A Graph Based Analysis of Tumour Vessel Networks

Magdy Saleh mksaleh@stanford.edu
Surag Nair surag@stanford.edu
Abhijeet Shenoi ashenoi@stanford.edu

Abstract

In this work, we use graph analysis techniques and concepts to carry out a comparative study between healthy tissue and two distinct types of tumours. Through an investigation of the structural properties of the tumour networks, we were able to ascertain key differences between the two tumour types. To further substantiate these differences, we carried out community detection and obtained localised communities in the LS174T tumour. We also performed network robustness tests which showed that the LS174T tumour is more robust to both targeted and random attacks when compared to the SW1222 tumour. Finally, to determine the most effective scheme for drug delivery we carried out influence maximisation, and showed that injecting drugs into less than 2% of the inlets in either tumour can cause drug delivery to over 80% of the tumour mass in the LS174T tumour and over 65% of the tumour mass in the SW1222 tumour.

1 Introduction

Small tumours survive by drawing on oxygen and nutrients supplied from the nearest existing vasculature. Once a tumour's size exceeds $1 - 2$ mm^3, it can no longer survive based on the diffusion limit of oxygen and nutrients. In order to continue growing and maintain its supply of oxygen and nutrients, an angiogenic process to build its own vasculature is triggered. Tumour vasculature consists of vessels recruited from pre-existing networks as well as angiogenic vessels, and differs significantly from healthy vasculature on both a micro (vessel properties) and a macro (network geometry) scale. [1]

Many of the most promising drugs designed to combat cancer have not found success due to the poor distribution of the drug within the tumour. Studies have shown that this poor perfusion can be attributed to the chaotic nature of tumour vasculature geometry. [2] Recent advancements in imaging technologies have allowed for the collection of datasets that model tumour vasculature as graphs. These techniques rely on staining blood vessels with bioflorescence and performing optical clearing, which makes the refractive index of all tissue very close to 1. The tumours are then imaged using Optical Projection Tomography, which relies on the bioflorescence, and the resulting signal received is used to reconstruct the vessel network structure with a high resolution, capturing vessels with diameters of less than $1\mu m$.

Using these graphs of vessels, where the edges are blood vessels and the nodes are points of splitting or joining of vessels, we look to apply network analysis tools to better understand the underlying characteristics of these networks. Specifically, we look to uncover structural differences between different tumour types as well as investigate how the structure of these graphs influence the effective spreading of drugs, which continues to be an unsolved problem in the field of oncology.

2 Related Work

There has been some previous work done in analyzing tumour vessels from a network perspective. Skinner et al. in 1990 reported higher frequency of branching within tumour vasculature compared to healthy vasculature [3]. These findings were expanded by Less et al. who directly studied the branching patterns of tumour vasculature and found two different types of branching patterns. The first was characterized by decreasing vessel diameter and length in successive generations of vessels. The second was characterized by higher fluctuations in both radius and length corresponding to higher degrees of branching. They reported the presence of loops in the vasculature and categorized them into "self-loops", which are loops between two nodes consisting of just two edges and "true loops" which are loops consisting of multiple nodes [4]. Pries et al. suggested based on the above properties that this can give rise to a "shunt problem" in the tumour vasculature, whereby short, low resistance paths divert blood away from longer paths [5].

This work is mainly a continuation of the findings of D'esposito et al. [6]. They identified two graph properties that contribute to drug distribution and propagation in two different tumour networks, LS174T and SW1222. The authors find that functional connectivity and redundancy contribute to lower drug perfusion. In this work we look at structural differences in the graphs in addition to the above metrics to further analyze the difference between the different tumour types.

3 Objectives

We outline the following aims for this work:

I. Understand the structural differences between the different tumour types: how do these tissues differ in their network topologies?

II. Investigate influence maximization: how many inlets of the network are required to target in order to reach a majority of the network?

III. Quantify robustness: how do the different networks withstand random and targeted attacks?

4 Data Sets

From the Walker-Samuel Lab at UCL, we obtained two different networks of colorectal adenocarcinoma tumours. The first network is that of an LS174T mouse model of a colorectal carcinoma. The second network is that of an SW1222 carcinoma mouse model. Structural properties of both networks, as well as the reference vasculature have been provided in Table 1. Additionally, we obtained a small 2-D network representing the vasculature of a rat mesentery. We used this network as

a proxy for healthy vasculature. The nodes correspond to points of splitting of vessels, and the edges represent the vessels themselves. These two tumour models have markedly different phenotypes: SW1222 tumours are well differentiated with a well-perfused vascular network, whereas LS174T tumours show moderate-to-poor differentiation and are comparatively less-well vascularized and perfused, with larger areas that have died due to lack of oxygen supply. [7] Based on simulations run by the Walker-Samuel lab, we possess scalar values on each of the edges that describe the simulated pressure, stress and flow.

Table 1: Graph properties of the data sets

Graph	Nodes	Edges	Clustering Coef.	Avg. Deg.	Edge Density
Mesentary	388	546	0.038	2.8093	0.728%
LS174T	16,354	20,377	0.024	2.4192	0.015%
ER LS174T	16,357	20,377	0.00014	2.4192	0.015%
SW1222	64,239	77,450	0.027	2.4113	0.004%
ER SW1222	64,239	77,450	0.00001	2.4113	0.004%

5 Data Preprocessing

The data we received contained some clear artifacts that we needed to correct for. First, there were multiple self edges, which are clearly erroneous as there can't be flow in a zero pressure gradient loop. Further, the graphs under consideration are directed, but the ordering of nodes in the produced edge lists does not accurately reflect the direction of flow. We corrected for this by ensuring that the start node for each edge was the node with the higher pressure.

Furthermore, there is an artifact in the data coming from the pressure simulation values associated with each node. There are multiple edges with non-zero flow but zero pressure drop across them caused by discretization errors. For the purposes of this project we omit these edges from our networks, being cognizant of the fact that this will have an impact on flow conservation. These were confirmed by the Walker-Samuel Lab as existing artifacts which were accounted for.

Moreover, there are pairs of nodes that share multiple edges, however, the pressure drops across these edges are sometimes not of the same magnitude. We use the mean of the pressure drop values in such cases.

6 Methods

In this section, we specify the relevant methods we have identified for analysing our graphs.

6.1 Clustering Coefficient

Clusters in vasculature networks denote nodes that are highly interconnected, as is often in the case in capillaries. The clustering coefficient c_n for a given node n is defined as:

$$c_n = \frac{2 \cdot t_n}{k_n(k_n - 1)}$$

where k_n is the degree of the node and t_n is the number of edges between the neighbours of n. The clustering coefficient of the graph is then the average of clustering coefficients over all nodes.

We find that the vasculature networks have high clustering coefficient when compared to an Erdös-Rényi graph due to the high edge density around the capillaries of the vasculature network.

6.2 Motif Discovery

We explore the motifs present in the graphs using the Exact Subgraph Enumeration (ESU) algorithm [8]. We first generate the set of non-isomorphic directed motifs having k nodes and then use the ESU algorithm to get counts of these motifs in our graph. We then compare these counts to those of the corresponding rewired networks. The rewiring method randomly alters the edges in the network while preserving the in and out degrees of each node. The statistical significance Z_i of motif i in the network is defined as:

$$Z_i = \frac{N_i^g - \bar{N}_i^r}{\text{std}(N_i^r)}$$

where N_i^g is the count of the motif in the real graph g, $\bar{N}_i^r$ and $\text{std}(N_i^r)$ are the average count and standard deviation of the motif count over multiple rewired graphs, respectively. In cases in which the count of the motif is the same value c in the real graph and in all rewired graphs, we set $Z_i = 0$.

Vasculature networks are networks of blood flow in the vessels, where the flow is induced by pressure difference between any two nodes. This precludes the presence of cycles in the graph. As a result, we expect motifs that don't have cycles to be highly enriched.

6.3 Community Detection

We performed community detection using the Leiden community detection algorithm from [9], while maximizing the modularity. The modularity, Q, is given by

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta_{ij}$$

However, our graph is directed, and so a slightly modified formulation is used.

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{2m} \right] \delta_{ij}$$

where k_i^{out} and k_j^{in} are the out and in degrees of nodes i and j respectively. We used an open source implementation of the algorithm which greedily maximizes the modularity, in an iterative fashion, much like the Louvain algorithm. However, unlike the Louvain algorithm, there is a guarantee that the generated communities are connected. The aim is to detect the presence of communities, and compare and contrast the characteristics of said communities in the two types of tumors. An additional exploration is done on a Mesentery Network to serve as a baseline.

6.4 Influence Maximisation

Lazy hill climbing was used to perform influence maximisation. Lazy hill climbing differs from vanilla hill climbing by reducing the number of influence evaluations needed. It does so by maintaining an ordered list of marginal gains, and a form of pruning to only evaluate a subset of the new influences. The algorithm can be summarised in the following steps:

1. Keep a list of marginal benefits δ_i from the previous iteration
2. Reevaluate δ_i only for top node
3. If the reevaluated δ_i is the largest marginal benefit, select this node. If not, reinsert it to maintain sorted order, and return to step 2.

We implemented lazy hill climbing from scratch, and were able to run influence maximisation on our networks. We restricted the set of possible nodes which could be selected by only considering inlets. An inlet is defined as a node with no incoming edges. The goal of running influence maximisation on these inlets is to determine which inlets supply a majority of blood to the tumour. This leads to far ranging implications when designing drug delivery systems.

6.5 Betweenness Centrality

Betweenness centrality was employed to conduct targeted attacks in both networks while measuring network robustness. To that end, we used the existing snap implementation. Betweenness Centrality aims to measure how prominent a node or an edge is, when considering all shortest paths between pairs of nodes in a network. For a node k, the betweenness centrality B_k can be written as:

$$B_k = \sum_{i,j} \frac{\mathcal{N}_{ijk}}{\mathcal{N}_{ij}}$$

where $\mathcal{N}_{ijk}$ is the number of shortest paths between nodes i and j which pass through k, and $\mathcal{N}_{ij}$ is the number of shortest paths between nodes i and j.

6.6 Network Robustness

We look at how the different networks react to nodes being deleted. We use the size of the largest weakly connected component as a proxy for network fragmentation.

We perform two kinds of attacks. The first is random attacks where we choose nodes at random and delete them from the graph as well as all the edges they connect to. The second kind of attack is targeted, where we used different metrics to order the nodes, including node betweeness centrality, flow, radius and degree.

7 Results

7.1 Degree Distribution

We first consider the degree distributions within each of the graphs. We observe, as seen in Figure 1a, that there is a high concentration of small degree nodes as expected, with a clear peak at 3. This is consistent with the relevant physiology, since bifurcations are the most prevalent branching pattern.

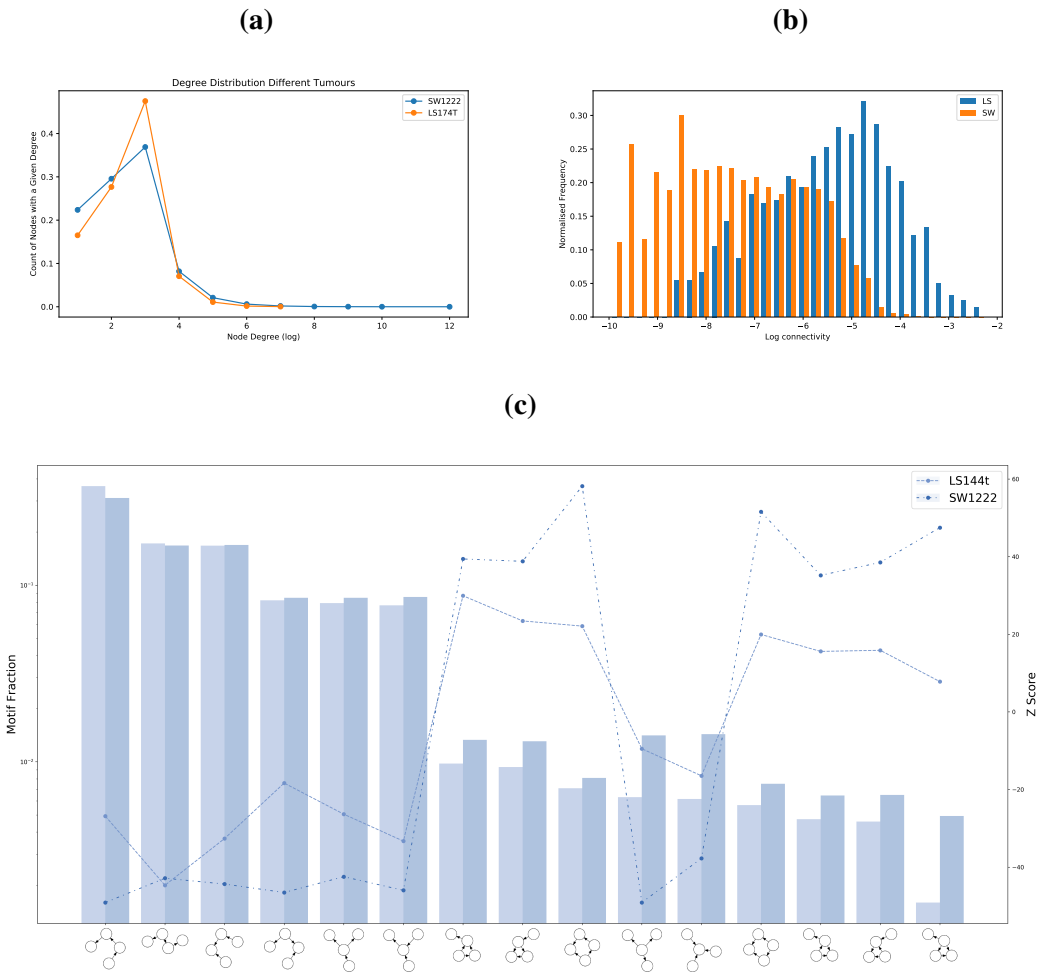

Figure 1: **(a)** Comparing the degree distribution between LS174T and SW1222 tumours. We see that there is a greater degree diversity in the SW1222 tumours, which is expected as they tend to have greater vascular density **(b)** Distribution of node connectivity for both tumours. We find that nodes in the SW1222 tumour have on average lower connectivity than the LS174T. **(c)** Motif distribution for the top 15 4-node motifs in LS174T and SW1222 tumours. We calculate motif fraction by dividing the count of a specific motif by the count of all motifs in the graph. We find that there are no motifs with cycles among the top motifs.

7.2 Motif Analysis

We analyzed the occurrence of motifs with 4 nodes in the different graphs as seen in Figure 1c. What we find, unsurprisingly is that by far the most common motifs are cycle free. Motifs representing successive nodes that connect to one another as a chain, bifurcations, and joining of veins are especially proliferate. However, what is not intuitive, is that these highly common motifs are in fact underrepresented - they are more common in randomly rewired null models. We believe that this cause by the fact that the nature of these tumor networks is such that there is a large number of motifs corresponding to bifurcations and combinations of two vessels at a given node. The count for these motifs is unlikely to reduce on rewiring, due to the lack of long range edges in the network.

We carried out an analysis of both the normalized motif counts, normalized for the total number of motifs in the graph, and the Z-scores. The normalized counts themselves are as expected, whereas the Z-scores follow from the argument detailed above. While there are no motifs with strongly connected cycles we find that the motifs that are over-represented in our network are those with weakly connected cycles.

We also find that the two networks have very similar structures, when considering motif distributions for three node motifs.

7.3 Community Detection

The community detection algorithm when run on the mesentery network yielded the results in Figure 2a. The nodes have been visualised using real world coordinates in 3D space, which were then projected onto a 2D plane. It is evident that communities consist of nodes in close physical proximity, in line with our expectation that within the capillary bed, nodes are likely to be highly clustered.

When running the algorithm on the LS174T tumour, we found that there exists a community structure, similar to what was observed in the mesentery network. We visualised the results at a node level, as well as at a community level. This can be seen in Figure 2b, Figure 2c, and Figure 2d.

The SW1222 tumour did not yield similar results. Communities were largely random, and showed radius, flow and pressure profiles which were nearly identical to the distributions of these values across the entire tumour.

We also carried out role discovery for all three networks. After obtaining 2 hop RolX features for each node, we ran k-means clustering to identify potential roles in the networks. Without any a-priori knowledge of possible roles, the results of role discovery were inconclusive.

7.4 Connectivity Analysis

In this analysis, we studied the distribution of nodes which can be reached and are reachable by a node in the graph. This gives us a notion of the connectivity of a graph– a graph with a higher average connectivity is more likely to have higher redundancy and hence more robustness. In Figure 1b, we computed the sum of sizes of the in-set and the out-set for each node divided by the total number of nodes in the network, and plot the distribution. The analysis revealed that the LS174T tumour has a significantly higher connectivity than the SW1222 tumour. LS174T has an average connectivity of 0.90% whereas SW1222 has an average connectivity of 0.14%.

7.5 Network Robustness

We proceeded to analyse the robustness of the vasculature networks when subjected to random and targeted attacks. This has potential applications in targeted radiation therapy for tumours— knowing the network structure of tumours can aid in determining the dosage of radiation required to break it down, and even decide which regions of the tumour need to be targeted. We consider the size of the largest weakly connected component as a surrogate for network robustness. Isolating weakly connected components breaks connectivity and hinders the growth of the tumour. We consider random attacks and targeted attacks on the LS174T, SW1222 tumours as

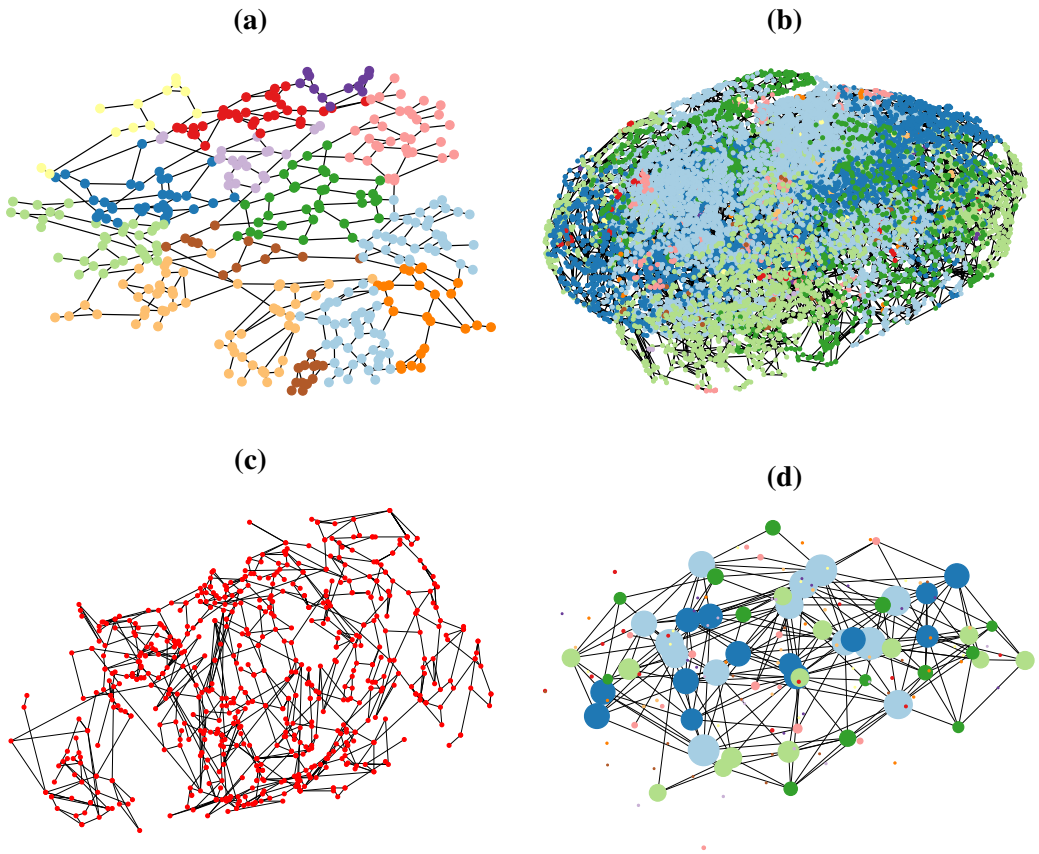

Figure 2: **(a)** Mesentery Network Communities **(b)** Community Detection results on the LS174T tumour. Each node is coloured by the community it belongs to. The nodes are shown in the figure as they are placed in 3D space. **(c)** The largest community found in the LS174T tumour, arranged with spatial coordinates **(d)** Contracted communities. Each node represents a community, and the size of each node is proportional to the size of that community

well as Erdös-Renyi graphs with the same number of nodes and edges. For targeted attacks, we compute the approximate node betweenness values, and then remove nodes in order of decreasing betweenness. The results are presented in Figure 3b.

We observe that in general, the LS174T network is more robust than SW1222. However, both the networks are less robust than the random Erdös-Renyi networks. As expected, the targeted attacks are more effective in disrupting the networks. The most notable observation is that disrupting only approximately 30% of the nodes in either LS174T or SW1222 brings down the size of the largest weakly connected component to 10% of the total nodes in the network. Moreover SW1222 is more susceptible to targeted attacks than LS174T. These results may have interesting implications for both radiation and drug based therapies.

7.6 Influence Maximization

Given the subtle structural differences we observed between the two tumour types, we turn to the second objective. We ran influence maximization using our own implementation of the Lazy Hill Climbing algorithm on both the LS174T and the

(a) **(b)**

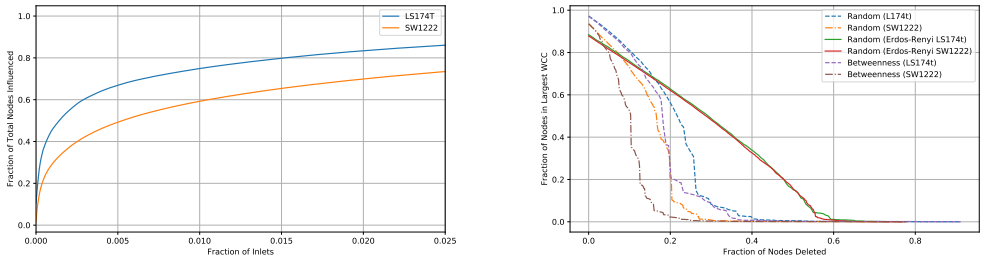

Figure 3: **(a)** Influence Maximization on the inlets when run on both the LS174T and SW1222 networks, using lazy hill climbing. Number of inlets chosen was set to 1795, which represents the total number of inlet nodes in the smaller network (LS174T). **(b)** Robustness analysis of the networks on LS174T, SW1222 and Erdös-Renyi networks to random and targeted attacks. In the targeted attack, the nodes are removed in decreasing order of betweenness.

SW1222 networks. This was motivated by the problem of effective drug transport in tumours. We consider the most influential k inlets to the network and evaluate the total fraction of nodes in the network that they influence. What we find is that the LS174T tumour requires fewer inlets to influence a higher proportion of the nodes as seen in Figure 3a.

If we consider only the top 2% most influential source nodes in each network we influence 83.4% of the nodes in the LS174T and 69.9% of nodes in the SW1222.

This is consistent with literature, which finds that the SW1222 tumours tend to have greater network homogeneity and so nodes deeper in the network are less dependant on single inlets, (i.e. inlet influence sets have much more overlap in SW1222 than those of the LS174T network). [6]

8 Discussion

In this work we set out to answer three main questions. First, we are interested in understanding the structural differences between the networks. What we find is that although both tumours differ in their spatial density of nodes and vasculature, a 3D-space agnostic network perspective fails to unearth any clear differences. We see that both networks present similar degree distributions (see Figure 1a) and motifs (see Figure 1c). The main structural difference we see concerned the edge density. We know that the SW1222 tumour is spatially more dense but given the prepossessing assumptions we made we find that there is no greater density of edges in the SW1222 tumour compared to the LS174T tumour, as seen in Table 1. LS174T is four times more edge-dense than SW1222. Furthermore we find that LS174T nodes tend to have a higher average connectivity.

Our results for structural differences come as a surprise since D'esposito et al.'s work on similar data sets found that when simulating drug distribution across LS174T tumours vs SW1222, the LS174T tumours had on average lower connectivity. [6] One possible reason for this discrepancy is the difference in the graph prepossessing,

since based on our assumptions we eliminated over 15% of the edges in both data sets due to zero flow regions and self edges.

We also consider the question of influence maximization. This is driven by the need for a better understanding of efficient drug delivery in tumours, which is a big challenge due to the chaotic and disorganized network morphology. We find that 2% of the source nodes influence over half the nodes for both networks. However, we understand the limitation of this approach since we are only considering the structure of the network and not the specific edge properties. The diluting effect of branching, wherein only a trace of the administered drug at an inlet would appear at every node, begs the question whether that quantity is at all sufficient.

Network robustness analysis represents an interesting way to consider how to target nodes for maximal damage. We considered the impact of different ways to target nodes, and found that removing nodes based on node betweenness was the most effective method. We also note that the LS174T tumour is more robust and therefore less susceptible to both random and targeted attacks than the SW1222 network. This is consistent with the work of El et al., where they measured the impact of administering higher doses of the same drug on impeding the growth of LS174T and SW1222 samples [10]. They found that the LS174T sample required three times as much radiotherapy as the SW1222 one, to achieve the same effect. The greater network robustness might be a contributing factor to this.

In conclusion, we present a novel analysis of this dataset. As far as we are aware, this is the first time that oncological vessel networks has been subjected to a rigorous graph analysis. By considering graph properties alone, we are able to correlate some of our results to published work. We believe that with greater advancement of blood vessel imaging, techniques presented here could have a direct impact on discovering novel therapies and assisting existing ones, specifically radiotherapy planning.

9 Acknowledgements

We would like to thank the Walker-Samuel Group for producing these datasets and making them accessible. We would also like to thank the CS224W staff for all the help and guidance through this project.

The code for this project can be found at: `https://github.com/magdyksaleh/cs224w_ssm`

References

[1] Rakesh K Jain. Delivery of molecular and cellular medicine to solid tumors, 12 2012.

[2] Janice A Nagy, Sung H. Chang, Shou Ching Shih, Ann M. Dvorak, and Harold F Dvorak. Heterogeneity of the tumor vasculature, 2010.

[3] Stewart A Skinner, Peter J M Tutton, and Paul E O'Brien. Microvascular Architecture of Experimental Colon Tumors in the Rat. *Cancer Research*, 50(8):2411 LP – 2417, 4 1990.

[4] Joanne R. Less, Thomas C. Skalak, Eva M. Sevick, and Rakesh K. Jain. Microvascular architecture in a mammary carcinoma: Branching patterns and vessel dimensions. *Cancer Res.*, 51(265):265–273, 1991.

[5] A R Pries, Michael Hopfner, Ferdinand le Noble, M. W. Dewhirst, and Timothy W. Secomb. The shunt problem: control of functional shunting in normal and tumour vasculature. *Nat Rev Cancer*, 10(8):587–593, 2010.

[6] Angela d'Esposito, Paul W. Sweeney, Morium Ali, Magdy Saleh, Rajiv Ramasawmy, Thomas A. Roberts, Giulia Agliardi, Adrien Desjardins, Mark F. Lythgoe, R. Barbara Pedley, Rebecca Shipley, and Simon Walker-Samuel. Computational fluid dynamics with imaging of cleared tissue and of in vivo perfusion predicts drug uptake and treatment responses in tumours. *Nature Biomedical Engineering*, 2(10):773–787, 10 2018.

[7] Angela D'Esposito, Daniil Nikitichev, Adrien Desjardins, Simon Walker-Samuel, and Mark F Lythgoe. Quantification of light attenuation in optically cleared mouse brains. *Journal of biomedical optics*, 20(8):80503, 8 2015.

[8] Sebastian Wernicke and Florian Rasche. Fanmod: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006.

[9] Vtraag. Leidenalg, 2016.

[10] Ethaar El, Uzma Qureshi, Jason L J Dearling, Geoffrey M Boxer, Innes Clatworthy, Amos A Folarin, Mathew P Robson, Sylvia Nagl, Moritz A Konerding, and R Barbara Pedley. Predicting Response to Radioimmunotherapy from the Tumor Microenvironment of Colorectal Carcinomas. *Cancer Res*, 67(24):11896–905, 2007.