Network Analysis of Saccharomyces cerevisiae

Florentin Anggraini Purnama¹, Widya Meiriska, Dimas Praja Purwa Aji, Devi Aprianti, Sparisoma Viridi²

Department of Physics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia

¹florenpan@gmail.com, ²dudung@gmail.com

Abstract. The pattern of yeast budding formation is essential in yeast colony expansion. By exploring how Saccharomyces cerevisiae as model cell grows, the distinction of pathological and non-pathological eukaryotic cells can be further understood. This research explores the use of FIJI software and network theory to characterize the formation pattern of Saccharomyces cerevisiae colony. By using speed distribution and degree distribution, three potential networks are compared: barbell network, relaxed caveman network, and triangular lattice network. Triangular lattice network emerges as the most suitable network to model yeast colony growth.

1. Introduction
There are pathogenic and non-pathogenic microorganisms for humans. Pathogenic microorganisms can cause the host to be ill, while non-pathogenic ones can be utilised to produce medicine, beverage, and other industrial goods. [1] There are many studies that focus on exploring further utilisation of these microorganisms through experiments. Computational modelling and simulation can contribute more to the existing study because of its economical nature and singular source.[2]

A research with simulation approach assumes the yeast cells as uniform circular particles in 2 dimensional plane that interact through gravitational force, normal force, and stokes force. From this approach, a simulation of yeast colony budding is produced that looks similar to the actual yeast growth, but the number of splitting and location of splitting in the simulation are still not very accurate because the splitting event is randomly generated.[3,4] This is then attempted to be corrected by using stochastic approach to determine the yeast growth area based on abundance of nutrients.[5]

A study that uses graph theory to characterise granular materials has been done to statical materials.[6] In the research, photoelastic disks are put into a box and are given force, causing them to interact with stress chains and emitting light. They are then assumed as particles in 2 dimensonal plane and modeled as triangular lattice graph that has pores to accommodate for force interaction between particles that behave as chains. Parameters that are successfully characterised are degree distribution, cluster size, and tree size distribution.

This paper explores the possibility of identifying other parameters of yeast cell colony growth by using graph theory and simple geometric means.

1.1. Network representation and metrics
Network or graph is a group of nodes that are connected by edges. Nodes are represented by points, while edges are represented by lines. One of the most usual ways to represent a network is by using an adjacency matrix.[7]
Figure 1. Undirected simple network

\[
A_{ij} = \begin{cases} 
1, & \text{if there is an edge between nodes } i \text{ and } j \\
0, & \text{if there is no edge between nodes } i \text{ and } j 
\end{cases}
\]

The adjacency matrix in equation (1) shows the connection between each node. When there is an edge between the nodes, the matrix element will have a value of 1. When there is no edge, the element will have a value of 0. This article focuses on the use of undirected simple network as shown in figure 1 to represent yeast colony growth.

There are a few network metrics that are used in this paper for comparison. Degree centrality characterises each node based on their position in a network by showing how much adjacent neighbor(s) a particular node has. In a network, two nodes are adjacent neighbors when they are connected by an edge. The degree can be calculated from adjacency matrix by summing up suitable elements. Equation (2) shows \( C_D(v_i) \) as the degree centrality of node \( i \) and \( \sum_j A_{ij} \) as the summation of number of all cells that are adjacent to node \( i \).

\[
C_D(v_i) = d_i = \sum_j A_{ij} \tag{2}
\]

Greedy vertex coloring is a variation of greedy algorithm that finds the amount of node colors so that there are no 2 adjacent nodes with the same color.

2. Methods

A video of *Saccharomyces cerevisiae* colony growth on agarose pads under a fluorescence microscope is used for observation. There are initially 2 separate colonies with 2 cells on top colony and 3 cells on bottom colony. The video is a time-lapse series of phase contrast and fluorescence images where frame is taken every three minutes.

2.1. Colony detection with FIJI

After being imported to FIJI[10], the video has 221 frames with the size of 946 × 720 px. A frame is 0.1 second of the video.[13] The colony growth is observed only on frame 0-137, where afterwards the colony is growing out of the observable area of the video. At frame 137, the cells are assumed to finish growing and cease splitting. It is also assumed that no cells die at the entire video. The video is optimized by removing noise and increasing the contrast between the cells and the background.
The cells of the yeast colony are assumed to be uniformly circular. By using straight tool on FIJI, the diameter of a cell is measured and determined as 50 px. The plugin TrackMate is used to track each cell’s trajectory for each frame. TrackMate’s tracker is able to detect a cell when the cell’s diameter is more than or equal to 30 px, thus making the cells that are not sufficiently big during the budding process virtually undetectable.

2.2. Analysis using NetworkX

After the tracking with FIJI’s TrackMate is complete, a network for each frame is defined. In each frame, the combinations of all possible cell pairs on the frame are reviewed. If the cell pair has distance of less than or equal to 50 px, an edge is defined between the two cells. This means if a frame has 50 cells or nodes, there will be \( \binom{50}{2} = 1225 \) possible node pairs that need to be checked for the condition. Python’s library NetworkX is then used to build three graph models to approximate the yeast colony growth.

\[
\binom{n}{r} = \frac{n!}{(n-r)!r!}
\]  

(3)
3. Results and Discussion

Before analysing the network, TrackMate’s reliability as an observation tool for the yeast colony growth must be justified. Curve fitting comparisons on cell count are done to find out the accuracy of the tracker count. The bottom colony of the yeast colony is selected for the count comparison.

3.1. Comparison of TrackMate count and manual count

From figure 6 and figure 7, it can be seen that the cell count by TrackMate are upper bounded the author’s manual count that accounts for the budding cells and lower bounded by manual count that doesn’t include the budding process. On figure 6, exponential function fitting are done on buds included data \((y = 5.0731 \times e^{0.0225x} \text{ with } R^2 = 0.994)\) and tracker count data \((y = 4.5784 \times e^{0.0221x} \text{ with } R^2 = 0.9933)\), while the buds not included data is fitted with first order polynomial \((y = 0.2602x + 0.0654 \text{ with } R^2 = 0.9721)\). From the fitting, it can be seen that the tracker count curve is the compression of the buds included curve.

On figure 7, second order polynomial are fitted to the buds included data \((y = 0.0052x^2 - 0.0839x + 6.6069 \text{ with } R^2 = 0.9976)\), tracker count data \((y = 0.0045x^2 - 0.0739x + 5.9465 \text{ with } R^2 = 0.9968)\), and the buds not included data \((y = 0.0007x^2 + 0.1674x + 1.9442 \text{ with } R^2 = 0.9813)\). From the fitting, it can be seen the tracker count curve is the compressed buds included curve that is translated downwards. The fitting with second order polynomial slightly increases the \(R^2\) value of all data.

Both figures show that the cell count by TrackMate is closer to when the cell that hasn’t bud off completely is counted as a cell already. This is related to the choice of tracker size on the TrackMate tracking process as can be seen in figure 3. It can also be inferred from the graphs that difference between actual cell count and the tracker count increases as the time increases.
3.2. Colony shape change with time

The colony shape change with time can be measured using eccentricity formula in equation (4). Eccentricity value ranges from 0 to 1, with 0 representing straight line and 1 representing uniform circle. The eccentricity is calculated for the colony in each frame.

\[ e = \frac{c}{a} \]  

(4)
Figure 8. Eccentricity of a colony

On figure 9, the top colony’s shape changes gradually from more elliptic into more stable circular shape. Its horizontal and vertical length also generally increases as time goes although it shows slight decrease at certain time. On figure 10, the bottom colony’s shape is much more fluctuative than the top colony’s shape. Where top colony shows gradual progression from ellipse to circle, the bottom colony shows sharp change from ellipse to circle and varies severely.

Figure 9. Plot of how top colony's eccentricity changes with time

Figure 10. Plot of how bottom colony's eccentricity changes with time

3.3. Degree distribution
A cell’s degree shows the number of neighboring cells a cell has in its proximity. Degree distribution is plot by counting number of cells that has certain number of neighbors. The distribution is drawn for each frame. In degree distribution graph, there is a clear pattern where the distribution peak moves from left to right from frame 0-137. Theoretically, if all yeast cells have identical size, then the maximum degree will be 6. However, because not all cells are identical in size, the degree can reach as high as 8.
3.4. Speed distribution

Each cell’s current speed is obtained by calculating the pythagorean distance from the cell’s current position to the cell’s previous position and dividing it by 1 (because 1 frame is the unit time in the video). The speed distribution is plot by counting the number of cells that are in the range of 0.25 px/frame steps. The speed distribution is plot for each frame. The highest speed is 4.81px/frame and the lowest speed is 0. After the distribution is plot from frame 0 to 137, a slight pattern is found where the distribution peak moves from left to right repeatedly from frame 0-65, frame 70-115, and frame 120-138.

3.5. Model

Three models are generated using Python’s library NetworkX. Each model is generated with both colonies containing identical number of cells. The generated model is visualised with spring layout. Spring layout assumes the interaction of 2 colonies as 2 particles that are connected with a spring, thereby making the distinction of two colonies clearer because it maximises the distance of two colonies while minimises the distance of members of a colony.[14] The barbell network model on
Figure 13 is a simple model where 2 identical complete networks are connected by an edge. A complete network is a network where each node on the network is connected to every other node.

Figure 13. Barbell network

Figure 14. Relaxed caveman network

Figure 15. Triangular lattice network
The relaxed caveman network on figure 14 is a network formed from the modification of a colony by removing 1 edge from each colony and using it to connect to the neighboring colony.[17] The triangular lattice network or penny network on figure 15 comes from the assumption that the cell members of the observed colony in 2 dimension will be identical in size and shape.[18] Just like coins arranged on a table, the assumption follows that the maximum degree a cell can have will be 6.

Even though at first glance the barbell network and relaxed caveman network are similar in shape to the actual yeast colony network, the parameters in table 1 shows that the triangular lattice network is structurally closest to the actual colony network.

Table 1. Parameter comparison between actual and generated graph

|                      | Actual network (bottom colony) | Barbell network | Relaxed caveman network | Triangular lattice network |
|----------------------|--------------------------------|----------------|-------------------------|----------------------------|
| Nodes                | 42                             | 42             | 42                      | 42                         |
| Edges                | 92                             | 861.5          | 861                     | 101                        |
| Average degree       | 4.3810                         | 10.25595       | 41                      | 4.8095                     |
| Greedy vertex coloring | 5                              | 21             | 42                      | 4                          |

4. Conclusions
FIJI’s TrackMate can be used as a tool to observe yeast colony growth even if there is some shifting from the actual cell growth. The most suitable graph model for the *Saccharomyces cerevisiae* colony observed is the triangular lattice network model. There are noticeable patterns on the degree distribution and speed distribution on actual cell network that are potential to be used as new parameters for cell growth modelling. Further testing on time series parameters needs to be done so that parameter accuracy can be increased.

Acknowledgements
The author would like to express gratitude to Dr. rer. nat Sparisoma Viridi and other authors who have been helping to generate ideas and guide the writing of this article. This paper is funded by Faculty of Mathematics and Natural Sciences, Bandung Institute of Technology.

References
[1] Liu G R and Liu M B 2003 Smoothed Particle Hydrodynamics *A Meshfree Particle Method* (World Scientific)
[2] Peck S L 2004 Simulation as experiment: A philosophical reassessment for biological modeling *Trends Ecol. Evol.* 19(10) 530-534
[3] Aprianti D, Khotimah S N and Viridi S 2016 Budding yeast colony growth study based on circular granular cell *J. Phys. Conf. Series* 739(1) doi:10.1088/1742-6596/739/1/012026
[4] Aprianti D, Haryanto F, Purqon A, Khotimah S N and Viridi S 2016 Study of budding yeast colony formation and its characterizations by using circular granular cell *J. Phys. Conf. Series* 694(1) 1-5
[5] Aji D P 2018 Simulasi Stokastik Pertumbuhan Sel dan Koloni Ragi dengan Model Partikel Granular Berbentuk Lingkaran untuk Kasus Peluang Pertumbuhan dan Kelahiran Bergantung Posisi (Master Thesis: Institut Teknologi Bandung)
[6] Wambaugh J F 2010 Simple models for granular force networks *Phys D Nonlinear Phenom.*
239(18) 1818-1826

[7] Newman M 2010 Networks: An Introduction (New York: Oxford University Press)
[8] John Pointon A 2019 An introduction to statistical physics for students (New York: Longman)
[9] Krane K S 2012 Modern Physics (Wiley)
[10] Schindelin J, Arganda-Carreras I, Frise E, et al 2012 Fiji: an open-source platform for biological-image analysis Nat. Methods 9(7) 676-682
[11] Tinevez J Y, Perry N, Schindelin J, et al 2017 TrackMate: An open and extensible platform for single-particle tracking Methods 115 80-90
[12] Buchler Lab 2009 Timelapse Movie of Fluorescent Yeast (USA: Duke University)
[13] Mazo-Vargas A, Park H, Aydin M and Buchler N E 2014 Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy Mol. Biol Cell. 25(22) 3699-3708
[14] Kobourov S G 2012 Spring Embedders and Force Directed Graph Drawing Algorithms http://arxiv.org/abs/1201.3011. Accessed January 3, 2019
[15] Ghosh A, Boyd S and Saberi A 2008 Minimizing Effective Resistance of a Graph SIAM Rev. 50(1) 37-66
[16] Bailey R F and Stevens B 2010 Hamiltonian Decompositions of Complete K-uniform Hypergraphs Discret Math. 310(22) 3088-3095
[17] Fortunato S 2010 Community detection in graphs Phys Rep. 486(3) 75-174
[18] Pisanski T Bridges between Geometry and Graph Theory In In Geometry at Work ed C.A. Gorini (MAA Notes 53: America) pp174-194