$$\langle X(t) \rangle = x_0 e^{-\alpha t}, \quad \mathrm{var} X(t) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}).$$

(22)

We see that we get transition probabilities of mean $x_0 e^{-\alpha t}$ and variance independent of x_0. Since all the transitions $p(x, t|x_0, 0)$ have the same variance (and are Gaussian) the 1-Wasserstein distance[46]

$$W_1(p(x, t|x_0, 0), p(x, \delta t|x_1, 0)) = |x_0 e^{-\alpha t} - x_1 e^{-\alpha t}|.$$

(23)

Finally,

$$\kappa(x_0, x_1) = 1 - \frac{W_1(p(x, t|x_0, 0), p(x, t|x_1, 0))}{|x_0 - x_1|} = 1 - e^{-\alpha t}.$$

(24)

Equation (24) illustrates the connection of fluctuation in a very simple explicit manner. Larger α corresponds to larger curvature κ and this corresponds to how quickly the systems returns to equilibrium, that is to the mean going to 0.

Convergence to Invariant Distribution. One can also see that relationship of robustness to the Ollivier-Ricci curvature in the following manner[18] dealing with Markov chains. The basic idea is that larger Ollivier-Ricci curvature indicates greater robustness via rate of convergence to the invariant (equilibrium) distribution. Specifically, suppose $\kappa(x, y) \geq k > 0$. Then there exists a unique invariant probability measure ν. Moreover, for any x,

$$W_1(\mu_x^{*t}, \nu) \leq \frac{W_1(\delta_x, \mu_x)}{k}(1 - k)^t.$$

(25)

Here,

$$\mu_x^{*t}(y) := \sum_{z \in X} \mu_x^{*(t-1)} \mu_z(y), \quad \mu_x^{*1} := \mu_x.$$

(26)

Note that $W_1(\delta_x, \mu_x)$ represents the jump of the random walk at x. On a connected graph X with diameter D (defined as the longest graph geodesic), this yields the following estimate for the mixing time:

$$\frac{1}{2}\sum_{y \in X}|\mu_x^{*t}(y) - \nu(y)| \leq D(1 - k)^t.$$

(27)

This example combined with the previous one provides further support that Ollivier-Ricci curvature can be employed as a natural proxy for robustness with the the distinct advantage of being easily computable.

Differential (Co-)Expression. We conclude with a simple computation of differential co-expression. Following previous work[20], differential co-expression was computed using the (non-transformed) sample correlation coefficient c_{xy} by first applying the Fisher z-transformation in order to stabilize variances due to population size:

$$z_{xy} = \frac{1}{2}\log\left(\frac{1 + c_{xy}}{1 - c_{xy}}\right).$$

(28)

If we let z_{xy}^T and z_{xy}^N denote the z-transformation for cancer and normal gene pairs, respectively, one can then compute the differential co-expression as

$$\Delta z_{xy} = \frac{z_{xy}^T - z_{xy}^N}{\sqrt{\frac{1}{N_T - 3} + \frac{1}{N_N - 3}}}$$

(29)

where N_T and N_N is the number of tumor and normal samples respectively. For differential expression values, we summed those co-differential values for a given gene's interaction defined by the underlying adjacency matrix. This was done in order to provide a fair comparison to the values computed by scalar curvature. We again note that complete information regarding this method and data can be found in previous work[20].

References

1. Albert, R. & Barabási, A. Statistical mechanics of complex networks. *Reviews of Modern Physics.* **74,** 47 (2002).
2. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall, 2006).

3. Barabasi, A. The network takeover. *Nature Physics*. **8,** 14–16 (2012).
4. West, J., Bianconi, G., Severini, S. & Teschendorff, A. Differential network entropy reveals cancer system hallmarks. *Scientific Reports*. **2,** (2012). doi: 10.1038/srep00802
5. Kitano, H. Cancer as a robust system: implications for anticancer therapy. *Nature Reviews Cancer*. **4,** 227–235 (2004).
6. Demetrius, L. Boltzmann, Darwin and directionality theory. *Physics Reports*. **530,** 1–85 (2013).
7. Demetrius, L. & Manke, T. Robustness and network evolution - an entropic principle. *Physica A*. **346,** 682–696 (2005).
8. DoCarmo, M. *Riemannian Geometry* (Birkhauser, 1992).
9. Lott, J. & Villani, C. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*. **169,** 903–991 (2009).
10. Tannenbaum, A. *et al.* Ricci curvature and robustness of cancer networks. http://arxiv.org/abs/1502.04512. (2015).
11. Teschendorff, A. & Severini, S. Increased entropy of signal transduction in the cancer metastasis phenotype. *BMC Systems Biology*. **4,** 104 (2010).
12. Rubner, Y., Tomasi, C. & Guibas, L. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*. **42,** 99–121 (2000).
13. Teschendorff, A., Sollich, P. & Kuehn, R. Signalling entropy: A novel network-theoretical framework for systems analysis and interpretation of functional comic data. *Methods* **67,** 282–293 (2014).
14. Banerji, C. *et al.* Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Scientific Reports* **3,** (2013). doi: 10.1038/srep03039
15. Csete, M. & Doyle, J. Reverse engineering of biological complexity. *Science* **295,** 1664–1669 (2002).
16. Bauer, F., Jost, J. & Liu, S. Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator. http://arxiv.org/abs/1105.3803 (2011).
17. Klartag, B., Kozma, G., Ralli, P. & Tetali, P. Discrete curvature and abelian groups. http://arxiv.org/abs/1501.00516 (2015).
18. Ollivier, Y. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*. **256,** 810–864 (2009).
19. Ollivier, Y. Ricci curvature of metric spaces. *C. R. Math. Acad. Sci. Paris*. **345,** 643–646 (2007).
20. Reznik, E. & Sander, C. Extensive decoupling of metabolic genes in cancer. *bioRxiv*. 008946, (2014).
21. Forbes, S. *et al.* COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Research*. (2010). doi: 10.1093/nar/gkq929
22. Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nature Biotechnology*. **31,** 419–425 (2013).
23. Villani, C. *Optimal Transport, Old and New*. (Springer-Verlag, 2008).
24. Villani, C. *Topics in Optimal Transportation*. (American Mathematical Society Publications, 2003).
25. Dekking, F., Kraaikamp, C., Lopuhaa, H. & Meester, L. *A Modern Introduction to Probability and Statistics: Understanding Why and How* (Springer Science & Business Media, 2005).
26. Ryland, G. L. *et al.* RNF43 is a tumour suppressor gene mutated in mucinous tumours of the ovary. *The Journal of Pathology*. 469–476 (2013). doi: 10.1002/path.4134
27. Ghibaudi, E. *et al.* Can estrogenic radicals, generated by lactoperoxidase, be involved in the molecular mechanism of breast carcinogenesis. *Redox Report*. **5,** 229–235 (2000).
28. Singh, B. & Bhat, H. K. Superoxide dismutase 3 is induced by antioxidants, inhibits oxidative DNA damage and is associated with inhibition of estrogen-induced breast cancer. *Carcinogenesis*. (2012). doi: 10.1093/carcin/bgs300
29. Son, J. *et al.* Glutamine supports pancreatic cancer growth through a Kras-regulated metabolic pathway. *Nature*. **496,** 101–105 (2013).
30. Brown, G. *et al.* The expression and prognostic significance of retinoic acid metabolising enzymes in colorectal cancer. *PloS one*. **9,** e907776 (2014).
31. Wang, C., Jonckheere, E. & Banirazi, R. Wireless network capacity versus Ollivier-Ricci curvature under Heat Diffusion (HD) protocol. Paper presented at IEEE American Control Conference. Portland, OR. June 04-06. 3536–3541 (2014). doi: 10.1109/ACC.2014.6858912
32. Billera, L., Holmes, S. & Vogtmann, K. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*. **27,** 733–767 (2001).
33. Sturm, K. Probability measures on metric spaces of nonpositive curvature. *Contemporary Mathematics*. **338,** 1–34 (2003).
34. Zairis, S., Khiabanian, H., Blumberg, A. & Rabadan, R. Moduli spaces of phylogenetic trees describing tumor evolutionary patterns. *Lecture Notes in Computer Science (LNCS)*. **8609,** 528–539 (2014).
35. Bauer, F. Normalized graph Laplacians for directed graphs. *Linear Algebra and its Applications*. **436,** 4193–4222 (2012).
36. Evans, L. C. Partial differential equations and Monge–Kantorovich mass transfer. *Current Developments in Mathematics*. 65–126 (1999).
37. Rachev, S. & Rüschendorf, L. Mass Transportation Problems, Vol. I and II. (Springer-Verlag, 1998).
38. Tannenbaum, E., Georgiou, T. & Tannenbaum, A. Signals and control aspects of optimal mass transport and the Boltzmann entropy. Paper presented at IEEE *Conference on Decision and Control*. Atlanta, GA. December 15-17. 1885–1890 (2010). doi: 10.1109/CDC.2010.5717821
39. Bonciocat, A. & Sturm, K. Mass transportation and rough curvature bounds for discrete spaces. *Journal of Functional Analysis*. **256,** 2944–2966 (2009).
40. Maas, J. Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis*. **261,** 2250–2292 (2011).
41. Chow, S., Huang, W., Li, Y. & Zhou, H. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Archive for Rational Mechanics and Analysis*. **203,** 969–1008 (2012).
42. McCann, R. A convexity principle for interacting gases. *Adv. Math.* **128,** 153–179 (1997).
43. Gromov, M. Hyperbolic groups. *Essays in Group Theory in Math. Sci. Res. Inst. Publ.* **8,** 75–263 (1987).
44. Demetrius, L., Gundlach, V. & Ochs, G. Complexity and demographic stability in population models. *Theoret. Population Biol.* **65,** 211–225 (2004).
45. Gardiner, C. *Handbook of Stochastic Methods, Third Edition*. (Springer, 2004).
46. Givens, C. & Shortt, R. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*. **31,** 231–240 (1984).

Acknowledgements

Author Contributions

R.S. analyzed the data, developed computer code for computing curvature and statistics, and contributed to the theory. T.G. helped in the formulation of the graph curvature approach. E.R. provided and analyzed data, and contributed in writing the biological analysis of the results. L.Z. and I.K. developed the code for the computation of graph curvature. Y.S. provided network data used in the study. A.T. developed the mathematical framework used in the present work, and oversaw the overall formulation and writing of the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sandhu, R. *et al.* Graph Curvature for Differentiating Cancer Networks. *Sci. Rep.* **5**, 12323; doi: 10.1038/srep12323 (2015).