A New Definition of Distance for Graphs

Project by: Farzan Vafa

Milton Academy, 180 Centre St., Milton, MA, 02186

Coach: Dr. Yong Lin

Department of Mathematics, Harvard University, Cambridge, MA, 02138

A New Definition of Distance for Graphs

Abstract

In this paper, we have determined that the normal definition of distance for graphs is inadequate. Thus, we have defined a new definition of distance for graphs, motivated from quantum mechanics, and we explore this new definition of distance using Mathematica and compare it to the standard one in various examples of graphs. The different types of graphs studied are linear graphs, grid graphs, and Erdős–Rényi graphs.

August, 2010

0.02 ------ 2

1. Introduction and Motivation

A graph G = (V(G), E(G)) is defined by the vertex-set V(G) and the edge-set E(G). For a general introduction to graphs, see [1]. Each edge is an undirected pair of two vertices. We will assume that there are no multiple edges between two vertices, no loops (a vertex that is connected to itself), and that the graph is connected. Such a graph is simple. For example, for the graph G = (V(G), E(G)) in Figure 1:

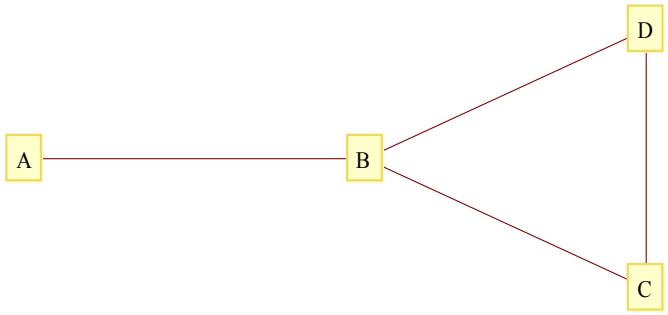

Fig. 1

V(G) = {A, B, C, D}

E(G) = { {A,B}, {B,C}, {B,D}, {C,D} }

Examples range from the collaboration graph to a Hollywood actor collaboration graph [1].

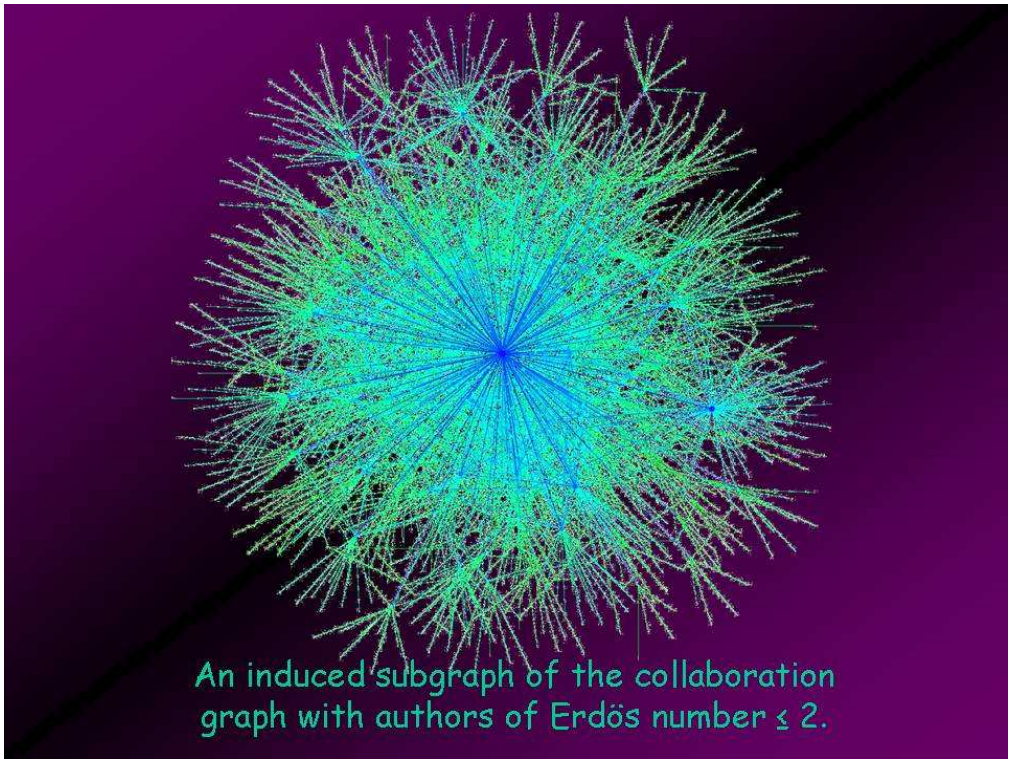

Fig. 2

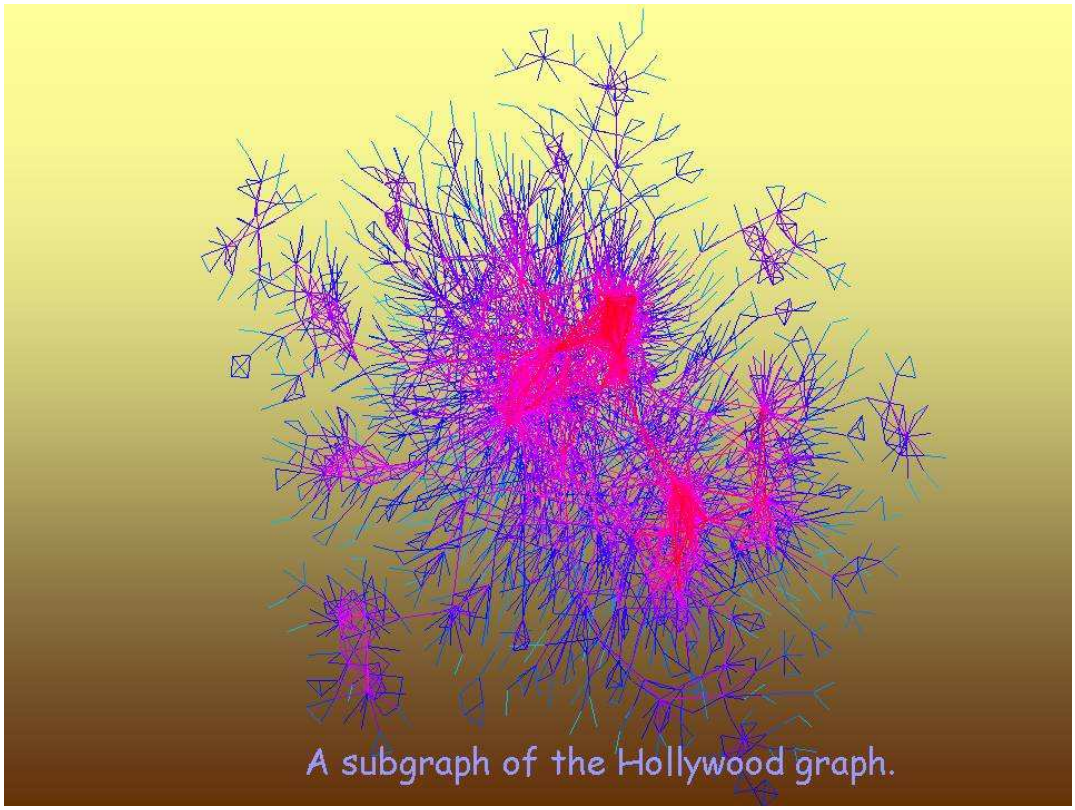

Fig. 3

Even space itself can be considered a graph of sorts, if one takes an infinite number of points and edges, with the points becoming closer to each other:

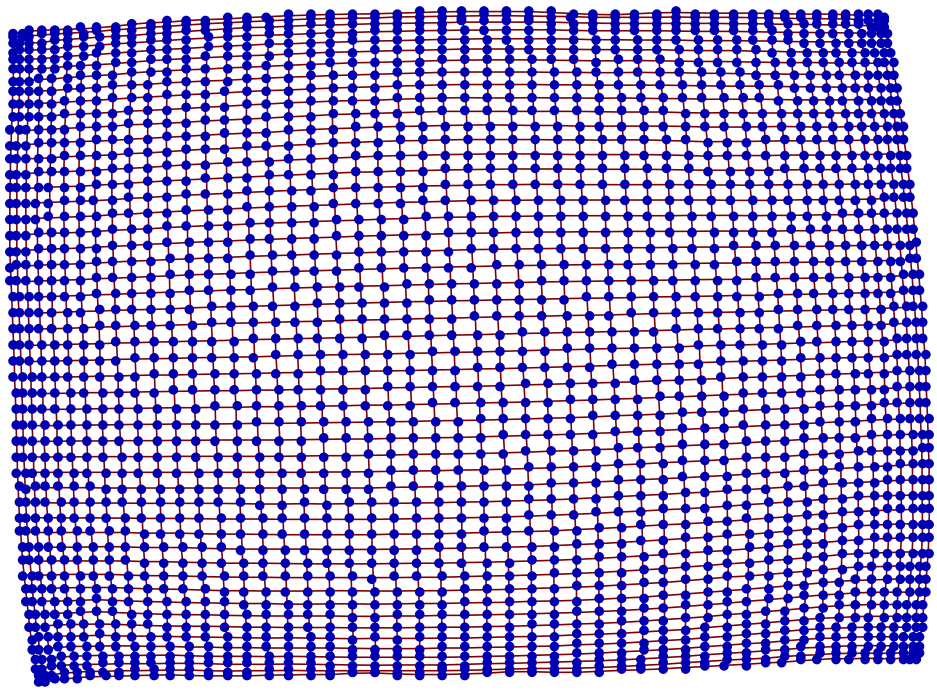

Fig. 4

A question that we can ask is what is the distance in a graph? Based on our intuition of the real world, the distance is merely the shortest path between the two points. But how does that apply to a graph? Traditionally, the distance between two vertices x and y of a graph is defined to be the shortest length of all of the paths from x to y [1]. A path p from x to y of length *l* defined as an ordered sequence of vertices from $v = y_0, y_1, y_2,\ldots,y_l$ that satisfy

$$\{y_i, y_{i+1}\} \in E(G) \quad \text{for } i = 0, 1,\ldots,l\text{-}1$$

and

$$x = y_0, \quad y = y_l.$$

We say that the length of p, $|p| = l$. Let $P_{(x,y)}$ be the set of all paths from x to y. The distance between x and y, d(x,y), is formally defined as follows:

$$d(x, y) = \inf |p|, \ \forall p \in P_{(x,y)}. \qquad (1)$$

For example, in the graph below:

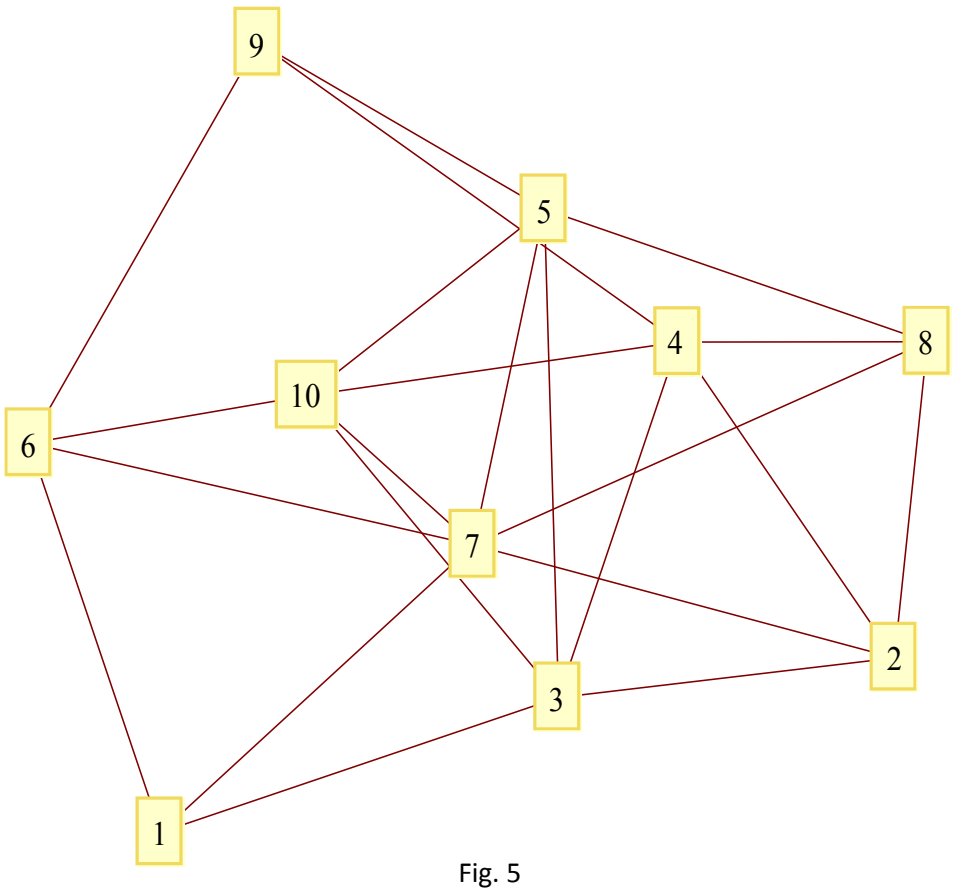

Fig. 5

using this definition of distance, we have d(1,6) = 1 and d(1,5) = 2.

A linear graph is a connected sequence of vertices. A linear graph of 8 vertices is depicted below:

Fig. 6

With this definition of distance, the distance between two vertices is the number of edges from one vertex to another. For example, for the distance between v_i and v_j, where v_i is the i-th vertex and v_j is the j-th vertex, the distance is simply $|j - i|$. For this case, this definition of distance agrees with the intuitive Euclidean notion of distance. But in the example of an (*m x n*) grid graph, where m is the number of rows and n is the number of columns, this definition of distance does not agree with the Euclidean one. For example, consider the *10 x 10* grid graph below:

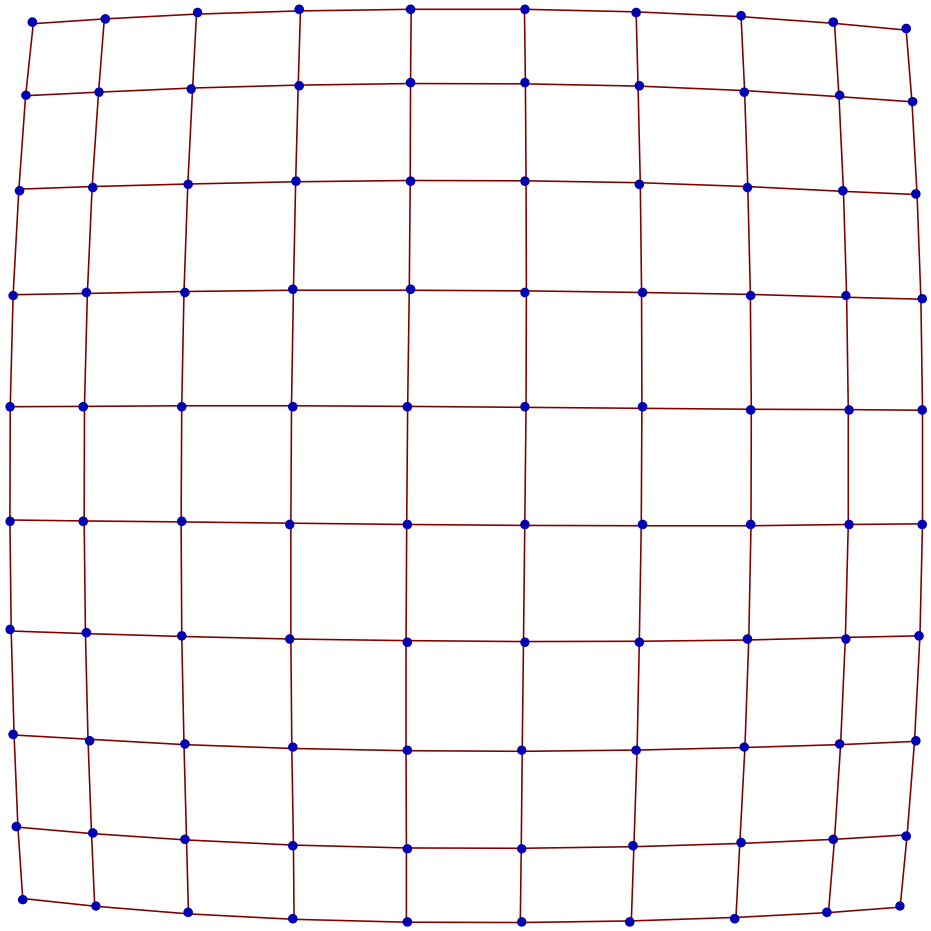

Fig 7.

The distance of the diagonally opposite vertices according to this distance is 9 + 9 = 18 (it is easy to prove that there are many minimal paths, all of which share this length). But the Euclidean definition gives us $10\sqrt{2}$ ~ 14.14. Which definition is better? The first definition, the one that I have defined, is clearly the natural definition of a graph. One might believe that the Euclidean definition of distance is even more natural. However, this definition of distance is not intrinsic because it depends on how the graph is drawn. Can one find an intrinsic definition of distance for a graph, which in this case is closer to the Euclidean definition rather than to the traditional definition? Viewing flat space as a graph with infinitely many vertices and edges, the Euclidean definition of distance is more natural. Thus, one would like to find a new, *intrinsic* definition of distance which is closer to the Euclidean one, at least for this case.

2. Motivated Definition of Distance from Quantum Mechanics

Consider a particle in one dimension (as in, in a line) with a double well potential. For example, consider the potential below:

$$V(x) = V \cdot \left(\frac{4x^2}{d^2} - 1\right)^2,$$

and the graph below:

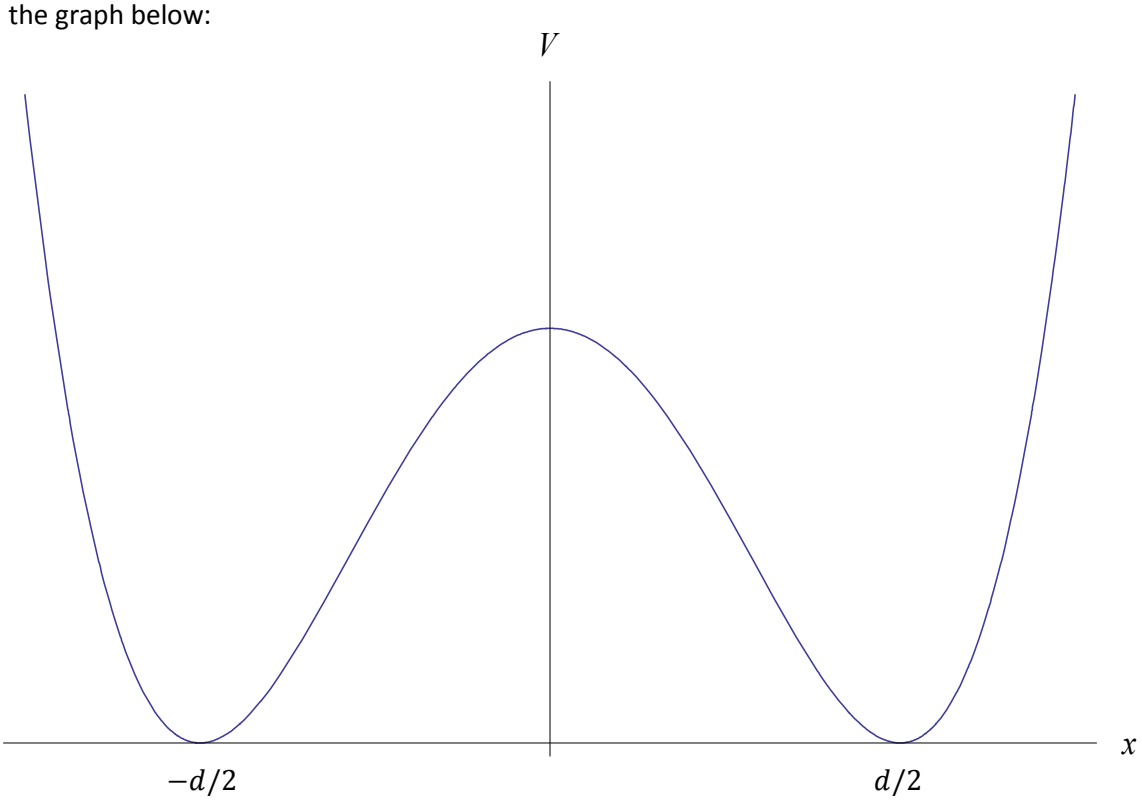

Fig. 8

In classical mechanics, the ground state is when the particle is at the bottom of one of the wells, at x = ± d/2. In particular, if it is at x = -d/2, it would permanently reside there. However, in quantum mechanics, the particle will always end up tunneling to the other potential, given enough time, and will oscillate back and forth between the two wells. For an introduction to quantum mechanics, see [2]. The tunneling time, the time it takes to make one oscillation, for $V \gg 1$ is given by:

$$T = e^{cd\sqrt{V}+O(\log(V))}, \tag{2}$$

where *d* is the distance, *V* the potential, and *c* a universal constant. This equation implies that

$$d \propto \frac{\ln T}{\sqrt{V}}. \tag{3}$$

Let us consider a simple model of quantum mechanics. The time evolution of a particle is given by the Hamiltonian matrix *H*. Each quantum state can be thought of as a vector in a vector space. The state of a particle is given by a vector v which depends on time according to

$$v(t) = e^{-iHt}v(0), \tag{4}$$

where *t* is the time and *v*(0) the initial state. For the double well, we can model the vector space as a two dimensional vector space, with the bottoms of the well forming a basis for the vector space. For our studies, we will only be looking at the two lowest states.

We represent the left ground state by the vector $\binom{1}{0}$, the right ground state by the vector $\binom{0}{1}$, and the matrix

$$H = \begin{pmatrix} -V & \varepsilon \\ \varepsilon & -V \end{pmatrix},$$

where $\varepsilon \ll V$.

The eigenvectors and eigenvalues for *H* are

$$\begin{cases} \binom{1}{1}, & E(1) = -V + \varepsilon \\ \binom{1}{-1}, & E(2) = -V - \varepsilon \end{cases}$$

Let the particle begin at $\binom{1}{0}$, at the bottom of the left well. Its state as a function of time is expressed as:

$$v(t) = e^{-iHt}\binom{1}{0}.$$

Using the eigenbasis, $v(t)$ can be written as

$$v(t) = e^{-iHt}\left[\frac{1}{2}\binom{1}{1} + \frac{1}{2}\binom{1}{-1}\right].$$

$v(t)$ can be simplified using its decompositions to the eigenvectors of *H* to

$$v(t) = e^{iVt}\binom{\cos \varepsilon t}{-i \sin \varepsilon t},$$

which implies that the particle oscillates between the two wells, with an oscillation time of $T = \frac{\pi}{\varepsilon}$. I will thus use this value of the time in my definition of distance. Therefore, using equation (3),

$$\mathrm{d} \propto \frac{\ln\left(\frac{\pi}{\varepsilon}\right)}{\sqrt{V}}. \tag{5}$$

Notice that ε is equal to half the difference of the eigenvalues of *H*:

$$\varepsilon = (E(2) - E(1))/2.$$

This relation implies that

$$\mathrm{d} \propto \frac{\ln\left(\frac{2\pi}{E(2)-E(1)}\right)}{\sqrt{V}}. \tag{6}$$

Here, we have assumed that there are only two states. In general, there will be many more states, and these states correspond to the two lowest eigenvectors and eigenvalues of *H*.

This relation suggests a new way of defining distance for a graph. Namely, we place wells at two vertices of the graph, compute the time to tunnel, and then use eq (6) to define the distance between them. We thus need to first set up the analog of quantum mechanics for a graph.

3. Quantum Mechanics on a Graph

Let us consider a simple graph G. We shall define *d(v)* to be the degree of vertex *v* (the number of vertices that v is connected to in G). Let the matrix ***L*** be defined as follows:

$$L(u,v) = \begin{cases} 1, & if\ u = v \\ -\frac{1}{\sqrt{d(u)d(v)}}, & if\ u\ and\ v\ are\ connected \\ 0, & otherwise. \end{cases} \tag{7}$$

The matrix ***L*** is the analog of the negative Laplacian, $-\Delta$ [3].

Our potential, *V*, is a diagonal matrix where the k-th diagonal element is the potential at the k-th vertex:

$$\boldsymbol{V} = \begin{pmatrix} V(1) & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & V(k) & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & V(n) \end{pmatrix}.$$

The Hamiltonian matrix H is defined as in the usual quantum mechanics by adding $\boldsymbol{H} = -\Delta + \boldsymbol{V}$.

For our definition of distance between vertices v_i and v_j, we will construct a double well with the property that the potential is zero everywhere but at v_i and v_j, the i-th and j-th vertex, which we set to be V(i) = V(j) = -V. In other words,

$$V(k) = \begin{cases} -V, & k = i \; or \; j \\ 0, & otherwise \end{cases} \tag{8}$$

Thus, the matrix v $\boldsymbol{V}$ looks like as follows:

$$\boldsymbol{V} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & V(i) = -V & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & V(j) = -V & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & V(n) \end{pmatrix}.$$

The Hamiltonian matrix H is defined as follows:

$$\boldsymbol{H} = -\boldsymbol{L} + \boldsymbol{V}. \tag{9}$$

The time T is defined as follows:

$$T = \frac{\pi}{E(2) - E(1)}, \tag{10}$$

where $E(1)$ and $E(2)$ are the two smallest eigenvalues of $\boldsymbol{H}$. This relation can be shown explicitly and will be done in the future.

And finally, the distance d_V is defined as follows:

$$d_V = \frac{\ln\left(\frac{2\pi}{E(2) - E(1)}\right)}{\sqrt{V}}. \tag{11}$$

0.02 ------ 10

Note that the definition of d_V depends on a positive real parameter $V \in R^+$. It would be interesting to study the limit of this definition as $V \to \infty$.

Our goal now is to explore using Mathematica the special properties that this definition of distance provides us.

4. Examples

In this section, we present several examples.

4.1 Linear Graph

Consider a ten vertex line graph, as in Figure 9 below:

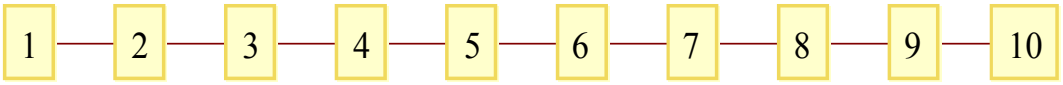

Fig. 9

Potentials of V = 5 are placed at the two ends, and the particle begins at the left end. When we plot the probability I.e. $|v(t)^2|$, we obtain the graph of:

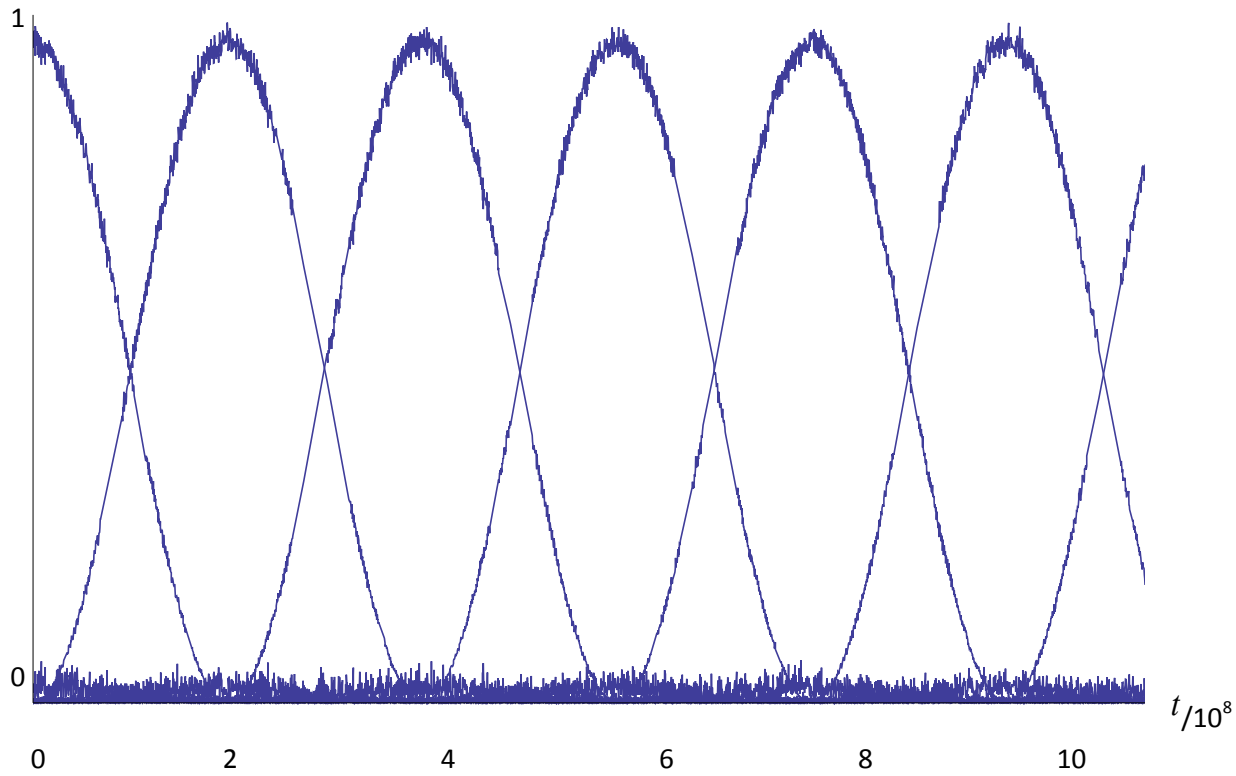

Fig. 10

We observe two curves corresponding to $|v|^2$ at the two ends of the graph that oscillate back and forth between the values of close to 0 and 1, and a bunch of tiny curves that oscillate between 0 and .1. We can deduce that the particle for the most part oscillates between the two wells, and always at any moment in time has a small chance of being elsewhere on the graph. The oscillation frequency, as shown in the graph, is around $3.5*10^8$, which is the same as the value from our definition of time in terms of the difference of the two eigenvalues given by eq. (10). This result supports our definition of time, and all of the line graphs that we have tested thus far support this definition of time. From quantum mechanics, we know this definition of time to be the true for the examples of a line and a grid graph.

For our next example, consider a 40 vertex line graph. We defined our double well potential to be not the entire line, but the 10 vertex line from the 16th vertex to the 25th vertex. We chose to embed our 10 vertex line graph in a 40 vertex line graph instead of creating a normal 10 vertex line graph because of the boundary effect. For the vertices that are close to the edge, unwanted boundary effects, just like in physics, appear to set in.

Figure 11 below represents our value of distance from the 16th vertex to every single other vertex up to the 25th vertex, with respect to the traditional distance:

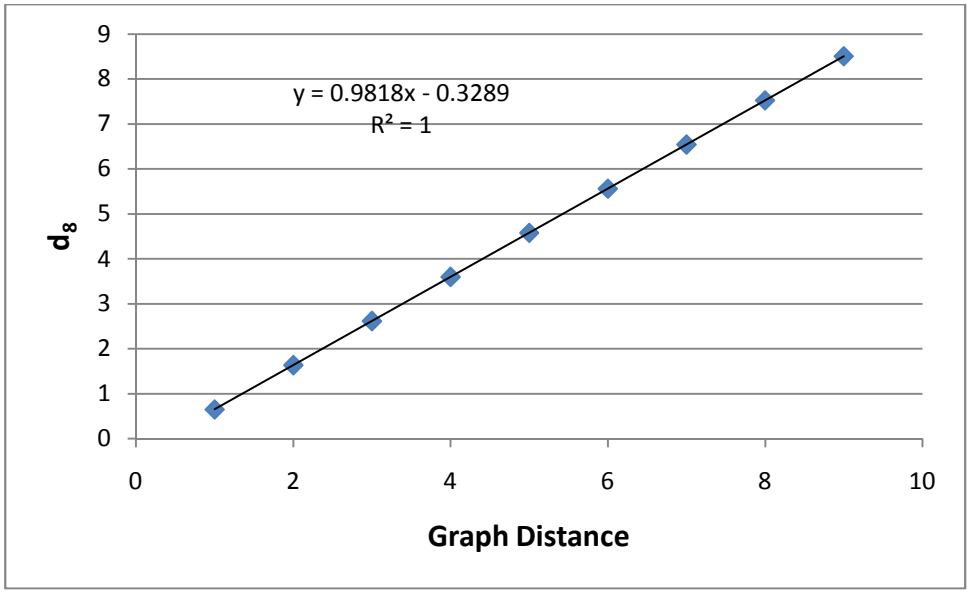

Fig. 11

Notice that our data conform to a linear trend with an R^2 value of 1. It is perfectly linear! It makes sense though that the slope is not exactly 1 and the graph is offset, because our definition of distance is approximate discrete lines, but the fact that the slope is constant affirms our suspicions that our definition resembles the Euclidean notion of distance.

If we look at the same graph, except with a potential of 4 instead of 8, we observe the same pattern: a slope of approximate 1, and an offset. The graph for the data is on the following page:

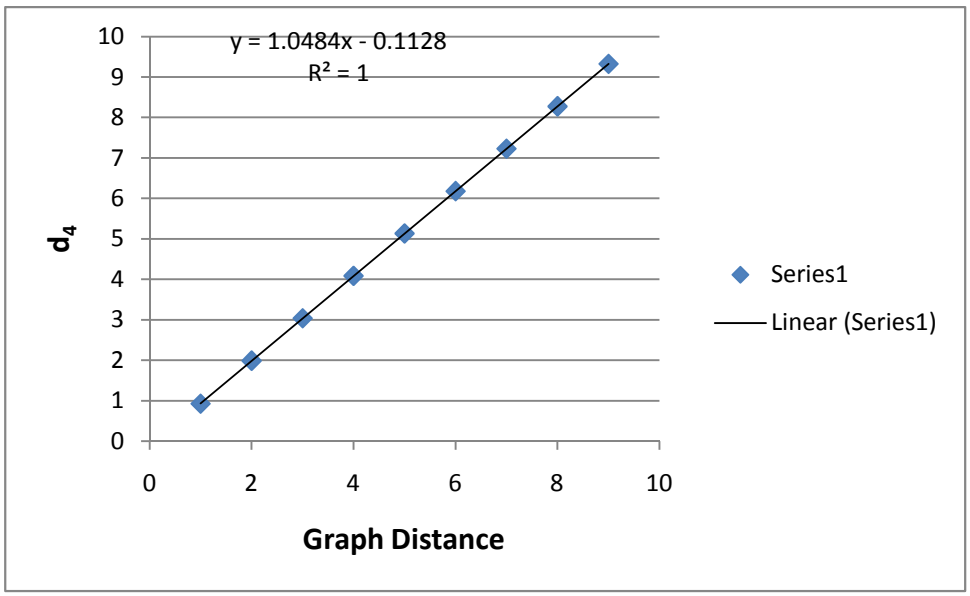

Fig. 12

Again the R^2 value is 1: a perfect linear trend.

4.2: Grid Graphs

For our next example, let us examine two grid graphs with potentials of 4 embedded in a 20 by 20 grid graph. One of them is a 6 by 6 in the exact center, and the other is a 4 by 4 not quite in the center. A sketch of it is given below:

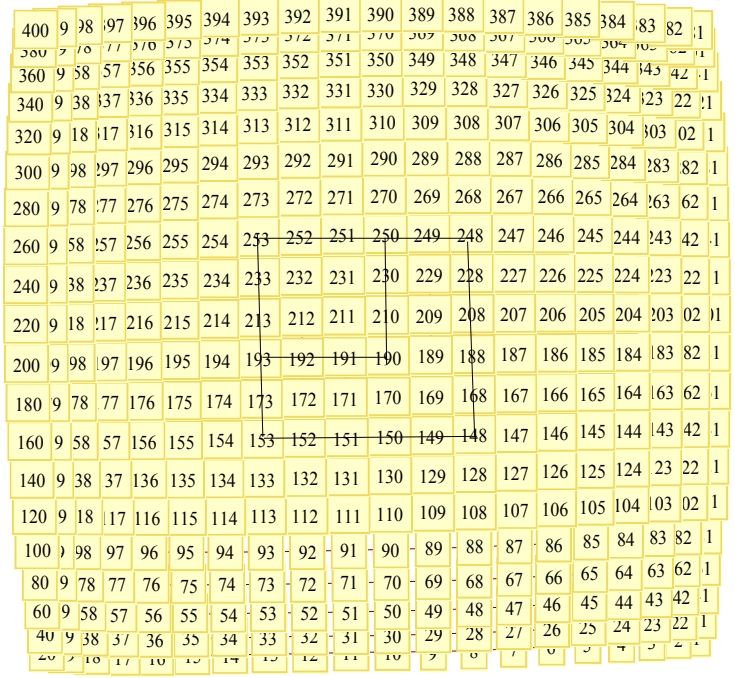

Fig. 13

We respectively placed the potentials at the diagonals of the two graphs and computed the distance. Our results are in Figure 14 below:

	Diagonal	Side
6 by 6 Grid Graph	10.986	6.772
4 by 4 Grid Graph	6.707	4.029

Fig 14

Since our grid graphs are squares, then according to Euclidean geometry, the ratio of a diagonal to its respective side is $\sqrt{2}$. Since each of our distances have a little shift, we decided to find the difference between the diagonals and the sides respectively, and then find the ratio of the diagonal difference to the side difference. The ratio, or slope in effect, should be $\sqrt{2}$ if it were the Euclidean definition. We obtained a slope of 1.56, which is quite close to 1.41. The normal graph definition of distance gives us the ratio of 2 instead, so our definition is a significant improvement over the traditional definition. Moreover, one expects that as the potential increases and the grid graph becomes bigger, the ratio approaches $\sqrt{2}$.

4.3: Erdős–Rényi Graphs

Erdős–Rényi graphs are defined by the number of vertices, and the probability of an edge existing between any two vertices [1]. An Erdős–Rényi graph G consisting of n vertices and a probability p is denoted as G(n, p). For example, the graph of a G(40, .2) is depicted below:

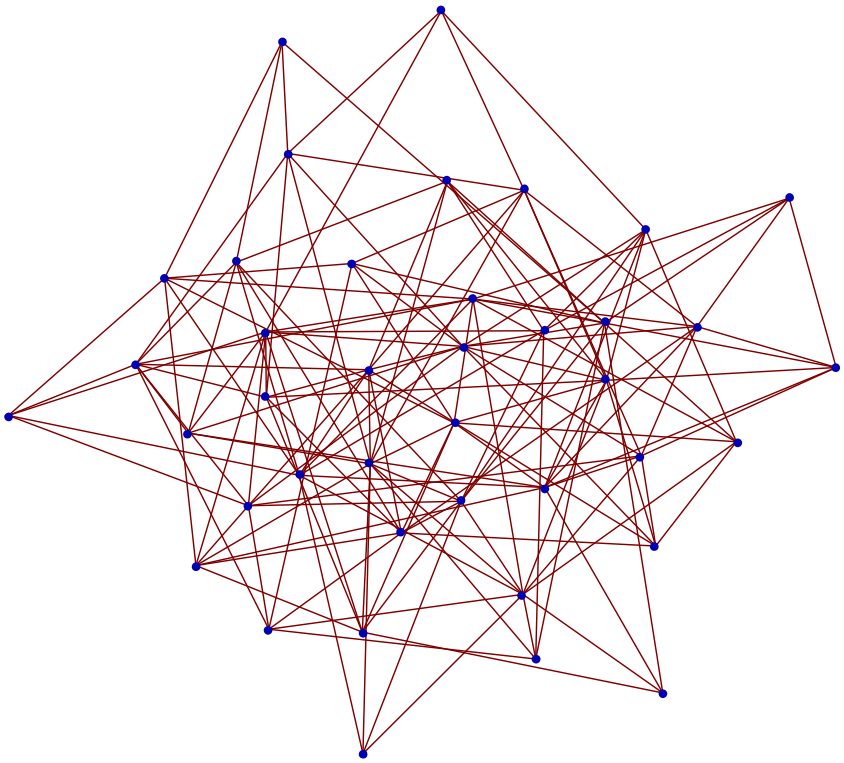

Fig. 15

Previously, we have explored examples of graphs where a good definition of distance is familiar to us. But for Erdős–Rényi graphs, or random graphs in general, there is no intuitive definition of distance other than the normal graph definition. Through our previous examples, we have gained confidence that our definition is more in line with the intuitive notion of distance. We will now thus explore the properties that this definition of distance provides us with random graphs and compare them with those of the conventional distance definition in graphs.

For our first example, we consider a specific G(250, .03), which is depicted below:

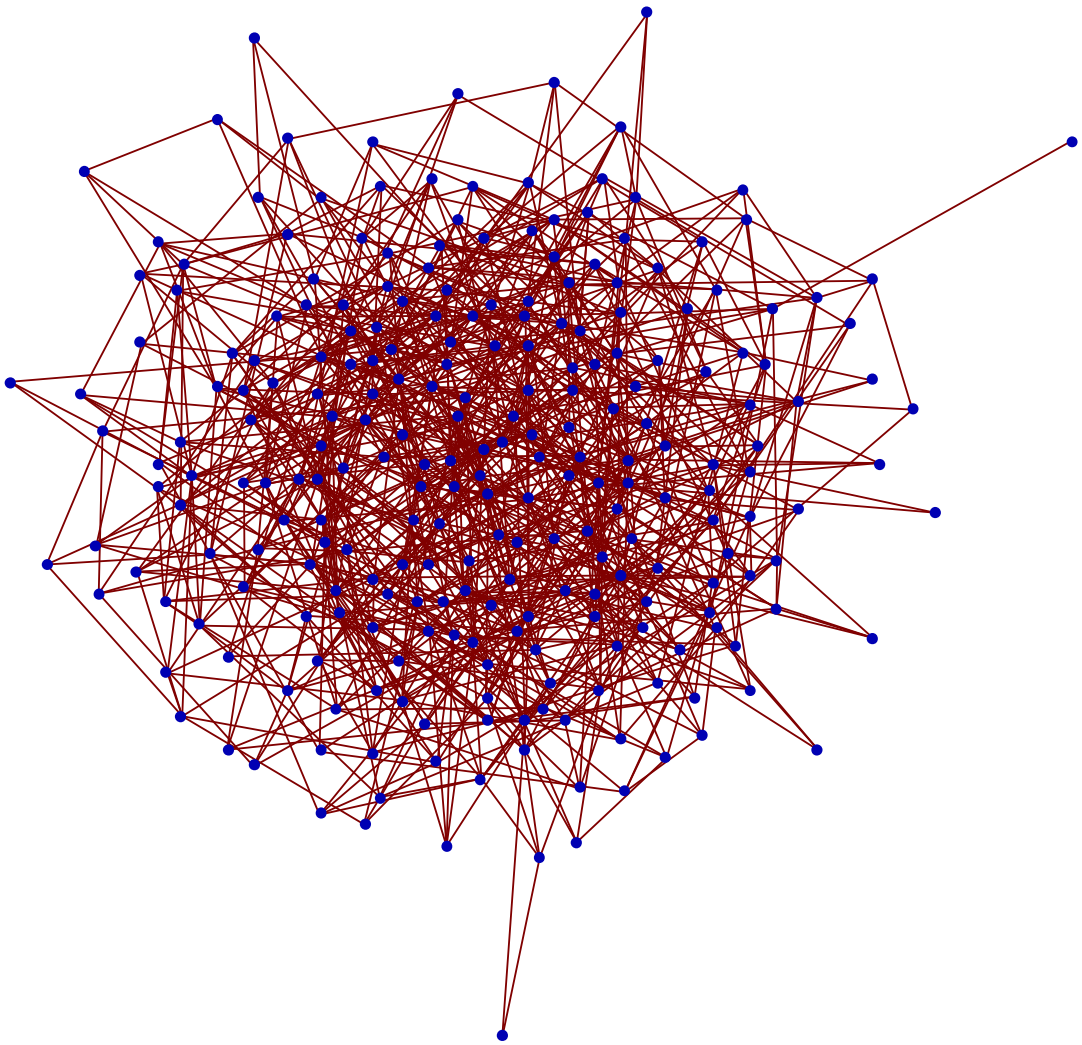

Fig. 16

For our computation of distance, we consider d_{100}. In other words, V = 100. The graph of the averages of some of the data is presented on the following page:

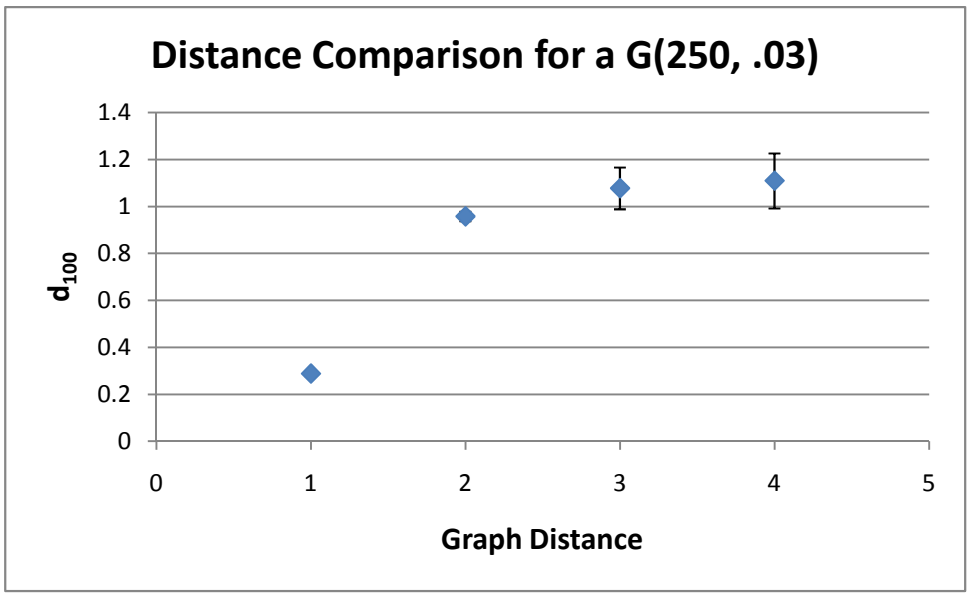

Fig. 17

(The data points for the graph above were computed by evaluating the distance many times for a given graph distance and then finding the average). We notice that from the graph distance of 2 onwards, there does not appear to be any trend, given the error bars. It seems that for the new definition the distance between any two vertices is for the most part equal, except for the adjacent vertices.

As another example, consider a specific graph G(250, .02) with the potential of V = 100 as before, which is depicted below:

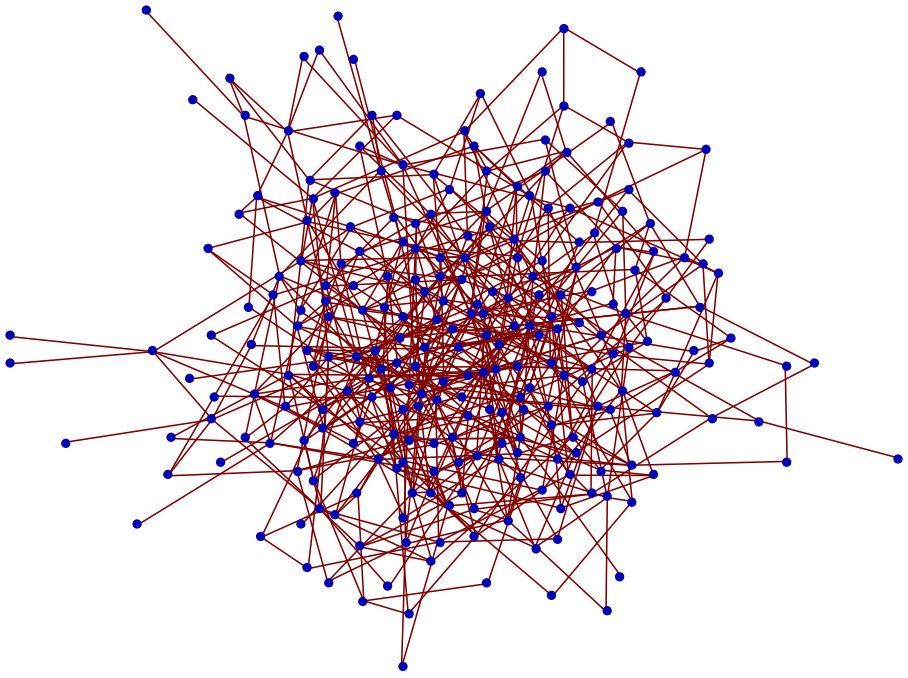

Fig. 18

The graph of the averages of some of the data is presented below:

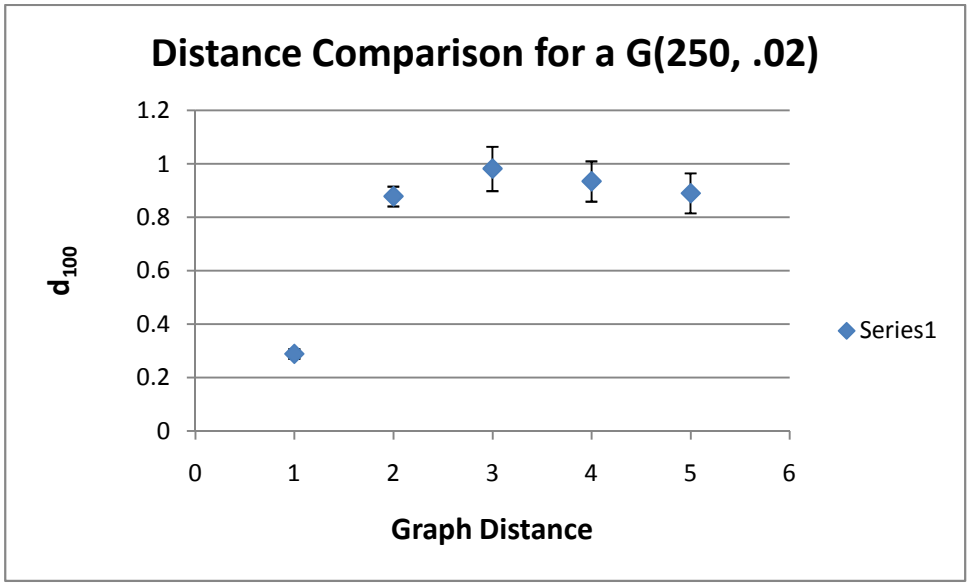

Fig. 19

As again, from the graph distance of 2 onwards, there does not appear to be any trend; all the distances are the same, given the error bars. It is quite amazing that for an Erdős–Rényi graph, the distances for points that do not share an edge does not depend on the graph distance! One would naively expect that the definition of our distance would vary monotonically with the graph distance. However, here observe that as the graph distance increases, ours does not.

5. Conclusion

While exploring different graphs, we noticed that there was a problem with the standard definition of distance for graphs. While the standard definition worked for linear graphs, it disagrees with the intuitive Euclidean distance for diagonals of grid graphs. It was always off by a factor of $\sqrt{2}/2$. And also, we know that in the real world, where space can be represented by an infinite cubic graph, the traditional definition of graph distance differs drastically from the Euclidean one. Thus, we decided to define our own distance in an intrinsic way that was closer to the Euclidean notion. We chose our definition using quantum mechanics because in quantum mechanics space is often discretized. And if it works in quantum mechanics, then that gives us a good intuition that it should work for graphs too. Through our examples, we have observed that our definition of distance works for linear graphs; that is, the property of linearity is conserved. It works much better for grid graphs than the standard definition.

However, there are limitations. In order to improve our values of distance, we have to consider much larger graphs and potentials. Our distance did not work too well for small graphs (but well for larger graphs), and the potential should in fact tend toward infinity in order for us to claim a truly valid distance (from quantum mechanics). It is hard to explore this numerically because Mathematica, the software that we used, simply lacks the necessary computational power and precision. Another

problem that we encountered was the boundary effect. Evaluations of distance near the edge were inaccurate due to boundary effects. We tried to avoid this problem by evaluation the distance as far away as from the boundary of the graph as possible. Also, our definition depends on the potential.

Further studies can be done to attempt to better explore the properties of our definition of distance. For example, we can consider large graphs and the limit $V \to \infty$. We have observed that for Erdős–Rényi graphs our definition of distance seems to be essentially the same for the vertices that are not adjacent to each other. It would be interesting to see if this can be proven in this limit. Another study that is needed is to determine whether our definition of distance upholds the convexity property of distance. Another interesting idea can be to explore the relation of our distance with the curvature of the graph, as defined in [4]. Clearly, there are many directions one can pursue exploring this definition of distance. We hope to have conveyed the naturalness and usefulness of this new definition of graph distance.

References

[1] Fan R.K., Chung, *Complex Graphs and Networks*, CBMS Regional Conference Series in Mathematics, Number 107, American Mathematical Society. (Images: http://math.ucsd.edu/~fan/graphs/gallery/)

[2] R. Liboff, *Introductory Quantum Mechanics*, Addison Wesley; 4th edition 2002.

[3] Fan R. K., Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, 1997, Number 92, American Mathematical Society.

[4] Y. Lin, L.Lu, and S.-T. Yau, *Ricci Curvature of Graphs*, 2010, submitted to Tohoku Mathematics Journal.

Acknowledgements

I would like to thank Dr. Y. Lin for his supervision of this project, and Prof. S.-T. Yau and Dr. G. Lippner for helpful comments.