Exploring Network Differences between Different Tumour Vessel Networks: Motif Analysis and More

Magdy Saleh
mksaleh@stanford.edu

Surag Nair
surag@stanford.edu

Abhijeet Shenoi
ashenoi@stanford.edu

1 Introduction

Small tumours survive by drawing on oxygen and nutrients supplied from the nearest existing vasculature. Once a tumours 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, the tumour starts an angiogenic process to build its own vasculature to sustain its supply of oxygen and nutrients. 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. Jain [2012]

Recent advancements in imaging technologies have allowed for the collection of the datasets that model the tumour vascualure as graphs. These techniques achieve a high degree of resolution, allowing for capturing of vessels with diameter of less than 1 μm. 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. We present the analysis of two distinct tumour types, and a small healthy vascular network.

2 Related Work

There has been some work done in analyzing tumour vessels from a network perspective. Skinner et al. [1990] reported higher frequency of branching within tumour vasculature compared to healthy vasculature. These findings were expanded by Less et al. [1991] 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. Pries et al. [2010] 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.

This work is mainly a continuation of the findings of d'Esposito et al. [2018]. They identified two graph properties that contribute to drug distribution and propagation in these different networks. They state that functional connectivity, which is the average ratio of the size of the OUT-set over the size of the IN-set per node, and redundancy, which is the average number of viable pathways going through a node, 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 cause of the different behaviours seen across the different tumour types.

3 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. The nodes correspond to points of splitting of vessels, and the edges represent the vessels themselves. These two tumour models have markedly different phenotypes as described in D'Esposito et al. [2015]: 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. 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 our data sets

Graph	Nodes	Edges	Clustering Coef.	Avg. Node Deg.
Mesentary	388	546	0.03780	2.8093
LS174T	18,093	26,118	0.02938	3.2090
SW1222	72,084	530,022	0.03110	3.4448

4 Data Preprocessing

The data we received contained some clear artifacts that we needed to correct for. Firstly, there were multiple self edges, which are clearly erroneous as there can't be flow in a zero pressure gradient loop. Additionally, there were multiple edges with non-zero flow but zero pressure drop, for which the flow had to be corrected as well and was set to 0. Lastly, we are dealing with directed graphs, but the ordering of the nodes in the edge lists we produced was not following pressure gradients to 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. This was confirmed by the Walker-Samuel Lab as existing artifacts which we need to account for.

5 Methods

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

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

5.2 Degree Distribution

We look at the distribution of degrees of the nodes in the network to better understand the structure of the network. Vasculature networks are very different from social and web networks, as there are very few or no nodes with extremely high degree. This is expected since vessels most commonly bifurcate and very rarely split into three or more vessels. A comparison of degree distributions across the two tumours and compared to a healthy mesentary network would shed light on aspects of oncological physiology.

5.3 Motif Discovery

We explore the motifs present in the graphs using the Exact Subgraph Enumeration (ESU) algorithm Wernicke and Rasche [2006]. 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 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.

5.4 Community Detection

We performed community detection using the Louvain community detection algorithm from Blondel et al. [2008], while maximizing the modularity. The modularity, Q, is given by

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

We used an open source implementation of the algorithm which greedily maximizes the modularity, in an iterative fashion. The algorithm consists of two phases, repeated until convergence. The first stage consists of assigning every node to its own community. Subsequently, every merging of two nodes into a single community is considered, and the one with the highest increase in modularity is selected. Once no more nodes are to be merged, phase 1 terminates. Phase 2 then combines all nodes of the same community into one super node, by aggregating edges. The resulting graph is then subjected to the phase 1 and the process is repeated.
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 Preliminary Results

6.1 Degree Distribution

We first consider the degree distributions within each of the graphs and observe, as seen in Figure 1, that there is a high concentration of small degree nodes as expected, with a clear peak at 3, which is consistent with the relevant physiology, since bifurcations are the most prevalent branching pattern. Since the data only counts nodes as branching points, we therefore see very few nodes of degree 2.

6.2 Motif Analysis

We analyzed the occurrence of motifs with 3 nodes in the different graphs as seen in Figure 2. What we find, unsurprisingly is that by far the most common motifs are loop free, specifically the motifs where either two veins join into a larger vein or there is an arterial bifurcation. 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 is caused due to a combination of two factors:

- 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.

- The random rewirings used to generate a null model result in a large increase in those motifs which do not have cycles, and almost completely eliminate the presence of motifs with cycles.

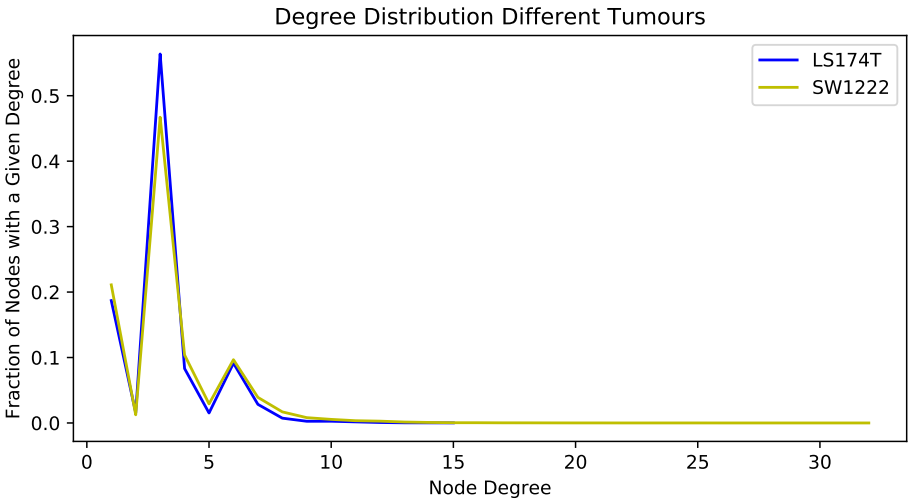

Figure 1: Comparing the degree distribution between LS174T and SW1222 tumours. We see that there is a greater degree diversity in the SW1222 tumours, which is expected as they tend to have greater vascular density

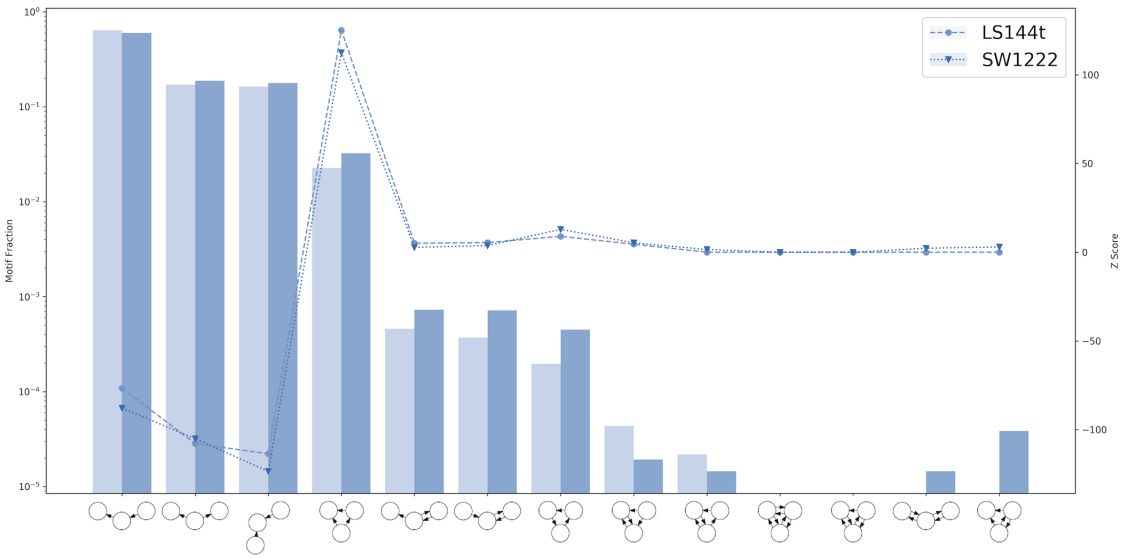

Figure 2: Motif distribution between LS174T and SW1222 tumours. We calculate motif fraction by dividing the count of a specific motif by the count of all motifs in the graph. We find that mainly acyclic motifs are prevalent.

We carried out an analysis of both the normalised motif counts, normalised 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. We see that there are low counts for motifs containing cycles, and high counts for motifs with a bifurcation structure. The null models in both cases have high counts for motifs without cycles, again from the argument above, and usually no counts for motifs with cycles. We also find that the two networks have very similar structures, when considering motif distributions for three node motifs.

6.3 Community Detection

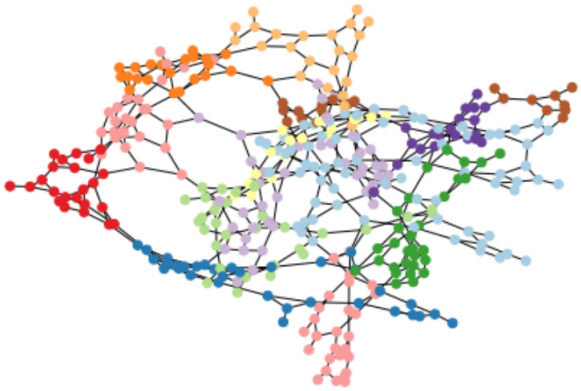

Figure 3: Mesentery Network Clustering

Finally, we used the Louvain algorithm, based on the implementation by Vtraag [2016] to perform community detection on our networks. This has proven to be effective when scaling to our larger networks. We visualize the mesentery network in Figure 3. We hypothesize that the communities represent nodes that are in close physical proximity representing capillary beds. We will confirm this by analyzing the spatial coordinates from the data set.

7 Future Work

Up to this point we have focused on building a sturdy data analysis pipeline for our graphs, which allows us to now easily explore different properties we are interested in. We have looked at basic properties of the graphs but we will now turn our attention to more interesting questions. We want to find out more about what hinders effective drug perfusion in these networks from a structural perspective. To do so we consider the following tasks:

- Visualize the larger networks based on actual spatial coordinates
- Shunt detection via in-set vs out-set comparison of the nodes in the graphs- analogous to the bowtie structure of the internet
- We also have data on the radius of each vessel in the network and plan to plot the average clustering coefficient as a function of radius to understand whether the clustering coefficient is driven mostly by capillaries (which have small radii)
- Based on our preliminary results, we will continue to search for clearer structural differences that are more nuanced. Once such metric we will consider is graph efficiency, which is a measure of how efficiently information (in our case mass) can flow through the network
- We will attempt to extract motifs with 4 and more nodes. Instead of counting all possible motifs, which is computationally expensive for a directed graph, we may reduce our search to a smaller set of predetermined nodes. Alternately, we may perform the analysis for the undirected version of the graph.
- We will extend the community detection analysis to explore whether the communities are located on the periphery of the network. This will help us understand if the center of the tumour is oxygen starved - a phenomenon that aggravates tumour growth.

References

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008 (10):P10008, 2008.

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. ISSN 1560-2281. doi: 10.1117/1.JBO.20.8.080503. URL http://www.ncbi.nlm.nih.gov/pubmed/26277988http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4568291.

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. ISSN 2157-846X. doi: 10.1038/s41551-018-0306-y. URL http://www.nature.com/articles/s41551-018-0306-y.

Rakesh K Jain. Delivery of molecular and cellular medicine to solid tumors, 12 2012. ISSN 0169409X. URL http://www.ncbi.nlm.nih.gov/pubmed/24511174http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3914635.

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. ISSN 0008-5472.

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. ISSN 1474-175X. doi: 10.1038/nrc2895. URL http://www.ncbi.nlm.nih.gov/pubmed/20631803.

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. URL http://cancerres.aacrjournals.org/content/50/8/2411.abstract.

Vtraag. Leidenalg, 2016. URL https://github.com/vtraag/leidenalg.

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