Chaos in Cancer: Analyzing Network Morphology to Predict Tumor Angiogenesis using Chaos Theory

Caroline Huang

BASEF 2022

Abbey Park High School

Introduction

Background on Angiogenesis

Human blood vessels build complex networks throughout the body that span a remarkable estimated total length of 100,000 km that, while necessary for homeostasis, can be exploited by various diseases. Angiogenesis is when new blood forms from pre-existing vessels through progressive sprouting and splitting to provide tissues with oxygen and nutrients. The physiological role of angiogenesis is pertinent to growth, development, and healing, as well as fundamentally linked to cardiovascular diseases, diabetes mellitus, and notably cancers, which was the focus of this study.

In cancer, angiogenesis provides the tumor with a supportive microenvironment required for cancer survival, malignancy, and metastasis. The architecture of tumor vasculature is structurally and functionally anomalous and characterized by chaotic excessive branching and shunts (Warren, 1979). This heterogeneity and poor perfusion, in a therapeutic sense, reduces the efficiency of both chemotherapy and radiotherapy while leading to the development of resistance and increased tumor aggressiveness after resorption (Demené, 2014). Presently, there are two prevalent strategies leveraging angiogenesis to combat tumors: A) Normalizing the cancerous vasculature to create a better physiological condition of the tumor to increase the sensitivity to cytotoxic therapies (Goel et al., 2012), and B) inhibiting angiogenesis to starve tumors by severing their sustained dependencies of neovasculature for growth (Folkman, 1971). However, both of these contexts require the anticipation of tumor progressions and response assessment for treatment discovery and planning relevant to a patient's particular case. Due to the dynamic morphological characteristic of tumor vascular branching, the inadequacies in current imaging techniques needed for precise vascular delineation, and the complexity of the tumor microenvironment interplay, the prediction of tumor angiogenesis remains a challenge.

Analysis through Network Science

Computational anatomy has become an emerging field merging applications of computation, mathematics, and data analytics for simulating biological structures and extracting anatomical coordinate systems at the morphome scale. These principles applied to angiogenesis can prove extremely powerful; a simplistic abstraction of such a biological network yields a numerical analysis of several morphological features simply through using properties and operations of graphs. For instance, expanding from an undirectional graph $G = (V, E)$ where V is a set of vertices, E is a set of edges, and G represents the topology of a given angiogenesis network, a simplicial generalization consisting of 1-simplices (edges) and 0-simplices (vertices) could elegantly represent the interaction of a proteome. Network models can also be modified for additional complexity in order to consider various lattice geometries in multiple

dimensions, weighted poly-values, unary and binary graph operations, include continuous properties visualization strategies, and facilitate greater management of significant control variables by reaching more exhaustive datasets. Furthermore, network features (such as number of nodes and branches) and properties (such as connectivity and clustering) have been studied extensively through various fields consisting of computational, graph, and discrete mathematics. Investigating angiogenesis using network science provides an original and comprehensive set of analytics to quantify otherwise qualitative observations using conventional segmentation methods.

Application of Chaos Theory

The unpredictability of cancer is a formidable challenge to developing personalized cures. To this author's knowledge, few applications of chaos mathematics have been discussed in the context of cancer and instead focus on cancer genomics (for example, to explain aneuploidy) done by Giam and Rancati, 2015, or cancer cell behavior (for example, to explain cancer cell migration themselves) done by Cucuianu, 1998. No application of chaos theory has been applied to tumor angiogenesis despite the process being a fundamental and influential factor in cancer evolution and volatility.

Chaos theory is an amalgamation of scientific theory and mathematics which examines the underlying set of patterns and deterministic laws that are highly sensitive to the initial conditions of a dynamical system (Šarloši, Bocko, and Surovec, 2014). Systems of deterministic chaos, which appear to be disorganized, highly unpredictable, and subject to vast variation, can in principle be predicted, hence the namesake "deterministic." For example, the strange attractor or Lyapunov exponents, mathematical representations of a dynamical system's unpredictability, of tumor angiogenesis has previously not been studied. Utilizing this novel approach to describe a system's inherent complexity and unpredictable behavior, deterministic chaos may uncover latent angiogenic behaviors and bolster the prediction of vasculature development.

Starting with modeling a given network to describe the tumor angiogenesis phase space, the objective of this study is to use Lyapunov Characteristic Exponents (LCEs) to measure the degree of chaos in a network. If the angiogenesis networks tend to be more chaotic than the healthy vasculature, this study seeks to identify what distinct forms and properties are demonstrated by angiogenesis chaos attractors and their relation to the physiological role of angiogenesis.

In summary, the computation of both the objective and subjective factors of tumor angiogenesis using chaos may prove to bridge gaps between qualitative topological observations and biochemical mechanisms to resolve the longstanding challenge of tumor vasculature unpredictability. Further, the analysis of deterministic chaos theory behind tumor angiogenesis may prove that the two models are

fundamentally akin, and lead to novel objective pattern recognition. This comparison will lend itself towards a multi-level understanding examining the unpredictability between morphology and proteins, genes, and disease in various biological networks. Such an approach will not only consolidate findings through a new visual and objective lens but also offer unique solutions in the form of a flexible tool in predicting angiogenic network behavior for therapeutic development and clinical diagnosis.

Question and Hypothesis

Can a computational network reconstruction of tumor angiogenesis mechanisms using coupled morphological, genomic, and physiological parameters when applied to chaos theory determine the predictability of tumor angiogenesis?

This study has been designed to achieve the following objectives:

I. Design a new experimental approach to distinguish healthy and tumorous vasculature with comprehensive preprocessing efforts to control significant variables

II. Build a novel set of network science-based analytics to visualize and investigate the features and properties of biological networks.

III. Determine if there is a significant positive correlation between tumor angiogenesis conceptualized as a dynamical system and deterministic chaos.

IV. Design an innovative algorithm to identify what, if any, distinct forms and properties are exhibited by tumor angiogenesis chaos attractors and their connection to the physiological functions of angiogenesis.

V. Generate an original manipulative pipeline assessing genomic factors and molecular mediators identified by network analysis to correlate with tumor angiogenesis for high-throughput drug discovery

VI. Describe the through case studies and in-vitro modeling of the chaotic morphology of biological networks in a tumor microenvironment

If tumorous and healthy vasculature imaging is analyzed using network science to conceptualize tumor angiogenesis through a dynamical system, then the characteristics of deterministic chaos systems can be applied to better understand the physiology of angiogenesis for therapeutic discovery because chaos theory demonstrates similar morphological patterns recognized in angiogenesis.

Methods and Materials

Please see Appendix B for tables, figures, and program code associated with Methods and Materials

Materials: Image and Data Acquisition

This study's model was developed based on breast cancers due to the accessibility of public labeled datasets and the ability to discern vasculature through mammography. Furthermore, breast vasculatures have been previously studied as indicators for other pathologies, such as cardiovascular diseases, expanding the applications of this study's novel network science-centric observations. In particular, the King Abdulaziz University Breast Cancer Mammogram Dataset was chosen due to images having been previously binned into folders using the BI-RADS (Breast Imaging Reporting and Data System) scheme, enabling the model's conclusions to be based on well-defined breast cancer diagnosis categories for mammogram screening [Figure B-4]. This model examines five of the seven BI-RADS categories (omitting category 0, images rendered incomplete, and 7, with known biopsy) of which increase consecutively in the probability of breast cancer diagnosis; the predictive values for each BI-RADS category is detailed in Appendix B [Table B-1].

Data Preprocessing: Image Segmentation, Skeletonization, and Network Analysis

Image segmentation was performed through various transformations in order to capture the intrinsic complexity and characteristic dense nodal structure of biological networks while controlling significant variables; this process is summarized in Appendix B [Table B-2].

Using ImageJ, the RGB image was converted into inverted grayscale to enhance the contrast of the image. Images whose contrast provided sufficient objective distinguishment between vasculature and background pixels were then met with a coupled homomorphic filter using a Fast Fourier Transform circular mask was applied to capture large areas while minimizing image noize from lighting discrepancies. Then edge segmentation using Gaussian filter (σ=2) observing the fading thin veins that would have otherwise been omitted in further binarization. Following this, the images were masked using adaptive thresholding under Gaussian blur and binarized by assigning a label of 1 or 0 to data points that met the threshold. Morphological skeletonization was then conducted to reduce binary objects to 1 pixel wide presentations, a method described by Zhang, 1984, that makes an adaptive number of recursive passes that erode and dilate pixels that do not break the connectivity of the object [Figure B-3]. The remaining images that did not pass the initial threshold for contrast were passed through the Angiogenesis Analyzer tool on ImageJ [Figure B-5].

Preliminary Network Analysis: Acquiring Network Features and Properties

In discrete mathematics, networks describe relationships between entities through vertices and edges and can be used to capture, communicate, and predict their interactions. Thus the topological features, network properties, and concepts behind graph theory are remarkably relevant towards tumor angiogenesis, and an overview of the novel network analysis variable types considered in this model are described below.

I. Nodes: Identified as pixels with 3 neighbors and a total number of nodes. When compared with the total number of segments, an efficiency index (the multiplicative inverse of the shortest path distance between the nodes) can be calculated.

II. Segment: Identified as edges that delineate two junctions. Variables such as the number of segments, number of master segments, and total length segment describe several morphological indexes. For example, the total length can be interpreted as the topological skeleton. On the other hand, the number of segments describes the connectivity of the graph a common metric used to quantify the "connectedness of a network." Connectivity can be calculated by the average connectivity of a graph G with n vertices $\bar{k}(G) = \frac{\Sigma_{u,v} KG(u,v)}{C(n,k)}$ where $k(u, v)$ is the minimum number of vertices whose deletion makes *v* unreachable from *u*, otherwise known as the maximum number of internally disjoint paths joining the two points.

III. Branch and Extremities: Branches are identified as edges that delineate a junction and extremity, while are the nodes which interrupt connectivity by terminating branches. Branches and extremities can describe skeletal branching patterns, and even in an undirected graph, provide an estimate of a network's arborescence.

IV. Junctions: Identified as a group of nodes and is the point where multiple segments converge. Junctions can also act as master junctions that link at least three master segments and are advantageous for describing a network's modularity.

V. Meshes: Identified as areas that are enclosed by segments. Meshes are useful in identifying the network's Covered Area (CA), Print Area (PA), and consequently the ratio of the two to understand the number of empty regions or "holes".

Secondary Network Analysis: Exploratory Data Analysis of Features and Properties

Upon computing network variables, a comprehensive set of analytics is conducted to investigate and visualize feature importance, variable correlation, network clustering using artificial intelligence and machine learning algorithms. These approaches, described below, are expected to i) elucidate trends

between network variables to uncover tendencies of biological networks and ii) define a metric of network variables that can be used to distinguish angiogenesis from healthy vasculature.

A. **Network Feature Importance:** Feature importance has twofold purposes. First, selecting appropriate features is pivotal for accurate predictive models as useless data results in biases that skew final results in machine learning. Secondly, through running a machine learning random forest tree model with characteristic y, a bar graph comparing the feature importance across all x characteristics describes which x variable is most correlated to y. Hence, by understanding the extent of causation between certain variables or the magnitude of cancer itself, one can identify the network features most characteristic of tumorous vasculature [Figure C-1].
B. **Network Variable Correlation:** To get a better sense of the relationship between the network variables themselves, a matrix of scatter plots between every pair of variables is generated. This original dataset-wide expansive lens allows one to observe general correlations of biological networks [Figure C-2].
C. **Space Representation:** Knowing network variables with feature importance in context, one can plot the 4 and 5 most significant variables in a standard space representation in a 2D [Figure C-3] or 3D [Figure C-4] graph by employing size and color definition. This is an effective initial and qualitative identification of possible correlations and clusters between the variables which have the greatest potential causation of the ultimate y variable, the magnitude of angiogenesis.
D. **K-Means Clustering:** K-Means is an iterative algorithm where each observation belongs to the cluster with the nearest mean (Centroids). To begin finding clusters, the optimal amount of clusters is determined by reducing the inertia as much as possible using K-Means. Clusters are defining groups among a set of objects with similar characteristics [Figure C-5]. Therefore K-Means Clustering competently uses network science to describe the number of categories in a data set which, for instance, could represent a categorical spectrum such as BI-RADS, thus representing a new method of describing the extent of cancerous morphology.
E. **Profile Description of Network Clustering:** Upon obtaining the dataset's clusters, this study proposes the novel practice of comparing each cluster's interaction with the dataset's variable using an intuitive polar coordinate plane. Consequently, researchers can observe the context of the network analysis variables and how they differ among the various clusters to describe the network characteristics of angiogenesis as opposed to healthy vasculature [Figure C-6] [Figure C-7] [Figure C-8].

Application of Chaos Theory

In summary, this study model the analyzed networks as dynamical systems to determine and employs concepts of Lyapunov Characteristic Exponents (LCEs) to assess the degree of chaos of the system and examines the properties in the context of both network science and physiology to determine which, if any, biological networks are chaotic. *This novel approach is further explained below, while the program can be found in Appendix B.*

At the base level, Chaos Theory simulations are based on dynamical systems where one or more variables change over time according to autonomous differential equations. The Lorenz attractor was chosen as the modeled dynamical system due to its origins in fluid dynamics, manipulative constants, and well-studied properties in the field of chaos mathematics. Tumor vascularization is conceptualized on a three-axis phase space data frame, x, y, and z where each point in the space is a unique state of the system and has its own rate of change represented by a vector using the Lorenz attractor described as follows:

$$\dot{x} = \frac{dx}{dt} = \sigma(y - x),$$

$$\dot{y} = \frac{dy}{dt} = x(p - z) - y$$

$$\dot{z} = \frac{dz}{dt} = xy - \beta z$$

From the above equations, $\dot{x}$ is the rate of change of x as time changes, $\dot{y}$ is the rate of change of y as time changes, and $\dot{z}$ is the rate of change as z change, creating three first-order differential equations with two non-linear equations (2nd and 3rd). Then, the constants can be traditionally described and interpreted in this study as the following:

α = sigma representing Prandtl number

p = rho representing the Rayleigh number divided by the critical Rayleigh number

β = beta represents a geometric factor

α = var x_1 highest feature importance

p = var x_2 second-highest feature importance

β = var x_3 third-highest feature important

*Where var $x_{1,2,3}$ are network analysis variables with a strong positive correlation to the magnitude of angiogenesis normalized using the MinMax scale * 10*

With three possible directions for which neighboring points can stretch or shrink, there are thus three LCEs. The Jacobian matrix is the linearized version of the 3D map at a point that indicates the mechanisms by which nearby points evolve forward along the orbit. This creates the Jacobian tangent space maps that evolve forward a pair of displacement vectors to monitor the divergence or convergence

of nearby orbits. The Jacobian of the Lorenz system is computed as the following so that local expansions are given by the behavior of $I + \Delta tJ$.

$$J(x,y,z) = Df(x,y,z) = \begin{pmatrix} -\sigma & \sigma & 0 \\ -x_3 + p & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{pmatrix}$$

A pull-back algorithm is implemented to estimate the three LCEs. The program carries an evolving 3D coordinate frame that is orthogonalized and normalized using the Gram-Schmidt-Scheme, attached to an orbit as it evolves. Thus, the accumulation of the stretching or shrinking of the coordinate frame vectors becomes the LCE estimates. To determine the evolution of the orbits, the program iterates through a continuous flow calculating the partial derivative of the current point using the Fourth-order Runge-Kutta integrator to estimate the next point of the ordinary differential equations, and so on. Finally, the fractal dimension from the LCEs is estimated using the Kaplan-Yorke conjecture to determine a system's measure of chaos. Additionally, a bifurcation diagram is plotted to describe the LCEs in context, and the volume contraction from the trace of the Jacobian (which theoretically should equal the sum of the three LCEs) is computed to measure the confidence of the program's LCEs estimates.

Protein Targeting and Molecular Docking with Statistical Tests

Bioactivity Data collection: The ChEMBL database web service package was used to retrieve bioactivity data, and for the protein derived from the network, analysis was queried [Figure B-6]. The target query search for compounds will come into contact and induce active and inhibitory responses. To analyze the bioactivity data for the specific targets for the specified target protein, each molecule was described by the Chembl ID molecule names, canonical smiles (describes the chemical connectivity of a molecular structure), standard values (potency of the molecule). Specific targets for the target protein were filtered using IC50 standard types for uniformity of data type. IC50 was chosen because it reflects the concentration of the drug required to produce a significant effect. The standard values for compound matching the protein were searched to remove missing data and binned as active (<1000 nM), inactive (>10,000 nM), or intermediate (1000-10,000 nM) and intermediate compounds were dropped. *See Aromatase_bioactivity_preprocessed_data.csv in the supplementary information for the resulting dataset.*

Exploratory Data Analysis: Lipinski descriptors were used to evaluate the drug-likeness of compounds based on the absorption, distribution, metabolism, and excretion known as the pharmacokinetic profile.

The Lipinski's Rule states the following, Molecular weight < 500 Dalton, Octanol-water partition coefficient (LogP) < 5, Hydrogen bond donors < 5, and Hydrogen bond acceptors < 10. Mann-Whitney U-Tests and statistical significance tests were applied for each molecule's Lipinski descriptors. The molecular descriptors, qualitative fingerprint-like descriptions of the compounds, were then calculated using the PaDEL database. Passing through each molecule's pIC50 value (log of IC50), the most effective compound was discerned through a machine learning random forest regression model with *100* estimators. The effects of the final compound were visualized in 3D modeling software through the manual manipulation of environmental presets on a diffusion-limited aggregation model. *See aromatase_06_bioactivity_data_3class_pIC50_pubchem_fp.csv for the resulting dataset.*

In vitro validation with physarum polycephalum

Species: Physarum polycephalum (P. Polycephlaum) is a popular organism that has been the subject of a number of computational experiments, in particular regarding its decentralized transportation of information and efficient pathfinding solving problems such as the Traveling Salesman Problem. It is a unicellular organism belonging to Amoebozoa and is most commonly investigated in its vegetative state as a plasmodium, a large mobile cell forming intricate cytoplasmic vein networks consisting of millions of nuclei and an intracellular cytoskeleton. A single sclerotium, the dormant stage of a starving plasmodium, was divided into multiple pieces and revived into active plasmodium using three water drops and a fresh food medium.

The species has fascinated scientists with its intrinsic decentralized network intelligence, not so much unlike that of angiogenesis being examined in this project. Thus, P. Polycephalum also being another biological network undergoing morphogenesis through the dynamic transformation of tubular polymer; exhibiting reactive behavior to adverse (modeling cancer-inhibiting tumor microenvironments), neutral (control), and generative growth environments (modeling cancer-inducing tumor microenvironments); having a rapid rate of growth, and finally being an accessible organism to culture; was selected as a model to validate this study.

Rearing conditions: The P. polycephalum was reared on a 1% agar medium using oat flakes. Slime molds were 1 week old when the experiment started. All experiments were carried out beneath an opaque cloth so growth occurred in the dark, at a temperature of approximately 20 degrees Celsius. Humidity was controlled by adding five drops of water a day and keeping the lid on the petri dish at all times. The only change in the controlled conditions occurred hourly at a maximum of 30 seconds when photos were taken

hourly using a self-programmed phone camera to document observations. The experiment was replicated through three trials.

Experimental Setup: A circular selection of P. Polycephalum of approximately 1 mm diameter was placed directly at the center of a circular arena that consisted of a 90 mm diameter petri dish filled with either a plain 1% agar (neutral environment, i.e., control treatment), a plain 1% agar mixed with sucrose (100 mM, nutritive environment), or a plain 1% agar mixed with salt (100 mM, adverse environment). P. polycephalum was also examined under a microscope to investigate how morphology influences the cellular level [Figure B-7].

Data and Results

Analysis of Network Data Model

From the feature importance conducted by the network data model, it was observed that the number of nodes, mean mesh size, and number of segments exhibited the greatest, second greatest, and third greatest significance to a given network's BIRADS label [Figure C-1]. The number of branches also demonstrated notable feature importance. These findings are consistent with finding from the network variable correlation, which demonstrates that the nodes, mesh size, and segments share a strong positive linear, strong negative linear, and strong positive linear correlation with the BIRADS value [Figure C-2].

Number of Nodes: The general trend of the number of nodes exhibited by the breast tissue vasculatures shows increases with a greater probability of cancer [Figure C-9] with a distribution that is skewed upwards. The high number of nodes can be reflective of anomalous sprouting which is a defining characteristic of angiogenesis. With the detection of hypoxia, capillary endothelial cells respond to angiogenic cues by extending filopodial protrusions, digesting the vascular basement membrane, and invading into the interstitial matrix as tip cells with the following stalk cells. The angiogenic sprouts then directly invade leading cells that developed filopodia-like protrusions characteristic of tip cells, following stalk cells exhibiting apical-basal polarity, and lumens and branches connecting back to the parent vessels (Nguyen et al., 2018). The high number of nodes thus explains the many heterogeneous vasculature morphogenesis, in addition to advantages such as increased permeability, perfusion, diffusion, and surface area.

Number of Segments: The number of segments across the BIRADS also appears to increase with a greater likelihood of angiogenesis and can describe the connectivity of a graph [Figure C-10].

Specifically, the distribution of the segments appears to demonstrate properties of scale-free networks due to the heavy-tailed, power-law degree distribution. These networks demonstrate preferential attachment, as for instance, in a scale-free network with failures in connectivity occurring at random, the network is still unlikely to lose its connectedness. Cancerous networks demonstrate a much greater relative commonness of vertices with a high degree of preferential attachment due to the cocktail of various pro- and antiangiogenesis growth factors interplaying between vasculature, tumor, and the body's immune response. This convoluted hierarchy allows for fault-tolerant behavior which would be advantageous for tumor survival and resilience.

Mean Mesh Size: The mean mesh size demonstrates a negative nonlinear relationship with the likelihood of cancer [Figure C-11]. Such a relationship is inferred to correspond to the process by which tumors generate a functional and mature vasculature. The process by which the angiogenic cells connect with each other to form a continuous endothelium, generates complex, tight junctions and creates loops that allow the blood to circulate through adhesion molecules. Then, the migration and proliferation of endothelial cells when inhibited, promote pericytes to form capillary walls and to stabilize the new vascular tubes.

Clustering

By plotting inertia, which measures how well a dataset was clustered by K-Means over cluster numbers, the number of groups with common characteristics found by the model was 3.28 [Figure C-5]. Considering there were four BIRADS labels used, of which BIRADS 3 and 4 both decrease a greater likelihood of benign mass over tumorous angiogenesis, the clustering was sufficiently accurate.

Using a polar coordinate system, one can compare the correlation of each model-identified cluster and the explanatory variables. When the model was manually set to four clusters, one cluster could be notably differentiated from the others and can be inferred as the model predicted BIRADS 5 cluster [Figure C-6]. A pie chart describing the proportions of each cluster to the whole relatively accurately confirms this inference to the real proportions of the 213 mammograms inputted [Figure C-8]. It is notable that in each polar coordinate system, the value for total branching length is equal to all clusters, indicating that its effect on BIRADS labeling is negligible.

Analysis of Chaos Theory and Tumor Angiogenesis

The deterministic chaos theory patterns of the proposed model can be observed to correspond with the following three principles describing the Lyapunov exponent, denoted by λ.

1. Stability: When $\lambda < 0$, the distance between coordinate trajectories will converge to zero. This corresponds to an autonomous dynamical system where any perturbation will eventually reach 0 and can be mathematically represented as $\frac{du}{dt} = f(u)$.
2. Periodic: When $\lambda = 0$, the distance will stay constant between coordinate trajectories. This corresponds to period vortex shedding of a black body at low Renoyldss numbers where if the perturbation is exactly in the time-evolution direction creating a constant perturbation flow one time-step ahead. This scenario can be mathematically represented as $\frac{du+\epsilon v}{dt} = f(u + \epsilon v)$ where $v = dt$.
3. Chaotic: When $\lambda > 0$, any distance between trajectories will increase exponentially, and the system's attractor will be chaotic. There exists perturbation in some direction that when perturbed will grow exponentially or exponential divergence from the unperturbed trajectories. This can be mathematically represented as $v(t) \sim e^{\lambda t}$ where $\lambda > 0$.

From the above, λ indicates the rate of change of which two initially close dynamics diverge if positive, or converge if negative in the phase space. Hence, in a dynamical system, such as in angiogenesis, the power of the equation serves as a robust indicator of how rapidly a complex system generated through several interdependent dynamics tends to run up to deterministic chaos. Additionally, it can be further observed that the number of dimensions considered in the applied equation corresponds to the number of Lyapunov exponents. The set of Lyapunov exponents can be regarded as the spectrum of Lyapunov Characteristic Exponents (LCEs) which compare the chaotic unpredictability and characterize the fractal dimension of the biological network. Thus generating a 3D plane to model the Lorenz chaotic equation using constants derived from the network feature importance variables, three Lyapunov exponents can thus be used to categorize breast cancer vasculature into four types of attractors: stable fixed point, stable limit cycle, stable 2D tori, and strange attractor as described by the Kaplan-Yorke conjecture.

Both of these chaos theory principles are remarkably evident through the conceptualization of the four BIRADS networks, demonstrating a steady-state flow field (negative exponent), constant periodic flow field (exponent equals 0), and chaotic flow field (positive exponent). The BIRADS 1

conceptualization, representing a negative test for breast cancer and healthy vasculature, has an LCE described by the following set:

$$\lambda_1, \lambda_2, \lambda_3 = -0.00024230742569512925, \ -0.1282739359573508, \ -19.392102248700645$$

Based on the signs of the Lyapunov exponents, the fractal dimension of the BIRADS 1 network is a stable fixed point, and this can be visualized by plotting the modified Lorenz equation [Figure C-12]. The fixed point can be visually observed in the graph, as all perturbations (represented as spirals) eventually reach 0, which is confirmed by the distances of the network's trajectories oscillating towards the origin [Figure C-12]. This study thus provides initial evidence that healthy breast vasculature, in particular a BIRADS 1 mammogram, demonstrates stable fixed point behavior. Consequently, stable fixed point equations can be leveraged to predict the vasculature of healthy breast vasculature.

However, the novelty of this model comes into greater use when describing challenging to predict dynamical systems, namely angiogenesis. To begin, the same logic can be applied to BIRADS 3, 4, and 5 where the LCEs can again be described by the following sets.

BIRADS 3

$$\lambda_1, \lambda_2, \lambda_3 = -0.0016021016772777763, \ -0.1642225377104428, \ -18.8461981282713$$

BIRADS 4

$$\lambda_1, \lambda_2, \lambda_3 = -0.7494673241944237, \ -0.0001127522908231926, \ -27.309638287683644$$

BIRADS 5

$$\lambda_1, \lambda_2, \lambda_3 = -1.4204153377552509, \ 0.00023199424653297385, \ -33.933559322701655$$

As the BIRADS categorization increases in the likelihood of cancer, the value of λ_2 nears becoming positive. In fact, one can observe an increase in the magnitude of chaos for the plots of BIRADS 3 and 4 through the more loosely wound and continuous spirals [Figure C-13] [Figure C-14]. It can then be determined that BIRADS 5, highly suggestive of malignancy (>95% of cancer), is a strange attractor. The BIRADS 5 LCE exhibits one positive Lyapunov exponent, indicating that the dynamic system is chaotic and the distance between the coordinates will exponentially increase [Figure C-15]. As with BIRADS 1, this study provides the first preliminary evidence which suggests that tumor angiogenesis is chaotic, and thus concepts relating to chaos theory can be applied to determine what underlying chaotic patterns are yet to be explored in angiogenesis. These applications are further discussed in the conclusion.

Aromatase Targeting and Molecular Docking with Statistical Tests

The model's bioactivity query for aromatase revealed Cytochrome P450 19A1 as a key single protein involved in breast cancer. 2817 molecules matched Cytochrome P450 19A1 and 2020 molecules that were either inactive or active. These molecules were compared using Lipinski's rule of five [Table C-17]. The following notable conclusions were made.

I. Active molecules demonstrated a higher pIC50 standard value, attesting to their potency as a drug candidate.
II. Active and inactive molecules share similar centers for molecular weight, however, active molecules have a smaller range describing a potential specificity for drug size.
III. The null hypothesis for LogP failed to be rejected, indicating that both active and inactive molecules share the same distribution. This indicates that both active and inactive drug candidates have the same permeability and ability to reach the target tissue in the body.
IV. Molecules targeting aromatase tend to accept more hydrogens than donate, thus it can be hypothesized that potential drug elements targeting aromatase generally will include electronegative chemicals, are soluble, and are less volatile.

Following the identification of the molecular descriptor, the random forest regression model was able to predict pIC50 values over experimental PIC50 with an r^2 of 0.512 [Figure C-18].

In-Vitro Validation

The qualitative observations of the change P. Polycephlaum morphogenesis in reaction to an inducing, neutral, or inhibiting environment aligns with the conclusions drawn from the network and chaos theory analysis of angiogenesis.

This is representative of solid tumors, which cannot grow beyond 10^6 cells in the absence of a blood supply. Thus this initial pre-vascular phase of growth is followed by a vascular phase in which tumor-induced angiogenesis is the limiting factor for further growth, effectively enabling malignant cells direct access to circulating in the bloodstream, which is observable when comparing the initial expansion between the adverse-neutral-generative cultures. While the generative cultures exhibit "malignancy" through prolific and dense globules of layered tubules surrounding food hubs, neutral cultures have more distinctive tendril pathways and large mesh areas, while adverse cultures seem to resort directly into the neovascular phase with weak islands [Figure C-19] [Figure C-21] [Figure C-23].

Endothelial cells derived from the successive differentiation of mesodermal cells into hemangioblasts create initial vascular structures called primitive blood islands. Remarkably these blood islands are topologically, functionally, and physiologically akin to P. Polycephalum islands. The

hemangioblasts from the center of the islands give rise to hemangioblasts stem cells while peripheral hemangioblasts differentiate into angioblasts, generating a morphology with a dense protruding leading edge leading the direction for precursors of mature endothelial cells. Fundamentally, this is also the process by which P. Polycephalum forages, through islands and tendrils. Vascular endothelial growth factor (VEGF) and a cascade of pro-and antiangiogenic agents influence angioblasts and newly formed endothelial cells to migrate on a matrix of collagen and hyaluronic, allowing for the fusion of blood islands and remodeling of tubular structures to form the first primitive vascular plexus (Michaelis, 2014). This is notable through comparing the notable difference in the network formation and delineation of the P. polycephalum cultures [Figure C-20] [Figure C-22] [Figure C-24]. Likewise, this prompts further research into whether one of the P. polycephalum growing environments can astutely model angiogenesis, or rather represent different stages in cycling angiogenesis undergoing various phases of growth rate.

The comparisons further exist at the molecular level. The cytoplasm of P. polycephalum consists of an ectoplasmic viscous phase consisting of actin and myosin forming the contractile walls of the veins, and an endoplasmic liquid phase characterized by different concentrations of fibrous protein. This is remarkably similar to the constant remodeling of the actin cytoskeleton by GTPase Cdc42 activated filopodia sensing the VEGF gradient, lamellipodia polymerization at the leading edge involving Rac and Arp2/3 complex driven protrusions, and the interplay between focal adhesion kinases and inside-out signaling which compute the directionality and the contraction of stress fibers for endothelial cell motility (Yang et al., 2017). Ultimately, both dynamic, actin-polymer-based systems can be represented by similar morphological and molecular pathway networks.

This experiment demonstrates that when coupled, computational analysis and in vitro modeling can astutely model the theoretical baseline of biological networks. While the cocktail of chemicals in the P. polycephalum's petri dish is immensely simplified from the tumor microenvironment, remarkable similitudes in behavior were distinguished through a much more cost and time-efficient methodology for validation purposes. In addition, other remarkable behaviors demonstrated by P. polycephalum, such as isolated island globules on petri dish lids, may uncover other latent rules behind angiogenesis [Figure C-25]. Thus, a compelling qualitative argument for the conceptualization of angiogenesis as a decentralized network, such as the morphology of P. polycephalum, can be made for simple yet novel in vitro modeling techniques.

Conclusion and Discussion

It was found that angiogenesis breast cancer vasculature exhibited a higher Lyapunov exponent than in nontumorous vasculature, confirming the hypothesis that angiogenesis architecture can be modeled by deterministic chaos; indeed, cancer is chaos. Vasculatures with a higher number of nodes, segments, and smaller mean mesh size were the most chaotic. It is inferred that these variables result from vascular sprouting and variation in vascular density and permeability, and tip cells formed by sprouting filopodia toward proangiogenic signal sources, and the interplay of pericytes and the endothelial wall, respectively. Furthermore, a model predicting the drug potency of aromatase target compounds with an r^2 value of 0.512 was designed and the in vitro experiments of P. polycephalum confirmed and compelled for further investigation of how the tumor microenvironment impacts the morphology of angiogenesis. In conclusion, the remarkable network analysis of chaos theory and its affinity to angiogenesis compel for further investigation, countless applications in predicting tumorous vasculature, drug discovery, and uncovering the chaotic rules that drive the biological networks in and around us. Below are the various applications and next steps for this project.

Clinical Diagnosis

Currently, angiogenesis prediction models most predominantly rely on neural networks based on test and training datasets and thus are limited to the morphologies described by the inputted data. These models either lack the scalability necessary to describe the various abnormalities of vessel branching with bias, or succumb to overgeneralization of data, leading to an inflexibility to accommodate vasculature heterogeneity resulting in ambiguous or inaccurately extrapolated conclusions (Allehaibi and Khan, 2021). Fortunately, studies of chaos theory by mathematicians have identified mathematical rules to determine the extent of unpredictability in a given system. To use an example demonstrated in this project, the Lyapunov exponent when assessing the chaotic magnitude of a system can be rearranged from $d_t = d_0 e^t$ into $predictability\ horizon\ \frac{1}{\lambda} ln \frac{a}{d_0}$. This mathematical represented uses d_0 as the initial error and a as the maximum allowed error, enabling clinicians to be confident in whether their inferences are relevant or not in highly volatile biological processes.

Bridging Biochemical and Mathematical Research for Novel Drug Discovery

Nearly all models employ a shape segmentation technique that is locally template-based. Thus considering angiogenesis growth is characterized by irregular outlines of which no discernable shape can be formulated, models are often unable to reliably distinguish large artifacts and salient features. These

spatial relationships, which are integral to pathological diagnostic interpretation, are difficult to quantify (Ing et al., 2017). Further, existing solutions often view tumor-induced angiogenesis solely through a morphological lens, failing to address the complex biomechanical and biochemical phenomena and their interactions across spatial and temporal scales, occurring between tumor and vessel polymers (Peirce, 2008). This static interpretation is problematic as angiogenesis is correlated with several complicated interactions with cell types, angiogenic factors, and components in the extracellular matrix, which constantly change the growth behavior of the tumor vasculature. Thus, there is an immense need for a novel approach to elucidate the connection between angiogenesis and molecular pathways through a multilevel pertinence to the proteomic, genetic, and pathological sphere.

The characterization of angiogenesis is a network type that bridges the gap between underlying biochemical mechanisms and mathematical morphological patterns. An exemplary example of this is the pursuit of an accurate model of the tumor microenvironment. Using the network analysis models proposed in this study, scientists can further investigate the percolation of tumor angiogenesis via methods used to analyze scale-free networks. For example, the critical probability p_c of the tumor microenvironment can be investigated using continuity, differentiation, spanning cluster density, and functional turning points. Thus it is believed that such a model could be employed to study the permeability of compounds for drug discovery.

Applications in Natural Decentralized Networks

Novel chaotic systems and network analysis methods such as the ones developed in this project are not exclusive to breast cancer vasculature. In fact, a large number of disorders has been linked to either poor or excessive levels of vessel growth including cancer, stroke, heart disease, and retinopathy. For example, the levels of endogenous revascularization after a brain or heart infarct are very limited while contrarily, excessive vessel formation can stimulate the growth of cancer or contribute to leaky blood vessels in retinopathy. In addition, biological networks in and around us, from our nervous system to the spreading of contagious diseases, may demonstrate undiscovered chaotic properties. Thus the investigation of network science and deterministic chaos in these systems may reveal underlying mechanisms for innovative solutions.

Future Actions

One next step for this project entails further validation of the model with coupled in-silico and in-vitro case studies to investigate the viability of chaos theory in a potential pharmaceutical setting. Additionally, the Python code used for applying deterministic chaos, network science, and drug screening,

will be converted into an open-source library. Hopefully, this will enable other scientists to explore and contribute to the novel applications of chaos theory in biological networks, while also promoting community biotechnology and the decentralization of research. Finally, the computational algorithms used in this project could be implemented in imaging hardware to develop a commercial scanning device robust in analyzing chaotic biological networks.

Acknowledgments

I would like to offer my special thanks to Mr. O'Byrne for fostering my curiosity for science, Dr. Brad Bass for his continuous mentorship, and Dr. Giannoula Lakka for her invaluable guidance as well as assistance in data collection.

References

Demené, C., Payen, T., Dizeux, A., Barrois, G., Gennisson, J. L., Bridal, L., & Tanter, M. (2019). 3-D Longitudinal Imaging of Tumor Angiogenesis in Mice in Vivo Using Ultrafast Doppler Tomography. *Ultrasound in medicine & biology*, *45*(5), 1284–1296. https://doi.org/10.1016/j.ultrasmedbio.2018.12.010

Hanahan D, Folkman J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell. 1996;86:353–364. doi: 10.1016/S0092-8674(00)80108-7. [PubMed] [CrossRef] [Google Scholar]

Goel, S., Wong, A. H., & Jain, R. K. (2012). Vascular normalization as a therapeutic strategy for malignant and nonmalignant disease. *Cold Spring Harbor perspectives in medicine*, *2*(3), a006486. https://doi.org/10.1101/cshperspect.a006486

Ing, N., Huang, F., Conley, A. *et al.* A novel machine learning approach reveals latent vascular phenotypes predictive of renal cancer outcome. *Sci Rep* 7, 13190 (2017). https://doi.org/10.1038/s41598-017-13196-4

Khalid Allehaibi, Yaser Daanial Khan, Sher Afzal Khan, "iTAGPred: A Two-Level Prediction Model for Identification of Angiogenesis and Tumor Angiogenesis Biomarkers", Applied Bionics and Biomechanics, vol. 2021, Article ID 2803147, 15 pages, 2021. https://doi.org/10.1155/2021/2803147

Michaelis U. R. (2014). Mechanisms of endothelial cell migration. Cellular and molecular life sciences : CMLS, 71(21), 4131–4148. https://doi.org/10.1007/s00018-014-1678-0

Peirce S. M. (2008). Computational and mathematical modeling of angiogenesis. Microcirculation (New York, N.Y. : 1994), 15(8), 739–751. https://doi.org/10.1080/10739680802220331

Pitulescu, M. E., Schmidt, I., Giaimo, B. D., Antoine, T., Berkenfeld, F., Ferrante, F., Park, H., Ehling, M., Biljes, D., Rocha, S. F., Langen, U. H., Stehling, M., Nagasawa, T., Ferrara, N., Borggrefe, T., & Adams, R. H. (2017). Dll4 and Notch signalling couples sprouting angiogenesis and artery formation. Nature cell biology, 19(8), 915–927. https://doi.org/10.1038/ncb3555

Šarloši, Juraj & Bocko, J. & Surovec, Robert. (2014). Deterministic Chaos. Procedia Engineering. 96. 10.1016/j.proeng.2014.12.116.

Warren BA (1979) The vascular morphology of tumors. In *Tumor Blood Circulation: Angiogenesis, Vascular Morphology and Blood Flow of Experimental and Human Tumors*, Peterson, H-I (ed) pp 1–47. CRC Press Inc.: Boca Raton, FL [Google Scholar]

Yang, X., Li, S., Zhong, J., Zhang, W., Hua, X., Li, B., & Sun, H. (2017). CD151 mediates netrin-1-induced angiogenesis through the Src-FAK-Paxillin pathway. Journal of cellular and molecular medicine, 21(1), 72–80. https://doi.org/10.1111/jcmm.12939