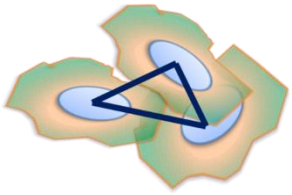

cytoNet: User Guide

cytoNet is a tool for the quantification of multicellular spatial organization using principles of graph theory. Taking fluorescence microscope images as input, cytoNet identifies cells, creates spatial network representations, and calculates a set of metrics derived from graph theory that describe the network structure of the local multicellular neighborhood – the cell community – of a cell of interest. Cell community metrics can be integrated with descriptors of cell phenotype, such as morphology and protein expression to provide a comprehensive description of single- and multiple-cell phenotype states.

This user guide provides step-by-step instructions on how to process images using cytoNet.

Select image files

Select one or more image files by clicking the 'Choose Files' button. Multiple files can be selected by: a) holding down the control key (command key in MacOS); b) clicking and dragging; or c) entering control-A (command-A in MacOS) to select all files in a directory or folder.

We provide two sample images for demonstration purposes. If you want to use the sample images, check the boxes next to them.

If you provide a grayscale image, cytoNet uses a default segmentation algorithm to identify cells and create a binary mask. The algorithm applied is locally adaptive thresholding followed by a watershed operation. If you require more sophisticated image segmentation, we recommend using programs like ilastik - http://ilastik.org/ or CellProfiler - http://cellprofiler.org/. Masks generated using these programs can also be selected as input to cytoNet.

Select layer number

Some image file formats like .tif support multiple images per file. In such cases, you may select the (1-based) indices of the particular images to be analyzed. Multiple indices can specified by using a dash to specify a range (e.g. 1-3) and commas to separate multiple indices and ranges (e.g. 1,2-4,7). Indices must be specified in strictly increasing order.

How to connect cells?

cytoNet represents cell communities using graphs. A graph is a mathematical structure consisting of a set of nodes and a set of edges where each edge connects two nodes. cytoNet creates a graph from a mask file where each foreground object in the mask file is represented by a unique node.

The existence of an edge between two nodes can be determined in one of two ways: a) objects being nearby, or b) objects touching each other.

Option 1: Centroid-distance method

If you choose this option, cytoNet places an edge between two nodes when the distance between the centroids of the corresponding foreground objects is less than a calculated threshold value. For each pair of foreground objects, a threshold is determined in a two-step process. First, the effective diameter of

each object is computed from its area by using the formula for circle area: $area = \pi\frac{d^2}{4}$. Second, the threshold is computed by multiplying the average effective diameter of the two objects by a scaling factor: $threshold = S \cdot \left(\frac{d_1+d_2}{2}\right)$ where S is a user-defined scaling factor. Larger values of S will result in more edges in the graph than smaller values (see Figure 1).

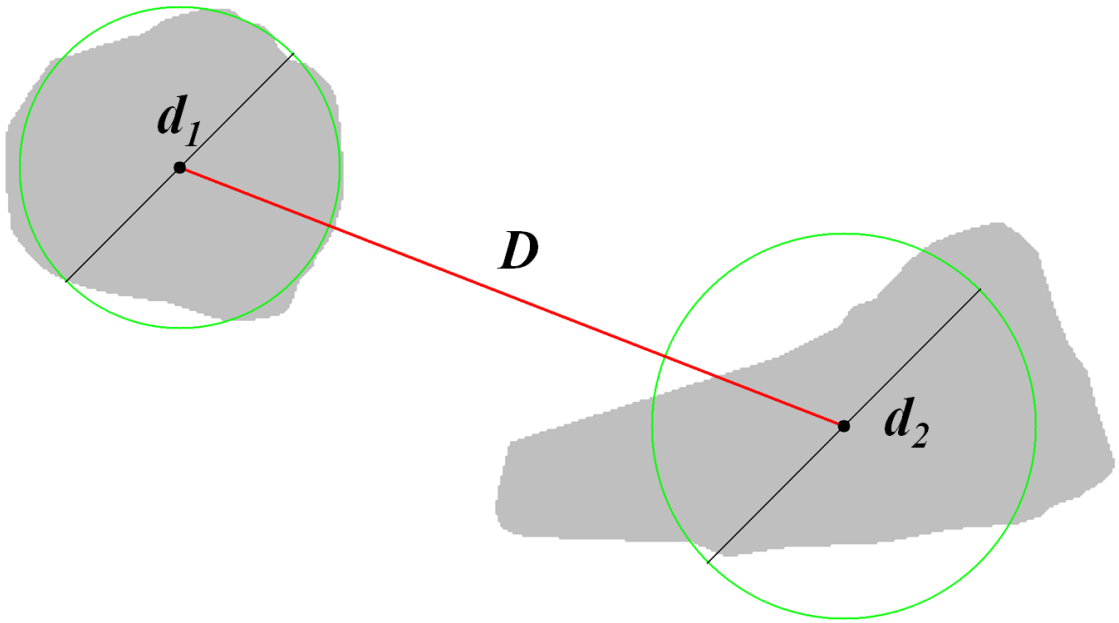

Figure 1. When edges are placed between nearby objects, an edge is placed between the two objects if $D < S \cdot \left(\frac{d_1+d_2}{2}\right)$ where S is a user-defined scaling parameter. The area enclosed by each green circle is equivalent to the area of its corresponding gray object.

We recommend choosing this option when processing images of nuclei or other markers that give the approximate location of the center of a cell.

Option 2: Border-overlap method

If you choose this option, cytoNet places edges between foreground objects when their borders overlap. Each object is slightly enlarged in order to calculate overlap with immediately adjacent objects. This overlap is measured, and if the number of overlapping pixels of two objects is greater than 0, then a graph edge is placed between the nodes corresponding to the objects (see Figure 2).

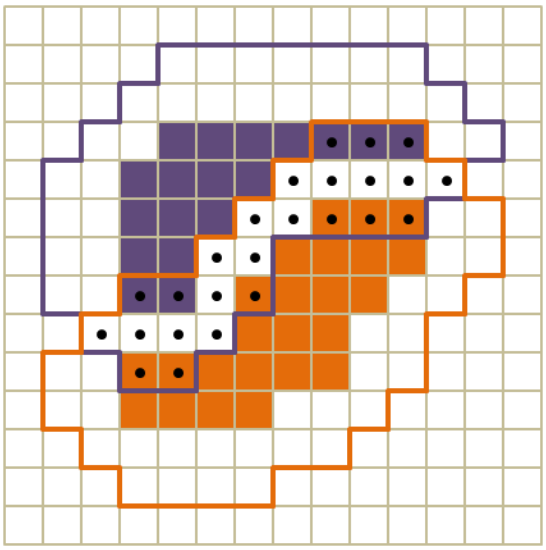

Figure 2. To determine if two objects touch (share parts of their borders), two objects (purple and orange squares) are expanded by two pixels each. If the number of squares in the intersection of their overlap (dotted squares) is greater than 0, then an edge between vertices representing the objects is placed in the graph.

The border-overlap approach is computationally slower. We recommend using this approach only when the entire cell can be identified, for instance in cells stained for cytoskeletal proteins or expressing cytoplasmic fluorophores.

Select scaling factor

If you choose the centroid-distance method, use the sliding bar to select the scaling factor. Increasing the scaling factor increases the likelihood of edges being placed between objects. See Figure 3 for visualization of graphs created with different scaling factors from the same image.

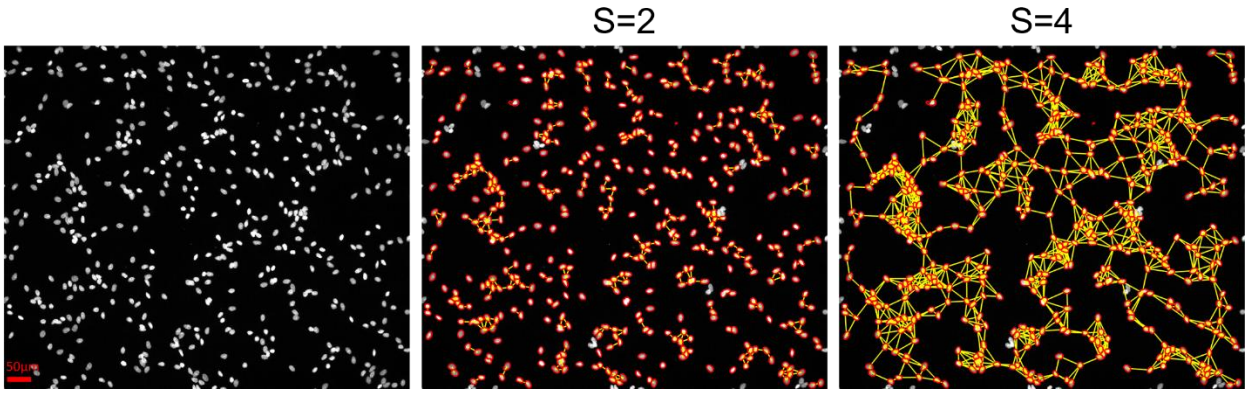

Figure 3. Immunofluorescence nucleus image with graph representations using scaling factors, S=2 and S=4. Number nodes in the graph (number of cells detected) = 464; Number of connections (shown in yellow) = 413 (S=2) and 1484 (S=4).

Output

cytoNet computes global and local network properties determined based on spatial proximity of objects in the image. Global metrics are tabulated in a file called 'GlobalMetrics.csv' for all images in the input folder. Local network metrics and morphology metrics, computed on a per-cell basis are tabulated in separate files for each image called 'LocalMetrics_filename.csv' and 'SingleCellMetrics_filename.csv' respectively, where filename is the original file name.

Processed images are also created for each image in the input folder, called 'filename-PROCESSED.tif' where the original image is overlaid with cell indices, object outlines (red) and spatial proximity edges (yellow). Cell indices displayed in the processed images are used in the first column of local metrics files. See the following section for a description of all the network metrics.

Description of Metrics

Local Network Metrics

(Computed on a per-cell basis; contained in the files 'LocalMetrics_filename.csv')

Degree

Number of neighbors one link away.

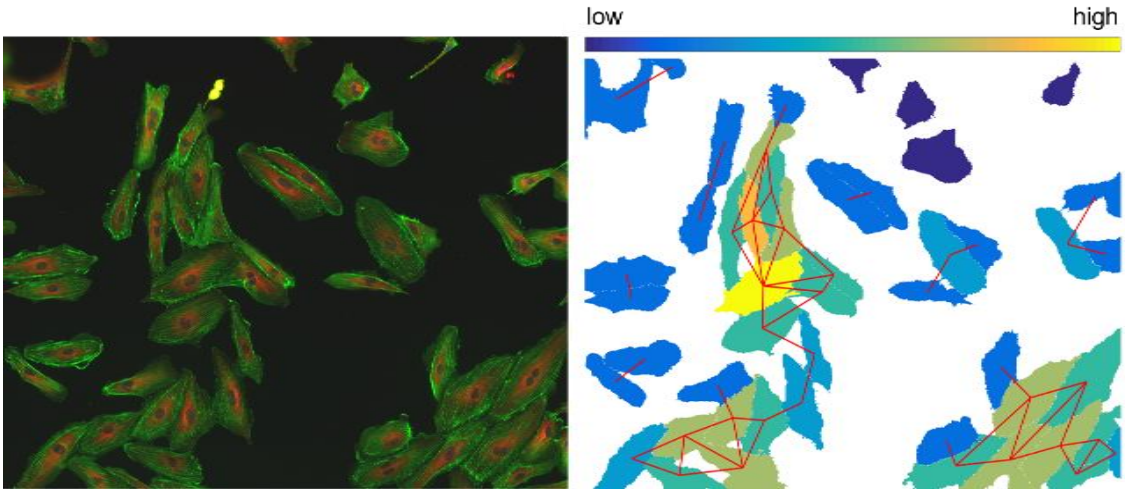

Illustration of degree in a culture of human umbilical vein endothelial cells. Graph edges are shown as red lines in the heatmap.

Average Neighbor Degree

Average of average degree for all neighboring nodes.

Clustering Coefficient

Number of connections in the local neighborhood of a node, divided by total possible connections in the neighborhood.

Local Efficiency

Network efficiency of local neighborhood of a node.

Closeness Centrality

Sum of the length of shortest paths between a node and all other nodes in the network.

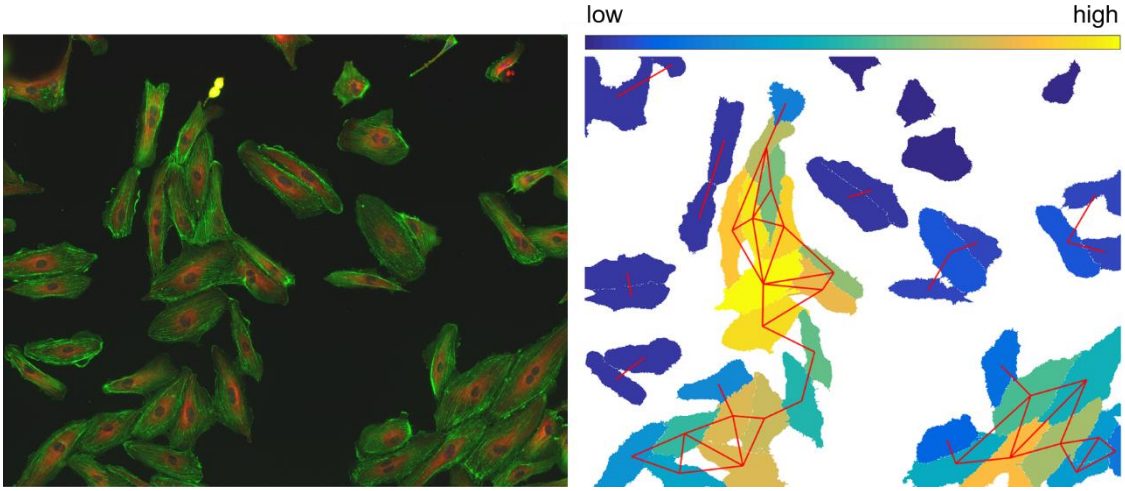

Illustration of closeness centrality in a culture of human umbilical vein endothelial cells (HUVECs). Cells at the center of colonies have high closeness centrality compared to cells at the edge of colonies or isolated cells.

Betweenness Centrality

Number of times a node occurs in the shortest path between two other nodes.

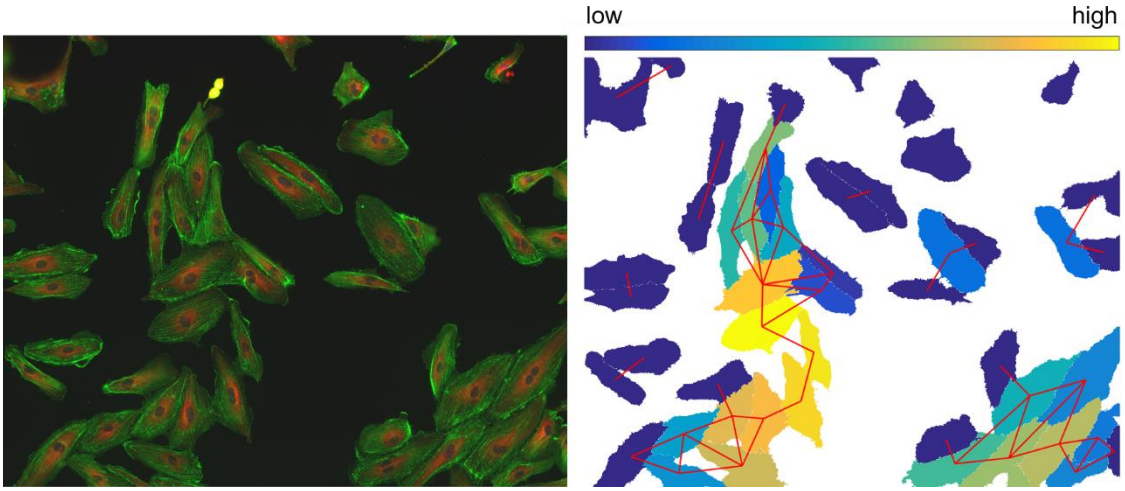

Illustration of betweenness centrality in a culture of human umbilical vein endothelial cells. Cells connecting different clusters have high betweenness centrality.

Global Network Metrics

(Contained in the file GlobalMetrics.csv)

Basic Network Parameters

Node Count (n)
Number of nodes (objects).

Edge Count (m)
Number of edges (connections).

Percent Area Cells
Fraction of total surface area covered by cells.

Average Degree
Average number of connections for a node in the network.

Degree-related Metrics

Network Density
Normalized version of average degree. Network density is a number between 0 and 1, and is the ratio of actual connections in the network to the total number of potential connections. A fully connected clique has a network density of 1.

$$Density = \frac{2m}{n(n-1)}$$

Variance in Degree
Variance of normalized node degree sequence

Network Heterogeneity
Reflects tendency of network to contain hub nodes

$$Heterogeneity = \frac{standard\ deviation(k_n)}{mean(k_n)}$$

Where k_n is the degree sequence.

Clustering Coefficient (normalized)
Clustering coefficient of a node is the number of connections in the local neighborhood of the node, divided by total possible connections in the neighborhood. The local neighborhood of a node comprises of all the nodes exactly 1 link away.

If the average clustering coefficient across all nodes in the network is C and the average clustering coefficient of 100 random graphs (generated through degree-preserving rewiring) is C_{rand}, then clustering coefficient (normalized) $= \frac{C}{C_{rand}}$

Average Neighbor Degree
Average degree of neighboring nodes, averaged across all nodes.

Variance in Neighbor Degree
Variance of the normalized average neighbor degree sequence.

Motif Counts

4-star Motif Count
Number of occurrences of motif with hub node and 3 spokes, normalized by total possible 4-tuples, ${}^{n}C_{4}$

5-star Motif Count
Number of occurrences of motif with hub node and 4 spokes, normalized by total possible 5-tuples, ${}^{n}C_{5}$

6-star Motif Count
Number of occurrences of motif with hub node and 5 spokes, normalized by total possible 6-tuples, ${}^{n}C_{6}$

Triangular Loop Count
Number of occurrences of motif with 3 connected nodes, normalized by total possible 3-tuples, ${}^{n}C_{3}$

Pair Node Count
Number of occurrences of motif with 2 nodes.

Isolated Node Count
Number of nodes with no neighbors.

Modularity Metrics

Number of Connected Components
A connected component is a collection of nodes in which no node is isolated. A highly connected graph has a small number of connected components.

Average Component Size (normalized)
Average number of nodes per connected component, divided by total number of nodes.

Variance in Component Size
Variance in normalized component size sequence.

Geodesics

Network Diameter
The largest distance between two nodes in terms of number of links.

Network Efficiency (normalized)
The shortest path between two nodes is the smallest number of links needed to travel from one node to the other. The reciprocal of the shortest path length is the efficiency. The average value of efficiency for the network is called network efficiency.

If network efficiency is E and the average network efficiency of 100 random graphs (generated through degree-preserving rewiring) is E_{rand}, then network efficiency (normalized) $= \frac{E}{E_{rand}}$.

Local Efficiency
Network efficiency of local neighborhood of a node, averaged across the network.

Assortativity
Pearson correlation coefficient of degrees between pairs of linked nodes. A highly assortative network is one where nodes with high degree tend to connect with other nodes with high degree.

Single-Cell Metrics

(Computed on a per-cell basis; contained in the file 'SingleCellMetrics_filename.csv')

Object Size

Total cell area in pixels.

Circularity

A metric of cell roundness, measured as the closeness to a perfect circle.

$$Circularity = \frac{4\pi . A_C}{P_C}$$

Where A_C = cell area, P_C = cell perimeter

Elongation

Shape factor measured as perimeter/area.

$$Elongation = \frac{P_C}{A_C}$$

Intensity

Average grayscale intensity of pixels within the cell, reported as a number between 0 and 1, where 0 indicates complete darkness and 1 indicates complete whiteness.

Note that the files containing local network metrics and single-cell metrics contain a column labeled 'Cell Index', listing the indices overlaid in the processed images. The indices in each of these files match each other.