The growth of minicircle networks on regular lattices

Y. Diao[†], K. Hinson[†] and J. Arsuaga[*]

[†]Department of Mathematics and Statistics
University of North Carolina at Charlotte
Charlotte, NC 28223
[*]Department of Mathematics
San Francisco State University
San Francisco, CA 94116

Abstract. The mitochondrial DNA of trypanosomes is organized into a network of topologically linked minicircles. In order to investigate how key topological properties of the network change with minicircle density the authors introduced, in an earlier study, a mathematical model in which randomly oriented minicircles were placed on the vertices of the simple square lattice. Using this model the authors rigorously showed that when the density of minicircles increases, percolation clusters form. For higher densities these percolation clusters are the backbones for networks of minicircles that saturate the entire lattice. An important relevant question is whether these findings are generally true. That is, whether these results are independent of the choice of the lattices on which the model is based. In this paper, we study two additional lattices (namely the honeycomb and the triangular lattices). These regular lattices are selected because they have been proposed for trypanosomes before and after replication. We compare our findings with our earlier results on the square lattice and show that the mathematical statements derived for the square lattice can be extended to these other lattices qualitatively. This finding suggests the universality of these properties. Furthermore, we performed a numerical study which provided data that are consistent with our theoretical analysis, and show that the effect of the choice of lattices on the key network topological characteristics is rather small.

AMS classification scheme numbers: Primary 57M25, secondary 92B99.

1. Introduction

Kinetoplast DNA is found in the mitochondria of trypanosomes, parasites that cause disease in humans, livestock, fishes and plants. Kinetoplast DNA is organized into a planar network of topologically linked minicircles (reviewed in [10]). In [2] it was estimated that before replication each minicircle is topologically linked to an average of three other minicircles (i.e. they have mean valence 3). In [3] it was estimated that the valence doubles to six after replication. These "optimal" arrangements of minicircles motivate this study.

In an earlier report, we introduced a model in which randomly oriented minicircles were placed on the vertices of the simple square lattice (Arsuaga et al [6]). Using this model, we rigorously showed that the formation of a network is an event that can be explained by the confinement of minicircles. We showed that upon increasing density minicircles undergo a critical transition at which they form percolating clusters. These percolating clusters serve as backbones for the formation of a unique network that naturally appears at a higher density. We estimated numerically the critical density at which minicircles percolate ($D \approx 0.637 \pm .001$) and the mean density at which the minicircles saturate ($D \approx 1.153 \pm .001$ for large grid sizes). In this model, a mean valence 3 network is already possible at a density of $D \approx .875$. At these low densities our model predicts a rather heterogeneous network structure. Interestingly our model also predicts that the average valence doubles when the number of minicircles doubles. A phenomenon that has been observed in the replication cycle of *Crithidia fasciculata* during which the number of minicircles doubles (from 5000 to 10000) [3] and the average number of minicircles topologically linked to a given single minicircle doubles. Based on these findings, we proposed that a large concentration of DNA minicircles in a confined volume and in the presence of type II topoisomerase (as found in vivo) naturally favor the formation of a network with very high probability and that a possible (evolutionary) pathway for network formation is through a percolation/saturation process by which long chains of minicircles are first formed (percolation) followed by the formation of entire two dimensional structures (saturation).

In [2] it was proposed that before DNA replication the average number of minicircles linked to a given minicircle is 3, and hence a more accurate model would be that of minicircles centered on a honeycomb lattice. After replication the triangular lattice would be the best representation of the topology of the network [3]. Furthermore one can conjecture that histone-like proteins present in the kinetoplast [5] interact in a way that induces the observed structures of minicircles. Therefore an important relevant question one has to answer is whether our initial findings are generally true. That is, are these results independent of the choice of the lattices on which the model is based? Intuitively, one may expect this to be the case, at least in a qualitative sense. Nonetheless, this needs some rigorous analysis. On the other hand, it is much less intuitive on how much an impact a different lattice choice in the model would cause quantitatively on the key topological characteristics of the minicircle network. This paper is aimed at answering this question. Motivated by the biological results we here study two additional lattices, namely the honeycomb and the triangular lattices. We show that the mathematical statements derived for the square lattice can be extended to these lattices qualitatively as well, suggesting the universality of these properties. Furthermore, we carried out extensive numerical

studies and compared the results among the three different lattices. These numerical studies provided data that are consistent with our theoretical analysis, and showed that the quantitative effect of the choice of difference lattices on the key network topological characteristics is in fact rather small.

2. Model descriptions and basic terminologies

2.1. The descriptions of the models

Let us first describe the three models that we will use for comparison purposes in this paper. The first model is the square lattice minicircle (SLM) model. In this model, we start with a square lattice grid and place a randomly oriented unit circle with its center at each grid point. The second model to be considered here is the triangular lattice minicircle (TLM) model and the last model is the hexagonal (or honeycomb) lattice minicircle (HLM) model. As is the case of the SLM model, we start with a triangular or hexagonal lattice grid and place a randomly oriented unit circle with its center at each grid point. In each model, the radius of the minicircles is fixed (here assumed to be 1), and the distance r between neighboring vertices in the lattice is variable. The density of the minicircles in a lattice grid is the number of vertices per unit area, since each lattice vertex is the center of a minicircle. In the SLM model, each minicircle has four immediate neighbors (whose distance to the said minicircle is r). In the TLM model, each minicircle has six immediate neighbors, while in the HLM each minicircle has three immediate neighbors. Of course the minicircles on the grid boundaries are the exceptions. Notice that the minicircle densities vary in the models for the same r. In fact, the density of minicircles is $1/r^2$ in a square lattice grid, $2/\sqrt{3}r^2$ in a triangular lattice grid, and $4/3\sqrt{3}r^2$ in a hexagonal lattice grid. Figure 1 illustrates the triangular and hexagonal lattices used in the paper.

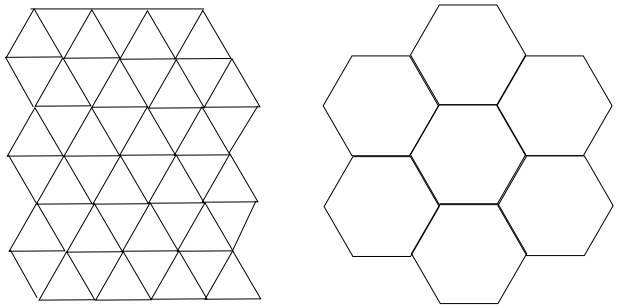

Figure 1. An illustration of a triangular lattice grid (left) and a hexagonal lattice grid (right) with the same r.

In the SLM model, the lattice grid on which our minicircles were arranged was taken to be in the shape of a square. However, this is clearly impossible for the triangular lattice and the hexagonal lattice. Instead, we will use a grid that looks more like a rhombus as shown in Figure 2. The *size* of a lattice grid is then defined as the total number of lattice points in the grid, which is the same as the number of minicircles in the grid if a minicircle is placed at each lattice point.

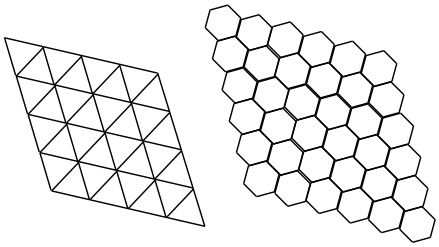

Figure 2. An illustration of the 4×4 triangular lattice rhombus grid and an 8×8 hexagonal lattice rhombus grid.

2.2. The main characteristics of interest

The three models presented here aim at modeling a monolayer of minicircles and are inspired by the minicircles found in Crithidia fasciculata. The minicircles are modeled by geometric circles of equal radius that are randomly oriented, i.e., the normal vectors of the minicircles are (independent) random vectors uniformly distributed on the unit sphere. In each model, the monolayer of minicircles is obtained by choosing a lattice grid (of some suitable size) first, then placing a minicircle at each lattice point in the grid (so that the center of the minicircle coincides with the lattice point). Figure 3 is an illustration of a minicircle grid in the hexagonal lattice.

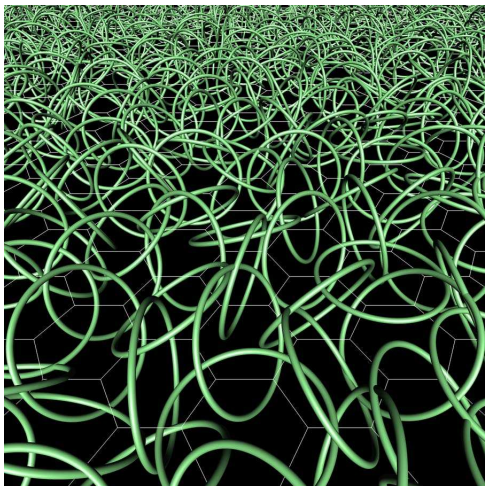

Figure 3. A grid of randomly oriented minicircles in the honeycomb lattice.

The key question one needs to address to characterize the formation of a network is how minicircles are linked to each other. Two minicircles are *linked* if one cannot separate them without breaking them first. More generally, a set of minicircles form a *linked cluster* if one cannot separate them into two (non-empty) parts without breaking any of the minicircles. The linking between two minicircles can be computed easily and details can be found in [6]. Given a square, triangular or hexagonal grid of minicircles, a linked cluster of minicircles that contains minicircles on the opposite boundaries of the grid is called a *percolating cluster*. We say that a grid of minicircles

forms a *minicircle network* when a percolating cluster exists in the grid. In the case that we are not certain whether the minicircles on the given lattice grid has formed a network, we will call the system a *minicircle grid* instead. The *critical percolation density*, or just *critical density*, is a positive value D_c with the following property: if the density $D > D_c$, then there exists a positive constant $\alpha > 0$ (which depends on D, of course), such that the probability for a minicircle grid to form a network is at least α, regardless of the dimension of the grid. On the other hand, if $D < D_c$, then the probability of the formation of a network goes to zero as the dimension of the grid goes to infinity. When most minicircles in a minicircle network fall into the same linked cluster, we say that the network *saturates*. Here 99% saturation level means that there exists a linked cluster in the minicircle network that contains at least 99% of the minicircles in the grid. In this study, a saturated minicircle network means a network with a 99% saturation level. For a given minicircle in the minicircle grid, its *valence* is defined as the number of minicircles that are linked to it directly. The *mean valence of a minicircle* is defined as the average valence of a minicircle over a set of different networks and the *average valence of a minicircle grid* (in which the minicircles are fixed) is the average of the valences of all minicircles in the grid.

2.3. Linking determination

The linking probability of two randomly oriented minicircles (which are of unit radius) can be computed by the following formula [7]:

$$p_r = \begin{cases} 1 - r/2, & 0 < r < 2, \\ 0, & r \geq 2, \end{cases} \tag{1}$$

where r is the distance between the centers of the two minicircles. In the case that the minicircles have been generated, one can check their linking directly using a simple geometric approach, see [6].

3. Theoretical results

There are three main theoretical results given in [6] for the SLM model, namely the existence of a (finite) critical percolation density, the inevitability of network saturation (as density goes to infinity) and the linearity between the network valence and the minicircle density. These results are all valid and the proofs are similar. One only needs some small modifications to accommodate the different lattice structures. For this reason, we will simply state these results in the following while refer our readers to [6] for the original proofs.

Theorem 1 *There exist critical percolation densities D_H and D_T for the honeycomb and triangular lattices respectively. That is, if $D > D_H$ (or D_T) then the minicircle network percolates with a positive probability. On the other hand, if $D < D_H$ (D_T) then as the size of the grid goes to infinity, the probability that a percolating cluster exists goes to zero. Furthermore, $1/2\sqrt{3} < D_T$ and $1/3\sqrt{3} < D_H$.*

The reason for $1/2\sqrt{3} < D_T$ and $1/3\sqrt{3} < D_H$ is that at the densities $1/2\sqrt{3}$ and $1/3\sqrt{3}$ respectively, the triangular and hexagonal lattice grids both have $r = 2$. Thus there will be no linking according to equation (1).

Theorem 2 *Under the TLM and HLM models, the probability for the minicircle grid to be completely saturated (*100% *saturation) goes to one as the minicircle density goes to infinity and the size of the grid stays fixed.*

Theorem 3 *Under the TLM and HLM models, the mean valence $E(V)$ of any given minicircle is of the order of $O(D)$. More specifically, there exist constants $b > a > 0$ such that $aD \leq E(V) \leq bD$ when D is large. Although $E(V)$ does not grow strictly linearly with D, the overall relation is a linear one.*

4. Numerical methods

For the sample generation, we used (rhombus shaped) grids of sizes from 100×100 to 1000×1000 for the case of triangular lattice and from 100×100 to 700×700 for the case of hexagonal lattice. Notice that at 1000×1000 for triangular lattice grid and 700×700 for the hexagonal lattice grid, the total numbers of minicircles in the two minicircle grids are roughly equal, hence the data obtained from each case are the same range for the comparison purposes. For each grid size, we generate 1000 minicircle lattice grids. The linking relation among the minicircles is (pairwise) determined and recorded for various r values. A snapshot of this process is illustrated in Figure 4 using the SLM model (for a grid of relatively small size). This clearly shows the minicircle grid has very little connection (in terms of linking) at low density, but progresses to a network with a percolating cluster at a higher density (usually higher than the critical percolation density), then becomes saturated as the density passes beyond a certain point.

5. Numerical Results

For comparison purposes, we will include results from the SLM model as well.

5.1. Comparison of the critical percolation densities

The numerical estimation of the critical percolation density for each model is illustrated in Figures 5 to 7. In the TLM model, we considered triangular lattice regions of size $N \times N$ for N ranging from 100 to 1,000, and in the HLM model we considered hexagonal lattice regions of size $N \times N$ for N ranging from 100 to 700. In both models we varied N at increments of 100, and in each case ran 1,000 independent trials. For each sample minicircle grid generated, its percolation density is obtained by gradually increasing its density (obtained by decreasing the r value where r is the distance between neighboring vertices on the lattice) until the first percolating cluster occurs. Computationally, it is easier to use the following equivalent approach. We fix $r = 1$ and gradually increase the minicircle radius until the first percolating cluster occurs. Let R be the corresponding radius when the first percolating cluster occurs. The corresponding percolation density is then computed by using the formulas $D_T = \frac{2}{\sqrt{3}}R^2$ for the TLM model and $D_H = \frac{4}{3\sqrt{3}}R^2$ for the HLM model. The critical percolation density is then estimated by averaging these percolation densities over the 1,000 samples. In each of Figures 5 to 7, the horizonal axis stands for the lattice size where 100 means the grid is of dimension 100×100, and so on. The vertical axis is the minicircle density. The error bars shown in the figures are the 95% standard errors.

Numerically, we estimated that $D_S \approx 0.637 \pm .001$, $D_T \approx 0.637 \pm .001$ and $D_H \approx 0.726 \pm .001$. Notice that D_S and D_T are almost identical numerically while D_H is slightly larger.

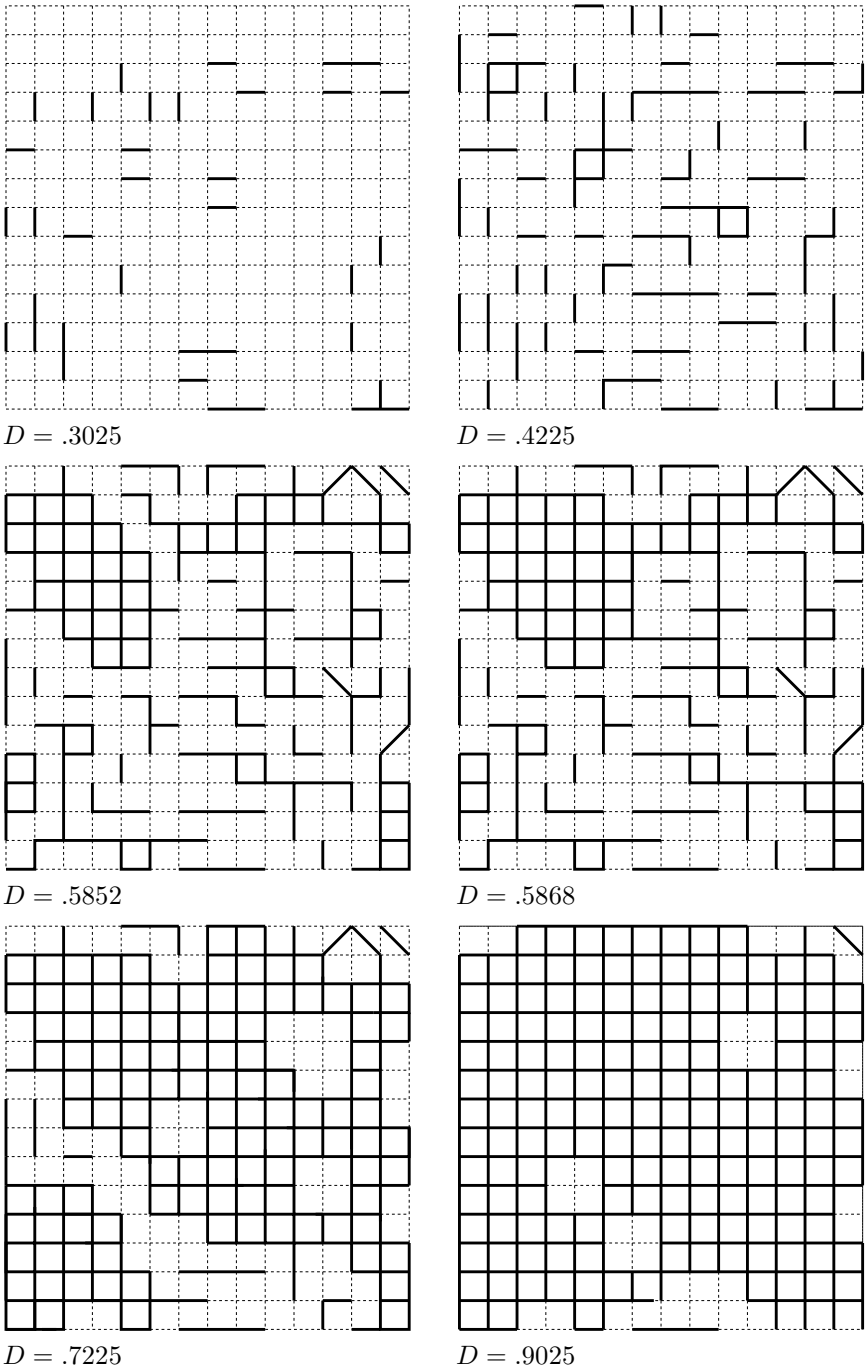

Figure 4. The progresses of the linked clusters as the density increases. The change from density $D = .5852$ to $D = .5868$ is when the minicircle grid becomes percolated. A solid line segment between two lattice points indicates that the two corresponding minicircles belong to the same linked cluster, however they may not be directly linked. Notice that $D_S \approx .637$ is not in the range $[.5852, .5868]$, an artifact due to the small grid size used in this illustration.

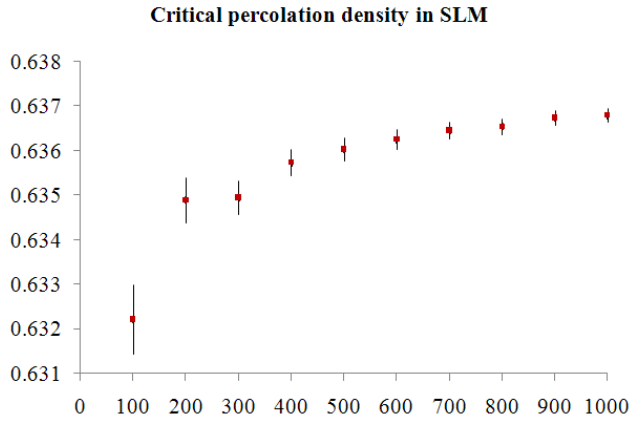

Figure 5. Estimation of percolation densities for the SLM as a function of the lattice size.

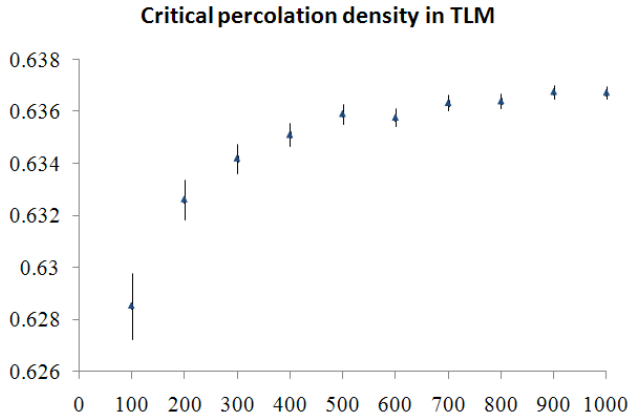

Figure 6. Estimation of percolation densities for the TLM as a function of the lattice size.

5.2. Comparison of the mean saturation densities

Next we carried out a similar study to estimate the density at which 99% saturation was attained in each of the models. As with the percolation study, we gradually increased the minicircle radius R until 99% of minicircles were found to be linked in a single cluster. The results are illustrated in Figures 8 to 10. As is the case of the critical percolation density graphs, in each of these figures, the horizonal axis stands for the lattice size where 100 means the grid is of dimension 100×100, and so on. The vertical axis is the minicircle density. The error bars shown in the graphs are the 95% standard errors.

Numerically, we estimated that the 99% saturation densities to range from $1.183 \pm .002$ (for grid size 100×100) to $1.153 \pm .001$ (for grid size 1000×1000)

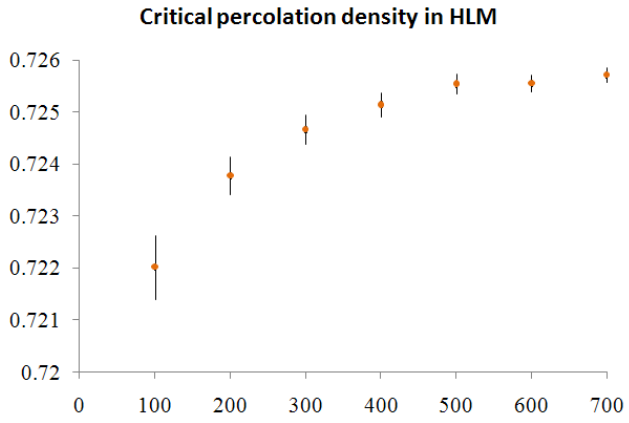

Figure 7. Estimation of percolation densities for the HLM as a function of the lattice size.

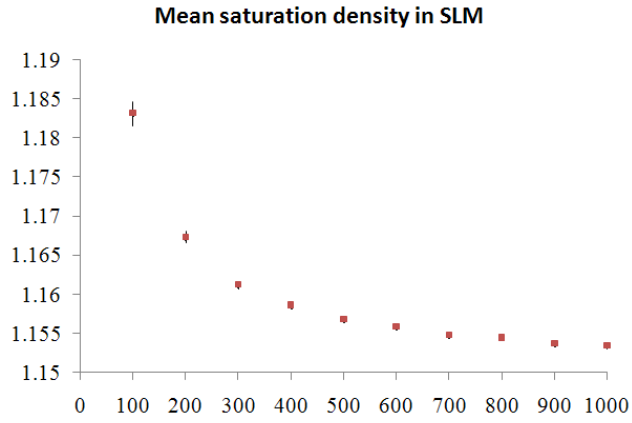

Figure 8. Estimation of mean saturation densities as a function of the lattice size in the SLM model.

in the SLM model, from $1.181 \pm .002$ (for grid size 100×100) to $1.157 \pm .001$ (for grid size 1000×1000) in the TLM model, and from $1.119 \pm .002$ (for grid size 100×100) to $1.094 \pm .001$ (for grid size 700×700) in the HLM model. Notice that in each case, the mean saturation decreases as the grid size increases and a horizontal asymptotic behavior is clearly seen.

5.3. Comparison of the mean valences

Our last set of numerical results concern the mean valence of the minicircle grids. Recall that the valence of a minicircle grid is the average of the valences of the minicircles in the grid (which is a random variable since the minicircles have random orientations). The mean valence of a minicircle grid at a given density can thus be

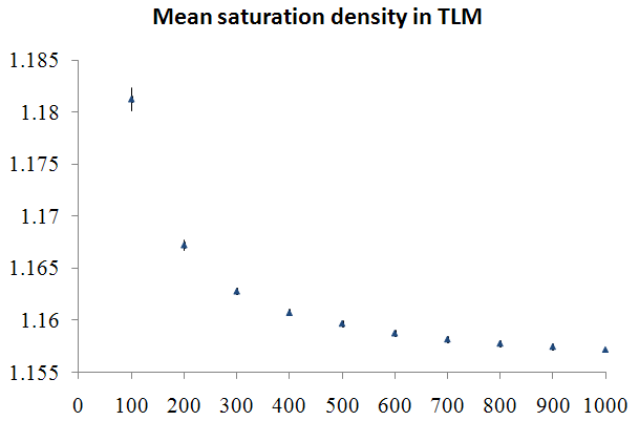

Figure 9. Estimation of mean saturation densities as a function of the lattice size in the TLM model.

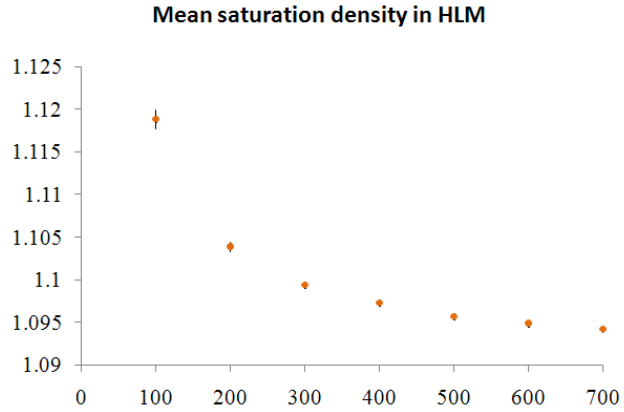

Figure 10. Estimation of mean saturation densities as a function of the lattice size in the HLM model.

numerically estimated by generating 1000 minicircle grids at the given density and then computing their average valence. By Theorem 3, we expect to observe a linear relation between the mean valence and the minicircle density in each model. The key question is how the models can alter the linear regression coefficients. The results are illustrated in Figures 11 to 13. In each of these figures, the horizontal axis is the minicircle density while the vertical axis stands for the mean valence. The regression line is shown in each case together with its corresponding regression line equation and the R^2 value of the fit. Notice how close the coefficients in the regression line equations are in all three models.

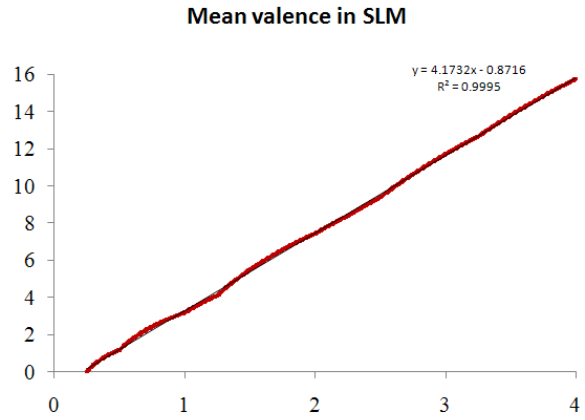

Figure 11. Estimation of mean valence as a function of the minicircle density in SLM.

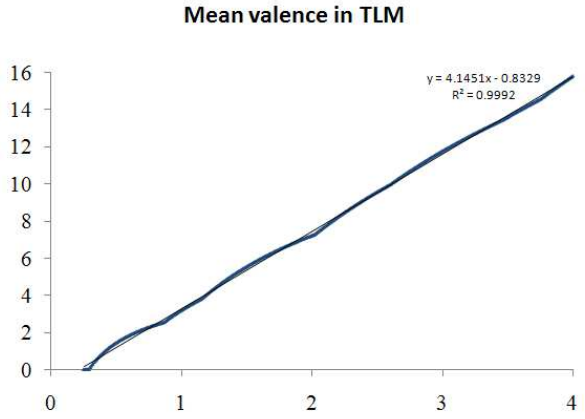

Figure 12. Estimation of mean valence as a function of the minicircle density in TLM.

6. Discussion

While one can intuitively expect that the results obtained in [6] to hold qualitatively under models based on different lattices (and we can indeed analytically prove these results as we indicated in Section 3), it is far from obvious how different lattices can change these results quantitatively. We addressed this question here by carrying out analytical and numerical studies similar to those in [6], but under two new models based on two different regular lattices.

First, in the case of the critical percolation density, the SLM model and TLM model produced very similar results: $D_S \approx 0.637 \pm .001$ and $D_T \approx 0.637 \pm .001$. The HLM model, on the other hand, has a slightly higher critical percolation density: $D_H \approx 0.726 \pm .001$. The likely reason behind these results is that a minicircle has only

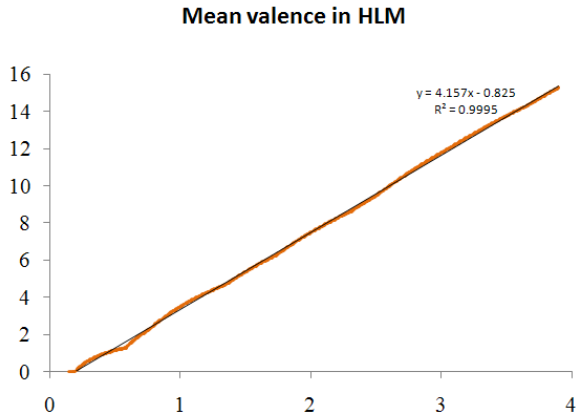

Figure 13. Estimation of mean valence as a function of the minicircle density in HLM.

three nearest neighbors in the HLM model, which is the fewest among any regular lattice. The surprising fact is that the critical percolation density stayed the same when this (connectivity) number changed from 4 (for the SLM model) to 6 (for the TLM model). We thus conjecture that even if a model is based on an irregular lattice (where lattice vertices do not have the same number of immediate neighbors), the critical percolation density will stay within a close range of the interval $[.637, .726]$, as long as the number of immediate neighbors of any lattice vertex is 3 or more, assuming that the distance between any two immediate neighboring vertices is the same.

Second, in the case of the mean saturation density, the situation is similar to the case of the critical percolation density. The SLM model and TLM model produced very similar results: from $1.183 \pm .002$ (for grid size 100×100) to $1.153 \pm .001$ (for grid size 1000×1000) in the SLM model, from $1.181 \pm .002$ (for grid size 100×100) to $1.157 \pm .001$ (for grid size 1000×1000) in the TLM model. It is interesting to notice here that, unlike the case of the critical percolation density, the HLM model actually has a lower mean saturation density: from $1.119 \pm .002$ (for grid size 100×100) to $1.094 \pm .001$ (for grid size 700×700). This fact suggests that once the density goes over some threshold, it is actually easier for the minicircles to fall into the same linked cluster in the HLM model, which must be due to the geometric shapes of the lattice. We also observe that in all cases the 99% saturation level is achieved at surprisingly low densities.

Finally, in the case of the mean valence, our simulation shows a remarkably similar and strong linear behavior in all cases. The linear regression fits all have R^2 values exceeding .999. Furthermore, the three linear regression line equations are $y = 4.1732x - .8736$, $y = 4.1451x - .8329$ and $y = 4.157x - .825$ for the SLM, TLM and HLM models respectively (where x is the density and y is the mean valence). The corresponding coefficients in all three cases agree up to the first digit after the decimal point. One may expect an even better agreement if a larger sample size or/and a longer range of density is used. Notice that in our linear regression we used density values up to 4, way beyond the 99% saturation level. It is clear that the density alone

becomes the dominating force in the case of mean valence, regardless of which lattice the model is based on.

Let us end this paper with an outlook into the future study of this subject. The simplicity of the models we have studied so far allowed us to carry out simulations and rigorous analyses. Of course, our models are quite limited at this time and are only intended for modeling minicircle networks similar to those observed in the mitochondrion of Crithidia fasciculata where minicircles are not supercoiled. This somewhat unusual situation for circular DNA is believed to be a key factor contributing to the formation of minicircle networks. In fact, the loss of supercoiling has been proposed as an essential step in the evolution of the kinetoplast network ([11]). On the other hand, experimental and theoretical studies by several authors have shown that supercoiling inhibits catenation (and promotes decatenation) [12, 13, 19]. In the context of our model one would predict an inhibition of network formation upon the addition of supercoiling. In any case more realistic models are needed for specific physical and biological systems. Towards building a more realistic model of the kinetoplast DNA the authors have recently investigated the effect of the tilting of minicircles in the formation of the network. The authors found that minicircle orientation is a key parameter in the formation of networks and may increase sharply percolation and saturation densities [1]. Also An extension of this work to a 3D setting may be applicable whenever the linking of a large number of topological circles occurs. For example, the so-called "olympic gels" in polymer science [4, 14, 15, 16, 17], the formation of topologically linked protein capsids [8, 18] and the formation of networks by type II topoisomerases [9].

Acknowledgement This work is partially supported by NSF grants DMS-0920887 (J. Arsuaga), DMS-0920880 (Y. Diao and K. Hinson), DMS-1016460 (Y. Diao) and NIH grant 2S06GM52588-12 (J. Arsuaga).

References

[1] Arsuaga J, Diao Y and Hinson K 2011, *Journal of Statistical Physics*, in press.
[2] Chen J, Rauch CA, White JH, Englund PT and Cozzarelli NR 1995, *Cell* **80**(1) 61–9.
[3] Chen J, Englund PT and Cozzarelli NR 1995, *EMBO J* **14**(24) 6339–47.
[4] De Gennes PG 1979, *Scaling Concepts in Polymer Physics*, Cornel University Press.
[5] Deng JS, Rubin RL, Lipscomb MF, Sontheimer RD and Gilliam JN 1984, *Am J Clin Pathol* **82**(4) 448–52.
[6] Diao Y, Hinson K, Kaplan R, Vazquez M and Arsuaga J 2011, *J Math Biology* DOI: 10.1007/s00285-011-0438-0.
[7] Diao Y and Rensburg Janse Van 1998, *Topology and Geometry in Polymer Science* (eds Whittington S G et al), IMA Volumes in Mathematics and its Applications **103** 79–88.
[8] Duda RL 1998, *Cell* **94**(1) 55–60.
[9] Kreuzer KN and Cozzarelli NR 1980, *Cell* **20**(1) 245–54.
[10] Liu B et al 2005, *Trends Parasitol* **21**(8) 363–9.
[11] Lukes et al 2002 *Eukaryotic Cell* **1** (4) 495–502.
[12] Marko J F 1999 *Phys. Rev. E*, **59**, 900–12.
[13] Martinez-Robles ML et al. 2009 *Nucleic Acids Res.*, **37**, 2882–93.
[14] Müller-Nedebock K K and Edwards S F 1998, *J Phys A Math Gen* **32** 3283–300.
[15] Pickett G T 2006, *Europhys Lett* **76** 616–22.
[16] Raphaël E, Gay C and de Gennes P G 1997, *J Stat Phys* **89** 111–8.
[17] Shafi KVPM et al 1999, *J Phys Chem B* **103**, 3358–60.
[18] Wikoff W R et al 2000, *Science* **289**(5487) 2129–33.
[19] Witz et al. 2010 *Nucleic Acids Res.*, **38**, 2119–33.