

## Research Article

# Quantification of Spatial Parameters in 3D Cellular Constructs Using Graph Theory

A. W. Lund,<sup>1</sup> C. C. Bilgin,<sup>2</sup> M. A. Hasan,<sup>2</sup> L. M. McKeen,<sup>1</sup> J. P. Stegemann,<sup>3</sup> B. Yener,<sup>2</sup> M. J. Zaki,<sup>2</sup> and G. E. Plopper<sup>1</sup>

<sup>1</sup>Department of Biology, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

<sup>2</sup>Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

<sup>3</sup>Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

Correspondence should be addressed to G. E. Plopper, ploppg@rpi.edu

Received 16 December 2008; Revised 22 June 2009; Accepted 16 August 2009

Recommended by Satoru Miyano

Multispectral three-dimensional (3D) imaging provides spatial information for biological structures that cannot be measured by traditional methods. This work presents a method of tracking 3D biological structures to quantify changes over time using graph theory. Cell-graphs were generated based on the pairwise distances, in 3D-Euclidean space, between nuclei during collagen I gel compaction. From these graphs quantitative features are extracted that measure both the global topography and the frequently occurring local structures of the “tissue constructs.” The feature trends can be controlled by manipulating compaction through cell density and are significant when compared to random graphs. This work presents a novel methodology to track a simple 3D biological event and quantitatively analyze the underlying structural change. Further application of this method will allow for the study of complex biological problems that require the quantification of temporal-spatial information in 3D and establish a new paradigm in understanding structure-function relationships.

Copyright © 2009 A. W. Lund et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Traditional reductionist biology has developed a wealth of knowledge concerning individual molecules and their function. This data set, however, has exceeded our capabilities for integration and analysis. Attempts to step outside these subcellular networks, into the equally complex world of three dimensional (3D) tissue development, are limited by our ability to deal with the spatial and temporal information that can be gathered in such studies [1]. Multispectral imaging has the potential to provide insight into the spatial orientation and dynamic organization of biological structures; however, we are still not equipped to quantitatively analyze and use all the data that is present in such images.

Graph theory has emerged as a method to characterize the structure of large complex networks leading to a better understanding of the dynamic interactions that exist between their components. A graph is a mathematical structure that represents the relationships between members of a given set and is depicted as discrete points (nodes) connected by

interactive links (edges) [2]. These graphs define pairwise interactions which can be either directed or undirected [1]. Both local and global topographical characteristics extracted from these graphs can define the network structure (topology) and relationships that exist within the node population [3]. Nodes with similar characteristics tend to cluster together and the pattern of this clustering provides information as to the shared properties, and therefore the function, of those individual nodes [4]. The graph theory approach permits users to test hypotheses that define the characteristics of node interaction, regardless of whether the relationships between nodes are well defined or not.

Graph theory has been applied to a variety of real networks from biology to social networking the internet and literature citations. Biological networks have common designs that are governed by simple and quantifiable organizing principles [1]. These principles are not random: evolution by natural evolution has consistently repeated efficient and functional patterns of organization. We can use our understanding of these principles to extract information

about the function of complex biological networks, from protein-protein interactions, cell signaling cascades, disease networks [5], transcriptional, metabolic [6, 7], and genetic regulatory systems [8], and brain connectivity [9–11]. These studies highlight the ability of graph theory to capture the modularity of biological networks and provide important insight into the function of these networks on a global level. Researchers have recently used graph theory to extract topological features from neural networks and compare connectivity between young and old patients. They were able to extract information from these graphs that suggest that the process of aging that occurs in the brain may actually be due to changes that are occurring in the modularity of their sparse networks [11]. While significant steps have been made, the key challenge still remains to understand the structure and dynamic interactions of living cells within the context of their physical environment including matrix architecture, cellular compartmentalization, and cytoskeletal affects that may restrict these subcellular interactions [1]. If we can understand the design principles that govern biological organization, then we can use those principles to inform and accelerate studies of tissue morphogenesis, development, differentiation, and organogenesis.

Natural and engineered tissues are a collection of cells arranged within a structural scaffold of extracellular matrix (ECM) proteins that provide biochemical and mechanical cues to direct the function of those cells. Tissue function is therefore dependent upon the spatiotemporal resolution of matrix proteins, cells, and signaling molecules. Understanding the interactions of this complex set of components over time requires novel techniques of extracting data from the intact tissue. In fact, the three dimensionality of tissue itself is critical for the maintenance of normal cellular function [12, 13]. Changes in cell shape [14], ECM organization [15], tissue geometry [16], and mechanics [17] affect gene expression profiles [18], differentiation [19], and cancer progression [20]. It is therefore critical to probe cellular function within native-like environments and treat these “tissues” as multiscale networks of interacting units.

This paper presents a novel application of graph theory to extract and quantify the tissue structure/function relationship. We construct cell-graphs that place cell nuclei as the nodes of the graphs and link those nuclei as a function of their pairwise distance based on the assumption that cells that are in physical contact (i.e., membrane proximity) interact. Previous work demonstrates that graph theory-based analysis of the locations of cell nuclei provides important information concerning the (dys)functional states of the tissue [21–23]. Cell-graphs and their variants (Hierarchical Cell-graphs and ECM-aware cell-graphs) based on pathological images of brain, breast, and bone cancer, respectively, model tissue structure to provide quantitative metrics that enhance diagnostic accuracy over that of other automated methods [24]. These results arose from graphs of 2D cell nuclei in histological images, suggesting that a wealth of knowledge has yet to be exploited in higher-resolution images. Here we extend graph-based analysis of 3D fluorescent images to track biological structures in a

dynamic fashion and extract data reflecting fundamental changes in tissue organization and structure.

## 2. Materials and Methods

**2.1. Cell Culture.** Cryopreserved hMSCs (Lonza) were grown according to manufacturer’s instructions. hMSCs were cultured in Dulbecco’s Modification of Eagle’s Medium 1x (DMEM) supplemented with 10% fetal bovine serum (FBS) and fungizone/penicillin/streptomycin (FPS) (10,000 units/mL). Medium was changed every three days and cultures were incubated at 37°C in a humidified atmosphere containing 95% air and 5% CO<sub>2</sub>. Cells were detached using trypsin-EDTA and passaged into fresh culture flasks upon reaching confluence. hMSCs were used between passages 6 and 8. In preparation for incorporation into 3D constructs, cells were washed with PBS, detached with trypsin-EDTA, collected, and counted. All reagents were purchased from Fisher Scientific unless otherwise noted.

**2.2. 3D Collagen I Culture.** Three-dimensional collagen I gels were prepared by mixing cells with the following reagents: DMEM (14%), FBS (10%), 5× Conc. DMEM (16%), 0.1N NaOH (10%), and 4 mg/mL collagen I (50%) (MP Biomedicals). The final collagen concentration is 2 mg/mL within each construct. Constructs of a volume of 1.0 mL were made in 12-well plates and the cellular density was kept at  $1.0 \times 10^6$  cells per mL ECM and  $3.0 \times 10^5$  cells per mL, for their respective experimental sets. The constructs were incubated at 37°C for 30 minutes, released from the wells, and incubated in DMEM for 24 hours. The floating gels were removed at from culture at each time point and imaged using a Nikon digital camera for compaction. Gel images were analyzed using Image J (NIH) and the percent change in volume was calculated. Triplicates of each sample were analyzed. Statistical analysis was performed using the student *t*-test; a *P* value less than .05 was considered significant.

**2.3. Fluorescence Imaging.** Collagen I constructs were prepared for fluorescence staining at hours 0, 1, 2, 4, 6, 10, 16, and 24. Constructs were washed with PBS followed by a 30-minute incubation at 4°C in 3% paraformaldehyde. The constructs were then washed with PBS and incubated at 4°C for 30 minutes in cell blocking solution (0.25% Tween20, 1% BSA in PBS). Collagen constructs were incubated for 10 minutes at room temperature in SYTOX Green Dye (Invitrogen) at a final concentration of 50 nM in cell blocking solution. The samples were stored at 4°C in PBS until imaging. 3D confocal images were taken using a Zeiss LSM 51 and Image J (NIH) was used to convert the files into .tiff format for segmentation.

**2.4. Image Segmentation Node Identification.** The 3D confocal images were segmented using the Otsu Thresholding algorithm [25]. The Otsu Thresholding algorithm assumes that there are two classes of pixels in the observed image and finds a threshold value that will automatically separate

TABLE 1: Description of graph metrics.

Number of nodes	The number of vertices in the graph, in this case cell nuclei as defined by image segmentation.
Number of edges	The number of links between the vertices of the graph, in this case the probability of cellular interaction based on distance.
Average degree	The number of neighbors/connections that a given node has averaged over all the nodes in the graph.
Clustering coefficient (C)	Is defined as $C_i = (2E_i)/(k(k+1))$ , where $k$ is the number of neighbors of node $v_i$ and $E_i$ is the number of existing links between $v_i$ 's neighbors. $C$ measures the ratio of existing links to possible links for each node.
Path length or hop count:	The shortest distance between two nodes reflective of the weight of each link.
Number of nodes	The number of vertices in the graph, in this case cell nuclei as defined by image segmentation.
Eccentricity and closeness	The maximum and average of the shortest lengths, respectively. The average of these two features provides global measure of average eccentricity and average closeness.
Diameter	Defined as the maximum eccentricity, greatest distance between two nodes in the graph.
Central points	Defined as nodes that have an eccentricity equal to the average eccentricity.
Hop plot value	Reflects the size of a neighborhood between any two nodes with a "hop", a single edge.
Hop plot exponent	Slope of the hop plot values as a function of $h$ in log-log scale.
Effective diameter	Defined as $\varepsilon = N^2/(N+2E)^{1/H}$ , where $N$ and $E$ are the number of nodes and edges, and $H$ is the hop plot exponent.
Giant connected component	Defined as the largest set of nodes that are reachable from each other.
Giant connected ratio	The ratio of the size of the giant connected component to the number of nodes in the graph.
Isolated node	A node with no edges and therefore a degree of 0.
End node	A node with one edge.

TABLE 2: Labeling scheme to capture local degree of subgraph networks.

Degree range	Label
0	0
1-2	1
3-4	2
5-6	3
7-8	4
9-10	5
13-14	7
15-16	8
17-19	9
20-22	10
23-26	11
More than 26	12

the foreground pixels from the background. The algorithm searches for the threshold value that minimizes the intraclass variance. After finding this optimal threshold, the intensity values of each pixel are compared against the threshold and the pixels with intensity values higher than the threshold assigned as foreground pixels. The foreground pixels that are touching each other are connected, and the center of mass of these nuclei is calculated to identify node (vertex) set for cell-graph generation.

**2.5. Cell-Graph Generation.** A graph is given by  $G = (V, E)$  where  $V$  is the vertex set of the graph and  $E$  is the edge (link) set of the graph. Cell-graphs are constructed for structural

modeling of tissue organization. The vertex set contains the nuclei, obtained from image segmentation as explained above. The edge set represents pairwise relationships between cells. We assume that a pair of cells interacts if they are in physical contact (e.g., cell membranes are touching or very close to each other). Thus, we define a distance-based edge function to model this hypothesis, and a link between a pair of two nodes is established when the Euclidean distance  $d$  between them is defined as

$$d(u, v) = \sqrt{(u_x - v_x)^2 + (u_y - v_y)^2 + (u_z - v_z)^2}, \quad (1)$$

where  $u_x$ ,  $u_y$ , and  $u_z$  are  $x$ ,  $y$ ,  $z$  coordinates of node  $u$ , respectively.

For each pair of vertices the pairwise Euclidean distance is calculated and a link is set when that distance is smaller than the linking (edge) threshold. This threshold value is determined by considering the nucleus-membrane ratio and estimating cell diameter. We have identified thresholds, corresponding to the approximate radii of spread mammalian cells, of 65, 70, and 75 microns (from a distortion free sphere representation of the cell membrane to observable distortion) and executed our algorithms for each.

**2.6. Cell-Graph Features.** Topographical properties for each cell graph were extracted and averaged over the entire graph to represent global measures of "tissue organization"; see Table 1. These computed properties or cell-graph metrics are well accepted in the internet topology generator and simulation literature (see <http://cat.inist.fr/?aModele=afficheN&cpsidt=14492873> and references therein) and used to (i) characterize graph topography and (ii) verify if two

graphs come from the same distribution. A rigorous explanation of these metrics and their significance is provided in our previous publications [23, 24].

**2.7. Computing Randomness of Tissue Organization.** The Cumulative Nearest Neighborhood Distribution (CNND) function [26] was used to generate random positions of nuclei and compared to experimental nuclei locations. The CNND function was calculated from the 3D geometry in our tissue samples defined by a 3D bounding box  $B$  that includes all the cells in a sample. For each cell we chose a uniformly random  $x, y, z$  coordinate in  $B$ . Given  $B$  the function takes distance threshold  $t$  as the input parameter, defined as follows:

$$G(t) = \frac{1}{n} \sum_{i=1}^n I(\gamma_i \leq t), \quad (2)$$

where  $I(\cdot)$  is an indicator function,  $\gamma_i$  is the distance of cell  $i$  from its nearest neighbor, and  $n$  is the total number of cells segmented in the image.

This function value represents the fraction of the cell counts for which their nearest neighbor falls within the threshold distance  $t$ , and as the threshold  $t$  increases, more cells include their nearest neighbor within the threshold. Thus, this function value always increases from 0 to 1.

The randomness of cell-cell relationships was tested by defining an edge function according to the Erdos-Renyi model, calculating the same set of features on our random graphs, and comparing the random graph metrics to that of the cell-graphs:  $G(N, p)$  where  $N$  nodes are linked by a probability of  $p$  [27]. Here the edge definition is independent of node location. A variation of the  $G(N, p)$  model, the  $G(N, M)$  model, chooses a graph at random from the set of all possible graphs with  $N$  nodes and  $M$  links. For each time point and cell density, we generated a cell-graph, calculated the number of nodes and links in that cell-graph, and then generated a random graph using the  $G(N, M)$  model. We generated 10 random graphs, calculated the metrics, and then averaged them. We compared the average values of these metrics to the cell-graph metrics.

**2.8. Voronoi-Delaunay Graph Modeling of Samples.** We compared the cell-graph technique to Voronoi diagrams and the Delaunay triangulation method [28–30]. On a sample tissue image, the Voronoi diagram partitions the image into convex polytopes (i.e., polygons in 2D and polyhedrons in 3D) such that each polytope contains exactly one cell (generating point) and every point in a given polytope is closer to its generating point than to any other generating point in the tissue (Voronoi cell). Delaunay triangulations were built by linking the generating points of neighboring Voronoi cells. We built Delaunay triangulations on cells identified in the segmentation step, then calculated the same set of metrics, and observed their changes over time.

**2.9. Subgraph Pattern Mining.** Classical graph mining [31] finds frequent subgraphs from a graph database,  $D$ . The

support of a subgraph  $g$  is defined as  $\sum_{d \in D} l(g, d)$ , where  $l(g, d) = 1$  if  $g$  is a subgraph of  $d$  and 0 otherwise. A subgraph is considered frequent if its support across the graph database is equal or higher than a user-defined support threshold.

For the cell graphs we sought subgraphs that are frequent with respect to their embedding count in one large graph. Support of a subgraph,  $g$ , is equal to its embedding count in one large graph,  $G$ , which is provided as input. We considered each cell graph in a separate graph mining session to find its frequent subgraphs. In each session, we used a common support threshold value [20] and used only the edge-disjoint embeddings (explained below) while counting the embedding. Frequent subgraphs for different cell graphs were processed further to find those with high support at one time point, but not in others. To do so, we modified the support count method of the graph mining algorithm in a DMTL library [32] which provides modular design and therefore simpler modification. In our support count subroutine, we used Ullman's algorithm [33] which returns the decision that a graph,  $g$ , is a subgraph of  $G$  and also provides all possible embeddings of  $g$  in  $G$ . Since subgraph mining is very costly (subgraph isomorphism test is  $NP$ -Hard), we restricted our mining process to consider subgraphs that have at most four edges.

For each subgraph we detected the frequency of its occurrence at each time point and for each cell density using edge-disjoint embeddings. Subgraph pattern size was restricted to an upper limit of four edges, and each vertex is labeled by an integer number corresponding to the proportional degree of that vertex to capture local network information. The labeling scheme is shown in Table 2. If every vertex in a graph has a unique label, the embedding count of a graph,  $g$ , in a larger graph,  $G$ , can be at most one. However, in our samples, multiple vertices may have the same label (Table 2); so the embedding count for a subgraph may be greater than one. To avoid unfairly elevated embedding counts caused by a subgraph where many of its vertices share the same labels we only considered edge-disjoint embeddings while counting the support of a pattern. Because finding the maximum value for the edge-disjoint embedding count from the total embedding count is  $NP$ -Hard, we used simple smallest degree heuristics to find an independent set [25] to obtain the edge-disjoint embedding support of a subgraph pattern. All the subgraphs that have an edge-disjoint embedding count more than the threshold value of 20 were returned by the algorithm.

### 3. Results

**3.1. Macro- and Microscale Compaction Occurs in the First 24 Hours of Collagen I Culture.** To provide proof of concept for our cell-graph method we encapsulated hMSC in 3D collagen I hydrogels at a concentration of  $3.0 \times 10^5$  cells/mL of collagen I. A hallmark of hMSC hydrogel culture is the cell dependent compaction event that begins immediately following encapsulation. Triplicate samples were fixed at hours: 0, 1, 2, 4, 6, 10, 16, and 24 postencapsulation. We imaged the entire collagen I gel after fixing, as shown in Figure 1(a).



TABLE 3: Quantitative metrics for  $3.0 \times 10^5$  cells/mL collagen I hydrogel 24-hour timecourse.

	Time (hours)						
	0	1	2	6	10	16	24
Average degree	1.18	1.53	2.45	3.62	5.56	9.40	14.40
Clustering Coefficient	0.26	0.37	0.47	0.56	0.63	0.59	0.61
Average eccentricity	1.11	2.08	2.16	4.27	13.52	14.38	14.24
Maximum eccentricity (diameter)	5.00	8.00	7.00	9.00	23.00	20.00	19.00
Minimum eccentricity (radius)	0.00	0.00	0.00	0.00	1.00	10.00	10.00
Average eccentricity 90	1.11	1.89	2.16	3.77	11.28	10.97	11.28
Maximum eccentricity 90	5.00	7.00	7.00	8.00	20.00	16.00	17.00
Minimum eccentricity 90	0.00	0.00	0.00	0.00	1.00	8.00	8.00
Average path length (closeness)	0.84	1.31	1.41	2.39	6.97	7.07	7.36
Hop plot exponent	0.20	0.49	0.41	0.70	1.08	1.51	1.46
Effective hop diameter	14410546.75	961.44	3191.99	164.88	26.67	10.34	11.29
Giant connected ratio	0.13	0.22	0.20	0.20	0.68	0.99	1.00
Number of connected components	35.00	34.00	35.00	26.00	12.00	3.00	1.00
Percentage of isolated points	0.38	0.32	0.22	0.05	0.02	0.01	0.00
Percentage of end points	0.28	0.19	0.17	0.12	0.05	0.02	0.01
Number of central points	23.00	23.00	21.00	9.00	11.00	1.00	14.00
Percentage of central points	0.38	0.32	0.22	0.05	0.05	0.00	0.03

TABLE 4: Quantitative metric statistics for different link thresholds for  $3.0 \times 10^5$  cells/mL collagen I hydrogel 24-hour timecourse.

	Time (hours)						
	0	1	2	6	10	16	24
Average degree	$1 \pm 0.21$	$0.9 \pm 0.19$	$1.05 \pm 0.18$	$1.54 \pm 0.31$	$1.85 \pm 0.35$	$2.6 \pm 0.46$	$4.15 \pm 0.67$
Clustering coefficient	$0.14 \pm 0.07$	$0.12 \pm 0.05$	$0.12 \pm 0.04$	$0.18 \pm 0.06$	$0.29 \pm 0.05$	$0.32 \pm 0.04$	$0.44 \pm 0.03$
Average eccentricity	$1.62 \pm 0.28$	$1.21 \pm 0.26$	$1.39 \pm 0.1$	$2.51 \pm 0.7$	$3.11 \pm 0.59$	$8.93 \pm 2.21$	$19.85 \pm 7.61$
Maximum eccentricity (diameter)	$7 \pm 0.64$	$5 \pm 0.05$	$5.33 \pm 0.64$	$7.78 \pm 1.69$	$10 \pm 1.26$	$15.44 \pm 2.36$	$29.58 \pm 7.93$
Minimum eccentricity (radius)	$0 \pm 0$	$0.33 \pm 0$	$0.33 \pm 0$	$0.11 \pm 0.19$	$1 \pm 0$	$1 \pm 0$	$4.25 \pm 2.47$
Average eccentricity 90	$1.57 \pm 0.29$	$1.17 \pm 0.25$	$1.37 \pm 0.12$	$2.39 \pm 0.59$	$2.87 \pm 0.48$	$7.57 \pm 1.8$	$16.64 \pm 6.55$
Maximum eccentricity 90	$6.78 \pm 0.67$	$4.78 \pm 0.63$	$5.22 \pm 0.5$	$7.55 \pm 1.32$	$9.11 \pm 0.96$	$14 \pm 2.06$	$27 \pm 7.12$
Minimum eccentricity 90	$0 \pm 0$	$0.33 \pm 0$	$0.33 \pm 0$	$0.11 \pm 0.19$	$1 \pm 0$	$1 \pm 0$	$3.5 \pm 1.89$
Average path length (closeness)	$1.07 \pm 0.17$	$0.85 \pm 0.15$	$0.97 \pm 0.09$	$1.56 \pm 0.34$	$1.82 \pm 0.29$	$4.35 \pm 0.84$	$10.22 \pm 3.81$
Hop plot exponent	$0.33 \pm 0.06$	NAN±	$0.36 \pm 0.03$	$0.49 \pm 0.08$	$0.52 \pm 0.09$	$0.7 \pm 0.08$	$1.15 \pm 0.11$
Effective hop diameter	1.2E10	2.1E15	1.4E6 ± 2E6	1.2E8 ± 1E8	2E4 ± 2.1E4	7E5 ± 7.7E5	46.52 ± 26.71
Giant connected ratio	$0.15 \pm 0.03$	$0.11 \pm 0.02$	$0.11 \pm 0$	$0.11 \pm 0.03$	$0.12 \pm 0.02$	$0.35 \pm 0.06$	$0.68 \pm 0.29$
Number of Connected Components	$45.78 \pm 4.88$	$43.67 \pm 4.59$	$47.33 \pm 4.91$	$54.55 \pm 11.21$	$46.78 \pm 10.36$	$37.56 \pm 7.63$	$19.75 \pm 12.82$
Percentage of isolated points	$0.42 \pm 0.04$	$0.45 \pm 0.06$	$0.39 \pm 0.08$	$0.23 \pm 0.07$	$0.2 \pm 0.06$	$0.16 \pm 0.03$	$0.04 \pm 0.02$
Percentage of end points	$0.28 \pm 0.04$	$0.32 \pm 0.02$	$0.32 \pm 0.02$	$0.34 \pm 0.04$	$0.25 \pm 0.02$	$0.22 \pm 0.04$	$0.1 \pm 0.04$
Number of central points	$31.55 \pm 4.12$	$26.67 \pm 2.9$	$28.22 \pm 3.15$	$31.11 \pm 7.42$	$25 \pm 5.56$	$24.78 \pm 4.25$	$10.84 \pm 6.92$
Percentage of central points	$0.42 \pm 0.04$	$0.41 \pm 0.03$	$0.34 \pm 0.04$	$0.24 \pm 0.06$	$0.18 \pm 0.03$	$0.19 \pm 0.03$	$0.04 \pm 0.02$
Number of nodes	$83.33 \pm 28.87$	$74.33 \pm 30.01$	$85.33 \pm 13.43$	$140.67 \pm 42.19$	$144 \pm 42.51$	$174 \pm 99.8$	$307 \pm 57.03$
Number of edges	$48.75 \pm 10.22$	$40.25 \pm 8.66$	$47.84 \pm 7.61$	$121.75 \pm 22.6$	$146.92 \pm 26.65$	$297.34 \pm 49.57$	$659.33 \pm 106.32$

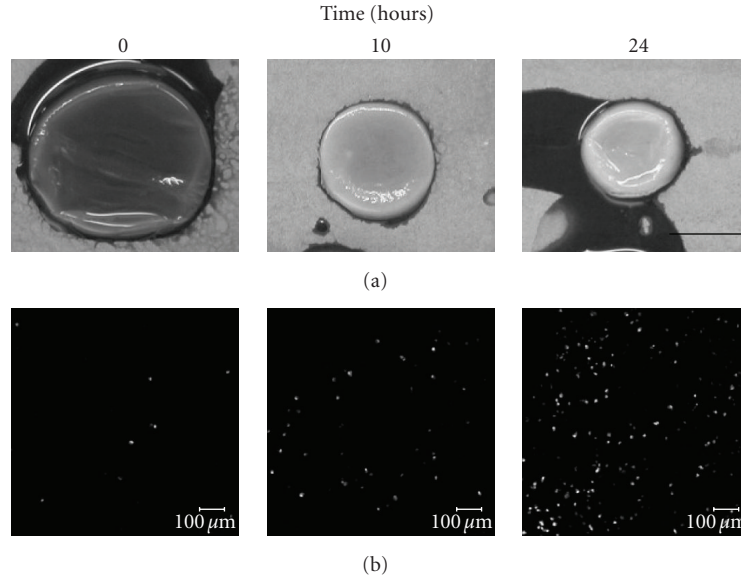


FIGURE 1: Macro- and micro scale compaction occurs over 24 hours of  $3.0 \times 10^5$  cell/mL gels. hMSCs at a cell density of  $3.0 \times 10^5$  cells/mL were embedded in 2 mg/mL collagen I gels. Collagen I gels were fixed and stained for nuclei with SYTOX Green (a) Macroscale images of hydrogel compaction (Scale bar = 1 cm). (b) Fluorescent image cross section of hMSC nuclei in collagen I gel (Scale bar = 100  $\mu$ m).

TABLE 5: Quantitative metrics for  $1 \times 10^6$  cells/mL collagen I hydrogel 24-hour timecourse.

	Time (hours)						
	0	1	2	6	10	16	24
Average degree	12.47	6.30	7.05	11.47	14.70	22.68	50.07
Clustering Coefficient	0.71	0.64	0.67	0.67	0.65	0.62	0.66
Average eccentricity	15.93	10.66	12.90	15.59	13.41	11.68	11.29
Maximum eccentricity (diameter)	21.67	18.33	21.00	21.67	17.33	15.00	14.67
Minimum eccentricity (radius)	1.00	0.67	1.00	8.00	6.33	8.00	7.67
Average eccentricity 90	12.28	8.56	11.08	12.51	10.82	9.35	9.02
Maximum eccentricity 90	19.33	16.33	18.67	18.67	14.67	12.67	12.33
Minimum eccentricity 90	1.00	0.67	1.00	6.33	5.00	7.00	6.00
Average path length (closeness)	7.84	5.23	6.65	7.58	7.00	6.11	5.85
Hop plot exponent	1.33	1.01	1.07	1.24	1.39	1.54	1.52
Effective hop diameter	12.47	35.15	22.77	16.26	10.99	9.44	9.12
Giant connected ratio	0.98	0.65	0.73	0.88	0.99	1.00	1.00
Number of connected components	4.00	14.67	11.33	3.33	3.00	1.00	1.00
Percentage of isolated points	0.00	0.02	0.02	0.00	0.01	0.00	0.00
Percentage of end points	0.01	0.09	0.04	0.01	0.01	0.00	0.00
Number of central points	4.67	10.33	10.33	5.33	9.67	13.33	34.67
Percentage of central points	0.01	0.06	0.05	0.02	0.02	0.02	0.02
Number of nodes	380.33	190.00	218.33	350.33	440.00	743.33	1458.67
Number of edges	2378.67	604.33	769.00	2049.33	3339.00	8454.33	36770.33

The collagen I hydrogels compacted to approximately 15% of their original volume over the 24-hour time course, with the largest change in volume occurring within the first six hours. At each time point the nuclei were imaged by confocal microscopy (10 $\times$ ) to reflect cellular organization within the 3D space of each hydrogel; see Figure 1(b). The number of nuclei captured within the 10 $\times$  field of vision increased over time as the hydrogels compacted.

*3.2. Graph Theory Applied to 3D Confocal Images Generates Distance-Dependent Cell-Graphs.* Cell-graphs, Figure 2(b), were generated from the 3D reconstruction of the fluorescent image z-stacks; see Figure 2(a). The 3D images were segmented, nodes identified, and links formed based upon their Euclidean distance as described in the methods. This method was able to successfully segment the 3D nuclear images and form cell-graphs that demonstrate the relationships between

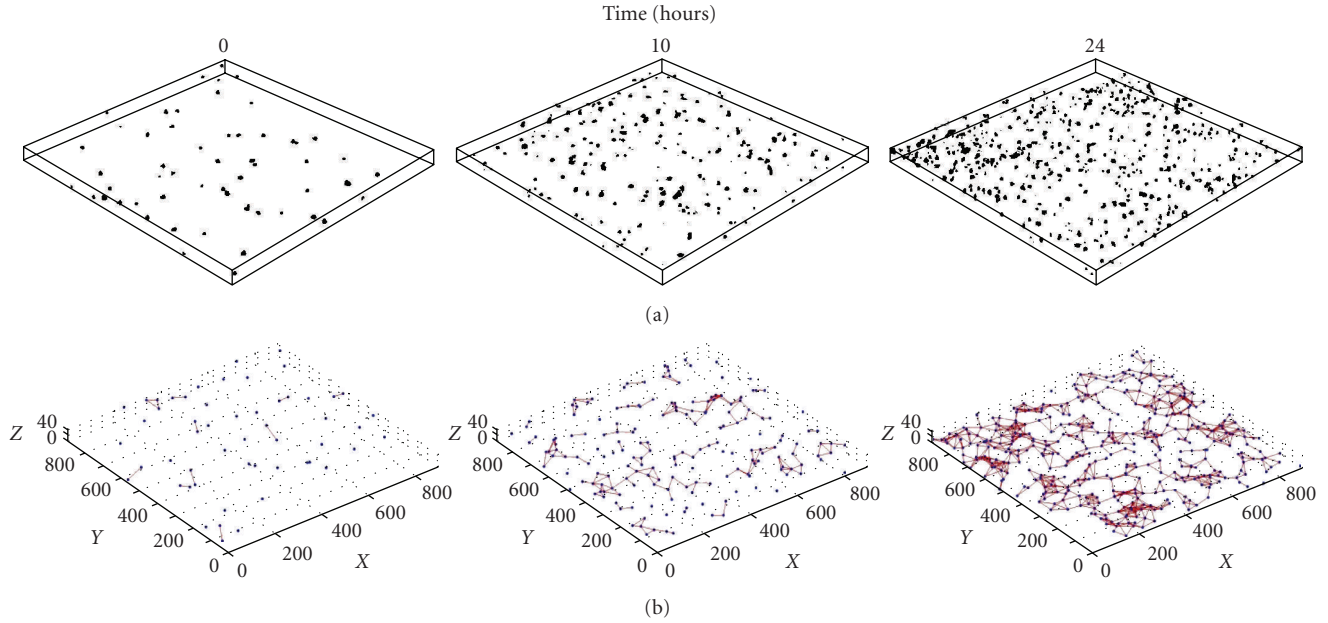


FIGURE 2: Graph theory applied to fluorescent confocal z-stack images to generate distance dependent cell-graphs for  $3.0 \times 10^5$  cell/mL gels. (a) 3D representation of fluorescent image z-stack. (b) 3D cell-graph generated from fluorescent image z-stack.

nodes over time. The number of edges, links, between cell nuclei increased throughout the time course; see Figure 2(b). This is consistent with the increased nuclear density observed in the fluorescent images.

**3.3. Quantitative Metrics Can Be Extracted from Cell-Graphs That Characterize 3D Organization Over Time.** The utility of these cell-graphs comes from the ability to extract meaningful and quantitative metrics to describe the relationships between nodes. The metrics (Table 1) were computed for the generated cell-graph of each z-stack over the time course. These metrics quantitatively provide information as to the 3D orientation and potential interactions of cells in culture; see Table 3. The metrics were dependent on the location of the cells relative to one another. The number of nodes increased with time which represents the number of nuclei and is again consistent with the visual data. The number of edges increased as the hydrogels underwent compaction and the distance between cells decreased. The giant connected ratio reached one by 16 hours in our collagen I hydrogel time course. The percent of isolated points and end point metrics both reached a value of zero by 16 hours. The average degree increased over the 24-hour period while the clustering coefficient had a sharp increase between 1 to 2 hours and leveled off by 10 hours. The diameter peaked at 10 hours and the number of connected components decreased to zero by 16 hours into the time course. Our results are stable for link thresholds between 65 to 75 microns as shown in Table 4.

**3.4. An Increased Cellular Density Increases the Rate of Hydrogel Compaction during the First 24 Hours.** To demonstrate the metrics' sensitivity to perturbations in the biological

system we repeated the experiment using a higher cell concentration ( $1.0 \times 10^6$  cell/mL of ECM). The  $1.0 \times 10^6$  cells/mL gels compacted to less than 10% of their original volume by hour 16; see Figure 3(a). These observations are reflected in the fluorescent images of nuclei over time; see Figure 3(b).

Cell-graphs, Figure 4(b), generated based upon the new 3D images, Figure 4(a), showed an increased connectivity, expressed as visible links at earlier time points. The transition to high connectivity occurred at six hours in the high cell density gels, while the low cell density samples transitioned at ten hours. Quantitative metrics from the high cell density gels are listed in Table 5. In the  $1.0 \times 10^6$  cell/mL gels the number of edges between the nodes increased over time, with the largest increase occurring between 10 and 16 hours. The giant connected ratio approached one beginning at 10 hours. The percent of isolated points and percent of end points reached zero at 10 hours. The average degree underwent the largest increase from 10 to 16 hours and peaked at 16 hours followed by a decrease of approximately 20% by 24 hours. The diameter peaked at 2 hours and decreased by hour 10. The number of connected components also reached one by hour 16.

**3.5. Increasing Cellular Density Leads to a Shift in the Metrics Transition over Time.** The sensitivity of our metrics to gross changes in biological function is a prerequisite to the extension of this method into more complex problems. Quantification of gel compaction revealed a relationship between collagen I compaction and cell density; see Figure 5. The  $1.0 \times 10^6$  cell/mL gels showed significant compaction within the first hour while the  $3.0 \times 10^5$  cell/mL samples did not significantly compact until the second hour ( $P < .05$ ).

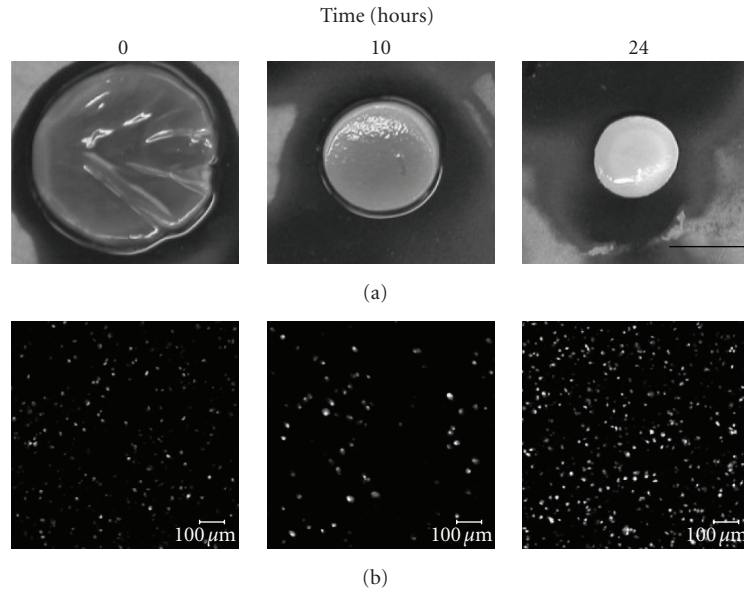


FIGURE 3: Changes in cell density alter compaction and nuclear organization. hMSC at a cell density of  $1 \times 10^6$  cells/mL embedded in 2 mg/mL collagen I gel. Collagen I gels were fixed and then stained for nuclei with SYTOX Green. (a) Macroscale images of hydrogel compaction (Scale bar = 1 cm). (b) Fluorescent image cross section of hMSC nuclei (Scale bar = 100 μm).

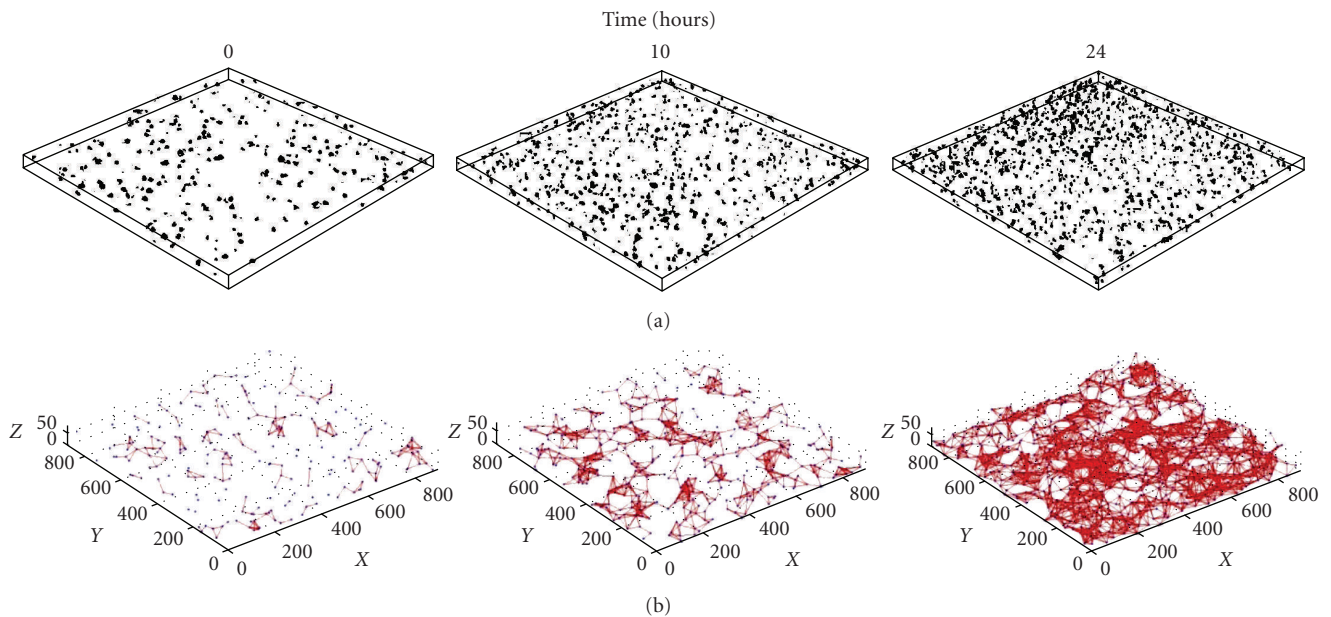


FIGURE 4: Graph theory applied to fluorescent confocal z-stacks from the  $1 \times 10^6$  cells/mL time course to generate distance dependent cell-graphs. (a) 3D representations of fluorescent image z-stacks of  $1 \times 10^6$  cell densities. (b) 3D cell-graph generated from fluorescent image z-stacks of  $1 \times 10^6$  cell density.

After 24 hours of culture the  $1.0 \times 10^6$  cell/mL gels reached a final volume of  $5.27\% \pm 0.76$  of the original while the  $3.0 \times 10^5$  cell/mL gels reached  $11.92\% \pm 2.86$ .

Figure 6 graphically represents the relationship between some of graph features in the two experimental sets. The average degree reached its maximum at an earlier time (16 hours) in the  $1.0 \times 10^6$  cell density gels than in the  $3.0 \times 10^5$  cell density gels (24 hours); see Figure 6(a). The clustering

coefficient also differed; see Figure 6(b). The diameter or maximum eccentricity had a large shift in its peak to an earlier time point (2 hours) in high density gels compared to low-density gels (10 hours); see Figure 6(c). The  $3.0 \times 10^5$  cell/mL graphs had a continuous increase in the number of edges during the 24 hours. The  $1.0 \times 10^6$  cell/mL graphs peak at an earlier time point of 16 hours with a large increase occurring between 10 and 16 hours; see Figure 6(d).



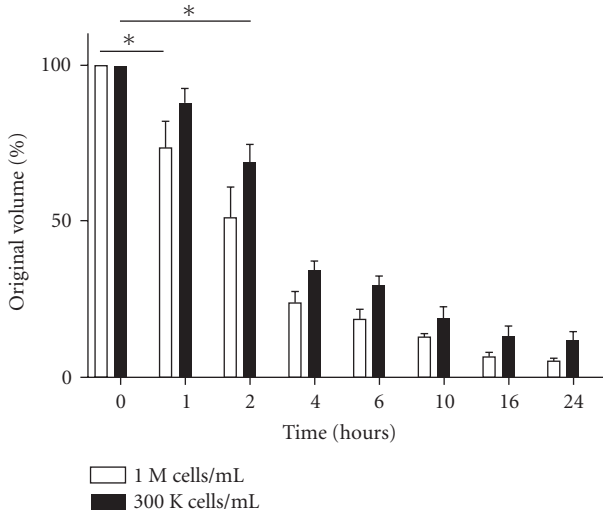


FIGURE 5: Collagen I hydrogels compact as a function of cellular density. The diameter of the collagen I gels at each time point was measured using Image J and plotted as a percentage of the original volume. Changes in cellular density resulted in significant changes in compaction over the observed 24 hours time course, where an increase in cell density correlated with an increase in gel compaction.

The giant connected ratio reached a value of one at an earlier time point when the cell density was increased to  $1.0 \times 10^6$  cell/mL from  $3.0 \times 10^5$  cell/mL; see Figure 6(e). The number of connected components decreased at a faster rate when the cell density was increased to  $1.0 \times 10^6$  cell/mL; see Figure 6(f).

**3.6. Cell-Graph Features Significantly Diverge from Those Extracted from Random Graphs.** To show that tissue has a structure which is different from the random organization of the same number of cells, we established two properties: (i) cells are not randomly scattered in 3D space, and (ii) cells interact in a nonrandom fashion. For the first property, we randomly distributed points representing the cells within 3D space with the same density and geometry seen in the images. The CNND function was implemented explained for both random scattering and for actual cell locations obtained from image segmentation. The results, Figure 7, indicate that cells in the compacting gels were not randomly scattered in the gels. Note that parameter  $t$  is a measure of distance and defines the neighborhood size such that as farther nodes are included to the neighborhood, the difference from random scattering diminishes. Similarly, small values of  $t$  are not informative since neighborhoods become sparse. However, Figure 8 demonstrates that for a wide range of  $t$  values the difference is significant and more importantly such difference is observed for smaller  $t$  values as time progresses and the gels compact.

To show that cell-cell interactions cannot be modeled by randomly picking pairwise interactions between cells, we defined an edge function according to Erdos-Renyi (ER) model and generated random graphs to distinguish their

features from those extracted from our cell-graphs. Consider three cases for generating random graphs: (i) cell locations are chosen randomly and the ER edge function is used, (ii) cell locations, obtained from image segmentation, are taken as node locations and ER edge function is used, and (iii) node locations are randomly chosen but the cell-graph edge function is used. Since case (ii) is subsumed in (i) (recall that ER model takes number of nodes  $n$  and edge probability  $p$  as its only parameters), we only analyzed case (i) and case (iii).

For case (i) random graphs were generated with equal numbers of nodes and edges as found in the cell-graphs at each given time point. We extracted the same features calculated for our cell-graphs from these random graphs. Some of the metrics are defined by the number of nodes and links present in the tissue (i.e., average degree) and therefore must be consistent between the cell and random graphs. The remaining values in our cell-graphs were significantly different from those found in the random graphs, indicating that the cell-graph metrics and random metrics did not come from the same distribution; see Figure 8. For example, the clustering coefficient of the random graphs for  $3.0 \times 10^5$  cell/mL samples fluctuated between 0.02 and 0.03, but gradually increased from 0.35 to 0.62 for the cell-graphs over the 24-hour time period. This difference captures the compaction of the tissue samples as the value gradually increases. Also, the giant connected component of the high density random graphs maintained a value of 1 throughout the 24 hour period, whereas the values for the cell-graphs steadily increased to 1, matching the compaction of the tissue samples. The average eccentricity increased throughout compaction in cell-graphs whereas in random graphs this metric was almost constant.

For case (iii) we also randomly distributed cells in a 3D space with the same density and geometry of the corresponding samples and used the cell-graph edge function to construct the cell-graphs. We compared the graph metrics on these random-cell-graphs with graphs from the 3D image segmentation. The cell-graphs defined a relationship (edge) between the nearest random nodes and the graph features correlated with many of the tissue cell-graph features. However, certain graph metrics such as the graph radius were different. In Figure 9, we plotted the Kolmogorov-Smirnov test for “radius” which indicates the distance of nodes to the central points. The red curve represents tissue cell-graphs while the blue curve corresponds to the random-cell-graphs.

**3.7. Cell-Graphs Outperform Voronoi Graphs in Modeling Gel Compaction.** In order to determine the relevance of our approach we assessed the performance of our cell graph method to the more traditional Delaunay triangulation model. The tissue was partitioned into Voronoi cells and the corresponding Delaunay graph was generated for each set of images. Graph features were extracted from these Delaunay graphs as shown in Table 6. These results suggested that Delaunay graphs did not exhibit the patterns that we observed with the cell-graph technique. For example, the

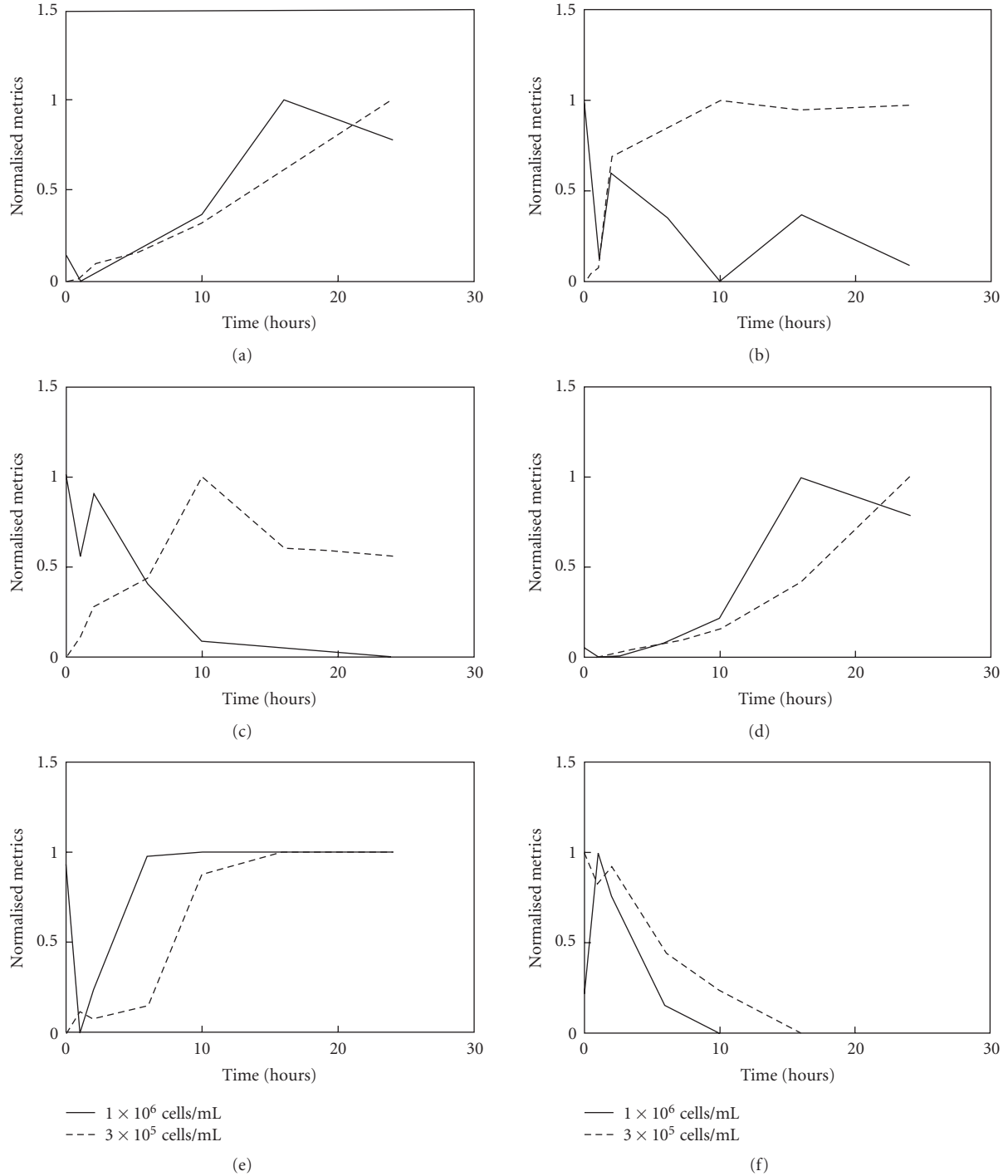


FIGURE 6: Cell-graph mining extracts metrics which represent the topological changes in 3D cell culture as a function of cell density. The effect of change in cell density from  $3.0 \times 10^5$  to  $1.0 \times 10^6$  on our graph features was plotted. (a) Average degree. (b) Clustering coefficient. (c) Diameter. (d) Number of edges. (e) Giant connected ratio. (f) Number of connected components over the 24-hour time course. The extracted metrics have been normalized from  $[0, 1]$ .

average degree of the Delaunay graphs over time deviated between 13.17 and 14.33 and average clustering coefficient varied between 0.45 and 0.49. The giant connected component ratio was 1 at all times and the diameter of the network stayed almost constant, deviating between 5.2 and 7.

**3.8. Significant Subgraphs Extracted from Cell-Graphs Determine Local Structural Motifs.** The features extracted from the tissue cell-graphs were representative of the global (average) structure of our “tissues.” To extract meaningful local structures and motifs we used subgraph mining techniques

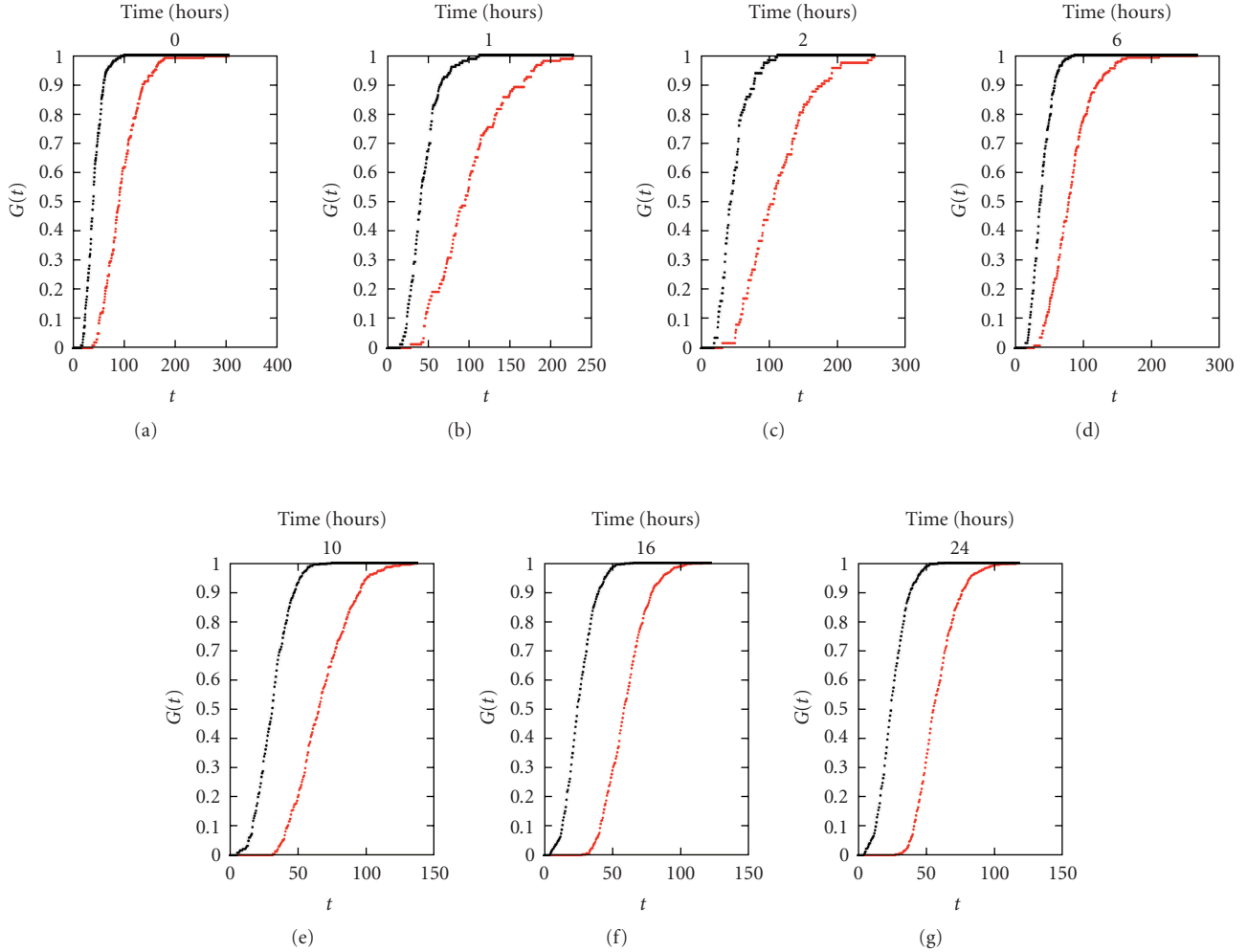


FIGURE 7: Point Pattern Analysis indicates that cells do not pick random locations. In order to show that cell locations obtained from segmented images are different from random scattering, we implemented CNND function  $G$ . The function  $G$  is plotted such that  $G(t)$  represents the fraction of the cell counts for which their nearest neighbor falls within the threshold distance  $t$ . The  $x$ -axis is the threshold value  $t$  in the distance unit and the  $y$ -axis is the value of  $G(t)$ . As the threshold  $t$  increases, more and more cells have their nearest neighbor within the threshold. So, this function value always increases from 0 and finally reaches 1. The black line is a random distribution and the red line is the location of the segmented, experimental nuclei. Note that the difference between random scattering and actual cell locations becomes clearer for smaller and smaller  $t$  values as the gels compact over time.

that identify patterns of node interaction and measure their frequency of occurrence over time. Thousands of subgraph patterns were generated that represent repeated local motifs. The frequency of common subgraphs was also measured in random graphs, to eliminate the possibility that the local structures extracted were simply structures present in any random organization of nodes and links. Distinguishable patterns were extracted that existed only in the cell-graphs, Figure 10(a), or existed only in the random graphs, Figure 10(b), indicating a significant difference in the underlying topography of the cell-graphs from that of the random graphs. The distribution of the differences between the frequencies at which the patterns were found in the random and cell-graphs is shown in Figure 10(c). The distribution of the histogram to the extremes indicated a significant

divergence from the patterns extracted from random graphs.

**3.9. Subgraph Mining Reveals Temporal Changes in Graph Patterning.** The appearance of specific subgraphs was observed at distinct time points and cellular densities. To reveal the time evolution of the significant subgraphs we calculated the number of patterns that first appeared in the  $3.0 \times 10^5$  cell/mL graphs at each time point as a function of their frequency at a given time point in the  $1.0 \times 10^6$  cell/mL graphs; see Table 7. Consistently, the first appearance of certain patterns at late time points in the  $3.0 \times 10^5$  cell/mL time series correlated with patterns frequently found at early time points in the  $1.0 \times 10^6$  cell/mL series. Some time points showed a very strong correlation, with a high number of

TABLE 6: Quantitative delaunay graph metrics for  $1 \times 10^6$  cells/mL collagen I hydrogel 24-hour timecourse.

	Time (hours)						
	0	1	2	6	10	16	24
Average degree	13.55	13.24	13.17	13.9	13.93	14.3	14.33
Clustering Coefficient	0.48	0.49	0.49	0.47	0.47	0.45	0.45
Average eccentricity	4.7	4.4	4.45	5.23	5.44	5.92	5.85
Maximum eccentricity (diameter)	5.8	5.4	5.2	6.4	6.8	7	7
Minimum eccentricity (radius)	0	0	0	0	0	0	0
Average eccentricity 90	4.04	3.85	3.88	4.45	4.71	5.03	5.01
Maximum eccentricity 90	5	4.6	4.6	5.4	5.8	6	6
Minimum eccentricity 90	0	0	0	0	0	0	0
Average path length (closeness)	3.01	2.85	2.89	3.38	3.58	3.9	3.93
Hop plot exponent	2.03	1.83	1.91	2.16	2.26	2.46	2.54
Effective hop diameter	3.94	4.02	3.91	4.43	4.71	4.84	4.83
Giant connected ratio	1	1	1	1	1	1	1
Number of connected components	2	2	2	2	2	2	2
Percentage of isolated points	0	0.01	0.01	0	0	0	0
Percentage of end points	0	0	0	0	0	0	0
Number of central points	1	1	1	1	1	1	1
Percentage of central points	0	0.01	0.01	0	0	0	0
Number of nodes	239.8	204	205	396	493.6	748.2	849.6
Number of edges	1628.6	1375.6	1371.2	2770	3438.4	5352.2	6100.8

TABLE 7: Time evolution of subgraphs.

		$1.0 \times 10^6$ cells/mL						
time (hrs)		0	1	2	6	10	16	24
$3.0 \times 10^5$ cells/mL	0	7	1	0	0	0	0	0
	1	3	0	0	0	0	0	0
	2	9	0	0	18	3	0	0
	6	26	0	10	5	42	0	0
	10	122	3	0	19	76	1	36
	16	143	2	9	46	<b>1868</b>	1303	1274
	24	8	0	0	0	457	842	515

similar patterns, specifically time 10 hours in the  $1.0 \times 10^6$  series and time 16 hours in the  $3.0 \times 10^5$  series (bold value in Table 7). Additionally, a transition was observed after this time point, indicated by the presence of  $3.0 \times 10^5$  subgraphs that appeared before, after and during respective  $1.0 \times 10^6$  subgraphs.

#### 4. Discussion

The extraction of nonrandom repeating motifs in biological structures presents a significant opportunity to define organizing design principles and use those principles to inform and accelerate natural and synthetic processes of tissue development. Here we present the spatial and temporal analysis of a simple biological observation, the compaction of collagen I hydrogels by hMSC. The progression of this event over time was dependent upon the cellular density of the starting material and represents a direct, cell-mediated

organizational event that can be observed through time and space. We observed that an increase in cell density caused an accelerated compaction event that resulted in a decrease in the final volume as compared to the original samples. We tracked this organization event using confocal microscopy and employed graph theory as a method to extract the 3D topography of our “tissue constructs” over the time course of the compaction event. As the hMSCs interact with and organize the naïve collagen matrix, we hypothesized that structural motifs would appear frequently and be evident at earlier time points in the dense samples. This predicts that cells interact with one another and their extracellular environment in nonrandom, defined patterns that we can quantify, and that these patterns are consistent in terms of their temporal organization.

To address this hypothesis we constructed cell-graphs from our 3D micrographs and extracted topographical features that assess the global 3D structure of the “tissue constructs” as they organize over time. Such techniques have



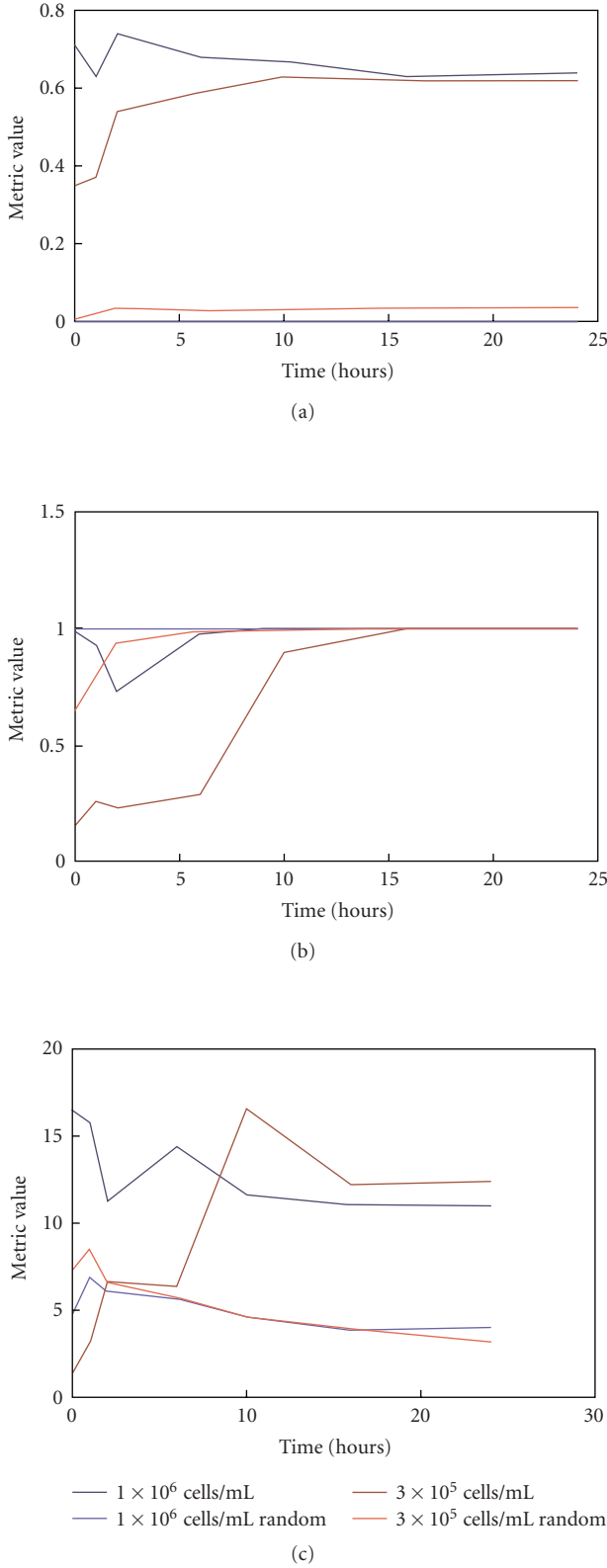


FIGURE 8: Cell-graph features are distinct from random graphs. Random graphs were generated keeping the number of nodes and links constant. Features were extracted from the random graphs and compared to features from cell-graphs to determine significance. (a) Clustering Coefficient. (b) Giant Connected Component. (c) Average Eccentricity.

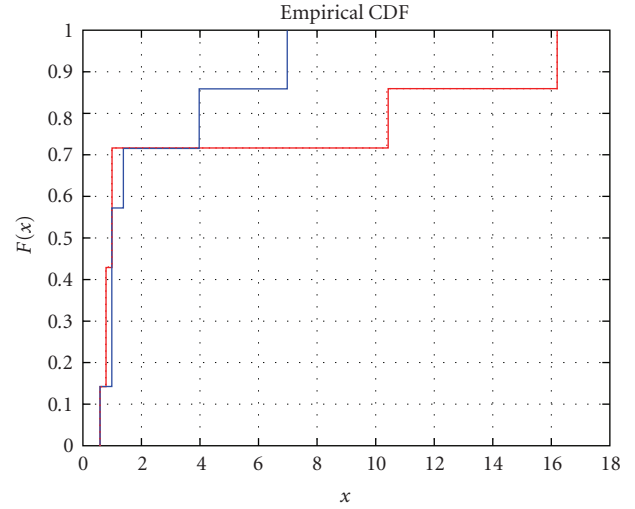


FIGURE 9: Kolmogorov-Smirnov test for graph radius. The plot (although the sample size is small) shows the difference between a cell graph with randomly scattered nodes and one with segmented nodes.

previously been reported to extract meaningful information concerning the (dys)functional state of 2D tissue and can play a significant role in diagnosis of cancer pathologies [21–24]. In addition, past and current work on the nature of complex biological networks has consistently benefited from the use of graph theory as a means of extracted underlying relationships between system components [1, 2, 5, 6, 8–11, 34, 35]. We employed this technique here to define distance-based 3D graphs of cell nuclei and found that significant trends do exist in the extracted features over time and that these features represent the connectivity, modularity, and relationships that exist within our tissue. We conclude that the cells within our tissue exist in dense communities that steadily converge over time and that each cell has a high interaction with an increasing number of neighbors as the tissues organize and compact. In addition, a decrease in cellular density resulted in a delay or shift in these feature trends that is consistent with visual observation of construct compaction over time.

The use of random graphs as a negative control was critical to ensure the extraction of meaningful and nonrandom metrics. The divergence of features extracted from our cell-graphs from those found in random graphs indicates that there exists structure within our “tissue” that is different from a random organization of the same number of cells and links. This is consistent with the theory that random networks do not accurately describe real networks, that networks are highly ordered structures, and that keys to their function lie within that inherent structure [36]. In fact, as the density of the experimental samples increases over time, the difference between the random graphs and cell-graphs becomes more and more remarkable. In a random graph at a density of  $1.0 \times 10^6$  cell/mL the graph is connected, and as the density increases, the diameter decreases as more links are added to the graph. Within our cell-graphs, however, at this

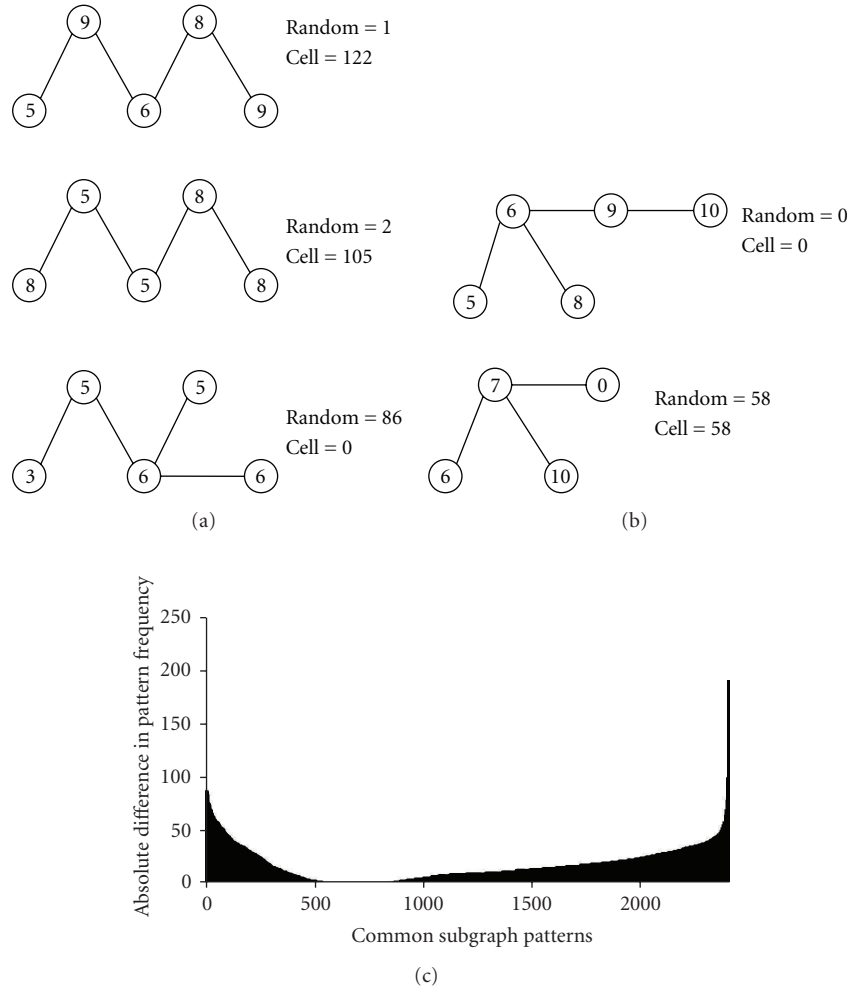


FIGURE 10: Common cell subgraph patterns are distinct from patterns extracted from random graphs. Subgraph patterns were identified that represented local graph topologies and compared to their frequency in random graphs. (a) Significant cell-graph patterns not found in random graphs. (b) Patterns found equally in random and cell-graphs and are therefore not significant. (c) Histogram of all patterns common among cell-graphs plotted as the absolute difference between frequencies in random versus cell-graphs. Patterns at the extremes indicate significant patterns.

same cell density, we see a steadily increasing diameter. This increase in diameter is due to the fact that the connectivity at time zero is sufficiently low to observe increases as the gels condense over time, unlike randomly generated graphs, and therefore the diameter of the entire graph responds by growing as the giant connected component size grows as well. The random graph is connected as a function of the number of nodes and links incorporated into its generation, but there is no true structural relationship evident, as there is no change observed in eccentricity or clustering coefficient. Instead what we see in our cell-graphs is a steadily emerging structural relationship between nodes that is quantified by changes in eccentricity and the clustering of the nodes.

We also used subgraph mining to extract a finer picture of the local structure present within our cell-graphs. Through a multilevel analysis of the cell-graphs we can project a quantitative, sensitive, and rigorous fine structure on the global features extracted above. Searching the micro level

space of the graph we find local structures that are not represented equally over the whole of the graph. In this way, functional components of the tissue, or modules, are identified. For example, a sample in which only a portion represents cancerous tissue might be averaged out and classified as inflamed or normal. However, through the use of subgraph mining, and local feature analysis, an area of densely connected cells would be identified as a specific tissue compartment with unique properties and lead to reclassification as cancer [23, 24, 37]. We identified subgraphs or motifs from the cell-graphs that were significantly different from those that were frequently extracted from random graphs. Again, this indicates that there is an inherent structure that is specific to our cell-graphs and is not characterized by a random distribution of nodes in space.

Over time these subgraphs became increasingly complex and incorporated nodes with higher degrees and therefore

more interaction with neighboring cells, consistent with the conclusion from the global features. In addition, the frequent subgraphs appeared at specific points in time, and not all subgraphs were present at all time points for all densities examined in this study. We hypothesize that the formation of distinct local structures occurs in a specific temporal order. The time evolution of our subgraphs indicates that there is a “route” that a population of cells takes while interacting with and organizing a naïve matrix, and if we change a fundamental property of that population, then we change the rate at which the “tissue” is able to complete that “route.” This hypothesis was confirmed by quantifying the frequency of subgraph occurrence as a function of time. When compared across the two densities, we observed that the appearance of new subgraphs in the less dense sample at late time points consistently correlated with high frequencies of those same patterns at earlier time points in the dense sample. We conclude that a population of cells follows a consistent organizing pattern over time, and while the rate of that organization can be altered by changes in the cell population, the route or order in which that organization takes place remains constant.

Point pattern analysis and its variants [31, 38] focus on spatial analysis of points (cells in our case), whereas to define the “tissue structure” one needs to model cell-cell (point-to-point) relationships. The most relevant work to the cell-graph approach is a Delaunay graph of cells obtained from a processed tissue image to compute graph-theoretical features for tissue quantification [28–30]. In this approach, nuclei are the vertices, the Voronoi diagram of the image is constituted, and its Delaunay triangulation is built. The Delaunay triangulation allows the existence of edges between only the adjacent vertices, such that only the relationships between closely located nuclei are represented. There are two main limitations of Delaunay graphs that cell-graphs successfully remedy. First, Delaunay graphs are restricted to planar graphs which are very limited in their structure and do not allow crossing of edges. There is no evidence to justify such a limitation in 3D tissue structural organization. Second, a Delaunay graph has a single connected component (i.e., the tissue is represented by a connected graph) which may not be a valid assumption for sparse tissues. Third, the girth of a Delaunay graph is always a triangle which is also a special case. Thus, the cell-graphs are a generalization of Delaunay graphs to arbitrary edge functions that can provide formulation of different cell-cell interactions and eliminate such constraints. Furthermore, our previous work [39] reports on a comprehensive comparison between graph-based metrics and spatial distribution-based approaches, showing that the former has greater predictive power when coupled to machine learning approaches.

Therefore, there are several advantages of our cell-graph approach to tissue modeling: First, cell-graphs enable us to benefit from well established principles of graph theory and provide a rich set of features defined precisely by these principles to be used as quantitative descriptor features. These features could be defined and computed locally from a single node’s point of view (e.g., number of its neighbors) or globally for the entire tissue sample (i.e., the shortest or

longest distance in the cell-graph between any two nodes). Second, cell-graphs benefit from sophisticated segmentation algorithms that can provide cell level attributes such as convexity, size, physical contact, and shape. to establish links between a pair of nodes. Furthermore, molecular details of a particular cell type (e.g., is it diseased?) do not need to be resolved, and a specific textural change in the image is not required. Third, cell-graphs provide a precise mathematical representation of cellular organization and the ECM that surrounds cells. If the images carry multichannel information by applying more sophisticated staining techniques (e.g., multispectral fluorescence imaging), it is possible to build cell-graphs that have different types of nodes corresponding to different types of cells that coexist (e.g., epithelial versus fibroblast) and other ECM entities (e.g., basement membrane underlying epithelial cell layers and blood vessels). With 3D images and 3D cell-graphs, such representation becomes more accurate and powerful.

## 5. Conclusions

The demonstration of higher order in biological structures is evidence of recurring principles that govern the design and therefore development of metabolic networks, protein-interactions, neural circuitry, and tissue. Cells interact with one another and their extracellular environment in non-random, defined patterns, and these patterns are consistent in terms of their spatial and temporal organization. Understanding the temporal order of this patterning provides organizing principles that can be used to help biologists and tissue engineers understand or optimize processes of natural and synthetic tissue development.

## Acknowledgments

This work was supported in part by the National Institutes of Health, Grant no. RO1 EB008016 and no. RO1 AR053231. The first and the second authors contributed equally to this work.

## References

- [1] A. L. Barabasi and Z. N. Oltvai, “Network biology: understanding the cell’s functional organization,” *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101–113, 2004.
- [2] A. Ma’ayan, “Insights into the organization of biochemical regulatory networks using graph theory analyses,” *The Journal of Biological Chemistry*, vol. 284, no. 9, pp. 5451–5455, 2009.
- [3] E. Pieroni, S. de La Fuente van Bentem, G. Mancosu, E. Capobianco, H. Hirt, and A. de La Fuente, “Protein networking: insights into global functional organization of proteomes,” *Proteomics*, vol. 8, no. 4, pp. 799–816, 2008.
- [4] J. Park and A. L. Barabasi, “Distribution of node characteristics in complex networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 46, pp. 17916–17920, 2007.
- [5] A. L. Barabasi, “Network medicine—From obesity to the “diseasome,”” *The New England Journal of Medicine*, vol. 357, no. 4, pp. 404–407, 2007.

- [6] D. S. Lee, J. Park, K. A. Kay, N. A. Christakis, Z. N. Oltvai, and A. L. Barabasi, "The implications of human metabolic network topology for disease comorbidity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 29, pp. 9880–9885, 2008.
- [7] M. D. Fricker, J. A. Lee, D. P. Bebbler, et al., "Imaging complex nutrient dynamics in mycelial networks," *Journal of Microscopy*, vol. 231, no. 2, pp. 317–331, 2008.
- [8] A. Rawat and Y. Deng, "Novel implementation of conditional co-regulation by graph theory to derive co-expressed genes from microarray data," *BMC Bioinformatics*, vol. 9, supplement 9, p. S7, 2008.
- [9] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, and C. J. Stam, "The application of graph theoretical analysis to complex networks in the brain," *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2317–2331, 2007.
- [10] C. J. Stam, W. de Haan, A. Daffertshofer, et al., "Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease," *Brain*, vol. 132, no. 1, pp. 213–224, 2009.
- [11] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, "Age-related changes in modular organization of human brain functional networks," *NeuroImage*, vol. 44, no. 3, pp. 715–723, 2009.
- [12] J. T. Erler and V. M. Weaver, "Three-dimensional context regulation of metastasis," *Clinical and Experimental Metastasis*, vol. 26, no. 1, pp. 35–49, 2008.
- [13] M. J. Bissell, D. C. Radisky, A. Rizki, V. M. Weaver, and O. W. Petersen, "The organizing principle: microenvironmental influences in the normal and malignant breast," *Differentiation*, vol. 70, no. 9–10, pp. 537–546, 2002.
- [14] C. M. Nelson, D. Khauv, M. J. Bissell, and D. C. Radisky, "Change in cell shape is required for matrix metalloproteinase-induced epithelial-mesenchymal transition of mammary epithelial cells," *Journal of Cellular Biochemistry*, vol. 105, no. 1, pp. 25–33, 2008.
- [15] C. M. Nelson and M. J. Bissell, "Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer," *Annual Review of Cell and Developmental Biology*, vol. 22, pp. 287–309, 2006.
- [16] C. M. Nelson, M. M. VanDuijn, J. L. Inman, D. A. Fletcher, and M. J. Bissell, "Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures," *Science*, vol. 314, no. 5797, pp. 298–300, 2006.
- [17] M. J. Paszek, N. Zahir, K. R. Johnson, et al., "Tensional homeostasis and the malignant phenotype," *Cancer Cell*, vol. 8, no. 3, pp. 241–254, 2005.
- [18] P. A. Kenny, G. Y. Lee, C. A. Myers, et al., "The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression," *Molecular Oncology*, vol. 1, no. 1, pp. 84–96, 2007.
- [19] A. W. Lund, J. P. Stegmann, and G. E. Plopper, "Inhibition of ERK promotes collagen gel compaction and fibrillogenesis to amplify the osteogenesis of human mesenchymal stem cells in three-dimensional collagen I culture," *Stem Cells and Development*, vol. 18, no. 2, pp. 331–341, 2009.
- [20] K. J. Martin, D. R. Patrick, M. J. Bissell, and M. V. Fournier, "Prognostic breast cancer signature identified from 3D culture model accurately predicts clinical outcome across independent datasets," *PLoS ONE*, vol. 3, no. 8, article e2994, 2008.
- [21] C. Demir, S. H. Gultekin, and B. Yener, "Learning the topological properties of brain tumors," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 3, pp. 262–269, 2005.
- [22] C. Demir, S. H. Gultekin, and B. Yener, "Augmented cell-graphs for automated cancer diagnosis," *Bioinformatics*, vol. 21, supplement 2, pp. ii7–ii12, 2005.
- [23] C. Gunduz, B. Yener, and S. H. Gultekin, "The cell graphs of cancer," *Bioinformatics*, vol. 20, supplement 1, p. i145, 2004.
- [24] C. Bilgin, C. Demir, C. Nagi, and B. Yener, "Cell-graph mining for breast tissue modeling and classification," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2007, pp. 5311–5314, 2007.
- [25] P. N. Klein and N. E. Young, "Approximation algorithms for NP-hard problems," in *Proceedings of the 37th IEEE Annual Symposium on Foundations of Computer Science*, vol. 1, Burlington, NC, USA, July 1996.
- [26] S. J. Eglen, D. D. Lofgreen, M. A. Raven, and B. E. Reese, "Analysis of spatial relationships in three dimensions: tools for the study of nerve cell patterning," *BMC Neuroscience*, vol. 9, no. 68, 2008.
- [27] P. Erdos and A. Reny, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, p. 290, 1959.
- [28] H. K. Choi, T. Jarkrans, E. Bengtsson, et al., "Image analysis based grading of bladder carcinoma. Comparison of object, texture and graph based methods and their reproducibility," *Analytical Cellular Pathology*, vol. 15, no. 1, pp. 1–18, 1997.
- [29] S. J. Keenan, J. Diamond, W. G. McCluggage, et al., "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)," *Journal of Pathology*, vol. 192, no. 3, pp. 351–362, 2000.
- [30] B. Weyn, G. van de Wouwer, S. Kumar-Singh, et al., "Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis," *Cytometry*, vol. 35, no. 1, pp. 23–29, 1999.
- [31] S. J. Eglen, D. D. Lofgreen, M. A. Raven, and B. E. Reese, "Analysis of spatial relationships in three dimensions: tools for the study of nerve cell patterning," *BMC Neuroscience*, vol. 9, article 68, 2008.
- [32] V. Chaoji, M. A. Hasan, S. Salem, and M. J. Zaki, "An integrated, generic approach to pattern mining: data mining template library," *Journal of Data Mining and Knowledge Discovery*, vol. 17, no. 3, pp. 457–495, 2008.
- [33] J. R. Ullmann, "An algorithm for subgraphisomorphism," *Journal of the Association for Computing Machinery*, vol. 23, no. 1, pp. 31–42, 1976.
- [34] P. Braun, E. Rietman, and M. Vidal, "Networking metabolites and diseases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 29, pp. 9849–9850, 2008.
- [35] C. Kaleta, F. Centler, P. di Fenizio, and P. Dittrich, "Phenotype prediction in regulated metabolic networks," *BMC Systems Biology*, vol. 2, article 37, 2008.
- [36] N. Gulbahce and S. Lehmann, "The art of community detection," *BioEssays*, vol. 30, no. 10, pp. 934–938, 2008.
- [37] C. Demir, S. H. Gultekin, and B. Yener, "Augmented cell-graphs for automated cancer diagnosis," *Bioinformatics*, vol. 21, supplement 2, pp. ii7–ii12, 2005.
- [38] B. N. Boots and A. Getis, *Point Pattern Analysis*, vol. 93, SAGE, Newbury Park, NJ, USA, 1988.
- [39] C. Demir, S. H. Gultekin, and B. Yener, "Learning the topological properties of brain tumors," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 3, pp. 262–269, 2005.



