The effect of volume exclusion on the formation of DNA minicircle networks

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

[†]Department of Mathematics and Statistics
University of North Carolina at Charlotte
Charlotte, NC 28223
[*]Department of Mathematics &
Department of Molecular and Cellular Biology
University of California at Davis
Davis, CA 95616

Abstract. The mitochondrial DNA of Trypanosomes, known as kinetoplast DNA (kDNA), is organized into several thousands of minicircles that are topologically linked, forming a large chainmail-like network. How and why minicircles form a network in some but not in other kinetoplastid organisms, remain unanswered questions. Motivated by these questions, in our earlier studies we introduced some simple analytical and numerical models to study networks of topologically linked minicircles. In these earlier studies, we used three key parameters, namely the mean minicircle valence (*i.e.* the number of minicircles topologically linked to any given minicircle in the network), the critical percolation and mean saturation densities to characterize the topological properties of the minicircle network and how these properties change with respect to the minicircle density changes. Using these models we showed, both theoretically and numerically, that high minicircle density leads to the formation of a linked minicircle network. Our previous studies however did not incorporate the DNA excluded volume due to electrostatic interactions and one would ask how these descriptors of the network change in this new scenario. We here characterize, using a numerical approach, the effects of volume exclusion on the properties of the network. We find that (1) the linking probability of two minicircles does not decrease linearly with the distance between the two minicircles, (2) the mean valence grows linearly with the density of minicircles and decreases with the thickness of the excluded volume, (3) the critical percolation and mean saturation densities grow linearly with the thickness of the excluded volume. Our results therefore validates our descriptors and suggest that the role of volume exclusion on the formation of kDNA networks is milder than what was initially proposed.

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

1. Introduction

Trypanosomatids are protozoa parasites that cause fatal diseases in humans and livestock with significant negative impact on the societal wellbeing and economies of many third world countries. A distinctive feature of these organisms is the organization of their mitochondrial DNA, called kinetoplast DNA (kDNA), into two types of circular structures: maxicircles and minicircles. Maxicircles contain the mitochondrial genes and minicircles encode transcripts that edit the product of the maxicircles [3]. The lengths of maxi and minicircles are different for different organisms. The length of maxicircles ranges between 20 and 40 kb and of minicircles between 1 and 2.5 kb. Typically, in a single cell, there are a few dozen maxicircles and a few thousand minicircles confined in a structure called the kinetoplast. While the topological structure of the kinetoplast DNA in free living organisms is mostly trivial, it becomes really intricate in parasitic species. In *Crithidia fasciculata*, minicircles are topologically linked into a planar, chain-mail like structure in which every pair of minicircles links to three or six of its nearest neighbors through the simplest kind of topological linkage: the Hopf link [4, 5]. The origin and maintenance of the network have remained mainly unexplored from a quantitatively view point.

In our earlier work, we presented a mathematical model to study how different biophysical factors of the network affect the its topological characteristics. In a series of studies we analyzed the effects of confinement, the minicircle orientation, relative positions of minicircles, and minicircle flexibility on network formation and on the topological characteristics of the networks [1, 2, 6, 7, 9]. Our results showed that, even at a relatively low minicircle density, slightly over $D \approx .637$, a percolating chain (cluster) already forms. By further increasing the confinement so that the minicircle density is increased to $D \approx 1.20$, a large percentage of minicircles fall into a single topologically linked chain, a phenomenon observed in *C. fasciculata*. Our results showed that the mean valence (*i.e.* the average number of minicircles topologically linked to a minicircle) increases linearly with the density of minicircles, a property in agreement with experimental data [4, 5]. These descriptors of the network (i.e. mean valence, percolation and saturation densities) were also found when the afore mentioned factors, however the values were shown to be different for the different conditions imposed.

In this work we study the impact of the effective helical diameter of the DNA molecule on the topology of the network. This work is motivated in part by the discussion in [4, 5] which suggested that the structure of the network was determined, at least partially, by the volume occupied by the DNA minicircles. To isolate the effects of DNA volume exclusion, we used our initially proposed model in which minicircles are randomly oriented with their centers placed on the vertices of the simple square lattice [7]. The incorporation of volume exclusion has some interesting impacts: (1) the linking probability of two minicircles changes in an nonlinear fashion with the distance between the centers of the two minicircles. A property that has been previously reported [10], (2) the mean valence grows linearly with the density of minicircles and decreases with the thickness of the excluded volume, (3) the critical percolation and mean saturation densities grow linearly with the thickness of the excluded volume.

Our results therefore show that the mean valence, percolation and saturation are good descriptors of the topology of the network since they can be estimated under a wide variety of conditions and at the same time are sensitive to the different restrictions

imposed on the network. The preserved linear relationship between minicircle density and mean valence in the presence of volume exclusion shows that volume exclusion is consistent with the experimentally observed properties of the network. However the linear response of these descriptors suggest that the volume exclusion of the DNA molecules has a rather limited impact on the topology of the network characteristics of the minicircle network.

2. Model descriptions and key topological characteristics of interest

To study the effect of minicircle volume exclusion on network formation we extend the square lattice minicircle (SLM) model developed in [7]. In the SLM model each DNA minicircle is modeled by a unit circle with its center uniquely occupying a vertex of the simple square lattice. The orientation of each minicircle is determined by its normal vector, which is uniformly distributed on the unit sphere. The density of minicircles is then increased by reducing the size of the lattice edges. To make the model incorporating volume exclusion computationally feasible we substituted unit circles by regular polygons. In this case, a regular polygon is one whose vertices are equally spaced along a circle. The number of edges of the polygon is an adjustable parameter, but the length of the polygon is the same as that of a unit circle (2π) so we can compare our results obtained here with our previously published results [7]. The volume exclusion effect was implemented as an impenetrable cylinder around each segment of diameter τ. In our studies the radius (thickness) of the cylinders ranged from 0.01 to 0.05. To facilitate the discussion of our results in this work we will use the term *minicircle* from this point on to refer to a regular polygon of a given thickness. Figure 1 is an illustration of two minicircle grids in the simple square lattice with different values of the parameter τ.

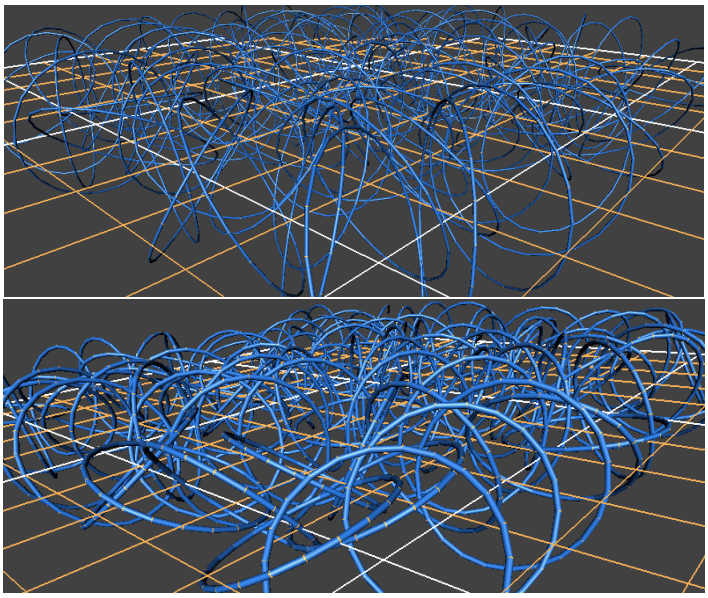

Figure 1. Minicircle grids modeled by regular polygons with thickness 0.0125 (top) and 0.03 (bottom).

Now let us briefly revisit the basic definitions needed to describe the topological properties of a minicircle network (for more details the reader is referred to [7]). A *minicircle grid* is a region of a planar lattice with minicircles at its vertices. Minicircle *density* is defined as the number of minicircles per unit area of the minicircle grid. If d is the distance between two adjacent grid points, then the density of the minicircle grid is given by $1/d^2$. The topology of the network is obtained by determining the topological link between all neighboring minicircle pairs. The *valence* of a minicircle is the number of minicircles that are topologically linked to it and the *mean valence* of a minicircle is its average valence over the entire ensemble of minicircle grids.

A linked cluster is a set of minicircles in which not any part can be separated from the rest without breaking any minicircles in the cluster. A linked cluster is a *percolating cluster* if the cluster reaches two opposite boundaries of the grid. An important topological characteristic of a minicircle network is its *critical percolation density* (or just *critical density*). This is a positive value D_c with the following property: if the density $D > D_c$, then there is a positive probability for the formation of a percolating cluster regardless the size of the grid and this probability rapidly increases as D increases; if $D < D_c$, then the probability for the formation of a percolating cluster goes to zero (also rapidly) as the size of the grid goes to infinity. Another important topological characteristic of a minicircle network is its mean *saturation density*. When a large percentage of minicircles in the network fall into the same linked cluster, we say that the minicircle network *saturates* and the corresponding density is called its saturation density. For example, a saturation at the 99% level means that at least 99% of the minicircles in the network belong to the same linked cluster. Saturation is given as a percentage to avoid artifacts introduced by single minicircles whose orientation may drive the saturation density artificially high. The average density at which the minicircle network reaches saturation at a given level is called its mean *saturation density*. In this paper, a 99% saturation level is used.

3. Generation of minicircle grids with volume occupying restriction

In our previous work we introduced algorithms for generating ensembles of minicircle grids and networks [1, 6, 7]. The addition of volume exclusion requires new algorithms since previous methods may generate conformations with overlapping thick minicircles which are invalid. Of course, one may attempt to use an accept/reject approach based on the previous algorithms. That is, minicircle grids are generated using previous methods, and the ones in which some thick minicircles overlap are discarded. Unfortunately, such overlaps occur with very high probability. For example, with a thickness of .0125, $100,000$ runs did not produce any valid 10×10 minicircle grids at density .55 or higher, and only a handful at density .45 and .5. For our purposes, we would need to use minicircle grids of dimension 100×100. Apparently, a simple accept/reject method is out of question. Our approach to get around this problem was to adopt a well established statistical method called *imputation method*. To apply this method we generated *virtual minicircle grids*. Notice that the linking of a minicircle to its neighbors follows a probability distribution and it is plausible that this local probability distribution is the dominate force in determining the topological properties of the minicircle grid. Thus a virtual minicircle grid can be generated based on these local (conditional) probabilities and these conditional linking probabilities (between neighboring minicircles) can be empirically estimated using small minicircle

grids ($n = 5$ or $n = 10$). The number of possible combinations of linked minicircles (and therefore the number of conditional probabilities) was extremely large, however most of them were very small and were neglected.

In order to empirically estimate the conditional probabilities, we generated large samples of 10×10 valid minicircle grids. As we pointed out in the above, this was also a very difficult task. To overcome this problem, we adopted a simulated annealing algorithm. In this algorithm a minicircle grid was first generated (which may or may not be valid) and an energy penalty, given by the number of intersecting minicircles, was associated to this grid. If the energy was positive, a fixed number of minicircles were randomly selected and their orientations changed (according to the uniform distribution). The new configuration was accepted/rejected according to the standard Monte-Carlo criterion. This process was continued until the minicircle grid reached a zero energy state, or was terminated when a preset number of steps had occurred without reaching a zero energy state. In this paper, this number was preset at 300. All the conditional probabilities were empirically estimated from samples of 10×10 minicircle grids and a total number of sampled grids, for each fixed τ and density, between $62,500$ to $100,000$. The reason for the smaller sample sizes was purely technical since the success rate for generating valid grids became very low for large thickness and density.

Once the conditional probabilities were estimated, large virtual minicircle grids were generated using the imputation method. We filled the entire grid by filling consecutive rows and columns with virtual minicircles. Specifically, virtual minicircles were placed starting at $(1,1)$ to $(1,n)$, then from $(2,1)$ to $(2,n)$, and so on, where (i,j) stands for the grid vertex on the i-th row and j-column in the grid. At any given point, when a virtual minicircle was to be placed, its linking with neighboring virtual minicircles already placed was determined based on the estimated conditional probability distribution for the specific conformation of these neighboring minicircles. We illustrate the algorithm with an example. Suppose that we are about to place the virtual minicircle at grid point $(2,3)$. At this point, all virtual minicircles on the first row and at $(2,1)$ and $(2,2)$ have been placed and we know how they are linked to each other. The immediate neighbors of $(2,3)$ (that precede it in our ordering) are $(2,2)$, $(1,2)$, $(1,3)$ and $(1,4)$. Suppose that $(2,2)$ is linked with $(1,2)$, $(1,3)$ is linked with $(1,4)$. Then the linking of $(2,3)$ with $(2,2)$, $(1,2)$, $(1,3)$ and $(1,4)$ is based on the empirical linking probability distribution conditioned on $(2,2)$ being linked with $(1,2)$ and $(1,3)$ being linked with $(1,4)$. Notice when a virtual minicircle not on the boundary is to be placed, that there are 4 neighboring minicircles already being placed, thus the total number of linking conditions to be considered is $2^8 = 256$, even if we only consider the linking among these minicircles. So these are rather large data sets and we shall not provide further details about them.

4. Numerical results

4.1. Linking probability between two neighboring minicircles

We first determined the linking probability of two adjacent minicircles as a function of the distance between their centers. For this analysis we first picked the number of edges representing the minicircle to be equal to 22. Results are as shown in Figure 2.

From previous studies [8] we know that the linking probability for geometric circles

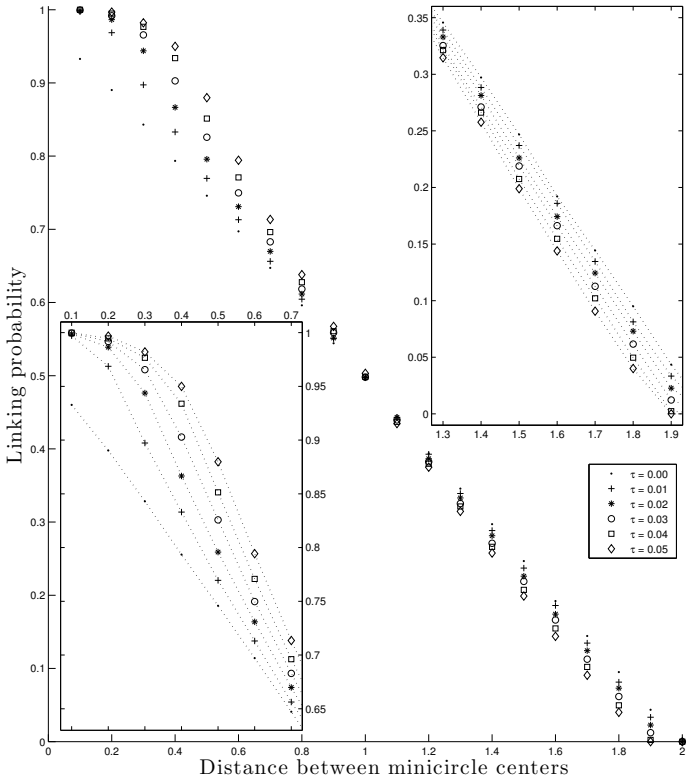

Figure 2. Estimation of the linking probabilities as a function of thickness of and distance between the centers of the polygons. The sample size for each data point is 10,000 and the sizes of the error bars are less than the sizes of the plotted data points. The linking between two minicircles is determined by computing the linking number between using the Gaussian integral formula.

without thickness is $1 - r/2$ when $r \leq 2$ and 0 when $r > 2$, where r is the distance between the centers of the circles. A rigorous formula for this probability seems to be out of reach when thickness is introduced. However, If one considers two minicircles (*i.e.* polygons of length 22) of radius τ some qualitative features of this probability are not that hard to observe and can be used to check the consistency of our numerical results. First, the linking probability of two minicircles without volume exclusion (*i.e.* $\tau = 0$) is closely approximated by the linking probability of geometric circles. But they are not identical since it can be rigorously shown that the linking probability of two such regular polygons does not completely vanish when $r = 2$; indeed our numerical simulations suggest it is about .00007. Second, as $r \to 0$, the linking probability will no longer approach 1. In fact the linking probability becomes zero when r is very small as the two minicircles would overlap each other and become an invalid pair. Similarly the linking probability is zero when $r > 2 - 2\tau$ since the minicircles would overlap. Lastly, at close distances (but 2τ away from zero) the linking probability of the polygons increases with the radii of the minicircles (Figure 2). This trend reverses for distances larger than 1, that is the linking probability of the minicircles decreases with their thickness τ. In fact, this trend becomes more apparent at distances close

to 2 (Figure 2).

To test whether the number of edges in the polygon had a significant impact on the numerical results of the model, we examined how linking probabilities between two thick regular polygons changed with the number of edges. A systematic testing showed that these probabilities had very small variations. The following table is a snapshot of this systematic study. It lists the estimated linking probabilities for 22, 30 and 40 edges and $\tau = .02$ at 4 different distances .2, .7, 1.5 and 1.9 (in that order). Thus we conclude that the number of edges in the polygon has little impact on the overall results of the SLMT model.

n	22	30	40	22	30	40
P(linking)	.986	.987	.986	.670	.672	.669
n	22	30	40	22	30	40
P(linking)	.229	.233	.230	.025	.027	.027

We therefore chose to use polygons of 22 edges and the following rescaling was performed to compare our results obtained here with our previously published results [7]. A 22-edged regular polygon was generated by first selecting a point uniformly selected on the unit circle. This point would completely determine the polygon since the vertices of the polygon are equally spaced on the unit circle. The arc length of a polygon generated this way has length $44\sin(\pi/22)$, which is different from the length of a unit circle (2π). So we re-scale the polygons by a factor of $\pi/(22\sin(\pi/22))$ so that the re-scaled polygons will have the same length as a unit circle.

4.2. Mean valence of minicircles

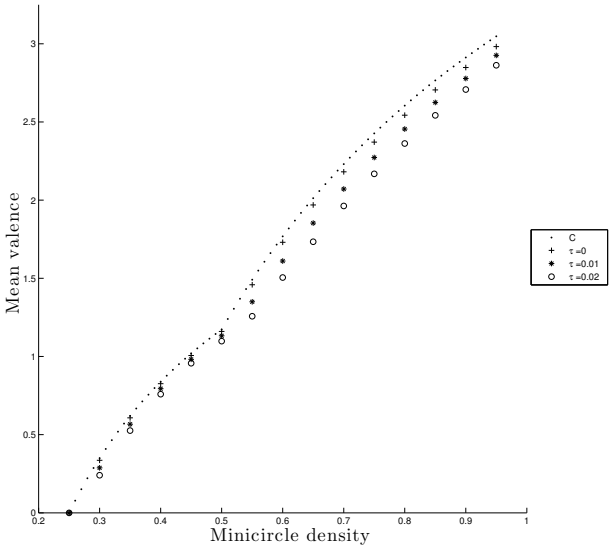

Figure 3. Estimation of mean valence as a function of the minicircle density with various thicknesses. Each data point is based on a sample of 10^6 7×7 minicircle grids. The sizes of the error bars are less than the sizes of the plotted data points.

In our previous studies we found, in agreement with experimental observations, that the mean valence of minicircles grows linearly with the density of minicircles. From a theoretical point of view, we know that this will not be the case once thickness is considered, since it introduces a constant upper bound to the minicircle density. However one may suspect that this linear behavior still holds true at relatively low densities. Furthermore, based on the results obtained above regarding the linking probability between thick minicircles, one would expect a lower mean valence with a higher thickness. This is clearly observed in our numerical results as shown in Figure 3. Notice also that for each given thickness, the mean valence does follow a mostly linear pattern with respect to the minicircle density.

4.3. Estimation of the critical percolation density

In our previous work where minicircles are modeled by unit circles, we established the existence of the critical percolation density rigorously. We also estimated numerically the critical percolation density to be $D \approx 0.637$ [7]. Here we estimated the critical percolation density as a function of the minicircle radius τ (denoted by D_τ). We found that for $\tau = 0$ the critical percolation density $D_0 \approx 0.642$. It is a well known fact that the percolation probability undergoes a rapid phase change when the density crosses the critical percolation density.

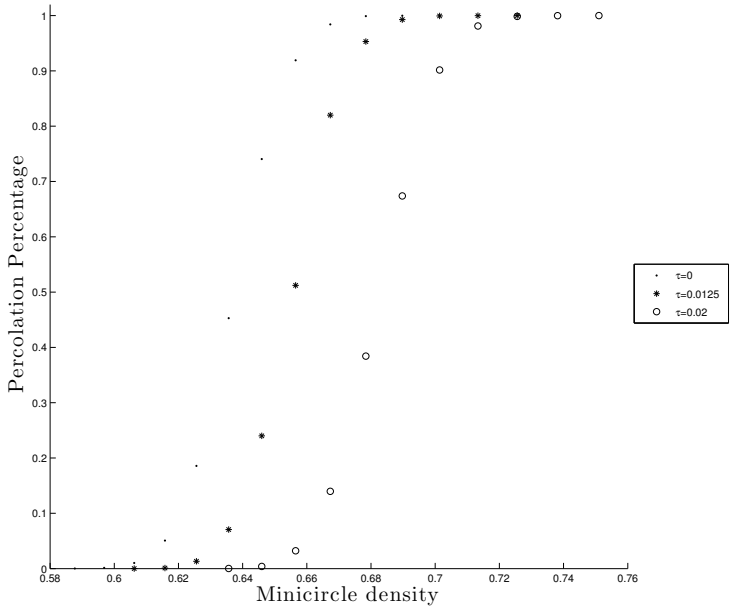

Figure 4. The phase changes in the percolation probability around the critical percolation densities for the cases of $\tau = 0$, $\tau = .0125$ and $\tau = .02$.

Figure 4 shows this phenomenon for the cases $\tau = 0$, $\tau = .0125$ and $\tau = .02$: the percolation probability changes quickly from near zero to near one once the density

goes beyond the critical percolation density increases as τ increases. The numerical results presented here are based on $10,000$ virtual minicircle grids of size 100×100. Furthermore we found that D_τ grows linearly with τ as shown in Figure 5 with a best linear fit equation of $y = 2.2689x + .6424$ (where x is the thickness τ and y is the critical percolation density) and $R^2 = .999$. We notice that within the range of the τ values ($\leq .05$) all critical percolation values D_τ are below 1.

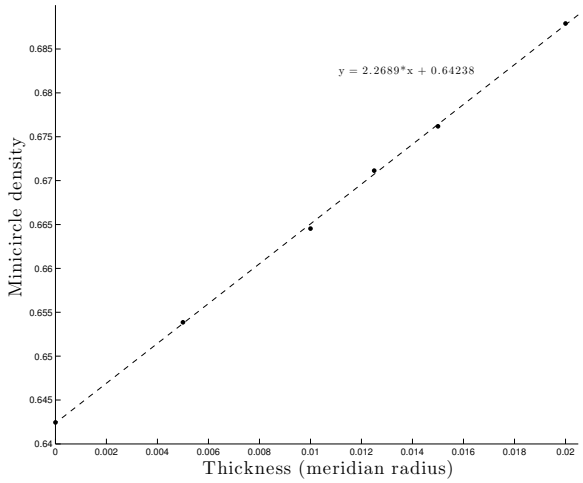

Figure 5. Estimation of critical percolation densities for various thicknesses.

4.4. Mean saturation densities

Next we estimated the mean density at which at least 99% of the minicircles are linked into one single unsplittable linked cluster and the results are shown in Figure 6.

In our earlier work [7] we showed that the probability for a minicircle network to become saturated increases exponentially to 1 with the density of minicircles at any given grid size. Our simulations studies showed that for grid sizes ranging from 100×100 to 1000×1000 the mean saturation density ranges from $1.183 \pm .002$ to $1.153 \pm .001$ and that it decreases as the grid size increases. In the case of the thick minicircles, since the linking probability between minicircles decreases when the thickness τ of the minicircle increases, we expect an increased mean saturation density with a larger τ value. This increased density may cause a small error in our calculations because at these larger densities non-neighboring minicircles may link. Thus, the estimated mean 99% saturation densities obtained here (as presented in Figure 6) are overestimates since we only considered linking between neighboring minicircles. Similarly to the critical percolation density results we observe a strong linear relation between the mean saturation density and the minicircle thickness τ. Each data point presented here is based on $10,000$ 100×100 minicircle grids. At $\tau = 0$ the mean saturation density is estimated at $D \approx 1.3511$ and it increases to $D \approx 1.5193$ as τ increases to .02. The best linear fit $y = 8.2905x + 1.3511$ (with x being the thickness and y being the mean saturation density) has a R^2 value of .999.

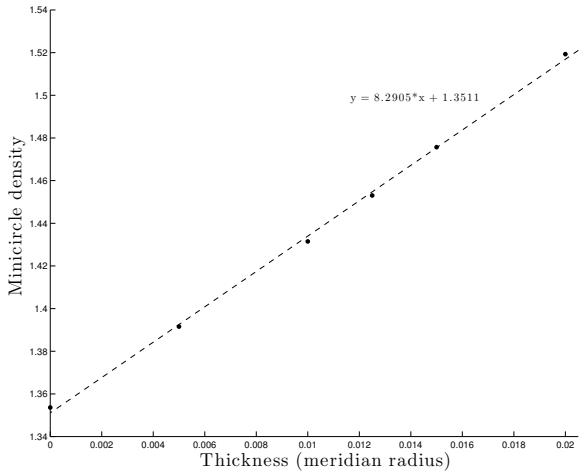

Figure 6. Mean saturation densities at different thicknesses.

5. Discussions and ending remarks

Trypanosomatida are characterized by a number of singular features including the organization of their mitochondrial DNA (kDNA) into large networks of minicircles. In [4, 5] the topology of the kDNA network for *C. fasciculata* was characterized and it was found that minicircles are organized into a planar topological network in which every pair of neighboring minicircles is singly linked and that in a non-replicating network, each minicircle is linked to an average of three or six other minicircles, depending on the cell's replication stage. These same studies proposed that DNA volume exclusion should play an important role determining the topological properties of the network.

In [7] we presented, to our knowledge, the first effort to quantitatively describe kDNA networks. There we showed: (1) the formation of a network of topologically linked DNA minicircles is an unavoidable event when minicircles are highly confined and in the presence of DNA strand passing agents (as found in the kinetoplast), (2) there is a possible evolutionary pathway in which the initial formation of a percolating network is followed by the formation of a saturating network , and (3) the mean value of the valence increases linearly with density. In a series of studies we showed that these topological properties of the network can be estimated when minicircles are positioned in other lattices different from the simple square lattice [6, 9], when minicircles are randomly oriented with certain bias [1], or when minicircles are modeled as flexible objects rather than rigid ones [2]. These studies however did not incorporate the effects of volume exclusion on the formation of the network.

To address this issue we had to introduce new algorithms because the computational complexity of the problem rapidly explodes. While we have not developed a way to test the accuracy and reliability of our new algorithms, the numerical results obtained are quite consistent and convincing. First, we notice that the empirical linking probabilities obtained using the two different methods are very similar to each other. Second, in the case of zero thickness, our estimated critical percolation density (using the imputation approach) is very close to the estimation

obtained using the directly generated minicircle grids. Third, it is possible to compare the minicircle grids to a standard square lattice bond percolation model where each pair of neighboring vertices may be connected by a bond with a probability p and this probability is independent to whether/how these vertices are connected to other vertices. It is known that the critical percolation probability is $p = 1/2$ in this case. In our case, the linking probabilities play a role somewhat similar to the probability p in the square lattice bond percolation model. The difference is that the linking probabilities are not all the same (each minicircle has 8 neighboring minicircles, 4 of which are at a distance farther away) and are not independent of each other. For comparison, we can define a "vertex" to be a grid point, but we say that two vertically or horizontally neighboring vertices are connected by a "bond" if the minicircles centered at these two grid points are either linked directly or linked together through a minicircle that is diagonally neighboring them. By generating large sample of 3×3 minicircle grids directly and compute the empirical linking probability of the minicircle centered at the middle grid point, one can verify that for the thicknesses we studied, the probability that the middle vertex is connected to a vertical (or horizontal) neighboring vertex by a bond in the above sense is always greater than $1/2$ when the minicircle density is close to 1, regardless how the other vertices are connected. This numerically proves that the critical percolation density is indeed less than 1 in all cases that we have covered.

Using these new algorithms in previous studies we were able to compute the critical percolation and saturation densities as well as the mean valence of a minicircle. Overall, our numerical results show that the three key topological characteristics did not change drastically when we added volume exclusion. Interestingly we found that both the critical percolation density and the mean saturation density increase linearly with the radius of the minicircles. The behavior of the mean valence is not that different form that observed previously without volume exclusion. However, at high densities we know the mean valence of the thick minicircles will behave very differently by a rigorous argument as the density is bounded above by a constant and therefore the linking probability becomes zero after the density passes that bound. Therefore we conclude that for minicircles with large radius of gyration such as geometric circles or regular polygons, volume exclusion only slightly increases the critical percolation density and mean saturation density. Finally, these results apply only when minicircles are relaxed. One can envision situations in which volume exclusion may still play a much more prominent role. This includes cases in which minicircles have specific orientations or shapes.

Acknowledgement This work was partially supported by NSF grants DMS-0920887 (J. Arsuaga and K. Hinson), by NIH 1R01GM109457 (J.Arsuaga), DMS-0920880 (Y. Diao and K. Hinson), DMS-1016460 (Y. Diao) and by Chinese NSF grant 91324201 (Y. Diao).

References

[1] J. Arsuaga, Y. Diao, and K. Hinson. The effect of angle restriction on the topological characteristics of minicircle networks. *J. Statist. Phys.*, 146(2):434–445, 2012.

[2] Javier Arsuaga, Yuanan Diao, Michele Klingbeil, and Victor Rodriguez. Properties of topological networks of flexible polygonal chains. *Molecular Based Mathematical Biology*, Submitted.

[3] Rob Benne, Janny Van Den Burg, Just PJ Brakenhoff, Paul Sloof, Jacques H Van Boom, and Marijke C Tromp. Major transcript of the frameshifted *coxll* gene from trypanosome

mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*, 46(6):819–826, 1986.

[4] J. Chen, P. T. Englund, and N. R. Cozzarelli. Changes in network topology during the replication of kinetoplast DNA. *EMBO J.*, 14(24):6339–6347, 1995.

[5] J. Chen, C. A. Rauch, J. H. White, P. T. Englund, and N. R. Cozzarelli. The topology of the kinetoplast DNA network. *Cell*, 80(1):61–69, 1995.

[6] Y. Diao, K. Hinson, and J. Arsuaga. The growth of minicircle networks on regular lattices. *J. Phys. A: Math. Theor*, 45:doi:10.1088/1751–8113/45/3/035004, 2012.

[7] Y. Diao, K. Hinson, R. Kaplan, M. Vazquez, and J. Arsuaga. The effects of minicircle density on the topological structure of the mitochondrial DNA from trypanosomes. *J. Math. Biol.*, 64(6):1087–1108, 2012.

[8] Y Diao and EJ Janse van Rensburg. Percolation of linked circles. In *Topology and geometry in polymer science*, pages 79–88. Springer, 1998.

[9] V Rodriguez, Y Diao, and J Arsuaga. Percolation phenomena in disordered topological networks. *Journal of Physics: Conference Series*, 454(1):012070, 2013.

[10] A. V. Vologodskii and V. Rybenkov. Simulation of DNA catenanes. *Phys. Chem. Chem. Phys.*, 11:10543–10552, 2009.