UNIVERSITY OF CALIFORNIA
Los Angeles

Using Graph Theory to Compute Sets of Cycles in Vascular Networks

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Molecular Biology

by

Adam Gomez

2019

ABSTRACT OF THE THESIS

Using Graph Theory to Compute Sets of Cycles
in Vascular Networks

by

Adam Gomez
Master of Science in Molecular Biology
University of California, Los Angeles, 2019
Professor Luisa M. Iruela-Arispe, Co-Chair
Professor Van Maurice Savage, Co-Chair

We use graph theory in conjunction with automated vessel data extraction software to identify and quantify looping structures in biological resource distribution networks. As a practical biomedical application, characterizations of looping structures may serve to non-invasively distinguish a pathological resource distribution network from a healthy one. A network with loops is structurally different from a network without loops and may result in a refined scaling exponent for metabolic rate. A refined scaling exponent would also have implications for aging, lifespan, and evolutionary development. Here we focus on developing mathematical tools to find looping structures in biological vascular networks. Looping structures in biological resource distribution networks can be extracted by using graph theory to quantify the cycle basis of the graph of the network. Algebraic ring sums are then used to quantify the total number of loops in the graph.

The thesis of Adam Gomez is approved.

Peter John Bradley

Jeffrey Aaron Long

Van Maurice Savage, Committee Co-Chair

Luisa M. Iruela-Arispe, Committee Co-Chair

University of California, Los Angeles

2019

In loving memory of my grandpa Lee Ahrens and all sufferers of congestive heart failure from arteriovenous fistula.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1144087.

The author is grateful for the guidance and contribution provided by fellow NSF GRFP recipient and colleague Andrés Ortiz-Muñoz.

CHAPTER 1

Introduction

In 1932 Klieber demonstrated an empirical relationship between body size and metabolic rate [5]. It wasnt until 1997 when West, Brown and Enquist (WBE) established a generalized model to account for the observed relationship between metabolism and mass [8]. Cardiovascular systems deliver essential resources to cells where the resources undergo chemical reactions to produce energy for cellular growth, maintenance, and division. The structure of these vascular networks directly regulates how resources acquired from the external environment are distributed to cells for metabolic processes. The link between body mass, metabolic rate, and vascular networks is formulated mathematically using branching structures along with measurements of vessel length and radii. WBE formally relates body mass (M) and metabolic rate (B) with a normalization constant (B0) and a scaling exponent (alpha) [8].

$$B = B_0 M^{\alpha} \tag{1.1}$$

As the mass of an organism scales linearly, if that organism's metabolic rate scales to the 3/4 power, it is following Kleiber's Law [5]. We refer to this property as allometric scaling, where an allometry is simply how characteristics of an organism change with its body size. The cardiovascular system is a resource distribution network that consists of branching vessels that perpetually diverge to form a measurable tree-like structure. Here I use a software tool called Angicart to make automated measurements of vascular networks. Angicart analyzes 3D radiographic images of blood vessels to reconstruct vascular networks [6]. So what happens when branching vessels converge with other branching vessels to generate loops or cycles in an otherwise divergent branching structure?

This work ties together vascular looping structures, organismal mass, and metabolism by considering resource distribution networks in the context of allometric scaling theory. The

study of biological systems requires new and improved methods for quantifying biological data. Resource distribution networks themselves, that is the structure and function of the network, determine how metabolic rate scales with organismal mass, so it is important to understand structure, function, behavior, and interactions of vessels throughout the network, namely loops. [7].

Building upon the ever-improving WBE model, can we use graph theory in conjunction with automated vessel data extraction software to accurately and efficiently identify, characterize, and quantify looping structures in biological resource distribution networks? As a practical biomedical application, characterizations of looping structures may serve to non-invasively distinguish a pathological resource distribution network from a healthy one. A network with loops is structurally different from a network without loops and may, therefore, result in a refined scaling exponent for metabolic rate depending on the loopiness of the network of the organism.

Loops may not only distinguish vascular type and health, but may also have implications for aging, lifespan, metabolic rate, and evolutionary development. Here we focus on developing the mathematical tools to find looping structures in biological vascular networks. I hypothesize that looping structures in biological resource distribution networks can be extracted by using graph theory to quantify the cycle basis of the graph of the network.

1.1 Why Loops?

One approach to establishing a metric that can be used to characterize vascular networks and predict other properties related to the network is to identify loci where branching structures of vessels that would otherwise never converge instead unite to form a looping structure. Identifying loops in vascular networks extracted from medical imaging modalities like MRI will allow us to contribute to modifications of WBE and distribution networking theory in general.

WBE metabolic scaling theory has been developed over time to become ever more respectful of biological systems. Assumptions and constraints are modified to optimize our

ability to accurately and consistently model the vascular system. We progressively modify the theory and assumptions of the model, not only to better understand human vascular resource-distribution networks, but also to learn about fundamentals of biological dynamics that affect disease, metabolism, lifespan, and evolutionary development in general. Im following in the footsteps of a long line of researchers that have successively contributed to WBE theory, making it more accurate. For example, finite size adjustments have been made to the theory [7], along with asymmetric scaling adjustments [1] by colleagues in my lab. One important factor of our theoretical understanding of the vascular network is that it is a branching structure. Over an averaged network, we assume that the architecture of a vascular network reflects a tree-like structure as it perpetually branches from the heart to the capillaries where oxygen exchange occurs and vessels are restricted in diameter by the diameter of red blood cells. There are few heterogeneous structures, like loops, which are averaged out over the primarily branching network. Naturally occurring anastomoses in vascular systems, by definition, connect two openings that would otherwise never have converged. This introduces changes in the dynamics and flow of resources that are being distributed through the network.

Although non-branching morphologies such as loops commonly occur in vascular systems around stroke sites [9] and tumors[4], they are rare in non-pathogenic human cardiovascular systems. Loops, therefore, can indicate pathologies, as in the case of arteriovenous fistulae, [3] where an artery becomes connected to a vein and can contribute to congestive heart failure. There are very few normal vascular loops that occur consistently across individual mammals. In the case of humans, isolated examples such as the Circle of Willis and the foramen ovale in the heart of a fetus can serve as important looping controls for healthy loops in the quest for identifying loops [3]. The work in this thesis focuses on developing mathematical and computational tools for identifying loops so that biomedical and basic science researchers can use the tools to investigate looping structures.

At first glance, identifying loops in a network might seem trivial to some, but it is deceptively complicated. Before loop analysis tools can be used for diagnostic purposes, it must be demonstrated that we are able to identify whether loops exist, how many loops

exist, where the loops exist, and what type of implications these looping structures have. For example, if a vascular structure that is normally divergent instead converges at points to form loops, distribution dynamics and blood flow are expected to differ. We have to be diligent to ensure that the loops we find are real, and that we do not overlook any multiplicative permutations of looping structures that might permeate throughout the network from one single anastomosis.

1.2 Metabolic Scaling Theory

We depend on the validity of WBE allometric scaling theory to inform our investigation into vascular resource distribution networks. The theory can be used to solve for unknowns provided there are some known variables. For example, we can determine metabolic rate given mass, scaling exponent, and a normalized constant. As we develop methods for quantifying looping structures, tools become available to improve scaling exponent estimations, and thus improve solutions for metabolic rate given mass. We can also modify the formulation to solve for many variables, like vessel volume, and flow dynamics of resources through the vascular network. For the world at large, this is important for medical and basic research endeavors aimed at measuring, understanding, predicting, diagnosing and designing biological systems with implications for metabolism, development, aging, and more.

The theory relies on eight assumptions [7], and primarily functions in the algebraic realm of mathematics, but requires knowledge of linear algebra and differential equations. To clarify the relationships between vascular networks, resource distribution, flow, scaling exponents, and loops within the network, we must defer to the assumptions of WBE. WBE assumes an averaged network, where network parameters are averages of the variation of the parameters in the biological networks [7]. Averaged heterogeneous parameters result in an estimated homogeneity that WBE assumes are homogenous throughout the resource distribution network. This has proven effective for analyzing and predicting features of networks, but is also an opportunity to modify the theory to more faithfully resemble biological systems.

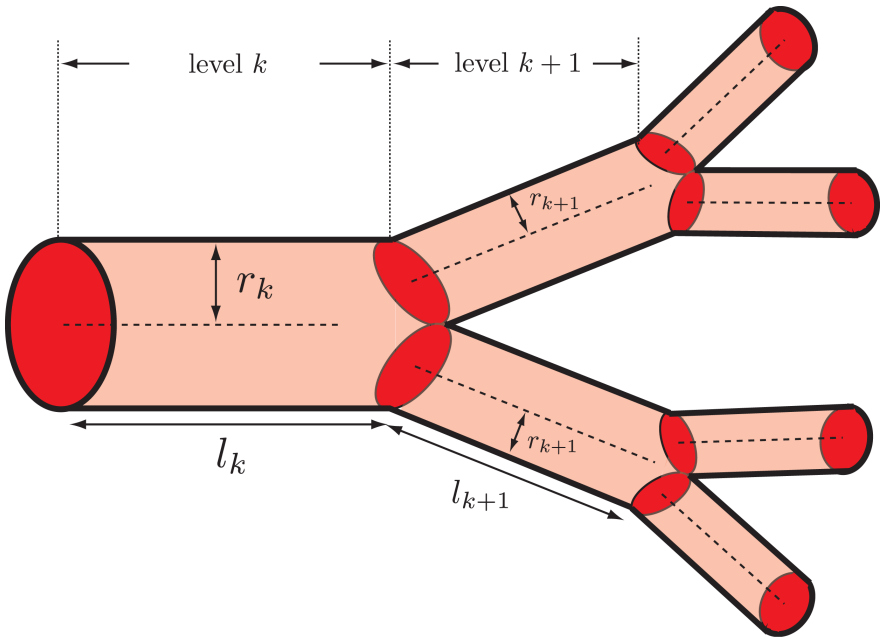

Figure 1.1: WBE network level labeling scheme[7].

WBE Assumption 1.) The distribution network determines the scaling relationship. This assumption is necessary for connecting looping structures of networks to metabolism. The structure of vascular resource distribution networks determines the scaling relationship between the mass of the organism and the metabolism of the organism. By establishing a method to quantify looping structures in naturally occurring biological systems, we can use loops to characterize and distinguish networks. This will be an important tool for researchers to use to modify the WBE model and improve predictions about fluid dynamics.

Assumption 2.) The distribution network is hierarchical. This assumption is necessary for understanding the labeling scheme that describes the branching structure of vascular networks that invokes parent-child relationships, where children are vessels that have divergently branched out from a parent vessel. Later we will see an analogue between hierarchical level as described here and length of the path of a graph as described by graph theory. The hierarchical levels here are defined as branch points where child/daughter vessels split off from parent vessels [Figure 1.1].

One example of how this assumption is not totally true lies in the differing number of levels from the heart (level 0) to the capillaries in the foot (level N) versus the number of

levels from the heart (level 0) to the capillaries in the coronary artery (N) [7]. A loop is formed if two branches that would otherwise continue to diverge at each level, as edges in a tree-like network do, instead converge at some level.

Assumption 3.) Vessels within the same level of the hierarchy are equivalent. This assumption is important for intuiting that the distance traveled by resources being transported through a network will be impacted by introducing a loop to an otherwise branching structure. Vessel length and, therefore, flow rate at a branch point would differ locally compared to a loop-free branch point at the same level.

Assumption 4.) The branching ratio is constant. This assumption is important for intuiting that branching is a process by which one vessel diverges into two vessels, while a loop results from two vessels converging into one. It has been shown that finite size corrections do affect the 3/4 scaling relationship in this case [7], an indication that investigating looping structures for affects on the scaling relationship may be in order.

Assumption 5.) The network is space filling. This assumption is important for understanding that loops in a network could have implications for evolutionary strategies to form space filling structures.

Assumption 6.) The energy loss of fluid flow through the network is minimized. This assumption is important for understanding that loops, again structures where two vessels converge into one vessel, might change the energy dissipated by both pulse reflection and viscous forces.

Assumption 7.) Capillary characteristics are the same across species. Unless we find that looping structures are plentiful among capillaries, this assumption stands. The tiny scale of capillaries would require a very large number of images to capture any ratio of loops among the terminal points.

Assumption 8.) Capillaries are the only exchange surfaces and thus directly relate blood flow rate to oxygen supply in tissues. Red blood cell diameter reflects capillary diameter, so more blood yields more oxygen and thus an increased metabolic rate. Changes in blood flow rate by loops may affect oxygen supply to tissues.

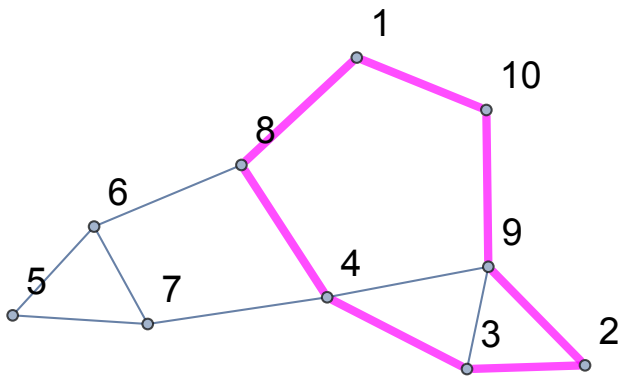

Figure 1.2: An example of a cycle in magenta.

1.3 Loops in a vascular network

Terminology used in graph theory is not standardized or consistent [2], so here I establish our uses of terminology. In our attempt to find loops among a tree-like resource distribution network, we defer to the language of graph theory. Most specifically, when we say we are looking for loops in the network, we mean that we are looking for cycles of the graph of the network. A cycle, by definition, is a non-trivial closed path [2]. That is to say, it consists of more than a single node, the edge/node steps terminate at the node of origin, there are no repeated vertices and there are no repeated edges [Figure 1.2].

We will find cycles by identifying acyclic subgraphs called spanning trees that generate a set of cycles called basis sets from which one can calculate the total number of cycles by implementing a ring sum.

In assumption 1 of WBE, we see that the network determines the scaling relationship between mass and metabolism. A network is a graph, which consist of 2 sets. One set consists of the vertices (V), or nodes, of the graph while the second set consists of the edges (E) of the graph. Vertices that are joined together by edges are called endpoints and every edge has either one or two endpoints. Of the three types of graphs to consider [2] we focus on simple graphs.

Simple graphs do not have self loops and do not have multiple edges [2]. We do not consider loopless graphs or general graphs here because loopless graphs contain multiple

edges and general graphs may contain self loops and/or multiple edges. We choose to focus on simple graphs for model simplicity, but iterations of this work that consider self loops and multiple edges are important future directions to explore. Most important to remember here is that we are using simple graphs to represent our vascular networks.

The WBE labeling scheme that describes the branching structure of vascular networks that invoke parent-child relationships is described in assumption 2. In graph theory, a walk from one vertex, V0 to another vertex, Vn, is defined as an alternating sequence between vertices and edges starting at V0 and ending at Vn. Because we are focusing on simple graphs, we can define a walk in terms of a sequence of vertices alone [2]. The number of steps in a walk sequence defines the length of the walk, so if a graph consists of one vertex and no edges, the length of the walk is zero and the graph is called a trivial walk. A walk length represented by the number 5 indicates 5 vertices from the vertex of origin. In the WBE model, V0 or level 0 represents the origin of the vascular network, or the heart. The level of the network, as defined by its number of branching points, or vertices, from V0, then reflects the length of the walk in graph theoretical terms. A path is a specific type of walk that does not contain repeated edges nor repeated vertices. As defined above, a cycle is a non-trivial, closed path. This definition constrains our considerations in terms of walk-type. Again, future work considering other types of walks, and therefore other types of looping structures beyond cycles, will further enlighten our work. Most important to remember here is that we are using paths and path length to map to WBE vessel levels starting at the heart in our vascular networks.

My work aims to find looping structures in vascular networks extracted from imaging modalities, but what do we mean by loops? A loop can be the shape you make with a shoelace where shoelaces overlap but do not become structurally integrated. A donut also forms a loop of sorts, but it is self-contained. We are not looking for either of those types of loops. We are looking for loops within in a branching network where two otherwise diverging branches instead converge to form a connected edge between the branches, integrating the flow of resources through the network. One approach to finding loops in a network is to find the proximity of vessel edges to one another by using the x, y, and z coordinates of image voxels in

a 3D network reproduced by angicart. Previous work done in our lab shows that background noise presents a challenge to verifying the proximity and/or integration of vessels to one another using this technique. It is difficult to discern whether we have identified the loops we are interested in or shoelace-type loops. Additionally, manual parameter adjustments are necessary for every network analyzed to find threshold values that combat noisiness.

CHAPTER 2

Main results

2.1 Method for calculating cycles in a graph

To computationally extract loops from a 3D network requires a formal definition of loops. In graph theory, cycles capture the looping structures we are interested in by definition. Finding cycles in networks is not novel in graph theory, but applying the powerful algebraic tool of ring sums to formally solve our biological problem is novel.

Because the graphs of vascular networks that we extract from 3D radiographic imaging only capture an inset-like sample from the entire vascular network, we extract a set of subgraphs that appear to be disconnected from one another. The subgraphs do all converge back to one parent node of origin at the aorta, but our imaging data only captures the collection of neighboring subgraphs within the limit of the boundaries of the image. This set of disconnected trees is called a forest. More accurately, it is a set of trees that touch every node of the subgraph it defines.

Our clever trick for finding cycles in a network begins with first identifying a connected graph that has no cycles, called a tree. Trees allow us to understand graphs structurally based on their varying equivalent characterizations, which enable a range of applications. For example our structural understanding of isomeric structures in physical chemistry, rooted trees in operating-system directories, and binary-search trees in information retrieval, results from employing varying characteristics of trees [2]. In the same way that we have first specified graph type and walk type, we must also specify what tree-characterizations we apply to understand looping structures in vascular resource distribution networks. Our structural understanding of a graph changes as we select varying characterizations of trees

within the graph. Of six of the most useful tree characterizations, the following statement is most relevant to deciphering our question about the number of cycles in a network: A tree (T) that has n vertices, does not contain cycles, but if any new edge (e) is added to the acyclic tree, exactly one cycle is formed. T + e has 1 cycle [2].

This particular characterization makes it possible to generate the fundamental set of cycles of the graph. One might wonder why we are so invested in finding a spanning forest that is comprised of acyclic spanning trees if the goal is to find cycles of the graph. Although it may seem counter intuitive, a spanning forest (F) is associated with something called a fundamental cycle, which is a cycle formed by adding a non-tree edge to a spanning tree of the graph. By adding any edge (e) from the graph G that is not included in the full spanning forest F, a fundamental cycle will be generated. The relative complement of a spanning forest F is the difference between the graph G and the spanning forest F such that G F generates the relative complement of F [2]. The significance of selecting an edge from the relative complement simply lies in the need to select an edge from the graph that is not already a part of the spanning forest. If we were to add an edge from the spanning forest to the spanning forest, we would not change the spanning forest. If we add an edge to the spanning forest that is otherwise not a part of the spanning forest, it must come from the part of the graph that excludes the spanning forest, hence, it must come from the difference between the graph G and the spanning forest F. Every non-tree edge that we add to the tree produces a fundamental cycle, of which there may be many. The set of all of the fundamental cycles is called the fundamental system of cycles. It is not enough to find a fundamental cycle of a graph, because we aim to find all of the cycles of a graph. We, therefore, seek to find the fundamental system of cycles of the graph because this set contains all of the fundamental cycles in the graph associated with the spanning forest. We find the fundamental system of cycles by subtracting F from G and adding each edge, one at a time, from the relative complement to F, to generate one cycle per edge added. Each edge added equals one cycle [Figure 2.1][Figure 2.2].

Fundamental cycles are unique because non-tree edges are part of a path in a simple graph [2]. It is still not enough to know the set of fundamental cycles of a graph to answer

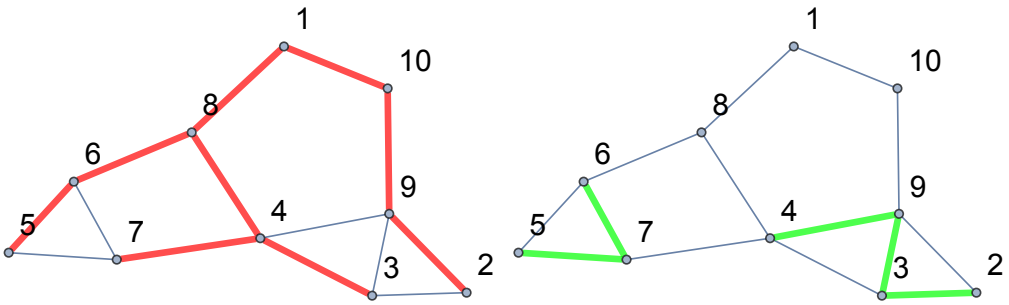

Figure 2.1: Spanning tree in red (left) and non-tree edges in green (right). When added to the spanning tree, each non-tree edge will generate a cycle.

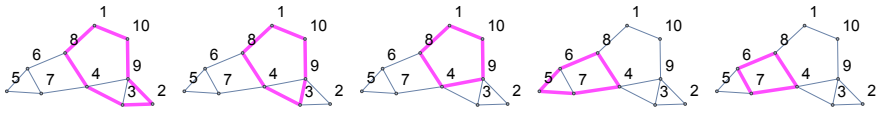

Figure 2.2: The set of cycles generated by each non-tree edge constitutes a fundamental system of cycles (in magenta).

our question. The set of all fundamental cycles only tells us about the cycles specifically associated with one spanning forest. From this point, we must consider the ring sum of the graph.[Figure 2.3]

In general there is more than 1 distinct spanning tree per network and every distinct spanning tree is going to give rise to a distinct cycle basis. A cycle basis set is not unique because there may be other spanning trees and, thus, other cycle basis sets. All of the possible cycles of a network, however, can be found using the cycle basis set obtained from any spanning tree of the network, where a cycle basis set is the set of all of the distinct and unique cycles of a spanning tree. In linear algebra, a basis is a set of vectors that are linearly independent and that can be used to generate any element of the vector space, so it is said to span the vector space [2][see appendix]. The ring sum operation is what allows us to generate any element of the vector space from the fundamental system of cycles. A ring sum operation is the sum of elements of a subset of edges of a graph, where a subset of edges can be trees, cycles, or disjoint cycles [2]. We consider the subsets of edges that generate the

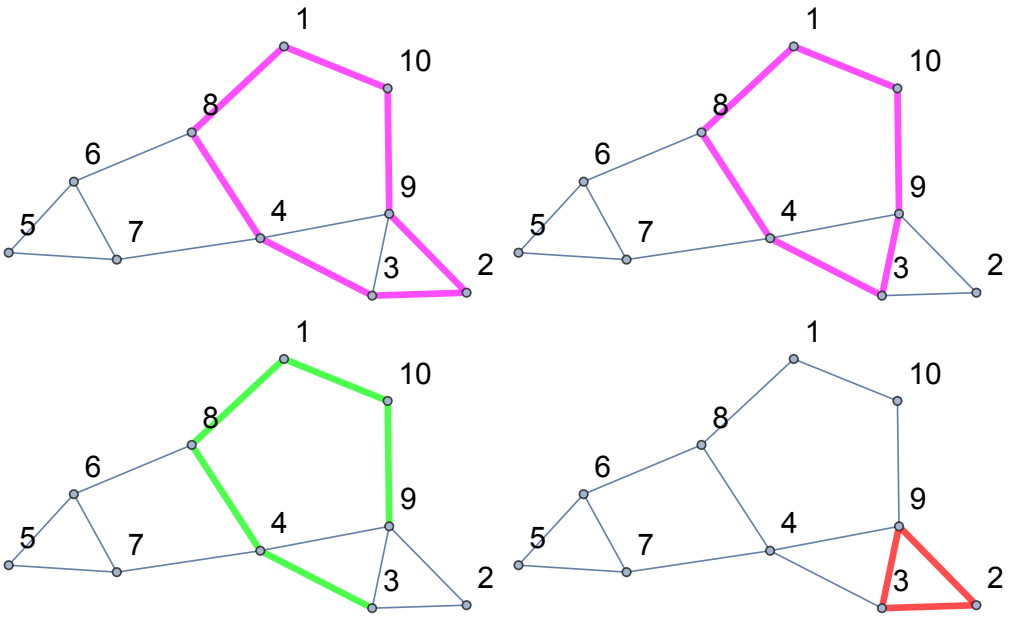

Figure 2.3: Two cycles in magenta(top), shared edges of cycle overlap in green (bottom left), and resulting ring sum of the two cycles in red (bottom right).

fundamental set of cycles.

Once we identified the characteristics of graphs, trees, and fundamental cycles most appropriate for identifying cycles in resource distribution networks, we defined the vector spaces associated with the graphs. The set of all subsets of edges of a graph is also a vector space over GF2 [see appendix], and is called the edge space of the graph. A cycle space is a subspace of the vector space associated with the graph and contains a null set, all cycles, and all unions of disjoint cycles [2]. Again, we are not considering disjoint cycles. For the ring sum, we consider the cycle space. We let T be a spanning tree of a connected graph G. Then the fundamental system of cycles associated with T is a basis for the cycle space We(G). A set B is a basis for V if the vectors in B span V and are linearly independent [see appendix]. A set of vectors spans V if every vector in V is expressible as a linear combination of vectors in S. Vectors are linearly independent if none of them is expressible as a linear combination of the remaining ones[2].

We extracted parent-child relationships from MRI data to construct a graph of the vascular network. We used Mathematica to find the spanning trees of the graph from which

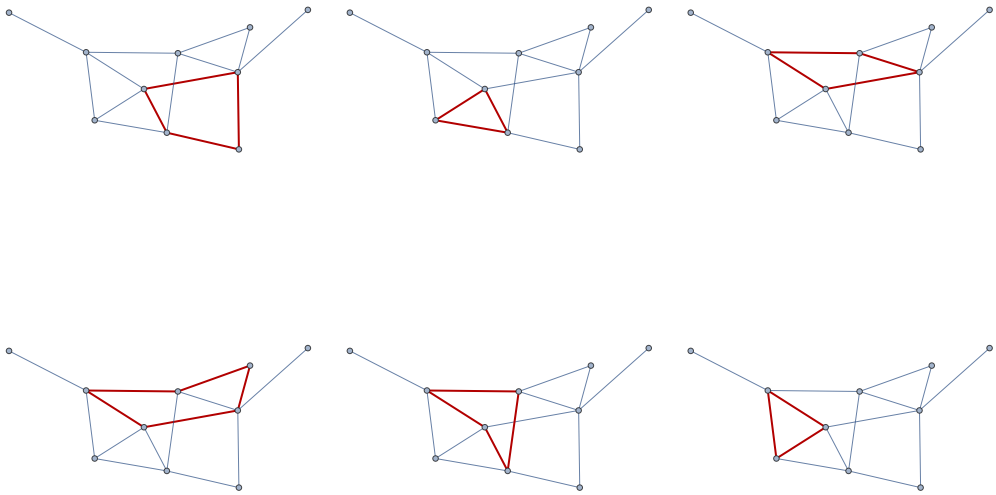

Figure 2.4: Random network with each cycle of a fundamental set of cycles in red.

we then calculated the cycle basis. To ensure that the cycle extraction method works, we generated random graphs in Mathematica, and extracted the fundamental systems of cycles.[Figure 2.4]

Our lab did not have access to datasets with biologically well known anastomoses, or loops, to use as controls. As our lab group waited to get access to additional data sets that could prove valuable as controls, such as data sets that include well known human vascular loops like the Circle of Willis, we processed data sets available to our lab. We looked for cycles in the vascular networks of stroke-induced mouse models including 15 microsphere injected mouse brains, 42 photothrombosis-model mouse brains, and 9 middle-cerebral-artery-occlusion-model mouse brains. I also looked for cycles in 1 human leg data set, and 1 zebrafish data set. Although loops are likely rare, I anticipated finding loops in the stroke model data sets. I did not find any loops in the data sets.

After gaining access to the BraVa database, we processed 53 human brain datasets that contained the Circle of Willis. None of the networks contained loops. Upon further investigation, we found that the group that produced the data manually eliminated items

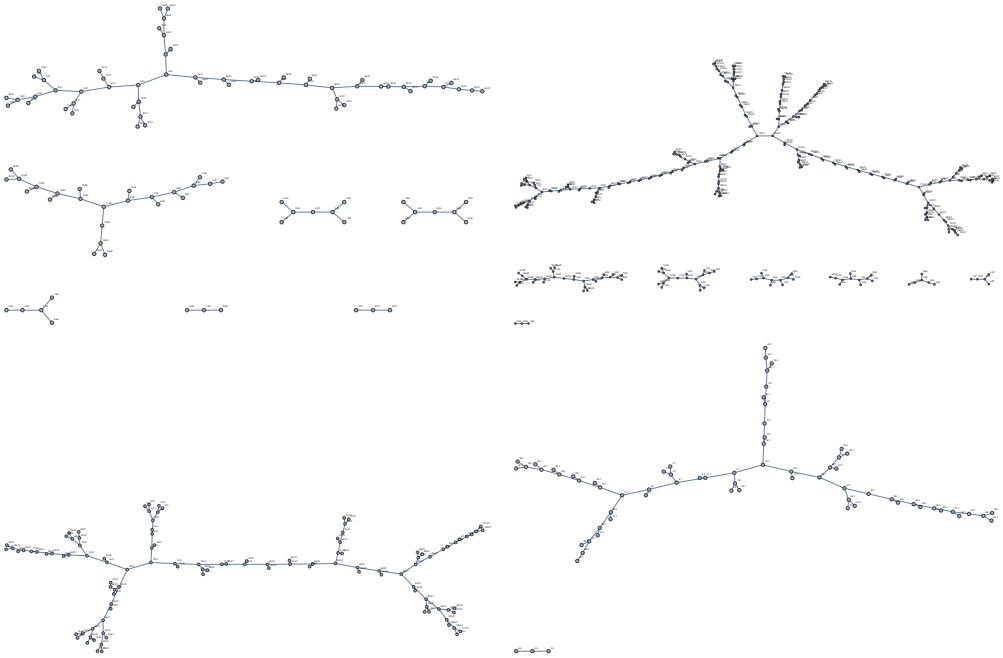

Figure 2.5: Representative examples of mouse brain stroke model data without loops. Photothrombosis (top) and middle cerebral artery occlusion (bottom).

considered noise from the data, including vasculature on the outside of the skull as well as the Circle of Willis. With the Circle of Willis eliminated, well established looping structures were excluded from the datasets, and were no longer valuable as controls.

We then aimed to verify that our mapping of parent-child relationships from angicart output data did not exclude looping structures. Angicart was originally written in Ocaml and was later translated to C++ in order to capture loops, but loops may not be captured in the parent-child relationships output by angicart. Loops could have been captured instead in vessel adjacency data also output by angicart. We used a very simple network image that included a loop by design, processed the image through angicart, and used the angicart output data to reconstruct a graph in Mathematica. The parent-child relationships from angicart output did not produce the correct graph with loops. We manually added the missing parent-child relationship to the Angicart output data, and were able to find the control loop that existed[Figure 2.6]. Similarly, we processed angicart vessel adjacency output data, and found that loops were identified, but were redundant.

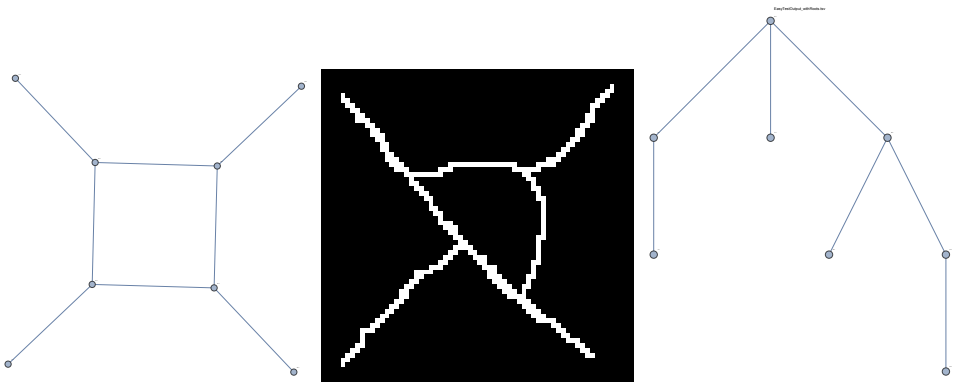

Figure 2.6: Image of network with loop by design (middle), reconstruction with corrected parent-child relationship containing loop (left), and reconstruction directly from angicart parent-child output data (right).

CHAPTER 3

Conclusion

Identifying and quantifying cycles in a network is a surprisingly difficult challenge that requires the application of discrete mathematics, most specifically, graph theory. I posit that it is difficult to find loops because cyclic structures materialize in a variety of forms that are easily overlooked without the consideration of cycles within cycles. Furthermore, data mapping from one computational source to another can be challenging. While the approach to finding cycles in a vascular network is sound theoretically, it requires further work for practical applications. Our mapping of the physical structure of a vascular network to a network graph relies on the assumption that the parent-child relationships in the angicart output data map directly to the graph theoretical paths taken along the graph of a network. This requires further investigation to validate. In lieu of parent-child relationships, vessel adjacency relationships from angicart output can also be used to produce graphs of networks from which cycles can be quantified. At this time, we experienced redundancy in vessel edges and cycles produced by the vessel adjacency output data. With further investigation, the source of redundancy can be identified and eliminated.

With tools to quantify loops in a vascular network, researchers can continue to modify the assumptions and constraints of metabolic scaling theory and generate predictions about biological resource distribution networks that more accurately represent empirical data. The technology and models we present here are scalable to investigate distribution networks from the micron level to an ecosystem level of organization, because we can use any imaging modality. Our work has shown that, given a complete description of parent-child relationships, we can find the fundamental system of cycles in a spanning tree of a vascular network and use the ring sum operation to calculate the total number of cycles. Angicart output data

does not provide the complete description of parent-child relationships necessary for direct mapping, but improved communication with angicart designers will reconcile any misunderstandings about output data. We hope that our work will result in better diagnostic and predictive techniques, both for biomedical applications and basic research [6].

3.1 Future work

Flow of resources through a vascular network is described by a Laplacian matrix and the equilibrium flow through the network is described by the matrix tree theorem. The matrix tree theorem gives the equilibrium solution in terms of the spanning trees of the network which is in turn related to the cycle basis of the network. The underlying approach to finding cycles in vascular networks can be implemented in a slightly different context to explore flow, and to possibly further modify WBE metabolic scaling theory.

CHAPTER 4

Appendix

4.1 Graph theory concepts

Definition 1. A *undirected graph* is a pair $\mathcal{G} = (N, E)$, where N is a set of *nodes*, and E is a set of *edges*, which are two-element subsets of N.

Definition 2. Let $\mathcal{G} = (N, E)$ be a graph and $\mathscr{C} = \{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$ be a sequence of edges. We say that $\mathscr{C}$ is a *cycle* if $y_i = x_{i+1}$ for $i = 1, \ldots, n-1$, and $y_n = x_1$. If no node or edge appears more than twice, with the exception of the first and last nodes, we say that the cycle is *simple*.

Definition 3. Let $\mathcal{G} = (N, E)$ be a graph and $\mathscr{F} \subseteq E$ be a subset of edges. We say that $\mathscr{F}$ is a *forest* if none of its subsets forms a cycle. If, in addition, every node of $\mathcal{G}$ is contained in some edge in $\mathscr{F}$ we say that $\mathscr{F}$ is a *spanning forest*.

4.2 Linear algebra concepts

Definition 4. $\mathcal{GF}(2) = (2, +, \cdot)$ is the two-element *field*, where $2 = \{0, 1\}$ is the *boolean set*, and $+$ and $\cdot$ are respectively the *addition* and *multiplication* operations defined in the following tables:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

For the general definition of a field, please see e.g [reference].

Definition 5. A *vector space* over a field $\mathcal{F}$ is a triple $\mathcal{V} = (V, +, \cdot)$, where V is a set of *vectors*, $+ : V \times V \to V$ is a binary operation called *vector addition*, and $\cdot : \mathcal{F} \times V \to V$ is a mapping called *scalar multiplication*, that satisfy the following properties for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars $a, b \in \mathcal{F}$:

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- There exists a vector $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$.
- There exists a vector $-\mathbf{u} \in V$ such that $(-\mathbf{u}) + \mathbf{u} = \mathbf{0}$
- $(a \cdot b) \cdot \mathbf{u} = a \cdot (b \cdot \mathbf{u})$
- $(a + b) \cdot \mathbf{u} = a \cdot \mathbf{u} + b \cdot \mathbf{u}$
- $a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + a \cdot \mathbf{v}$
- $1 \cdot \mathbf{u} = \mathbf{u}$

Definition 6. Let $\mathcal{V} = (V, +, \cdot)$ be a vector space over a field $\mathcal{F}$ and $W = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V$ be a set of vectors. If:

$$a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \ldots + a_n \cdot \mathbf{v}_n = \mathbf{0}$$

implies that $a_1 = a_2 = \ldots = a_n = 0$ we say W is *linearly independent*. If for every vector $\mathbf{w} \in V$ we can find scalars $a_1, \ldots, a_n \in \mathcal{F}$ such that:

$$a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \ldots + a_n \cdot \mathbf{v}_n = \mathbf{w},$$

we say that W *spans* $\mathcal{V}$. If W is linearly independent and spans $\mathcal{V}$ we say that W is a *basis* for $\mathcal{V}$.

4.3 Vector spaces from graphs

Definition 7. Let $\mathcal{G} = (N, E)$ be a graph. We define the *edge space* of $\mathcal{G}$, denoted with $W_E(G) = \mathcal{GF}(2)^E$, as the vector space with dimension $|E|$ over the field $\mathcal{GF}(2)$

Notice that the elements of the edge space can be seen as the subsets of E. Furthermore, vector addition corresponds to the *symmetric difference* of sets, which we describe below

$$A \triangle B = (A - B) \cup (B - A), \quad A, B \subseteq E.$$

Definition 8. Let $\mathcal{G} = (N, E)$ be a graph and $W_E(\mathcal{G})$ its edge space. The *cycle space* of $\mathcal{G}$, denoted by $W_C(\mathcal{G})$, is the subspace of $W_E(\mathcal{G})$ spanned by the simple cycles of $\mathcal{G}$. A *cycle basis* is a basis for the space $W_C(\mathcal{G})$.

Notice that, in general, the cycle space will contain subgraphs that consist of disjoint cycles. Since we are interested only on those subgraphs that consist of a single simple cycle, our algorithm for generating the set of such cycles from the cycle basis will have to exclude subgraphs that consist of disjoint cycles.

REFERENCES

[1] Alexander Byers Brummer, Van M Savage, and Brian J Enquist. A general model for metabolic scaling in self-similar asymmetric networks. *PLoS computational biology*, 13(3):e1005394, 2017.

[2] Jonathan L Gross and Jay Yellen. *Graph theory and its applications*. Chapman and Hall/CRC, 2005.

[3] Barbara A Gylys and Mary Ellen Wedding. *Medical terminology systems: a body systems approach*. FA Davis, 2017.

[4] Alexander B Herman, Van M Savage, and Geoffrey B West. A quantitative theory of solid tumor growth, metabolic rate and vascularization. *PloS one*, 6(9):e22973, 2011.

[5] M Kleiber. Body size and metabolism. *ENE*, 1(9), 1932.

[6] Mitchell G Newberry, Daniel B Ennis, and Van M Savage. Testing foundations of biological scaling theory using automated measurements of vascular networks. *PLoS computational biology*, 11(8):e1004455, 2015.

[7] Van M Savage, Eric J Deeds, and Walter Fontana. Sizing up allometric scaling theory. *PLoS computational biology*, 4(9):e1000171, 2008.

[8] Geoffrey B West, James H Brown, and Brian J Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122–126, 1997.

[9] Susan N Wright, Peter Kochunov, Fernando Mut, Maurizio Bergamino, Kerry M Brown, John C Mazziotta, Arthur W Toga, Juan R Cebral, and Giorgio A Ascoli. Digital reconstruction and morphometric analysis of human brain arterial vasculature from magnetic resonance angiography. *Neuroimage*, 82:170–181, 2013.