Cell Differentiation Processes as Spatial Networks: identifying four-dimensional structure in embryogenesis

Bradly Alicea[1,2], Richard Gordon[3,4]
[1] Orthogonal Research Lab, Champaign, IL USA [2] OpenWorm Foundation, Boston, MA USA [3] Gulf Specimen Marine Laboratory and Aquarium, Panacea, FL USA [4] C.S. Mott Center for Human Growth & Development, Department of Obstetrics & Gynecology, Wayne State University, Detroit, MI USA

KEYWORDS: Complex Networks, Embryogenesis, Computational Biology

ABSTRACT

There are two basic forms of animal development: mosaic (as found in roundworms and sea squirts) and regulative (as found in amphibians and mammals). How might one distinguish between each type of development? Using a four-dimensional spatial representation (*x,y,z,t*), major features of the developmental process are revealed. To establish the role of mosaic mechanisms, we can map the cell division process to a computational representation of *C. elegans* embryogenesis using a directed, acyclic graph (DAG) and a differentiation code. Mosaic development is identifiable by observing spatial localization of progenitor and descendent cells. This three-dimensional compartmentalization should be consistent with nesting in the cell lineage tree. Regulative development should demonstrate spatial "smearing", or deviations from the compartmentalization of the mosaic process. In a complementary manner, complex network statistics should confirm this by providing an approximation of embryo geometry. Characterizing the spatial organization and geometry of embryos in this way allows for heuristic indicators of developmental patterns both within and between organisms.

INTRODUCTION

There are two basic forms of animal development: mosaic (as found in roundworms and sea squirts) and regulative (as found in amphibians and mammals). How might one distinguish between each type of development? Using a four-dimensional spatial representation (*x,y,z,t*), and quantitative analyses of the data, major features of the developmental process are revealed. These major features include the regularity of cellular movements and geometric properties of the differentiation process [1-3].

Using a directed, acyclic graph (DAG) and a differentiation code, we can map the cell division process to a computational representation of embryogenesis. The network topology of a lineage tree [4] and spatial location of cell nuclei [5] allow for the prediction of major features that constitute each type of developmental process in a model organism (the nematode *C. elegans*). Mosaic development is identifiable by observing spatial localization of progenitor cells and their descendent cells in a given subnetwork. This three-dimensional compartmentalization should be consistent with nesting in the cell lineage tree. Conversely, regulative development should demonstrate spatial smearing that is indicative of deviations from the compartmentalization observed in the mosaic process.

The smearing or cell mixing signature can be detected in two ways. The first way is to test for the presence of intercalation in the early embryo [6]. Intercalation occurs as groups of cells shift their geometry and orientation in preparation for tissue induction. In our three-dimensional embryo space, changes in the variance along one dimension are accounted for by changes along another dimension. The other means of detection involves examining the distance between mother

and daughter cell nuclei. While mosaic embryo cells are known to engage in cell focusing (or changes in position) during the differentiation process [7], the frequency of longer to shorter migrations are of particular interest. When the cell migrations in a specific group of cells are uniform, we hypothesize that the likelihood of autonomous specification is high. By contrast, a few long-distance migrations in a background of shorter migrations suggests that there are regulative mechanisms at work.

We can also employ complex networks to generate degree distributions and connectivity patterns in both lineage and physical space based on spatial location and the differentiation code. In mosaic organisms each cell may be a one-cell tissue [1, 3], whereas in regulating embryos each tissue has many cells, that, however, tend not to mix across tissue boundaries [8]. Cell mixing varies considerably between regulating embryos of different species [9]. However, if we disregard intra-tissue cell mixing, these regulating embryos may be essentially the same as mosaic embryos, as is suggested by their coarser grained fate maps [10]. Verification is possible by checking if adjacent pairs of single-cell tissues in mosaic animals are homologous with corresponding adjacent pairs of multiple-cell tissues in regulating embryos.

Diffusion in the early embryo can be informative in terms of how cells move and ultimately position themselves for differentiation during embryogenesis [11]. Thus, further confirmation of regulative developmental mechanisms will exist through the discovery of superdiffusive signatures [12]. This form of random walk behavior may allow us to distinguish between smearing due to regulative cues themselves and cell migrations after differentiation. While the existence of nested sets in the spatial domain do not completely predict mosaic development, the lack of long-distance migration and/or induced fates provides us with a heuristic indicator of differences in developmental types.

In this paper, we will examine the early (pre-200 minute) embryo at three stages: 4 level (16-cell condition), 6 level (64-cell condition), and 7 level (128-cell condition). Four complementary analyses (visualization, examination and classification of subtree variation, analysis of cell intercalation, and aggregate cell migration analysis) are conducted that will lead up to a bidirectional network analysis. These complex networks will provide a window into the quantitative relationships between spatial structure and cellular differentiation over time.

METHODS

Distance Metric

To calculate a relational distance between three-dimensional cell positions, a Euclidean distance metric (D) was used. This distance was then normalized in the form of an index (d_i), which takes all distances as a proportion of the maximum distance between cells in the embryo. These metrics are defined mathematically as

$$D = \sqrt{x^2} + y^2 + z^2 \quad \textbf{[1]}$$

$$d_i = \left(\frac{D_i}{\max(D_i)}\right) \quad \textbf{[2]}$$

A symmetric matrix indexed relational distances is then used to construct an unordered network of all nodes in the lineage tree. This matrix can then be partitioned by subtree or level.

Bipartite and Octopartite Subtrees

The bipartite lineage tree can be defined as a tree split into two non-reticulating subtrees, originating at cell AB and cell P1. The octopartite lineage tree can be defined as a tree split into eight non-reticulating subtrees, originating at cells ABpl, ABpr, ABar, ABal, MS, E, C, and P3.

3-D Cell Positions

Three-dimensional cell positions were obtained from primary datasets [13, 14]. Data collected from the nuclear positions of 261 embryos. Cell positions were registered using the lineage tracking software Starry Night, and nuclei were marked using a GFP+ signal, Distances between cells calculated using a Euclidean distance and defined as "embryo units (AU)". Full details on the dataset can be found in [15].

3-Dimensional Graphs and Network Visualization

All 3-dimensional graphs are created using MATLAB. The associated code can be found in a project-related Jupyter Notebook, located at Figshare, doi:10.6084/m9.figshare.4667848. All network visualizations and associated graph analyses are produced using Gephi 0.9.0. Circular graphs are produced by using the Circular Layout plug-in for Gephi (https://marketplace.gephi.org/plugin/circular-layout/).

Clique Analysis

All clique analyses and associated network statistics are generated using Gephi 0.9.0. The clique analysis is conducted using the 7 level dataset. Network cliques are subsets of nodes in an undirected network that produces a completely connected subgraph of a pre-determined size. A clique size of 5 was chosen through a pre-evaluation of the data using clique sizes ranging from 3 to 8, and a size of 5 yielded the largest number or cliques (n=117).

Data

All raw data used in this study are available at Github. Lineage Tree Raw Data: https://github.com/balicea/DevoWorm/tree/master/Lineage%20Tree%20DB, Differentiation Tree Raw Data: https://github.com/balicea/DevoWorm/tree/master/Differentiation%20Tree%20Dataset. All processed data (used in the analyses) are available from the Open Science Framework: https://osf.io/q9jvb/

RESULTS

To test these hypotheses, we extracted differentiation codes from the lineage tree at 4 level, 6 level, and 7 level stages. This provided us with a sample size of 30, 114, and 230, respectively. In the 16-cell condition, we compare the components of a bipartite lineage tree originating at the 2-cell embryo, one rooted at AB and the other rooted at P1. In the 64- and 128-cell condition, we compare components of a bipartite lineage tree originating at each cell of the 8-cell embryo.

Five analyses were conducted to demonstrate the mosaic developmental signal in the early stages of a *C. elegans* embryo. The first is to visualize the three-dimensional position of cell nuclei according to their differentiation code. This code corresponds to *n*-partite components (discrete

subtrees) of a lineage tree. While the 4 level condition remained a bipartite tree, the 64 and 7 level conditions were characterized as octopartite trees. These visualizations demonstrate separability between subtrees. To quantify and further investigate this spatial segregation, we use a k-means classification method and compare these categories with the differentiation code classification.

To better understand the resulting overlap in categories, we use an intercalation analysis to find signatures of biological intercalation. While biological intercalation commonly occurs in mosaic development, it can often mimic our criterion of regulative development. Another phenomenon that can be confused with regulative development are ballistic cell migrations. To test for cell migrations, we use an exponential cell migration analysis to detect quasi-stochastic cell migrations that occur according to an exponential distribution.

Finally, we use a connectivity measure to build an undirected network of relationships between cells. The structure of this network will provide information about interactions between cells. Clustering within this network topology will demonstrate local autonomous specification, while random structure or scale-free connectivity might indicate a more regulative process. Taken collectively, these analyses will allow us to distinguish between deterministically specified cell fates and cell fates that result from regulative influences.

Visualization of Bipartite Differentiation Trees

The first analysis is to demonstrate that dividing the lineage tree up in a bipartite manner reveals spatial segregation amongst cells in each subtree. The spatial locations of cells by these subtrees is shown in Figures 1, 2, and 3. Figure 1 shows cells from the 4 level tree, Figure 2 from the 6 level tree, and Figure 3 from the 7 level tree.

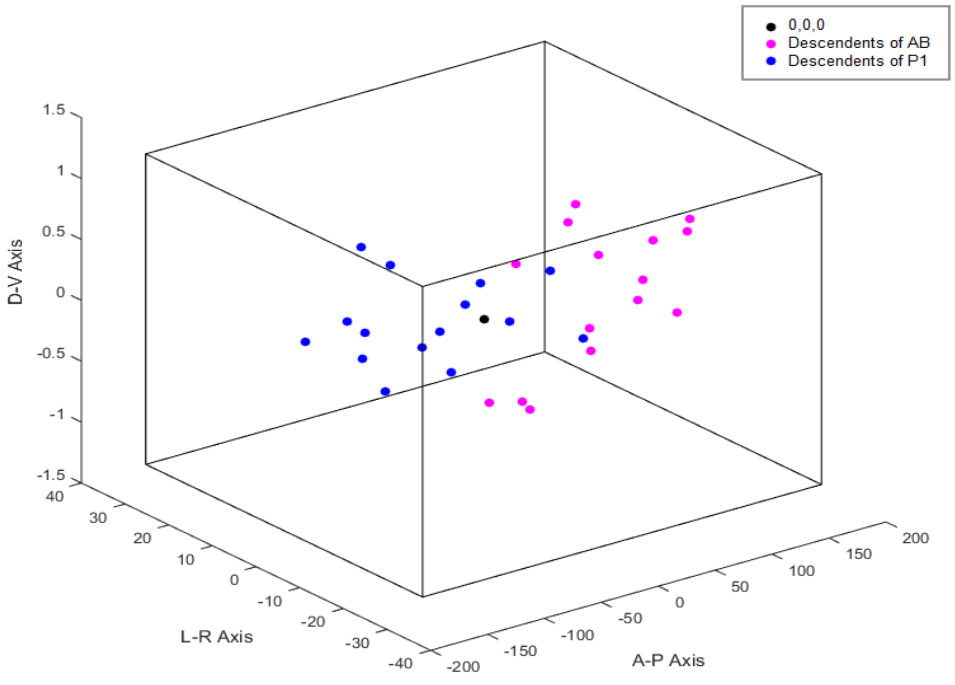

Figure 1. Visualization of spatial segregation by differentiation code for the 16-cell condition (N = 30), embedded in a bounding box (black) that represents the extent of embryo space.

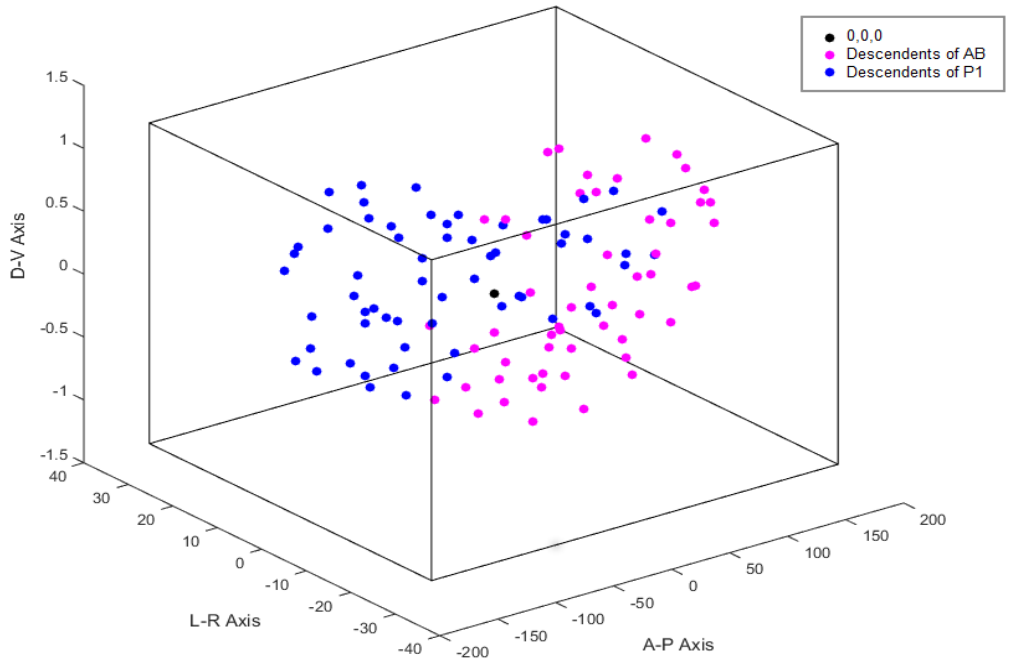

Figure 2. Visualization of spatial segregation by differentiation code for the 64-cell condition (N = 114), embedded in a bounding box (black) that represents the extent of embryo space.

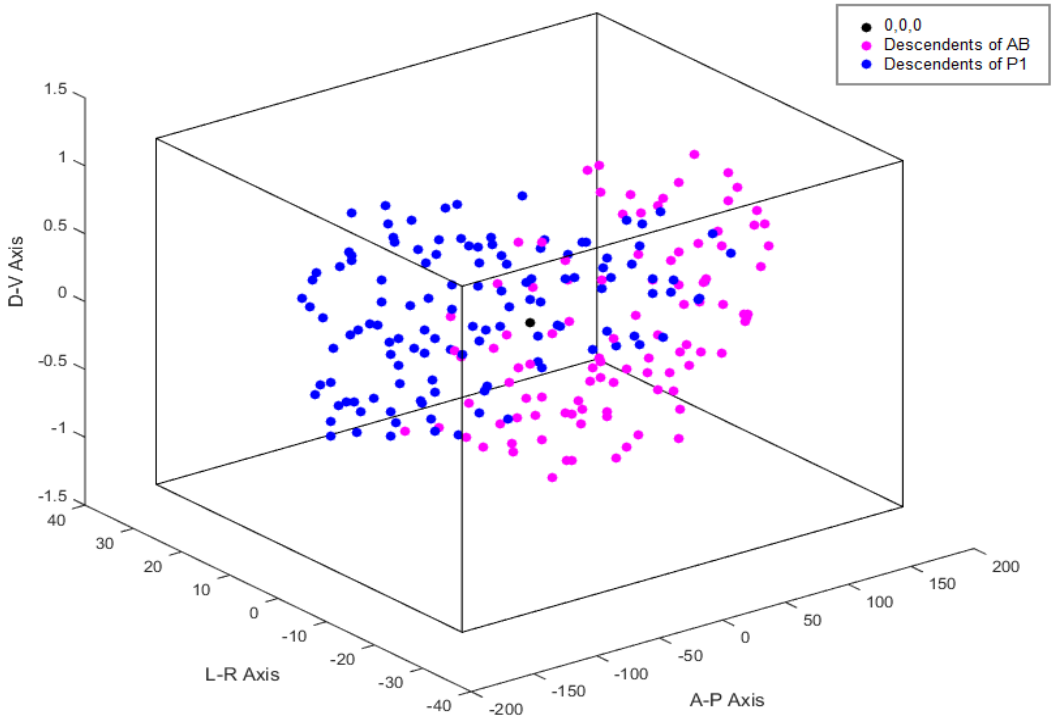

Figure 3. Visualization of spatial segregation by differentiation code for the 128-cell condition (N = 224), embedded in a bounding box (black) that represents the extent of embryo space.

While the spatial segregation of cells decreases with the birth of new cells, in the 7 level example (Figure 3) there are clusters of cells that are descended exclusively from each subtree. Figure 4 shows the results of Figures 1 through 3 on the basis of comparing two major axes of variance (anterior-posterior and left-right) for the eight major subtrees in the lineage tree. Each major subtree descends from a single cell in the 8-cell embryo. For the two examples in Figure 4, variation along the anterior-posterior (A-P) dimension is compared to variation along the left-right (L-R) dimension for six of the eight octopartite components of the lineage tree. Data from the 6 level stage are shown.

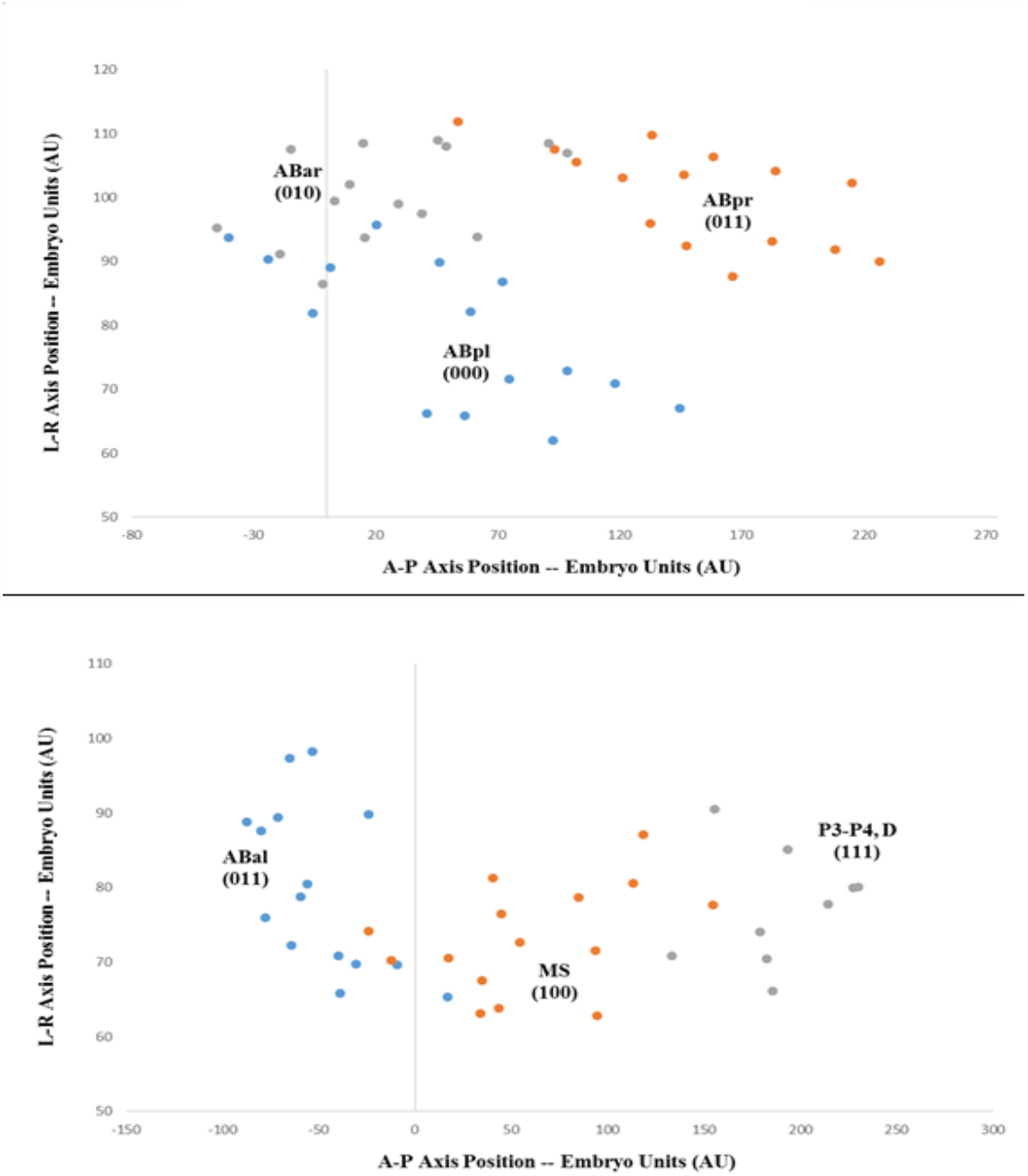

Figure 4. Anterior-posterior axis versus left-right axis variation for various octopartite subtrees for a 6 level tree (fixed axial plane). TOP: ABar (010), ABpl (000), ABpr (011). BOTTOM: ABal (011), MS (100), P3/P4/D (111). All identities: Lineage subtree (Differentiation subtree).

We can learn some basic relational lessons from this qualitative set of comparisons. The left panel of Figure 4 shows spatial separability along the left-right (L-R) axis, while the right panel of Figure 4 demonstrates spatial separability long the anterior-posterior (A-P) axis.

_k_-means Classification of Subtrees

A more quantitative way to measure the degree of spatial integration across development is to biologically-based categories with the outcome of an unsupervised classification process. A *k*-means cluster analysis based on the *x,y,z* position of the cells is used to demonstrate this for both the bipartite (16-, 64-, and 7 levels) and octopartite case (64- and 7 level). As we can see in Table 1, the *k*-means groups are highly predictive of lineage subtrees (G1 predicts AB at a rate of 0.8, and G2 predicts P1 at a rate of 0.8) in the 16 cell condition, but become less predictive as the tree grows in size. Even for the 128 cell condition, the k-means groups are moderately predictive (0.56 for G1 and 0.64 for G2).

Table 1. Comparison of *k*-means cluster analysis (*k*=2) versus each bipartite subtree for 16, 64, and 128 cell conditions. G1-G2 represents the groups generated by the *k*-means analysis, and the sublineage groups are identified by their lineage tree nomenclature.

16 cell condition	**G1**	**G2**
AB	0.80	0.20
P1	0.20	0.80
64 cell condition	**G1**	**G2**
AB	0.67	0.33
P1	0.20	0.80
128 cell condition	**G1**	**G2**
AB	0.56	0.44
P1	0.36	0.64

In Tables 2 and 3, *k*-means generated groups are even less likely to predict the position of octopartite sublineages. There are, however, specific groups that are moderately predictive (0.4 to 0.7) of sublineage. For Table 2 (the 64 cell condition), these include group 4 as a predictor of cells in the C and P3 sublineages and group 8 as a predictor of the ABar sublineage. For Table 3 (the 128 cell condition), these include group 1 as a predictor of cells in the C sublineage and group 5 as a predictor of the ABal sublineage.

Intercalation Analysis

The intercalation analysis tests for uniform changes over time in spatial position amongst cell nuclei in a certain differentiation code subtree. Biological intercalation occurs by a set of cells (e.g. a tissue) narrowing along one anatomical axis while simultaneously expanding along an orthogonal axis. We test for this by making pairwise comparisons of correlation in variance for each differentiation code category along the A-P (*x*), L-R (*y*), and D-V (*z*) axes (Table 4). This results in the following cross-corrleations: *x*-*y*, *y*-*z*, and *x*-*z*.

For the 4 level condition, these cross-correlations were obtained for the two subtrees of the bipartite lineage tree. This results in cells of the AB sublineage exhibiting a weak positive correlation amongst the *x*-*y* comparisons, but an anti-correlation for the *y*-*z* and *x*-*z* comparisons. For the 6 level condition, the cross-correlations were obtained for the eight subtrees of the octopartite lineage tree. In the case of the *x*-*y* comparison, this results in moderately-to-strong negative correlations for the ABpl, ABpr, and ABal sublineages and no moderately-to-weak positive correlations. For the *y*-*z* comparison, this results in moderately-to-strong negative correlations for the MS and E sublineages and moderately-to-weak positive correlations for the ABar sublineage. The *x*-*z* comparison, this results in no moderately-to-strong negative correlations but moderately-to-weak positive correlations for the E and P3 sublineages. These results mostly hold for the 7 level condition, despite the addition of data from another round of cell division.

Table 2. Comparison of *k*-means cluster analysis (*k*=8) versus each octopartite subtree for 64-cell condition (114 node tree with 6 levels). G1-G8 represents the groups generated by the *k*-means analysis, and the sublineage groups are identified by their lineage tree nomenclature.

	G1	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**
ABpl	0.00	0.13	0.27	0.00	0.00	0.13	0.27	0.20
ABpr	0.20	0.00	0.07	0.27	0.00	0.33	0.13	0.00
ABar	0.00	0.13	0.33	0.00	0.00	0.00	0.13	0.40
ABal	0.00	0.33	0.00	0.00	0.60	0.00	0.00	0.07
MS	0.00	0.13	0.4	0.07	0.00	0.13	0.2	0.07
E	0.00	0.00	0.13	0.33	0.00	0.27	0.27	0.00
C	0.27	0.00	0.00	0.53	0.00	0.07	0.13	0.00
P3	0.33	0.00	0.00	0.56	0.00	0.11	0.00	0.00

Table 3. Comparison of *k*-means cluster analysis versus each subtree for 128-cell condition (230 node tree with 7 levels). G1-G8 represents the groups generated by the *k*-means analysis, and the sublineage groups are identified by their lineage tree nomenclature.

	G1	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**
ABpl	0	0.20	0.29	0.06	0.03	0	0	0.16
ABpr	0.23	0.03	0	0.32	0	0.06	0.35	0
ABar	0	0.32	0.06	0	0.03	0.26	0.13	0.19
ABal	0	0.13	0	0	0.52	0	0	0.35
MS	0	0.13	0.26	0.03	0	0.32	0.16	0.10
E	0.08	0	0.24	0.36	0	0.16	0.16	0
C	0.56	0	0	0.33	0	0	0.11	0
P3	0.24	0	0.18	0.35	0	0	0.24	0

Aggregate Cell Migration Analysis

To understand the role of diffusion in the early *C. elegans* embryo, we calculate the three-dimensional distance between all mother-daughter cell pairs at different stages (16-cell condition, 64-cell condition, and 128-cell condition) of early embryogenesis. Table 5 shows the results of this analysis. The distances between mother-daughter pairs are ranked according to distance. A power law exponent (**α**) is then calculated for the 4, 6, and 7 level conditions. Negative values of **α** greater than 1.0 represent a superdiffusive effect. The results in Table 5 show no trend towards a superdiffusive process in time as the embryo gains cells but a slight evidence for a superdiffusive process in space as one moves from the A-P axis to the L-R axis and then to the D-V axis.

Table 4. Correlations of positional variance between three pairs of spatial dimensions (X-Y, Y-Z, and X-Z) and within differentiation code categories (4, 6, and 7 level).

4 level	**X-Y**	**Y-Z**	**X-Z**
AB (0)	0.34	-0.46	-0.43
P1 (1)	0.44	0.56	0.83
6 level	**X-Y**	**Y-Z**	**X-Z**
ABpl (0)	-0.71	-0.34	-0.02
ABpr (1)	-0.64	0.34	0.20
ABar (2)	0.47	0.87	0.39
ABal (3)	-0.55	-0.43	0.48
MS (4)	0.39	-0.84	-0.37
E (5)	-0.11	-0.63	0.58
C (6)	-0.27	-0.31	0.26
P3 (7)	0.15	-0.44	0.68
7 level	**X-Y**	**Y-Z**	**X-Z**
ABpl (0)	-0.56	-0.03	-0.24
ABpr (1)	-0.67	0.28	0.14
ABar (2)	0.36	0.72	0.37
ABal (3)	-0.25	-0.37	0.21
MS (4)	0.22	-0.68	-0.31
E (5)	-0.11	-0.33	0.56
C (6)	-0.27	0.06	-0.14
P3 (7)	-0.17	-0.29	0.60

This result is consistent with the lineage tree being organized along the A-P axis, as it exhibits the most regularity in terms of positioning. While this matches a trend towards decreasing

size along each of these axes, the nature of this analysis pulls out the rare large diffusion of a daughter cell along that axis.

Table 5. Power law exponent (α) for 16-cell, 64-cell, and 128-cell embryos for each spatial axis.

	α		
	16-cell	**64-cell**	**128-cell**
A-P axis	-0.8	-0.67	-0.78
L-R axis	-1.08	-0.91	-0.93
D-V axis	-1.15	-1.05	-1.05

Network Analysis

The first analysis involved constructing a complex network out of a lineage tree by comparing the three-dimensional position of each node in a pairwise fashion. This comparison is done by calculating a Euclidean distance between each pair of cells, and then normalizing to the maximum value (see Methods). These comparisons were stratified by both Euclidean distance threshold and a distinction between bipartite subtrees.

The first analysis involves examining the potential interactions between cells at different distances in the lineage tree. For example, comparisons can be restricted to all cell pairs at the same level in the lineage tree, cells that are only one level away, cells two levels away, and cells that are 3-7 levels away. This provides a baseline for subsequent analyses which smear the time interval component of the lineage tree. Supplemental Figure 1 shows a series of Circular graphs for network topologies for all potential interactions between pairs of cells at different levels of the lineage tree (calculated for cells at different differentiation tree depths). Networks based on data from 7 level condition (N=224).

Table 6 shows that there is diffusion across space as the subtrees grow in terms of depth based on an analysis of inter- and intra-subtree connections shown in Supplemental Figure 1. As several rounds of proliferation separate cell pairs in the same sublineage, they tend to spread out either axially or concentrically. This diffusion according to developmental time is best explained as a shift from intra-subtree to inter-subtree above-threshold connectivity as cellular relationships become separated over time (e.g. lineage tree depth).

Table 6. Pairwise comparisons for a 7 level tree where embryo space distances were calculated for cells at different differentiation tree depths. Threshold for distance metric was 0.75. Intra (AB) represents all pairwise comparisons where cells are descended from AB, Intra (P1) represents all pairwise comparisons where cells are descended from P1.

Distance	**Intra (AB)**	**Intra (P1)**	**Inter**
0	0.57	0.31	0.12
1	0.43	0.35	0.22
2	0.27	0.28	0.45
3-7	0.13	0.13	0.74

Figures 5, 6, and 7 show circular network topologies for the 16-, 64-, and 7 level conditions, respectively. Each bipartite sublineage is color-coded: the nodes and intra-subtree connections for AB are shown in yellow, while the nodes and intra-subtree connections for P1 are shown in purple.

Connections that span the subtrees are shown in gray, while an analysis of these network topologies are shown in Figure 7.

Table 7 connects Figures 1, 2, and 3 to Figures 5, 6, and 7 by showing the network summary statistics for different distance thresholds for the 16-cell, 64-cell, and 128-cell embryos. The two main results are that network modularity increases with relative distance, and that the number of communities increases dramatically as the relative distance threshold increases (and relative distance decreases). While these measures are interrelated, the community metric in particular provides strong evidence that dense patches of local connectivity becomes apparent at very limited spatial scales (5% of the entire embryo).

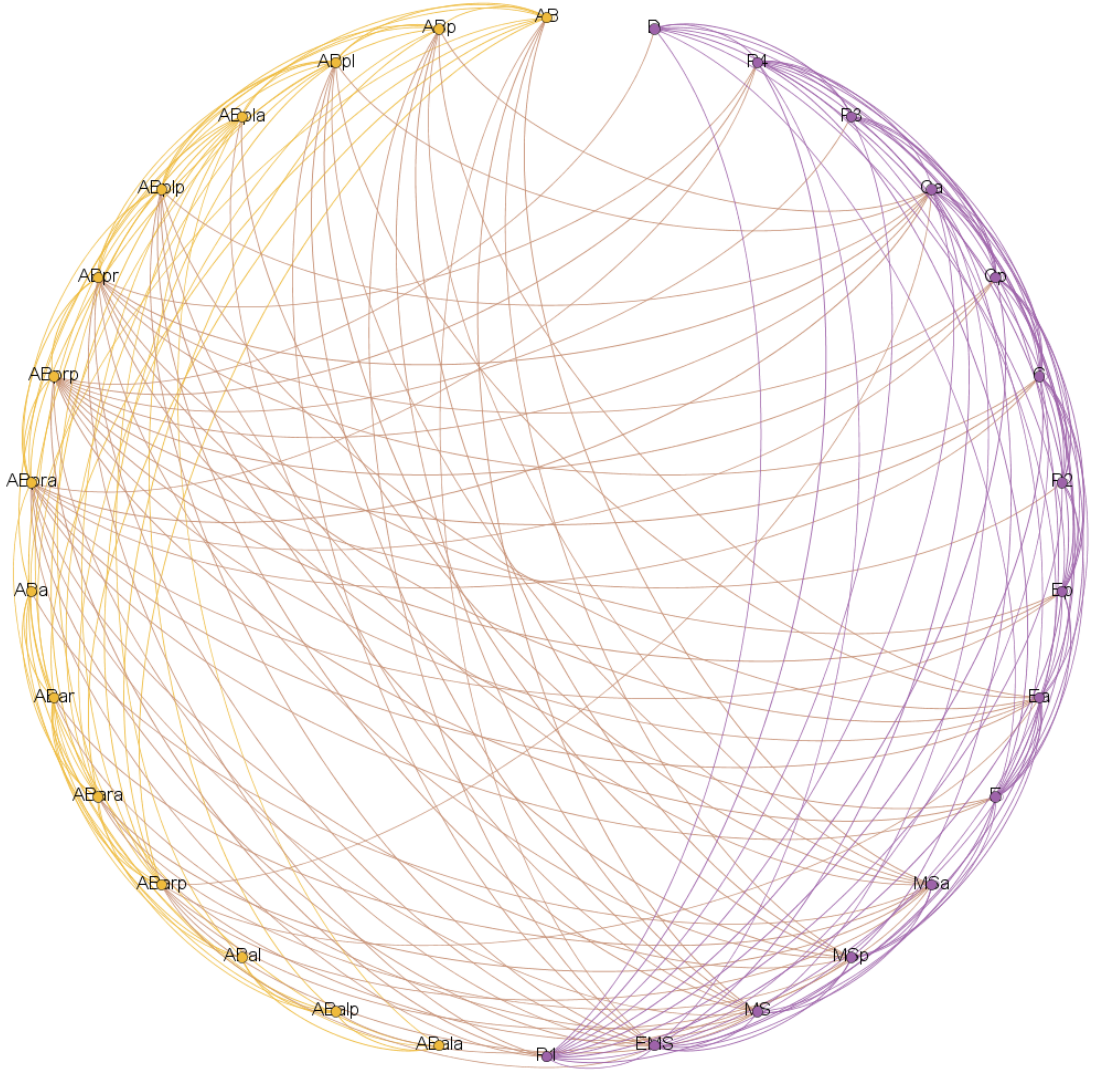

Figure 5. Circular layout of network topology with descendants of the smaller 2-cell division on the left and the larger 2-cell division on the right. Undirected complex network for all cells (represented as nodes) in 4 level condition (N = 30). Threshold is 0.75. Yellow and purple arcs are intra-sublineage connections (smaller and larger, respectively), while the light red arcs are inter-sublineage connections.

Table 7. Network summary statistics for different terminal cell sizes (16, 64, and 128) and thresholds (0.25 to 0.95).

		Average Degree	Network Diameter	Graph Density	Modularity*	Number of Communities	Clustering Coefficient
16	**0.25**	29.13	2	1.005	0	1	0.93
	0.50	23.6	3	0.814	0	1	0.85
	0.75	14.4	5	0.497	0.34	2	0.78
	0.85	8.53	8	0.294	0.46	3	0.71
	0.95	-	-	-	-	-	-
64	**0.25**	103.79	2	0.92	0	1	0.94
	0.50	81.88	3	0.73	0	1	0.87
	0.75	45.51	5	0.41	0.35	2	0.80
	0.85	25.91	9	0.23	0.54	3	0.76
	0.95	5.6	14	0.05	0.8	11	0.74
128	**0.25**	-	-	-	-	-	-
	0.50	145.8	3	0.654	0.226	2	0.85
	0.75	77.58	6	0.348	0.372	2	0.79
	0.85	41.37	10	0.185	0.560	3	0.74
	0.95	7.83	20	0.035	0.834	9	0.72

* resolution = 2.0.

Table 8. Pattern of intra-subtree versus inter-subtree connectivity patterns for three sizes of embryo distance network. Subtrees are based on the descendent of the smaller (subtree 0) and larger (subtree 1) cells of the 2-cell embryo.

Number of Terminal Cells	**16**	**64**	**128**
Distance Threshold	**0.75**	**0.95**	**0.95**
Proportion of Intra-subtree Connections (total)	0.68	0.79	0.74
Proportion of Inter-subtree Connections	0.32	0.21	0.26
Proportion of Intra-subtree Connections (AB)	0.31	0.34	0.39
Proportion of Intra-subtree Connections (P1)	0.37	0.46	0.35

To gain a better understanding of how geometric and spatial heterogeneity of the early embryo relate to cell lineage, we conduct two analysis related to connectivity. As connectivity is defined as potential interaction between cells within a given Euclidean distance, both analysis utilized data from the 128-cell condition at a threshold of 0.75 (25% of the maximum distance across the embryo). The results of these analyses are shown in Figures 5 (4 level), 6 (6 level), and 7 (7 level). These circular network plots demonstrate rich local and cross-connectivity between

each subtree. Table 8 shows the proportion of connections between cells within and between each subtree in the bipartite lineage tree for a given distance threshold.

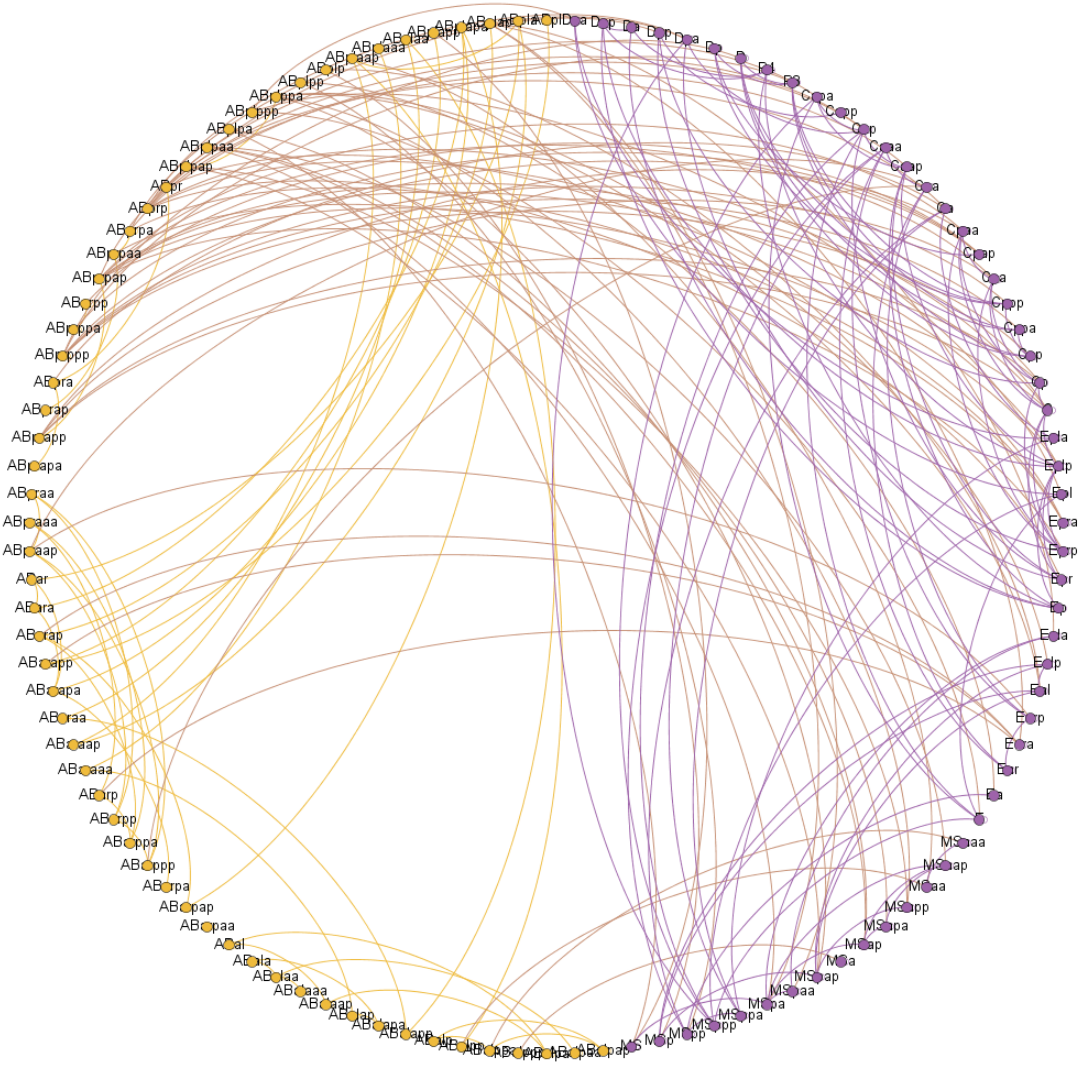

Figure 6. Circular layout of network topology with descendants of the smaller 2-cell division on the left and the larger 2-cell division on the right. Undirected complex network for all cells (represented as nodes) in the 6 level condition (N = 114). Threshold is 0.95. Yellow and purple arcs are intra-sublineage connections (smaller and larger, respectively), while the light red arcs are inter-sublineage connections.

From this analysis of connections based on Euclidean distance (Table 8), roughly 70-80% of all connections across early embryogenetic time are within a subtree. This is consistent with the 3-D spatial location plots shown in Figures 1, 2, and 3. The only aberration in this pattern is in the 6 level embryo, where the number of intra-connections within the P1 subtree rise to 46% at the expense of inter-subtree connections. The P1 subtree contains all non-AB sublineages, which give rise to a number of specialized fates.

Table 9 shows the results of a clique analysis for the octopartite linage tree. A clique analysis for cliques of size 5 yielded 117 unique cliques. A majority of the nodes amongst all of the cliques found came from just two octopartite sublineages (C and ABpr). Furthermore, 75.2% of all cliques generated exhibited overlap between each of the bipartite subtrees (AB and P1). This suggests that there is strong inter-subtree connectivity and thus local structure in the midline of the embryo, in particular involving cells from the C and ABpr sublineages.

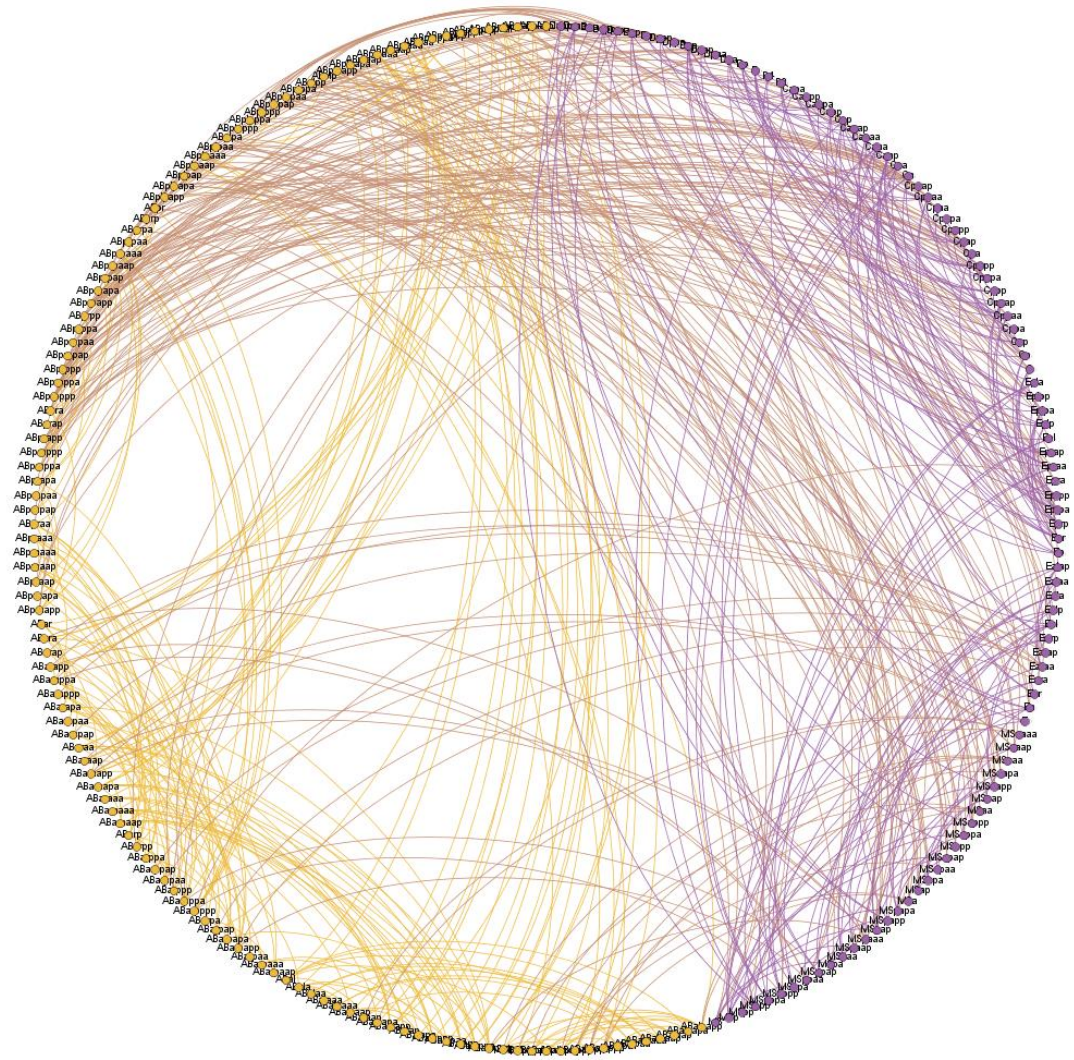

Figure 7. Circular layout of network topology with descendants of the smaller 2-cell division on the left and the larger 2-cell division on the right. Undirected complex network for all cells (represented as nodes) in the 7 level condition (N = 224). Threshold is 0.95. Yellow and purple arcs are intra-sublineage connections (smaller and larger, respectively), while the light red arcs are inter-sublineage connections.

Supplementary Figure 2 shows a heat map that displays the frequency of nodes in each 8-cell subtree category for all cliques of size 5 (N = 117). This shows the relative heterogeneity of cliques with respect to cells from different sublineages. For example, sublineage ABal has the

fewest cells amongst the 117 identified cliques, and cells that are contributed by ABal tend to belong to cliques containing other cells from ABal. While ABal cells keep to themselves, sublineages such as ABpr and C are more social, showing up in most of the 117 identified cliques and contain cells from a wide variety of other sublineages.

Table 9. Clique analysis for a 7 level network topology.

	Sublineage							
	ABpl	**ABpr**	**ABar**	**ABal**	**MS**	**E**	**C**	**P3**
Number of members of sublineage across all cliques generated	42	151	68	34	35	58	159	38
	AB				**P1**			
Number of cliques with overlap	75.2%							
Number of cliques without overlap	15.4%				9.4%			

Another way to assess connectivity is by conducting a rank-order analysis of connectivity. Cells are ranked by their relative connectivity (number of connections with other cells) from most connected to least. In Figure 8, this relationship is observed to range from linear to nonlinear. Nonlinear relationships are indicative of particularly sparse or dense areas of cellular differentiation. Figure 8 shows a slight nonlinear trend across the AB sublineage, a more pronounced nonlinear trend in the P1 sublineage, and a linear trend for interactions between the AB and P1 sublineages. This reflects the geometry of the embryo, and suggests that the P1 sublineage is more spatially heterogeneous than the AB sublineage overall.

Stratifying the data by lineage tree depth also suggests that certain sets of closely-related cells (and thus specific locations in the embryo) are the source of most of this nonlinearity. Figure 9 demonstrates this by visualizing the rank-order relationships in each of the levels (7) in a 7 level lineage tree. Consistently, the descendent cells of MS and ABpr serve as least-connected outliers at their respective level. These cells represent 16 of the 17 labelled cells in Figure 9.

DISCUSSION

In this paper, we offer a statistical approach to understanding spatial diversity in the embryo. While the Sulston lineage tree and technologies such as lineage tracking offer spatial information, the ability to draw interferences from this information is somewhat limited. This is particularly true for space-time processes at various scales that unfold over time. The statistical and network approaches presented here also provide a means to examine the relationships between subtrees and other structural partitions of the embryo. An accounting of the relative distance between cells (e.g. Euclidean distance metric) allows us to construct an interactome between cells, which represents the possible signaling mechanisms and associated relationships between nearby cells.

What does the network analysis tell us about the developmental process? Taken together with the *k*-means analysis, we can see that cells in *Caenorhabditis elegans* embryos only partially segregate by lineage and sublineage. While there are hubs which serve to connect primarily cells of the same sublineage (e.g. AB), other nodes with an enriched number of connections are heterogeneous, and serve to connect (or share a common spatial location with) cells from different sublineages or locations in the lineage tree.

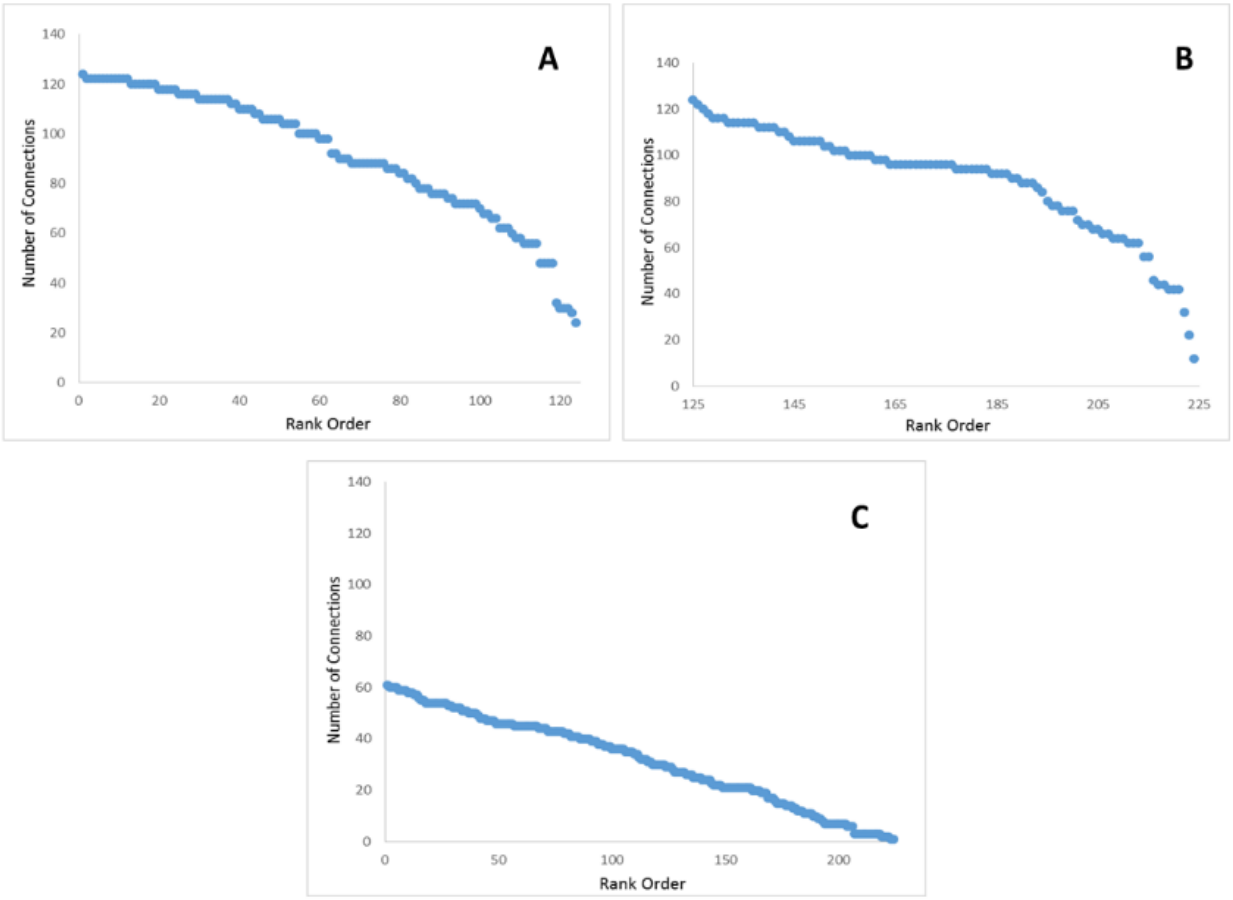

Figure 8. A rank-order analysis of connectivity in AB subtree (A), P1 subtree (B), and potential interactions occurring between the AB and P1 subtrees (C).

Cells that are colocated, or in this case have connections, with cells of a different sublineage may serve as weak connectors in a network representation. By capturing their position in the middle of the embryogenetic process, we can see how a given sublineage can be founded in a single spatial location and later spread out across the Nematode body, performing either specialized or diverse functions depending upon its autonomously-defined identity. In this case, weak connectors serve to indirectly link different populations of cells in a way that would not be possible in a regulative developmental system. In such a system, a cell from one sublineage in close proximity to cells of another sublineage would be induced to the fate of the latter sublineage. This is the advantage of using this type of analysis in *Caenorhabditis elegans*: we can keep track of the ultimate fate of a given cell based on its lineage identity. Therefore, no matter its temporary

developmental position, we can make the rather stark distinction between a mosaic and regulative process.

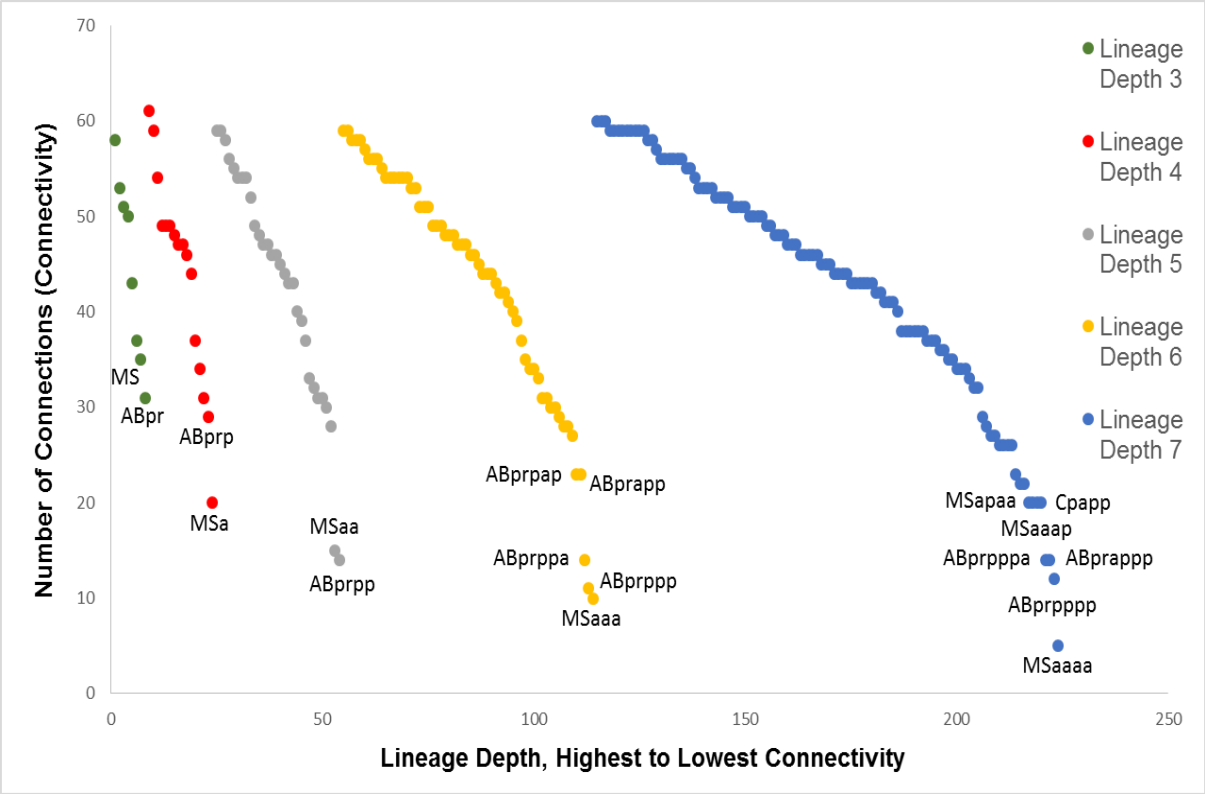

Figure 9. A rank-order analysis of 128-cell condition (N = 224), stratified by lineage depth (ranging from 3-7). Outlier cells are labeled for each series.

One intriguing outcome, suggested from the clique analysis, is that specific sublineage groups can coexist in a spatial region of limited size. For example, some cliques might contain several closely-related cells (based on their position in the lineage tree), and a few cells that are from a separate sublineage. Whether this has any consequence for future differentiation is an open question. However, if the mosaic mode of development dominates, we should expect these interloping cells to remain autonomous and be defined by their lineage tree location rather than by their spatial position (e.g. the cells they interact with).

Broader Implications

Previous network approaches to development focus on either the level of gene and protein expression or the level of phenotypic modules [16]. With respect to *C. elegans*, there have been a number of functional association network studies at the molecular level (gene, protein, and interactome data) and the phenotypic level [17].

With respect to the development of tissues, Green et.al [18] use a network reduction method to examine the gonad attempt to make a connection between protein expression and phenotypic modules in early development [19]. A similar method called the scale-invariant power law approach [20] has also been used to study self-organization in developing embryos. As their approach focuses on the entirely of the early embryo, these authors treat macroscopic cellular

proliferation as a scale-invariant process. In this study, approximations of cell positions and distances were used to estimate the geometry of the early embryo, with statistical variation expressed in the non-AB sublineages. This is similar to our results, as the non-AB lineages represent

A network approach is also used to examine the role of modularity and evolvability in anatomical evolution. In Esteve-Altava et.al [21], a network is drawn between landmarks to examine the relative locations between muscles and how this structure changes across closely-related species and evolutionary time. The relationships in evolutionary time can demonstrate the evolvability (or evolutionary capacity) of a specific phenotypic configuration [22]. This type of structural approach provides information regarding spatial organization in the embryo, which may or may not be similar at different spatial scales. Scale-invariance is an important but understudied feature of early embryogenesis, and may exist in the early embryo as a self-regulating system of positional information [23]. It also provides empirical detail to a model of the cybernetic embryo [24]. In such models, networks can reveal potential functional relationships and mechanisms.

While existing network models have made strides in uncovering both the structural and functional detail of the early embryo, a connectivity map has not been applied to a precise positional map of developmental cells that map to the lineage tree. In doing so, we reveal potential cell-cell interactions at multiple scales, including commonly-recognized regionalization patterns [7, 25]. In this paper, we construct such a map, and treat the cells of a pre-hatch embryo lineage tree as a scalable interactome. This allow us to explore the process of differentiation at several scales, in addition to treating the lineage tree as a series of interconnected subnetworks.

ACKNOWLEDGEMENTS

We would like to acknowledge Drs. Zhirong Zhao and Stephen Larson for providing data on the *C. elegans* embryo. Thanks also go to members of the OpenWorm Foundation and the Network Science community for scientific feedback.

REFERENCES

[1] Gordon, R. (1999). The Hierarchical Genome and Differentiation Waves: Novel Unification of Development, Genetics and Evolution. Singapore & London, World Scientific & Imperial College Press.

[2] Alicea, B., McGrew, S., Gordon, R., Larson, S., Warrington, T., and Watts, M. (2014). DevoWorm: differentiation waves and computation in *C. elegans* embryogenesis. *bioRxiv*, doi:10.1101/009993.

[3] Gordon, N.K. and Gordon, R. (2016). Embryogenesis Explained. Singapore, World Scientific Publishing Company.

[4] Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology*, 100(1), 64-119.

[5] Bao, Z.R., Murray, J.I., Boyle, T., Ooi, S.L., Sandel, M.J., and R.H. Waterston (2006). Automated cell lineage tracing in *Caenorhabditis elegans*. *PNAS*, 103(8), 2707-2712.

[6] Walck-Shannon, E. and Hardin, J. (2014). Cell intercalation from top to bottom. *Nature Reviews Molecular Cell Biology*, 15, 34-48.

[7] [26] Schnabel, R., Bischoff, M., Hintze, A., Schulz, A.K., Hejnol, A., Meinhard, H., and Hutter, H. (2006). Global cell sorting in the *C. elegans* embryo defines a new mechanism for pattern formation. *Developmental Biology*, 294(2), 418-431.

[8] Fagotto, F. (2014). The cellular basis of tissue separation. *Development*, 141(17), 3303-3318.

[9] Wetts, R., and Fraser, S.E. (1989). Slow intermixing of cells during *Xenopus* embryogenesis contributes to the consistency of the blastomere fate map. *Development*, 105(1), 9-15.

[10] Nishida, H., and Stach, T. (2014). Cell lineages and fate maps in tunicates: conservation and modification. *Zoological Science*, 31(10), 645-652.

[11] Woods, M.L., Carmona-Fontaine, C., Barnes, C.P., Couzin, I.D., Mayor, R., and Page, K.M. (2014). Directional Collective Cell Migration Emerges as a Property of Cell Interactions. *PLoS One*, 9(9), e104969.

[12] Fedotov, S., Tan, A., and Zubarev, A. (2015). Persistent random walk of cells involving anomalous effects and random death. *Physical Review E*, 91(4), 042124.

[13] Bao, Z.R., Zhao, Z.Y., Boyle, T.J., Murray, J.I., and Waterston, R.H. (2008). Control of cell cycle timing during *C. elegans* embryogenesis. *Developmental Biology*, 318(1), 65–72.

[14] Murray, J.I., Boyle, T.J., Preston, E., Vafeados, D., Mericle, B,, Weisdepp, P., Zhao, Z., Bao, Z., Boeck, M., and Waterston, R.H. (2012). Multidimensional regulation of gene expression in the *C. elegans* embryo. *Genome Research*, 22(7), 1282-1294.

[15] Alicea, B. and Gordon, R. (2016). Quantifying Mosaic Development: Towards an Evo-Devo Postmodern Synthesis of the Evolution of Development Via Differentiation Trees of Embryos. *Biology*, 5(3), 33.

[16] Gunsalus, K.C. and Rhrissorrakrai, K. (2011). Networks in *Caenorhabditis elegans*. *Current Opinion in Genetics & Development*, 21, 787–798.

[17] Sonnichsen, B., Koski, L., Walsh, A., Marshall, P., Neumann, F., Brehm, M., Alleaume, A-M., Bettencourt, P., Cassin, E., Hewitson, M., Holz, C., Khan, M., Lazik, S., Martin, C., Nitzsche, B., Ruer, M., Stamford, J., Winzi, M., Heinkel, R., Roder, M., Finell, J., Hantsch, H., Jones, S.J.M., Jones, M., Piano, F., Gunsalus, K.C., Oegema, K., Gonczy, P., Coulson, A., Hyman, A.A., and Echeverri, C.J. (2005). Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature*, 434, 462-469.

[18] Green, R.A., Kao, H.L., Audhya, A., Arur, S., Mayers, J.R., Fridolfsson, H.N., Schulman, M., Schloissnig, S., Niessen, S., Laband, K., Wang, S., Starr, D.A., Hyman, A.A., Schedl, T., Desai, A., Piano, F., Gunsalus, K.C., Oegema, K. (2011). A high-resolution *C. elegans* essential gene network based on phenotypic profiling of a complex tissue. *Cell*, 145, 470-482.

[19] Gunsalus, K.C., Ge, H., Schetter, A.J., Goldberg, D.S., Han, J.D., Hao, T., Berriz, G.F., Bertin, N., Huang, J., Chuang, L.S., Li, N., Mani, R., Hyman, A.A., Sonnichsen, B., Echeverri, C.J., Roth, F.P., Vidal, M., and Piano, F. (2005). Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature*, 436, 861-865.

[20] Tiraihi, A., Tiraihi, M., and Tiraihi, T. (2011). Self-organization of developing embryo using scale-invariant approach. *Theoretical Biology and Medical Modelling*, 8, 17.

[21] Esteve-Altava, B., Diogo, R., Smith, C., Boughner, J.C., and Rasskin-Gutman, D. (2015). Anatomical networks reveal the musculoskeletal modularity of the human head. *Scientific Reports*, 5, 8298.

[22] Esteve-Altava, B., Boughner, J.C., Diogo, R., Villmoare, B.A., and Rasskin-Gutman, D. (2015). Anatomical Network Analysis shows decoupling of modular lability and complexity in the evolution of the primate skull. *PLoS One*, 10(5), e0127653.

[23] Lokeshwar, B.L. and Nanjundiah, V. (1980/1981). The scale-invariance of spatial patterning in a developing system. *Wilhelm Roux's Archives of Developmental Biology*, 190(6), 361-364.

[24] Stone, R. and Gordon R. (2017). Cybernetic embryo. In “Biocommunication” Gordon R, Seckbach J, eds., pgs. 111-164. London: World Scientific Publishing.

[25] Schnabel, R. (1996). Pattern formation: regional specification in the early *C. elegans* embryo. *Bioessays*, 18(7), 591-594.