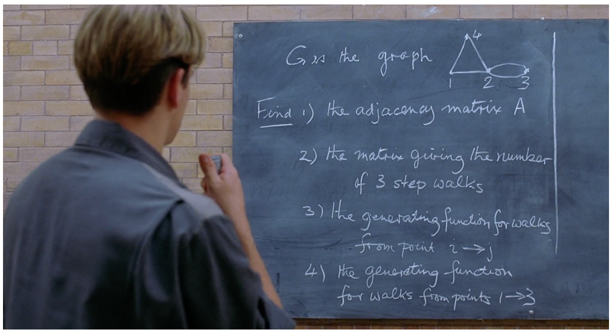
# On the Will Hunting Problem: Linear Algebra meets Graph Theory

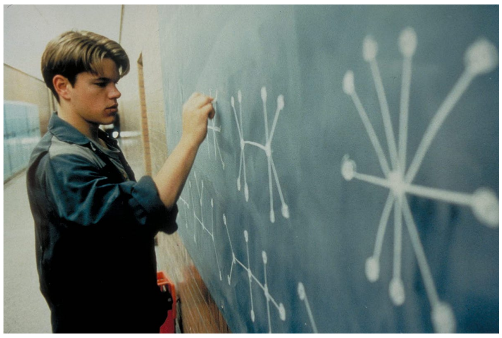
Author: Bsc. in Physics Humberto Segura Guzman
Date: April 2, 2026

## 1 Introduction

The year is 1997, something strange happened in Hollywood, they produced and released a film named *The Good Will Hunting* about a fictional brilliant mathematician named Will Hunting, interpreted by Matt Damon and co-starred with Robin Williams. In the film professor Gerard Lambeau interpreted by Stellan Skarsgård a Fields Medal mathematician is finishing a class at MIT and says the following *" I am also putting an advanced Fourier system on the main hallway chalkboard. I am hoping that one of you might prove it by the end of the semester. The first person to do so will not only be in my good graces, but go on to fame and fortune by having their accomplishment recorded and their name printed in the auspicious MIT Tech. Former winners include Nobel Laurets, word renowned astrophysicist, Field Medal winners and lowly MIT professors."* . The problem the professor refers to is Fig.1a, this essay is dedicated to solve that problem about a link between Linear Algebra and Graph Theory. The film presents more problems being the most famous the second problem to appear about irreducible homeomorphically trees but many people have already posted videos solving the problem.




**(a)** Problem 1 that appears in the movie about Linear Algebra and Graph Theory.



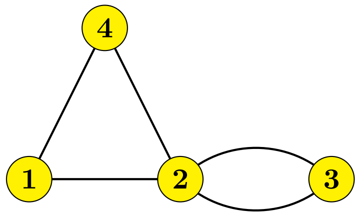
**(b)** Problem 2 that appears in the movie about irreducible homeomorphic trees.

**Figure 1:** Screenshots from *Good Will Hunting* (directed by Gus Van Sant, produced by Lawrence Bender, 1997, Miramax Films).

When I first saw the movie, my main intrigue was how Will was able to solve the problem, what mathematics he used, and what was the answer to the questions proposed on the chalkboard, as the film gives us the answer but not how Will arrived at those results. This essay is the perfect opportunity to fulfill this long-lasting intrigue —Yes, I saw it many years ago— and probably show you how interesting this problem really is. Also, this problem is not like it was never solved before; a class of 2003 at Harvard University gave the answers [1], but I mainly used it to ensure I did not mess up somewhere and provide a correct solution.

## 2 Problem statement.

**G**




**Given graph G, find**

1. The adjacency matrix, A
2. The matrix giving the number of 3 step walks
3. The generating function for walks from $i \to j$
4. The generating function for walks from $1 \to 3$.

## Part I

The problem starts with a undirected graph $G$ in which the edges connect only vertices at distance 1. To solve the first point on the list, it is necessary to know what an adjacency matrix is. The following definition is a hybrid one that I constructed and best explains what this matrix is [2].

**Definition**

Let $G = (V, E)$ be a finite graph where $V$ is a set of $n$ vertices $\{v_1, v_2, \ldots, v_m\}$ and E a set of $m$ edges $\{e_1, e_2, \ldots, e_m\}$. The adjacency matrix $\boldsymbol{A}$ is a $n \times n$ matrix whose entry $A_{ij}$ is $k$ if there exist $k$ edges between node $i$ and $j$.

So for example, in the graph $G$ we are already provided with labels for the nodes. Let us take node 1 and we see an edge connecting node 1 and 2, this means that the entry $A_{12}$ is 1 and by symmetry the entry $A_{21}$ is also 1. Now for the entry $A_{13}$ we note that there is no connection between node 1 and 3 so the entries $A_{13}, A_{31}$ should be equal to 0. Another set of zero entries are the diagonal elements $A_{11}, A_{22}, A_{33}, A_{44}$ as there is no edge beginning and finishing in the same node. Now we note that between vertex 2 and 3 there exist two edges so the entries $A_{23}, A_{32}$ is equal to 2. By now it should be straightforward to get the remaining values of the adjacency matrix so the solution to part I is

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \tag{1}$$

## Part II

The problem is to find the matrix giving the 3-step walks. To be clear, a one-step walk is traversing just one edge from node $i$ to node $j$, given that there exists at least one edge between this pair of nodes. A two-step walk would be traversing from node $i$ to $j$ passing through an intermediate node $k$, with $k$ connected by at least one edge to each of the nodes $i, j$; this also includes paths like $i \to k \to i$ because the graph is not directed, so this walk is also possible. But how can we use the matrix we constructed? The key insight is to focus on a substructure, isolating the nodes $i, j$ and the nodes between them; as an example, let us look at the node structure depicted in Fig.2.

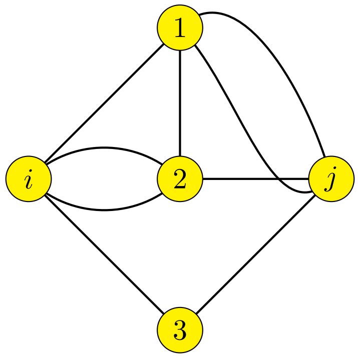



**Figure 2:** Substructure connecting nodes $i$ and $j$ through intermediate vertices.

I labeled the intermediate nodes in this convenient manner because it will be better to keep track of the indices this way. Let us count the number of two step walks form node $i$ to $j$, there is 1 edge between node $i$ and 1, then there are 2 edges form node 1 to $j$, so the count is on $1 \cdot 2 = 2$. Then we note there exist 2 edges between node $i$ and 2 and then 1 node form node 2 to $j$ and this contributes $2 \cdot 1 = 2$ to the total of two step walks. In the example graph there is a total of $1 \cdot 2 + 2 \cdot 1 + 1 \cdot 1 = 5$ two step walks from $i$ to $j$ but wait, the number of edges is encoded in the entries of our adjacency matrix $\boldsymbol{A}$. This is why we constructed this matrix in the first place, as the counting of two step walks in this substructure is given by $A_{i1}A_{1j} + A_{i2}A_{2j} + A_{i3}A_{3j}$, and if we have $k$ intermediate nodes the number of two step walks form node $i$ to $j$ $w_2^{(ij)}$ is equal to [3]

$$w_2^{(ij)} = \sum_{n=1}^{k} A_{in}A_{nj} \tag{2}$$

But wait again, we recognize that structure, is matrix multiplication of the matrix with itself. The number of two step walks in graph $G$ must be $\boldsymbol{A}^2$. Note that we covered three subtleties here. As shown in Fig. 2 the count of the connection between node 1 and 2 is excluded for the path $i \to j$ as this belongs to the entry $A^2_{i2}$ not to $A^2_{ij}$, also if the entry let us say $A_{il}$ is zero this does not contribute to the two step walk from $i$ to $j$. Finally this also covers the case $A^2_{ii}$ because as previously stated a two step walk is also the path $i \to k \to i$. To check if we are in the correct path let us now construct a substructure of nodes $i$ to $j$ now with and additional column of intermediate nodes as given by Fig.3

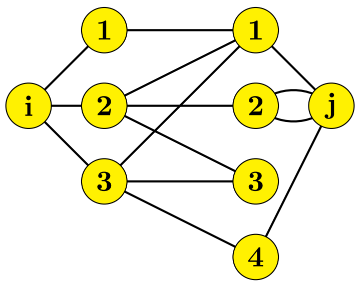



**Figure 3:** Substructure connecting nodes $i$ and $j$ with an additional layer of intermediate vertices, representing three-step walks between them.

The total number of three step walks form $i$ to $j$ is given by $A_{i1}A_{11}A_{1j} + A_{i2}A_{21}A_{1j} + A_{i2}A_{22}A_{2j} + A_{i3}A_{31}A_{1j} + A_{i3}A_{33}A_{3j} + A_{i3}A_{34}A_{4j} = 6$ and if we were to have $k_1$ intermediate nodes in the first column and $k_2$ intermediate nodes in the second column then the number of three step walks $w_3^{(ij)}$ is

$$w_3^{(ij)} = \sum_{n=1}^{k_1}\sum_{t=1}^{k_2} A_{in}A_{nt}A_{tj} \tag{3}$$

In our graph $G$ the cube of the adjacency matrix $\mathbf{A}^3$ then encodes all the three step walks for all the nodes. We of course can generalize this for $s$ step walks from $i$ to $j$ and the number will be given by $A_{ij}^s$, but I will use a similar argument to Fermat when he claimed to have solved Fermat's last theorem *the proof is too long to fit in this essay.* Then answer to part II is

$$A^3 = \begin{pmatrix} 2 & 7 & 2 & 3 \\ 7 & 2 & 12 & 7 \\ 2 & 12 & 0 & 2 \\ 3 & 7 & 2 & 2 \end{pmatrix} \tag{4}$$

# Part III

This part involves a mathematical tool that is widely used in combinatorics, generating functions. The idea of this special type of functions —polynomials are the most used but one can use exponential functions— is that if we have an infinite sequence of counts $\{a_0, a_1, a_2, a_3, \dots\}$, we can store the terms of this sequence in the coefficients of a polynomial $f(z)$ such that [4]

$$f(z) = \sum_{n=1}^{\infty} a_n z^n = a_0 z^0 + a_1 z^1 + a_2 z^2 + \cdots$$

The $z$ symbol and its powers acts as formal indicators to encode what term in the sequence we are referring to and the coefficient the corresponding value. A concrete example is take the power series of the function $\frac{1}{1-z}$, you should get that

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \dots$$

So $\frac{1}{1-x}$ is the generating function for the sequence $\{1, 1, 1, 1, \dots\}$ where this sequence counts the number of ways to choose items from a single type when order does not matter and repetitions are allowed. To get the generating function for walks from $i \to j$ we already know the sequence that counts the $s$ step walks is stored in the powers of the corresponding entry in the adjacency matrix $\{A_{ij}, A_{ij}^2, A_{ij}^3, \dots\}$. So let us use this sequence as coefficients for our generating function for each entry $i, j$

$$f_{ij}(z) = \sum_{n=0}^{\infty} A_{ij}^n z^n$$

Next noting that $\sum_{m=0}^{\infty} x^n = (1-x)^{-1}$ we can get a closed form for the generating function for walks from $i$ to $j$

$$f_{ij}(z) = \left(\sum_{n=1}^{\infty} (zA)^n\right)_{ij} = (I - zA)^{-1} = \frac{\mathrm{adj}(I - zA)_{ij}}{\det(I - zA)} \tag{5}$$

where $\det(I - aA)$ is the determinant and $\mathrm{adj}(I - zA)_{ij}$ the adjugate or transpose of the cofactor matrix. At this step the mathematicians will jump off their seats and question about the convergence of the sum because we use the physics criteria of assuming something converges until it doesn't. In our setting, we do not assign a specific value to $z$, so the series $\sum_{n=0}^{\infty}(zA)^n$ is understood as a formal power series. Therefore, the identity $\sum_{n=0}^{\infty}(zA)^n = (I - zA)^{-1}$ holds algebraically, without appealing to convergence in $\mathbf{R}$ or $\mathbf{C}$.

# Part IV

At this point we surpassed the difficult part of the problem. This last part is a specific case of Eq.(5). At last the tedious homework about calculating fifty determinants and a hundred inverses comes in handy. I will show just the relevant results and you should arrive at the same answers

$$\det(I-zA)=\det\begin{vmatrix}1 & -z & 0 & -z\\ -z & 1 & -2z & -z\\ 0 & -2z & 1 & 0\\ -z & -z & 0 & 1\end{vmatrix}=4z^4-2z^3-7z^2+1$$

$$\operatorname{adj}(I-zA)_{13}=\begin{pmatrix}
+\begin{vmatrix}1 & -2z & -z\\ -2z & 1 & 0\\ -z & 0 & 1\end{vmatrix} & -\begin{vmatrix}-z & -2z & -z\\ 0 & 1 & 0\\ -z & 0 & 1\end{vmatrix} & +\begin{vmatrix}-z & 1 & -z\\ 0 & -2z & 0\\ -z & -z & 1\end{vmatrix} & -\begin{vmatrix}-z & 1 & -2z\\ 0 & -2z & 1\\ -z & -z & 0\end{vmatrix}\\
-\begin{vmatrix}-z & 0 & -z\\ -2z & 1 & 0\\ -z & 0 & 1\end{vmatrix} & +\begin{vmatrix}1 & 0 & -z\\ 0 & 1 & 0\\ -z & 0 & 1\end{vmatrix} & -\begin{vmatrix}1 & -z & -z\\ 0 & -2z & 0\\ -z & -z & 1\end{vmatrix} & +\begin{vmatrix}1 & -z & 0\\ 0 & -2z & 1\\ -z & -z & 0\end{vmatrix}\\
+\begin{vmatrix}-z & 0 & -z\\ 1 & -2z & -z\\ -z & 0 & 1\end{vmatrix} & -\begin{vmatrix}1 & 0 & -z\\ -z & -2z & -z\\ -z & 0 & 1\end{vmatrix} & +\begin{vmatrix}1 & -z & -z\\ -z & 1 & -z\\ -z & -z & 1\end{vmatrix} & -\begin{vmatrix}1 & -z & 0\\ -z & 1 & -2z\\ -z & -z & 0\end{vmatrix}\\
-\begin{vmatrix}-z & 0 & -z\\ 1 & -2z & -z\\ -2z & 1 & 0\end{vmatrix} & +\begin{vmatrix}1 & 0 & -z\\ -z & -2z & -z\\ 0 & 1 & 0\end{vmatrix} & -\begin{vmatrix}1 & -z & -z\\ -z & 1 & -z\\ 0 & -2z & 0\end{vmatrix} & +\begin{vmatrix}1 & -z & 0\\ -z & 1 & -2z\\ 0 & -2z & 1\end{vmatrix}
\end{pmatrix}^{T}_{13}$$

Hence

$$f_{13}=\frac{2z^3+2z^2}{4z^4-2z^3-7z^2+1}=2z^2+2z^3+14z^4+18z^5+94z^6+146z^7+\cdots \tag{6}$$

You can check in the movie that Eq.(6) is the answer Will gets. For the sake of completion, here are more generating functions

$$f_{23}=\frac{2z-2z^3}{4z^4-2z^3-7z^2+1}=2z+12z^3+4z^4+76z^5+52z^6+492z^7+\cdots$$

$$f_{24}=\frac{z^2+z}{4z^4-2z^3-7z^2+1}=z+z^2+7z^3+9z^4+47z^5+73z^6+319z^7+\cdots$$

$$f_{34}=\frac{2z^3+2z^2}{4z^4-2z^3-7z^2+1}=2z^2+2z^3+14z^4+18z^5+94z^6+146z^7+\cdots$$

$$f_{14}=\frac{-4z^3+z^2+z}{4z^4-2z^3-7z^2+1}=z+z^2+3z^3+9z^4+19z^5+65z^6+139z^7+\ldots$$

## Conclusion

The problem is now finished, but this is only the beginning of a whole active branch of mathematical research. As was treated in the essay, the graph $G$ was a relatively simple one, now we may ask what would happen if we have restrictions on the way we can traverse the graph, meaning there is the possibility of going through one edge and not being able to return the previous node. Another layer of complexity is what happens if now the edges have some weight such that the cost of traversing one edge is not one but a natural number $n$, what is the optimal way to traverse the graph from node $i$ to $j$ such that the sum of these weights is minimum. The last part is directly in touch with the *Traveling Salesman Problem* in computer science [5], but also in modern theory of network communications [6]. If you are more interested in probability, now one can postulate that to pass one edge you throw a dice and if it lands on a number less than three you cannot pass that edge and have to use another edge, so now what would your average distance walked be for the given graph. As you can see, this problem only opens the door for new and exciting areas of discrete mathematics, and I hope this essay inspires you to possibly unravel all these mysteries.

# References


[1] O. Knill. *The Good Will Hunting Problem.* Fall 2003, Harvard University, Linear Algebra and Differential Equations. 2003. URL: `https://people.math.harvard.edu/archive/21b_fall_03/goodwill/`.

[2] Sebastian M. Cioabă and M. Ram Murty. *A First Course in Graph Theory and Combinatorics.* Texts and Readings in Mathematics. Singapore: Springer Nature, 2022. ISBN: 9811913358.

[3] S. Axler. *Linear Algebra Done Right.* 3rd. Springer, 2015. ISBN: 3319307657.

[4] S. Ross. *A First Course in Probability, Global Edition.* 10th. Harlow, United Kingdom: Pearson Education Canada, 2019. ISBN: 1292269200.

[5] D. L. Applegate et al. *The Traveling Salesman Problem: A Computational Study.* Princeton Series in Applied Mathematics. Princeton, NJ: Princeton University Press, 2006. ISBN: 9780691129938.

[6] K. Srinivasan. *Mathematical Foundations of Computer Networking.* Boston, MA, USA: Addison-Wesley, 2012. ISBN: 0321792106.