Character Social Network Graph Analysis of Official Touhou Project Games, Music, and Print Works

Christopher Hendrik Gunawan - 13525065
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail: chico.gunawan@gmail.com , 13525065@std.stei.itb.ac.id

***Abstract*—The Touhou Project series has a large cast of characters. While there has been documentation of the characters and their connections as well as graphs that visualize them, there has been no graph that includes the entire cast and their canonical connections. This paper aims to create a simple graph containing the entire cast from the PC-98 games, the windows games, the music albums, and the print works, and their connections based on canon. The result is a graph with 153 vertices and 240 edges. It has a mean degree of 3.137, a median degree of 2, a modus degree of 1, and a max degree of 23. The average path length of the largest connected component is 4.203, with the longest path length being 10. The vertex closest to all other vertices in the largest connected component has a longest path length of 5.**

Keywords—Touhou; graph; character; connection; connected component

I. INTRODUCTION

Touhou Project is a series of bullet hell shooting games made by Team Shanghai Alice, founded by ZUN, featuring an large cast of characters. As of June 2026, there are 34 games (6 PC-98 and 14 Windows games), 12 music albums, and 22 print works. Along with them are hundreds of characters. All of these characters exist in the same universe and have connections with each other, scattered across all of ZUN's works. A part of the fan community takes an interest in piecing all the connections, making wikis and even social network graphs. However, most if not all the graphs are either incomplete, outdated, or contain non-canonical connections. Thus, I have took it upon myself to attempt at constructing a simple graph of the characters in all of ZUN's works. After constructing the graph, I analyze the connection graph to observe the properties of the entire graph as well as individual nodes.

Besides benefiting the Touhou community, this paper also showcases the relevance of network graphs in analyzing social connections of a population. While the population I chose to perform a network graph analysis on is fictional, the same method can be applied on real populations, which may give way to useful insights.

II. THEORY

A. Vertex

A vertex is a point in a graph.

B. Edge

An edge is a line that connects two vertices.

C. Ring Edge

A ring edge is an edge that connects a vertex to itself [1].

D. Graph

A graph is a collection of nodes and edges [1].

E. Undirected Graph

An undirected graph is a graph with edges which have no orientation [1].

F. Simple Graph

A simple graph is an undirected graph with no ring edges [1].

G. Network Graph

A network graph is a graph in which the vertices represent people and the edges represent connections between people [1].

H. Adjacency

When two vertices are connected by an edge, they are said to be adjacent [1].

I. Isolated Vertex

An isolated vertex is a vertex that has no connecting edges [1].

J. Degree

Degrees represent the amount of edges connected to a vertex [1]. In the case of a network graph, it represents the amount of connections a person has.

K. Path

A path is a row of vertices and edges that connect to each other forming a chain between the source vertex and target vertex [1].

L. Cycle

A cycle is a path that starts and ends at the same vertex [1].

M. Connected

Two vertices are connected if there exists a path that goes from one vertex to the other [1].

N. Connected Graph

Two graphs are connected if there exists a set of paths for all elements from one graph to connect to all elements from the other graph [1].

O. Disconnected Graph

Two graphs are disconnected if there exists a vertex from one graph that doesn't have a path to a vertex from another graph [1].

P. Subgraph

A subgraph is a graph which contains a subset of the vertices and edges of the original graph [1].

Q. Weights

A weight is a number that is assigned to an edge. A weighted graph is a graph in which edges have weights [1].

R. Connected Component

The amount of connected components in a graph is equal to the amount of disconnected subgraphs [1].

S. Spanning Subgraph

A spanning subgraph is a subgraph that contains all the vertices of the original graph [1].

T. Complete Graph

A complete graph is a graph in which every vertex is adjacent to every other vertex [1].

U. Isomorphism

Two graphs are isomorphic with each other if they contain the same nodes and edges, regardless of how they are laid out visually [2], as can be seen by the following example:

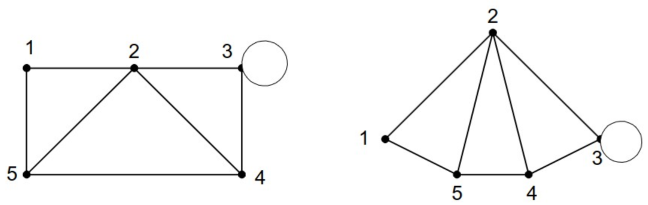

Fig. 1. The two graphs above are isomorphic with each other.

Source: https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2025-2026/21-Graf-Bagian2-2026.pdf

V. Planar Graph

A planar graph a graph which is isomorphic with a plane graph [2]. The planar graph itself does not need to be a plane graph.

W. Plane Graph

A plane graph is a graph in which no edges intersect [2].

X. Kuratowski's Theorem

Kuratowski's Theorem states that if a graph is a Kuratowski graph, is isomorphic with a Kuratowski graph, or contains a subgraph which is or is isomorphic with a Kuratowski graph, then the graph is not planar [2].

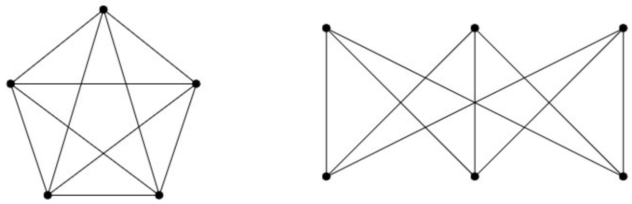

Fig. 2. The two graphs above are Kuratowski graphs. The left Kuratowski graph is K_5 and the right Kuratowski graph is $K_{3,3}$.

Source: https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2025-2026/21-Graf-Bagian2-2026.pdf

Y. Chromatic Number

The chromatic number of a graph is the minimum amount of colors needed to draw the graph in such a way that no two vertices of the same color are adjacent [3]. For a complete graph with N vertices, N colors are required [3].

III. Methodology

A. Graphing Tool

In order to be able to visualize the graph, I need a graphing tool. After looking at a bunch of graphing tools, I decided to choose gephi as the graphing tool to visualize the graph. Besides the basic visualization feature, it also has a layout, statistics, and filter feature which will all be useful during the analysis process.

B. Constraints

Since I am making a network graph of fictional characters, I need to set some rules to decide which characters to add as vertices and what interactions to include as edges. This step is very important, since it will affect the final graph.

I set two rules for the characters who will be added as vertices. First of all, the character must have a canon name.

This excludes characters whose canon names are unknown, like the fortune teller in Forbidden Scrollery. Next, the character must have at least one line of dialogue. This excludes characters who never speak, like Kasen's animals in Wild and Horned Hermit. These two rules narrow down the amount of characters to process while keeping the graph's properties mostly unaffected. This is due to most unnamed characters being background characters and characters without dialogue not having many connections due to their lack of interaction with anyone.

As for the edges, the rules I set are more vague as there is little to no objective rule that can be made without causing the graph to include or exclude connections that are nonsensical. I add an edge if the two characters have interacted with each other multiple times beyond that of which is used as a means of tying the story to the gameplay. This excludes most fighting in games, since those interactions tie to gameplay. I also add an edge if the two characters are implied to have interacted with each other in the past. For example, two characters who were friends before the events of the story.

Besides rules for deciding whether to add a vertex/edge, I also have rules on when to add edges. For each new vertex I add, I only add edges from that vertex to already existing vertices. In practice, this means characters that have connections with other characters later on in the list of characters do not receive an edge until said other character has been added.

C. ID Organization

In order to keep track of the vertices and edges more easily, it should be organized. This is why instead of using the graph editor which creates automatically creates unchangeable node IDs, I used Comma Separated Values (CSV) files to add my vertices and edges. I use a four digit ID, allocating the thousandths and hundredths place for works and the tenths and ones for characters. I start with an ID of 101 (work 01, character 01) in order to make the numbers more intuitive to look at. This gives me a capacity of 99 works, each work with a capacity of 99 characters. This is more than enough to account for Touhou Project's large amount of works, some with a large cast of new characters.

D. Data Sifting

With the existence of a lot of Touhou Project works, I looked into ways to sift through all the data quickly while avoiding taking too many shortcuts. To get the characters and their connections, I start with Touhou wiki, a wiki site maintained by the community, containing all works up to date. However, just reading the wiki could lead to me missing connections or adding connections that are too weak. Therefore, I also look into the wiki's sources, which include the game text files, taken from the thpatch site, and the print works. Fortunately, the game text files and the print works are all available for free, which allows me to check all sources.

Some of the sources I checked include game endings, which considered to be heavy spoilers by the community. So this paragraph is a warning that besides the basic game and print work spoilers, some of the explanations for the connections in the edges CSV file may also include game endings.

For this paper, I assume that a PC-98 character and their Windows namesake are the same character. I also assume that all scenarios and endings are cannon.

I started by sifting through the PC-98 games in order. Most of the characters there do not show up in the Windows games, but some do.

After that, I sifted through the Windows games. This came with the challenge of deciding whether a connection should be deemed strong enough to be an edge. To do this, I did some guesswork using context clues to determine the likelyhood of a pair of characters to count as a connection.

After the game characters and connections have been added, I went on to the music. Out of all the music works, only the collection features stories with two new characters, Maribel Hearn and Renko Usami.

After those two characters and their connections were added, only the print works were left. In it were a couple characters who mostly have connections with characters from works I previously added to the list.

E. Explanations

Since the process of deciding connections to add as edges is subjective, I took it upon myself to provide some documentation stating my reasoning for adding each edge. However, some edges may be documented more poorly in some cases, mainly in ones that have a vague source location or multiple source locations.

F. Analysis

After the graph had been completed, it was time to continue to the analysis phase. To analyse the properties of the graph and its vertices, I used gephi's filters and statistics.

IV. Result

The CSVs used for making the graph can be access via the following GitHub link: github.com/chrishengun123/Touhou-Project-Network-Graph-Data

The graph contains 153 vertices and 240 edges. Due to the large amount of vertices, I found no way to show the full graph in an image without losing some important information like vertex labels and edges. The following image is my attempt at displaying the full graph:

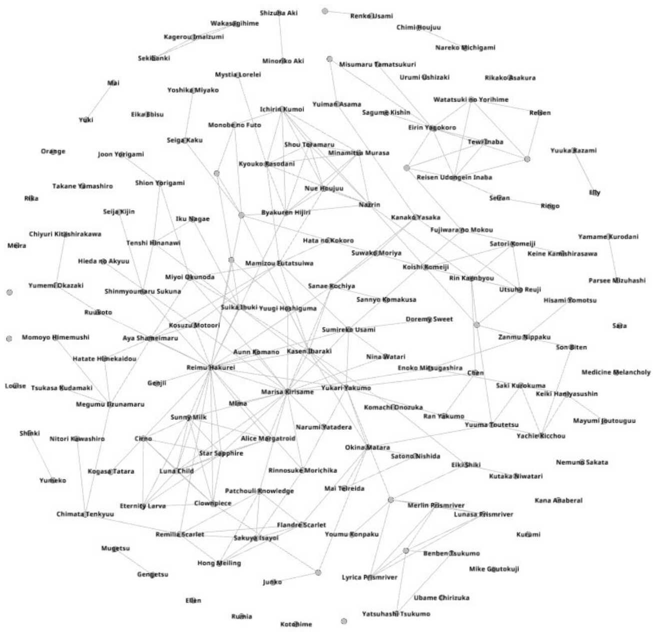

Fig. 3. The graph above shows all the nodes and edges.

And this following image is my attempt at displaying the largest connected component:

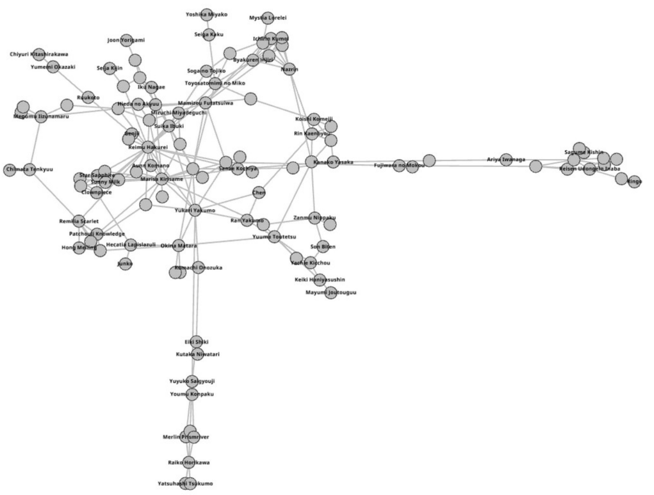

Fig. 4. The subgraph above shows all the nodes and edges of the largest connected component.

A. Connected Components

The graph is made up of 31 connected components. The high amount of connected components is caused by there being a lot of characters who do not have any connections, which in the case of the graph, means having a degree of 0. If we exclude those characters we are left with 10 connected components.

These are the 10 connected components which contain more than one vertex:

- The largest connected component, containing 113 vertices, 74% of the entire graph
- Elly and Yuuka Kazami, who had some connection in Lotus Land Story
- Muugetsu and Gengetsu, who are siblings
- Yuki and Mai, who had some connection in Mystic Square
- Yumeko and Shinki, who are servant and master
- Shizuha Aki and Minoriko Aki, who are siblings
- Yamame Kurodani and Parsee Mizuhashi, who are both in Former Hell
- Wakasagihime, Sekibanki, and Kagerou Imaizumi, who are all part of the Grassroots Youkai Network
- Chimi Houjuu and Nareko Michigami, who are from the latest Touhou game
- Maribel Hearn and Renko Usami, who are both in the outside world

B. Degree Analysis

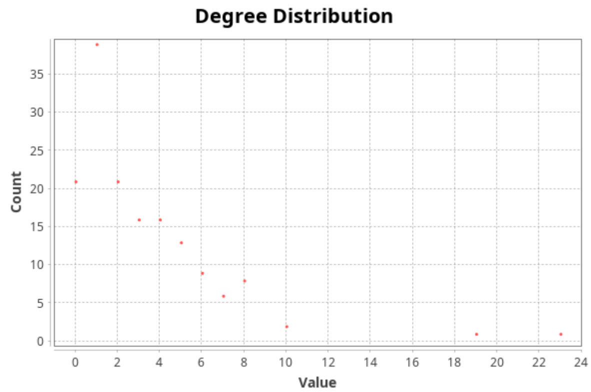

Fig. 5. The distribution above shows the distribution of degrees in the graph. Its x axis represents the degree value and its y axis represents the amount of vertices.

According to gephi's statistics, the mean average degree of the entire network graph is 3.137. However, the median degree of the graph is 2 and the majority of characters have a degree of 1. As shown by the degree distribution, the discrepancy between the mean and median is caused by the distribution being skewed to the right by characters with high degrees.

The characters with degrees 10 or higher are:

- Reimu Hakurei: 23 degrees
- Marisa Kirisame: 19 degrees
- Yukari Yakumo: 10 degrees
- Kasen Ibaraki: 10 degrees

Reimu and Marisa's high degrees are caused by them being the main characters in most of the games and print works.

Yukari's high degree comes from her being connected to important figures of Gensokyo, such as Reimu, Okina, Kasen, and Yuyuko, as well as having a shikigami that has her own shikigami.

Kasen's high degree comes from her being the protagonist in Wild and Horned Hermit, as well as having a past of being affiliated with important figures of Gensokyo.

Without Reimu and Marisa, the mean average degree would be 2.636.

Without any characters with degrees 10 or higher, the mean average degree would be 2.47.

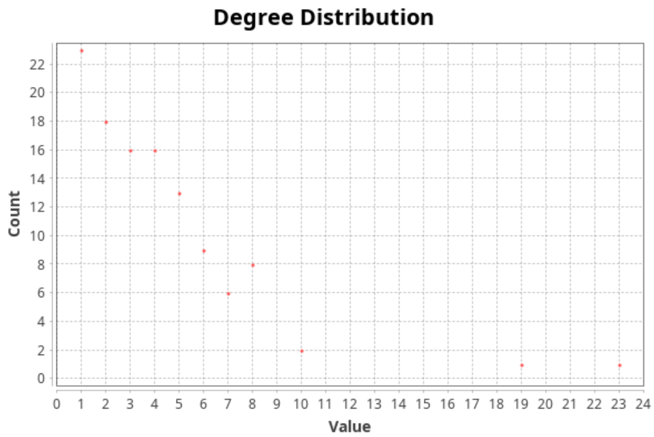

Fig. 6. The distribution above shows the distribution of degrees in the largest connected component. Its x axis represents the degree value and its y axis represents the amount of vertices.

In the largest connected component, the mean degree is 4.053, the median degree is 3, and the modus degree is still 1.

C. Paths

Within the largest connected component, the average path length is 4.203. This means that on average, a pair of characters in said connected component are separated by a chain of 4 connections.

The longest optimal path between any two nodes is 10 steps. There are six pairs of vertices that have path lengths of 10 steps, which are Ringo with Benben Tsukumo, Ringo with Yatsuhashi Tsukumo, Ringo with Mayumi Joutouguu, Reisen with Benben Tsukumo, Reisen with Yatsuhashi Tsukumo, and Reisen with Mayumi Joutouguu.

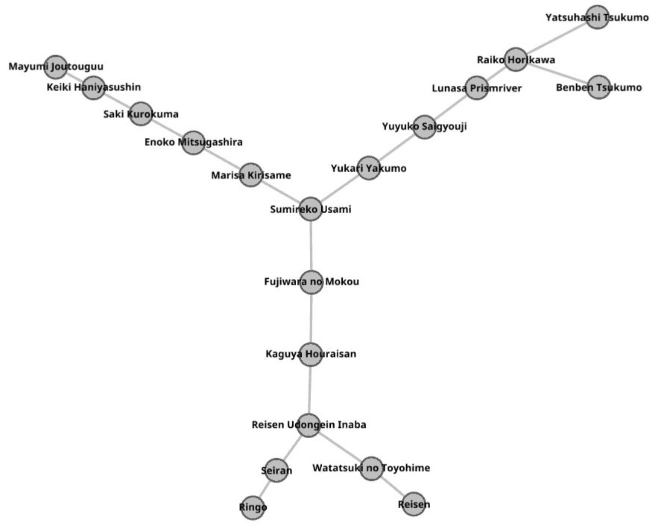

Fig. 7. The subgraph above shows the paths of the six aforementioned pairs.

The paths of all six pair go through a common subpath located in between the vertices of each pair, which is the following subpath: Reisen Udongein Inaba, Kaguya Houraisan, Fujiwara no Mokou, Sumireko Usami.

To connect to Ringo, the middle subpath must be extended with the following subpath: Reisen Udongein Inaba, Seiran, Ringo.

To connect to Reisen, the middle subpath must be extended with the following subpath: Reisen Udongein Inaba, Watatsuki no Toyohime, Reisen. By looking at vertices adjacent to this subpath, it's found that Watatsuki no Toyohime can be swapped with Watatsuki no Yorihime, due to both of them having the same adjacent vertices, which are Reisen Udongein Inaba, Eirin Yagokoro, and Reisen.

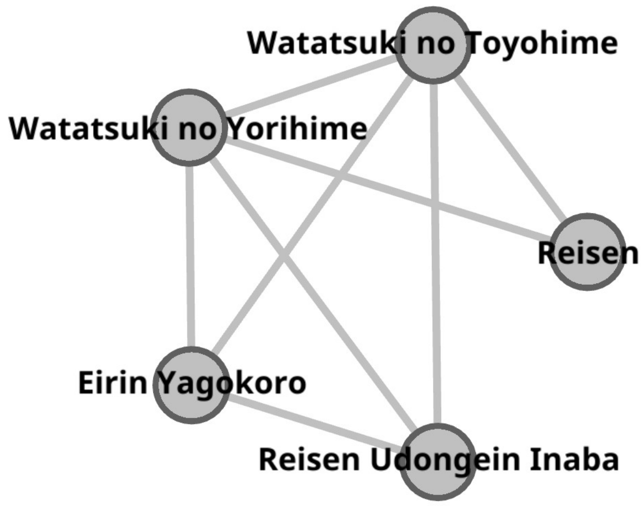

Fig. 8. The subgraph above shows all the vertices that are adjacent to Watatsuki no Toyohime or Watatsuki no Yorihime, which happen to be the same vertices.

To connect to Mayumi, the middle subpath must be extended with the following subpath: Sumireko Usami, Marisa

Kirisame, Enoko Mitsugashira, Saki Kurokuma, Keiki Haniyasushin, Mayumi Jouhoutuu.

To connect to Benben and Yatsuhashi, the middle subpath must be extended with the following subpath: Sumireko Usami, Yukari Yakumo, Yuyuko Saigyouji, Lunasa Prismriver, Raiko Horikawa. From there, Raiko is adjacent to both Benben and Yatsuhashi. By looking at vertices adjacent to this subpath, it's found that Lunasa Prismriver can be swapped with Merlin Prismriver and Lyrica Prismriver, due to all three of them having the same adjacent vertices, those being Yuyuko and Raiko.

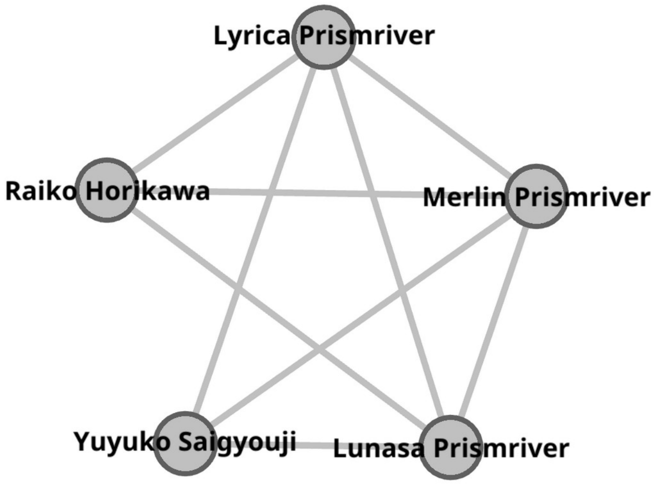

Fig. 9. The subgraph above shows all the vertices that are adjacent to Lunasa Prismriver, Merlin Prismriver, and Lyrica Prismriver, which happen to be the same vertices.

The vertex with the shortest optimal path to any other node in the biggest connected component is Sumireko Usami with a length of 5. In practice, it means that if someone were to establish a connection with someone in the Touhou cast with the purpose of being connected with the majority of the cast, the best option is Sumireko, as it would only take at most a chain of 6 connections to get in touch with anyone in the large network.

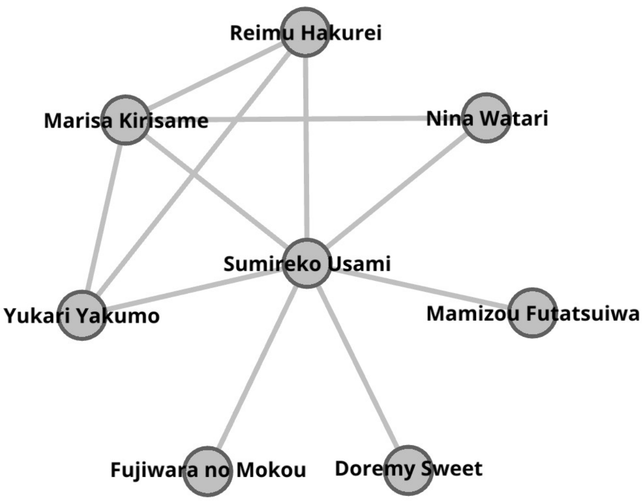

Fig. 10. The subgraph above shows Sumireko Usami's direct connections.

The reason Sumireko has this property can be explained by looking at her direct connections. She has 7 connections, well above the mean and median average. She's connected to Reimu, Marisa, Yukari, and Mamizou, whose degrees are 23, 19, 10, and 8, respectively. She's connected to Mokou, who's in one of the two only paths linking the subgraph containing Eientei's residents to the rest of the connected component.

Using Gephi's neighbors network tool, it's possible to see how many characters are in a certain distance to Sumireko by adjusting the depth. These are the amount of characters within Sumireko's reach at certain depths:

- Depth 1: 7 characters
- Depth 2: 42 characters
- Depth 3: 85 characters
- Depth 4: 108 characters
- Depth 5: 113 characters

This means that sumireko can reach to more than half of the entire Touhou Project cast within three connections.

Following Sumireko are 7 vertices which have a maximum optimal path length of 6. These vertices are Reimu Hakurei, Marisa Kirisame, Yukari Yakumo, Mamizou Futatsuiwa, Fujiwara no Mokou, Nina Watari, and Doremy Sweet. All of these characters are connected to Sumireko, as when a vertex adjacent to her is chosen, the vertex will be one step closer to some vertices but one step further from others as well.

D. Planar Test

The graph is non-planar because it contains a Kuratowski graph. There is a K_5 subgraph consisting of Hong Meiling, Patchouli Knowledge, Sakuya Izayoi, Remilia Scarlet, and Flandre Scarlet, as shown in the image below.

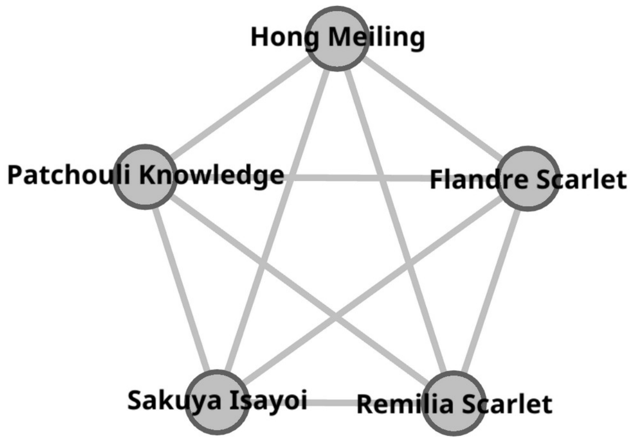

Fig. 11. The subgraph above shows a K_5 Kuratowski graph.

E. Chromatic Number

The chromatic number of the graph is at least 6. This is caused by the existence of a perfect subgraph with 6 vertices, which are Cirno, Star Sapphire, Luna Child, Sunny Milk, Clownpiece, and Eternity Larva. This means that in best case scenario, if all the characters were to be grouped in such a way that no two characters with a connection are in the same group, there would be 6 groups, each group containing one of the 6 characters in the aforementioned subgraph.

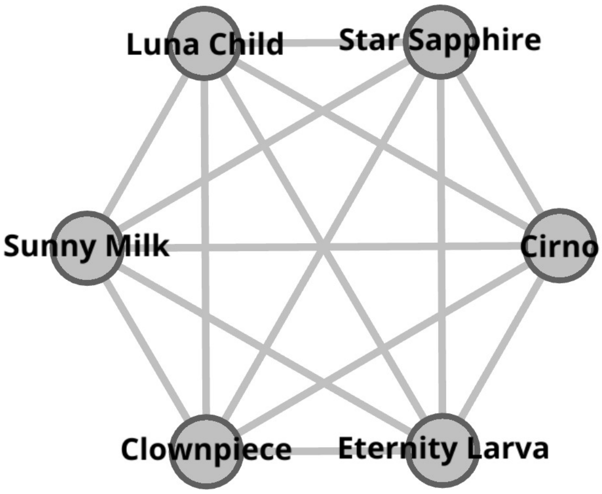

Fig. 12. The subgraph above shows a perfect graph containing 6 nodes.

V. Conclusion

In the world of the Touhou Project series, 74% of all characters are directly or indirectly connected with each other. Those 113 characters form the main network. On average, each character in the main network has 3 to 4 connections, with the majority having only 1. Reimu has the highest number of connections, followed by Marisa. The minimum amount of steps needed to get from any character to any other character in the main network is 10 steps. In order to minimize the amount of steps needed to reach someone in the worst case scenario, start from Sumireko, who can get to anyone else within at most 5 steps.

VI. Reccomendation

As I've mentioned before, the vertices and edges files may not be fully complete and lack proper documentation. Future papers should recheck the data and modify it should there be any mistakes. As of the time of this paper, the graph is also undirected and unweighted. Making the graph directed and weighted could result in the discovery of more insights.

References

[1] M. Rinaldi, "Graf (Bag.1)," Rinaldi Munir, 2026. https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2025-2026/20-Graf-Bagian1-2026.pdf

[2] M. Rinaldi, "Graf (Bag.2)," Rinaldi Munir, 2026. https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2025-2026/21-Graf-Bagian2-2026.pdf

[3] M. Rinaldi, "Graf (Bag.3)," Rinaldi Munir, 2026. https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2025-2026/22-Graf-Bagian3-2026.pdf

[4] Ō. Jun'ya, Touhou Project series, 1997-2026.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Jakarta, 19 June 2026

Christopher Hendrik Gunawan / 13525065