Topological properties of complementarity spatial graph models

Pinqi Guo

Delft University of Technology

TU Delft

Topological properties of complementarity spatial graph models

by

Pinqi Guo

to obtain the degree of Master of Science
in Electrical Engineering
Track Wireless Communication and Sensing
at the Delft University of Technology,
to be defended publicly on Wednesday March 12, 2025 at 10:00 AM.

Student number:	Pinqi Guo	5656540
Project duration:	January, 2024 – March, 2025	
Thesis committee:	Dr. Maksim Kitsak,	TU Delft, chair
	Prof. dr. ir. Rob Kooij,	TU Delft
	Dr. Sebastian Feld,	TU Delft

Cover : https://pixabay.com/zh/

TU Delft

Preface

Time has passed quickly, and I am now ready to finish my Master's program in Electrical Engineering, specializing in Wireless Communication and Sensing. Looking back on these two and a half years of my Master's journey at TU Delft, I have experienced and gained a lot. When I officially joined the Network Architectures and Services (NAS) group in January 2024 to start my master's thesis, I knew it would be a rare and valuable journey.

During my thesis period, my daily supervisor, Professor Maksim Kitsak, provided me with great guidance and support, serving as a mentor throughout my academic journey. Specifically, whether it was offering solution ideas when I encountered problems, providing encouragement and support at critical moments, or giving timely feedback on each part of my thesis, his guidance has been invaluable in helping me navigate challenges and make steady progress. Meanwhile, PhD candidates Elizaveta Evmenova and Zhihao Qiu selflessly dedicated their time to my daily discussions and study sessions, offering me timely feedback and academic advice in their spare time. In addition, I would like to express my gratitude to Dr. Gabriel Budel, for his guidance during the preparatory phase of my thesis. Furthermore, I am grateful to Professor Rob Kooij and the other professors of the NAS group for their feedback during my midterm presentation, as well as to all the PhD students and staff of the NAS group for their support. This one-year experience has been an invaluable asset in my life. Finally, I would also like to thank Professor Sebastian Feld for kindly accepting to be a member of my thesis committee.

Lastly, I would like to express my gratitude to everyone who has supported me over the past two and a half years. A special thanks goes to my girlfriend, Ziyang Ye, and my family for always being there for me, providing the motivation I needed to keep moving forward. I am truly grateful for all that I have and look forward to a bright tomorrow and a promising future.

Pinqi Guo
Delft, March 2025

Abstract

In network science, numerous studies based on the complementarity principle have emerged since 2018 [1]. Unfortunately, theoretical foundations of complementarity are still in their infancy. Recently, a synthetic complementarity-based model called Complementarity Random Hyperbolic Graph (CRHG) has been proposed. CRHG model was assumed to explain the topological properties of real complementarity-driven networks. In other words, this model could serve as a foundation for studying complementarity mechanisms in networks.The main goal of this thesis is to address the knowledge gap: although a complementarity-based network model has been proposed in previous studies, its topological properties have not been systematically examined. To fill in this gap, this work systematically studies complementarity network models (CRHG, GCRHG) and documents their topological properties. Moreover, we interpret the topological properties as a function of the network model parameters. We find that the CRHG model exhibits three fundamental properties: Scale-Free property, Small-World property, and Non-vanishing bipartite clustering. It indicates that it is a unique combination compared to other synthetic models. Its unique complementary connectivity mechanism makes it particularly effective for modelling complex networks formed by the complementarity mechanism. Furthermore, we also study and investigate a generalized synthetic model called Generalized Complementarity Random Hyperbolic Graph (GCRHG). We measure and analyze its clustering and bipartite clustering properties. We find that this model allows for smooth turning between similarity and complementarity. Overall, we document and interpret the topological properties of simulations for complementarity-based spatial graph models. Additionally, we conduct partial simulation verification of the theoretical topological properties of synthetic complementarity-based models, providing a reference for their future development and applications.

Contents

1 Introduction

Since the beginning of the 21st century, with the development of data science and the rise of data mining techniques, network science has emerged, and the network characteristics exhibited by various types of real networks have increasingly aroused people's attention. Networks are characterized by topological properties, also known as structural characteristics. These properties include among others, the Small-World property, Scale-Free degree distributions, and strong clustering. We will discuss the topological properties of interest in more detail in Section 2.3. Essentially, the various mechanisms that different networks exhibit are closely related to the distribution and arrangement of nodes and links within the networks.

In real world networks, many topological properties have been observed. To study and understand these characteristics, numerous synthetic network models have been proposed. More specifically, the simulation results computed by these mathematical models using different combinations of input parameters not only help us study the established topological properties observed in real networks but also inspire us to formulate conjectures regarding previously unknown characteristics. G. Budel and M. Kitsak have proposed a complementarity-based network model, but the topological properties of this model have not been studied systematically. In this thesis, we aim to fill in the knowledge gap by studying the topological properties of the complementarity-based synthetic network model proposed by G. Budel and M. Kitsak [2].

1.1. Network: definitions and examples

A network is a structure that consists of a set of nodes connected by links. Many complex systems surrounding us can be represented as complex networks, abstracting elements of the systems as nodes and interactions between them as links.

Examples of complex networks include social networks [3] where nodes are individuals and links are social interactions; communication networks [4] where nodes are computing devices exchanging data and links indicate the data transfer relationship between them; and biological networks [5] where nodes are molecules and links correspond to physical interactions between them. These networks originate from datasets across various scientific fields.

1.2. Motivation and Objectives

Motivation:

Real networks possess several common topological properties: the Small-World property, Scale-Free degree distribution, and strong clustering. A large number of synthetic network models were proposed to explain the emergence of these properties.

Complementarity has recently been identified as an important mechanism for network formation. In simple terms, two nodes with complementary properties have higher chances to be connected. In [2], G. Budel and M. Kitsak have proposed a complementarity-based network model, but the topological properties of this model have not been studied systematically.

Objectives:

The main objective of this thesis is to address the knowledge gap: although a complementarity-based network model has been proposed in previous studies, its topological properties have not been systematically examined. To fill in this gap, this work systematically studies complementarity network models (CRHG, GCRHG) and documents their topological properties. The following interesting properties are considered :

(1) What is the scaling of network-based distance in a graph created by the model?
(2) What is the degree distribution of the node degree in the graph created by the model?
(3) What is the density of triangular subgraphs in the graph created by the model?
(4) What is the density of quadrilateral subgraphs in the graph created by the model?

The outcomes:

(1) The code for the generation of the complementarity network
(2) The code to measure network-based distances, degree distribution, and the densities of triangles and quadrilaterals in graphs created by synthetic network model
(3) The results of the measurement and the interpretation of the topological properties (1)-(4) as a function of the network model parameters.
(4) The results of the measurement and the interpretation of the topological properties (1)-(4) as a function of the generalized network model parameters.

1.3. Summary

The structure of this thesis is as follows:

Chapter 2 briefly introduces key network topology metrics, essential network topological properties, and classical network models along with their properties. This chapter aims to establish the background required for the research presented in the subsequent chapters.

Chapter 3 describes the concepts of similarity and complementarity, highlighting the motivation for studying the Complementarity Random Hyperbolic Graph (CRHG) model.

Chapter 4 presents the Complementarity Random Hyperbolic Graph (CRHG) model and its properties. It begins with a formal definition of the CRHG model. Then, we measure and interpret the topological properties as functions of the network model parameters. Finally, the

chapter includes an intermediate summary, discussing the position of the CRHG model in relation to other network models.

Chapter 5 presents the Generalized Complementarity Random Hyperbolic Graph (GCRHG) model and its properties. The chapter starts with a formal definition of the GCRHG model and proceeds to analyse and interpret its topological properties as functions of the network model parameters.

Chapter 6 summarizes the main achievements and discusses the findings. Additionally, the chapter provides a brief outlook on potential future research directions.

2

Background

In this chapter, we begin with the basic definition of the graph. Next, we introduce the related topological metrics. We then define and briefly explain several common topological properties, such as the Small-World property, Scale-Free property, Strong clustering coefficient and Strong bipartite clustering, which can be quantified using these topological metrics. Finally, we explore network modelling by presenting several classic synthetic models that facilitate the study and explanation of these topological properties.

2.1. Graph Theory

The interacting relationships between the components and structures form a **network** in complex systems [6]. A **graph** is a mathematical representation of a network. In graph theory, it is a structure that consists of a set of nodes connected by a set of links [7]. In more detail, the set of nodes is denoted as $\mathcal{N}$, with the number of nodes represented as $N = |\mathcal{N}|$. The set of links is denoted as $\mathcal{L}$, with the number of links represented as $L = |\mathcal{L}|$, see Fig. 2.1 (a). It is a toy graph consisting of $N = 5$ nodes and $L = 5$ links. Meanwhile, we can also describe the graph G using its adjacency matrix $A_{N\times N}$ as shown in Fig. 2.1 (b). The adjacency matrix of an undirected graph is a square $n \times n$ matrix A such that its element $A_{ij} = 1$ if there is a link between node i and node j, else $A_{ij} = 0$. The term "undirected" will be formally defined later. The rows and columns of the adjacency matrix $A_{N\times N}$ correspond to the labels of nodes in Fig. 2.1 (a).

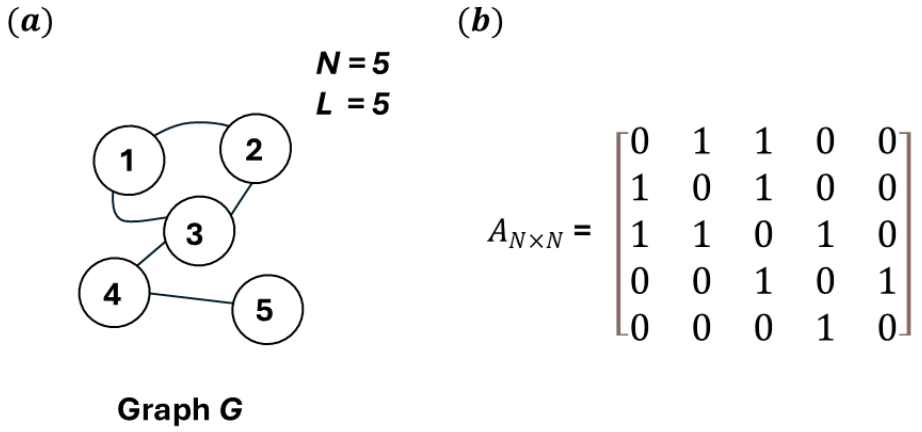

Figure 2.1: A toy graph consisting $N = 5$ nodes and $L = 5$ links with its adjacency matrix $A_{N\times N}$. The labels of nodes are the numbers in nodes

In this thesis, we focus exclusively on **connected**, **undirected**, **unweighted** and **simple** graphs. This implies that the graphs satisfy the following properties:

(1) **Connected:** In such a graph, there is a path between any pair of its nodes [8].
(2) **Undirected:** The links have no direction in such a graph.
(3) **Unweighted:** The links have no weights in such a graph.
(4) **Simple:** There are no self-loops and multiple links between the nodes in such a graph.

2.2. Topological metrics for networks

In this section, we introduce the basic topological metrics of the graphs and summarize them in Table. 2.1. We have provided the definitions of the number of nodes N and the number of links L in Section 2.1.

Mathematical notation	Metrics
N	number of nodes
L	number of links
ρ	diameter
$\langle d \rangle$	average shortest path length
k	degree
k_{min}	minimum degree
k_{max}	maximum degree
$\langle k \rangle$	average degree
$Pr[D = k]$	degree distribution
C_3	clustering coefficient
C_4	bipartite clustering coefficient

Table 2.1: Topological metrics used in this thesis and their mathematical notations

2.2.1. Graph Diameter and Average shortest path length

A **path** is a route that we can take from one node to another, passing through a set of intermediate nodes without revisiting any node. To quantify the path, a **hopcount** of a path P is defined as the number of links in P [7]. Specially, one example of the path of length n from node i_0 to node i_n in an unweighted and undirected graph G is shown in Fig. 2.2 (a) with a purple line. In Fig. 2.2 (b), the red line represents the **shortest path** from node i_0 to node i_n, which is $P_{i_0 \to i_n}$ with the minimum hopcount.

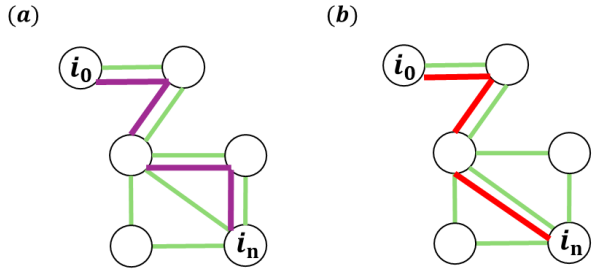

Figure 2.2: Toy example path and shortest path (between node i_0 and node i_n): The purple line is one path example between node i_0 and node i_n; the red line is the shortest path example between node i_0 and node i_n

The **diameter** ρ of graph G is defined as the hopcount of the longest shortest path in G [7]. It represents the shortest path between the two nodes farthest apart in graph G. The **average shortest path length** $\langle d \rangle$ is defined as the average hopcount of the shortest paths between all pairs of nodes.

2.2.2. Degree and Degree Distribution

The **degree** $k_i \in [0, N-1]$ of a node i in a graph G is the number of its neighbours [7] or, equivalently the number of adjacent links. The node i is disconnected from the rest of the graph when $k_i = 0$. In this thesis, we focus on connected graphs; thus, the range of k_i is transformed into $k_i \in [1, N-1]$. We denote the minimum and maximum nodes degrees in graph G as $k_{min} = \min_{i \in G} k_i$ and $k_{max} = \max_{i \in G} k_i$, respectively. A fundamental relationship between node degrees and the number of links L is:

$$\sum_{i=1}^{N} k_i = 2L \tag{2.1}$$

which states that it equals twice the number of links in the graph. For example, when calculating the total degree of two nodes connected by a single link, this link is counted twice (once for each node). Using Eq.(2.1), we can express the average graph degree $\langle k \rangle$ as:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N} \tag{2.2}$$

The **degree distribution** $Pr[D = k]$ of a network is defined as the probability a randomly chosen node has degree $D = k$. In other words, it shows the probability that a randomly chosen

node has exactly k links [6]. Given a graph G, one can computer $Pr[D = k]$ as:

$$Pr[D = k] = \frac{N_k}{N} \tag{2.3}$$

where N_k is the number of nodes with degree k.

The degree distribution $Pr[D = k]$ plays a crucial role in networks because many network properties depend on it, including robustness and epidemic spreading [6].

2.2.3. Clustering Coefficient, C_3

In a graph G, the **clustering coefficient** [9] quantifies the extent to which the neighbours of a given node are interconnected [10].

The local clustering coefficient $C_3(i)$ of node i in a graph G is the ratio of the number of links y between its k_i neighbours over the total possible connections $\frac{k_i(k_i-1)}{2}$ [7]:

$$C_3(i) = \frac{2y}{k_i(k_i-1)}, C_3(i) \in [0, 1] \tag{2.4}$$

$C_3(i)$ is defined to be equal to zero for all nodes with $k_i = 0$ or $k_i = 1$. $C_3(i)$ is equal to one only when the neighbours of node i are completely connected. The local clustering coefficient provides a quantitative measure of the local density of connections in the neighbourhood of the node i. Figure. 2.3 illustrates an example of the calculation of the local clustering coefficient $C_3(i)$. For node i, there exists only a single connection among its three neighbours. The maximum possible number of connections between these three neighbours is three. The $C_3(i)$ for node i is $\frac{1}{3}$ in this context.

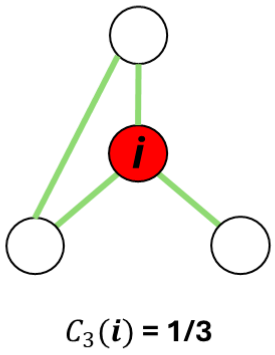

Figure 2.3: An example of computing $C_3(i)$ for node i

As shown in the above example, node i and its neighbours form a closed triangle. The clustering measures the density of triangles.

For the entire graph, the average clustering coefficient $\bar{C}_3$ of the graph G is defined as:

$$\bar{C}_3 = \frac{1}{N} \sum_{i=1}^{N} C_3(i), \tag{2.5}$$

where $\bar{C}_3$ is equal to zero when there are no triangles in G. In contrast, $\bar{C}_3$ is equal to one when G is fully connected.

The average local clustering coefficient $\bar{C}_3(k)$ of nodes with degree k is defined as:

$$\bar{C}_3(k) = \frac{1}{N_k} \sum_{i \in \{\text{nodes with degree k}\}} C_3(i) \tag{2.6}$$

where N_k is the number of nodes with degree k in a graph. It describes the relationship between the clustering coefficients and the degrees of the nodes in a graph.

2.2.4. Bipartite Clustering Coefficient, C_4

There are situations where the clustering coefficient is not useful, such as when its value is consistently zero. One such example is bipartite networks, a class of networks consisting of two distinct sets U and V, with connections existing only between nodes from different sets [11]. In bipartite networks, triangles are absent because no two nodes within the same set are connected, making the clustering coefficient C_3 always equal to zero. Zhang et al. [12] introduced the **bipartite clustering coefficient**, which can be used to measure the density of quadrilateral in a graph G, and is inspired by the conventional clustering coefficient C_3.

The local bipartite clustering coefficient $C_4(i)$ of node i in a graph G is the ratio of the number of common neighbouring nodes of all its neighbouring pairs over the number of all neighbouring nodes of all neighbouring pairs [11] as:

$$C_4(i) = \frac{\sum_{j \neq e} (m_{je} - 1)}{\sum_{j \neq e} [k_j + k_e - m_{je} - 1]}, C_4(i) \in [0, 1] \tag{2.7}$$

where $m_{je} - 1$ is the total number of common neighbours between node j and node e expect node i itself. The entire numerator of the equation represents the sum of $m_{je} - 1$, iterated over all pairs of neighbours j, e of node i. k_j and k_e are the degrees of node j and node e. The entire denominator of the equation represents the number of all potential quadrilaterals that include node i. Here, "potential" refers to both existing and possible quadrilaterals. To explain $C_4(i)$ more clearly, we suppose that each of the neighbours j and e of node i is connected to only one other node besides i as shown in Fig. 2.4.

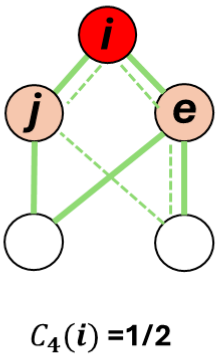

Figure 2.4: One example of computing $C_4(i)$ for node i, where the nodes j and e are the neighbours of node i

when j and e have exactly one common neighbour besides i, as depicted in Fig. 2.4. A quadrilateral is formed by node i, node j, node e, and the common neighbour shared by node j and node e. The dotted line in the figure shows another possible quadrilateral that contains node i. The $C_4(i)$ for node i is calculated as $\frac{1}{2}$.

The average bipartite clustering coefficient C_4 of the graph G follows the Eq.(2.8).

$$\bar{C}_4 = \frac{1}{N}\sum_{i=1}^{N} C_4(i) \tag{2.8}$$

where $\bar{C}_4$ is equal to zero when there is no quadrilateral in G. In contrast, $\bar{C}_4$ is equal to one when every local bipartite clustering coefficient $C_4(i) = 1$.

The average local bipartite clustering coefficient $\bar{C}_4(k)$ of nodes with degree k is defined as:

$$\bar{C}_4(k) = \frac{1}{N_k}\sum_{i\in\{\text{nodes with degree k}\}} C_4(i) \tag{2.9}$$

where N_k is the number of nodes with degree k in a graph. It describes the relationship between the bipartite clustering coefficients and the degrees of the nodes in a graph.

2.3. Topological properties of networks

Real networks have been documented to exhibit common topological properties that can be quantified using specific topological metrics. In this section, we will apply these metrics to define and explain some of these properties. The following important topological properties will be addressed: Small-World property, Scale-Free property, strong clustering coefficient, and strong bipartite clustering.

2.3.1. Small-World Property

The Small-World property refers to a phenomenon that typical paths are small even in large size networks. This can be described that although many networks are large in size, a relatively short path often connects any two nodes. One of the most well-known manifestations of this property was first mentioned as the concept of ***six degrees of separation*** [13], which suggests that any two people, wherever in the world, are typically linked by a chain of about six acquaintances. The term Small-World property was first introduced in the context of random networks by I. de Sola Pool and M. Kochen [14]. Building on their theoretical work, J. Travers and S. Milgram conducted the first empirical study of the Small-World phenomenon, famously illustrating it through experiments involving human social networks [13, 15].

In terms of theoretical under pinning, Erdős et.al. [16] proved that the diameter ρ between any two nodes in a random graph scales as the logarithm of N [6, 17]. This laid the theoretical foundation for the short path property in random graphs. Furthermore, Watts et.al.[9] introduced the small-world model based on the idea that networks in reality are in between random graphs and regular lattices. This model retains the high clustering of the regular lattice and introduces the short path property of the random graph [9, 18].

As we now understand, graphs with the Small-World property exhibit one prominent characteristic: The average shortest path length $\langle d \rangle$ typically scales as the logarithm of N as [6], [9]:

$$\langle d \rangle \sim \ln N \tag{2.10}$$

Here, the short-path feature of the Small-World property is reflected not only in the fact that the absolute value of $\langle d \rangle$ is small compared to the size of the network N, but also in the observation that $\langle d \rangle$ grows very slowly with N.

2.3.2. Scale-Free Property

Most real networks are characterized by power-law degree distribution:

$$Pr[D = k] \sim k^{-\gamma} \tag{2.11}$$

with $\gamma \in (2, 3)$. They are called Scale-Free networks [19, 20] because the standard deviation σ_k in node degree is infinite. In other words, the "scale" in this context is quantified by the standard deviation of the degree distribution in a graph. However, when the standard deviation $\sigma_k = \infty$, this "scale" fails to capture the variability in the degree distribution, called "scale-free".

Indeed, the moments of the degree distribution have a decisive impact on standard deviation σ_k:

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2 \tag{2.12}$$

here, the n^{th} moment of the degree distribution is defined as [10]:

$$\langle k^n \rangle \equiv \sum_{K=1}^{\infty} k^n Pr[D = k] \approx \int_{k_{min}}^{k_{max}} k^n Pr[D = k] dk \tag{2.13}$$

where the first moment is $\langle k \rangle$ when $n = 1$. The second moment is $\langle k^2 \rangle$ when $n = 2$. As $n \geq 2$, the $\langle k^n \rangle$ can be represented as:

$$\langle k^n \rangle = \int_{k_{min}}^{k_{max}} k^n Pr[D = k] dk = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n - \gamma + 1} \tag{2.14}$$

here, we first assume that $k_{max} \to \infty$ while k_{min} remains fixed for a large size network. Under this assumption, the range of values for $\langle k^n \rangle$ depends on the coefficient $n - \gamma + 1$. When $n \geq 2$ and $\gamma \in (2, 3)$, the coefficient $n - \gamma + 1$ is always greater than zero. Therefore, $\langle k^n \rangle \to \infty$ as $k_{max} \to \infty$, leading to the second moment of the degree distribution $\langle k^2 \rangle$ being much larger than the first moment $\langle k \rangle$. Consequently, σ_k^2 becomes infinite, which implies that σ_k is also infinite. In other words, the network is scale-free, as it lacks an intrinsic scale. When $\gamma \leq 2$, the conditions for a scale-free network no longer hold. In particular, for a scale-free network, the relationship between k_{max} and k_{min} is as follows:

$$k_{max} = k_{min} N^{\frac{1}{\gamma - 1}} \tag{2.15}$$

when $\gamma \leq 2$, k_{max} grows faster than N. This scenario is not possible in a network due to $k_i \in [0, N - 1]$. In the case $\gamma \geq 3$, it is hard to distinguish a scale-free network from a random network because documenting the scale-free nature of a network in this context requires large networks for sufficient scaling, which is rarely achievable.

An example of a degree distribution $Pr[D = k]$ measured on a real-world network is shown in Fig. 2.5. The figure illustrates a Scale-Free real network following a power-law degree distribution with $\gamma \in (2, 3)$.

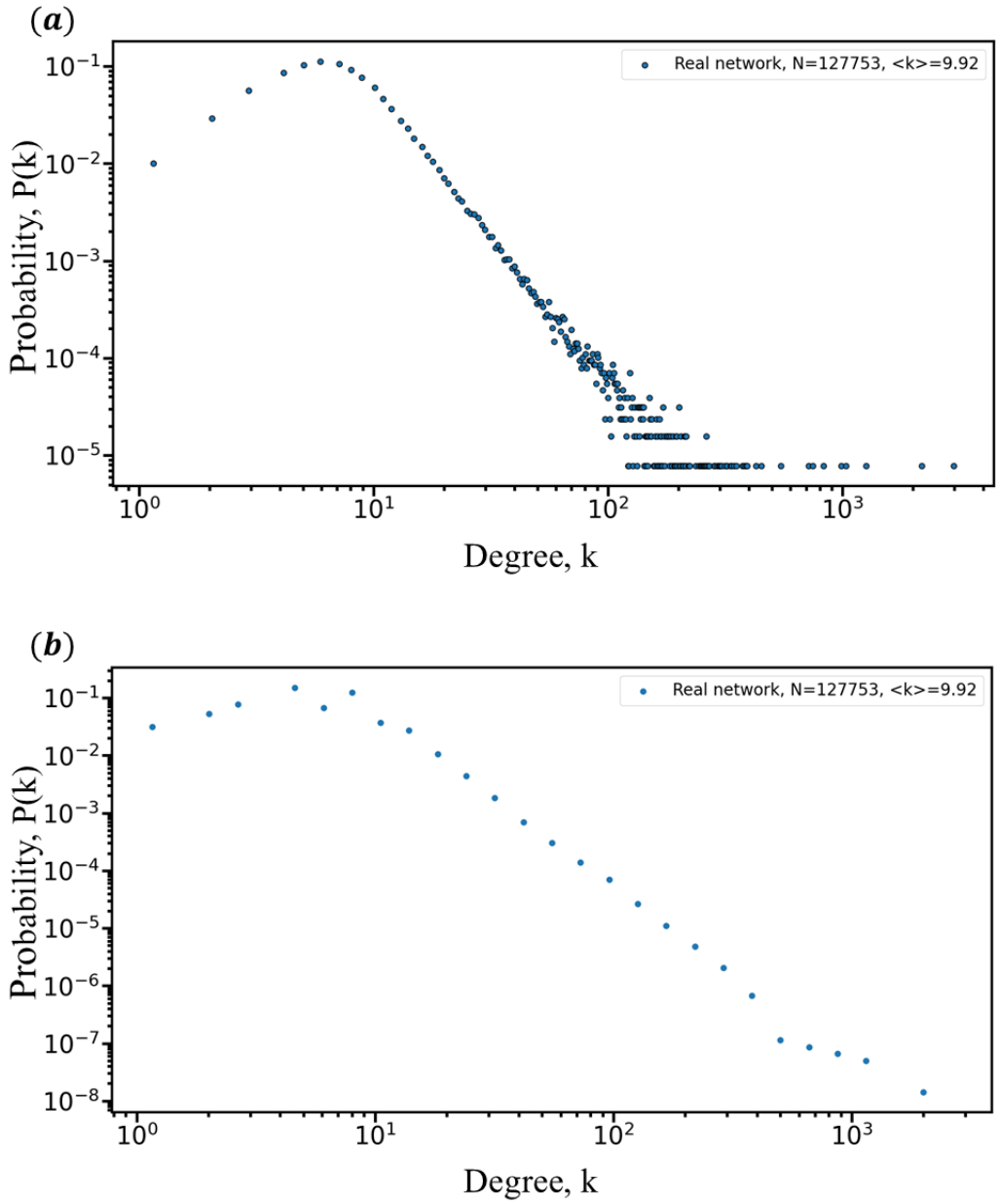

Figure 2.5: Degree distribution $Pr[D = k]$ of a real-world network with $N = 127753$, $\langle k \rangle = 9.92$. We use the linear binning in panel (a) and logarithmic binning [21] in panel (b).

2.3.3. Strong Clustering Coefficient and Strong Bipartite Clustering

In network science, clustering coefficients describe the tendency of nodes in a graph to form closely connected groups. These groups are typically represented as closed triangles. Clustering measures the density of such triangles within the entire graph. To quantify clustering in a graph, the local clustering coefficient $C_3(i)$ is used to describe the clustering of a single node, while the average local clustering coefficient $\bar{C}_3(k)$ is used to describe the clustering of nodes with the same degree k. Additionally, the average clustering coefficient $\bar{C}_3$ reflects the clustering of the entire graph and can be used to capture the trend of clustering as the size of the network N increases. On this basis, **strong clustering** emphasizes the case where connections in a graph form a highly cohesive structure, usually dominated by a large number of closed triangles.

Similarly, all coefficients in bipartite clustering serve the same purpose. The only difference is that bipartite clustering measures the density of quadrilaterals. Consequently, **strong bipartite clustering** refers to the scenario where the clustering of the entire graph is dominated by closed quadrilaterals rather than closed triangles. In the synthetic models mentioned in Section 2.4, **"strong"** implies:

(1) The clustering coefficient, the bipartite clustering coefficient does not vanish as the network size N increases.

(2) The clustering coefficient, the bipartite clustering coefficient remains significantly higher than expected for rewired (Degree-Preserving Randomization [22]) graphs with similar N and L as the network size N increases.

As illustrated in Fig. 2.6, this example demonstrates the measurement of C_3 as the network size N increases in the Generalized Complementarity Random Hyperbolic Graph (GCRHG) model, as discussed in Chapter 5. We can observe that the clustering coefficients C_3 are much higher than those of the rewired (Degree-Preserving Randomization [22]) graphs, and these coefficients do not disappear as the network size N increases. This means that the graph created by the GCRHG model with these parameters has strong clustering coefficient.

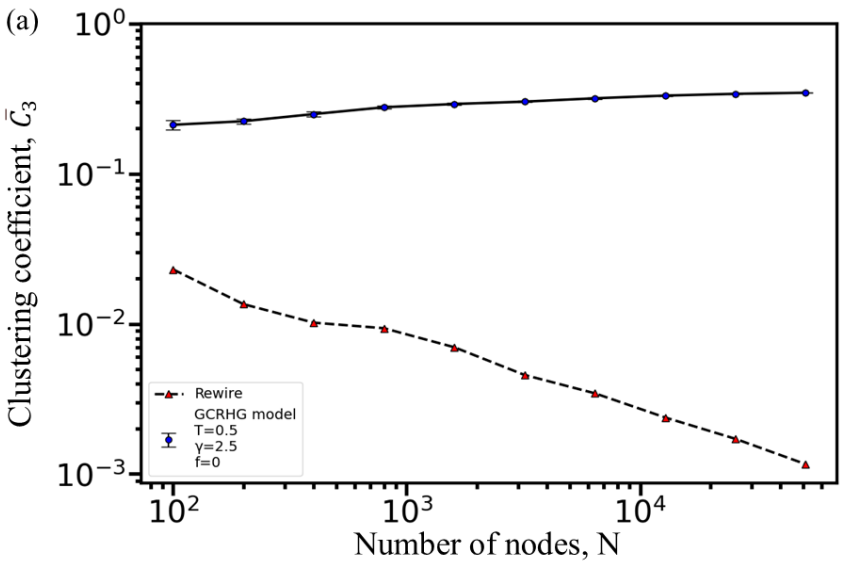

Figure 2.6: As shown in Fig. 5.3 (a), trends of C_3 for networks with strong clustering: The black line reflects the float of the C_3 in networks with strong clustering when the size N of the network increases. The black dotted line represents the C_3 values for the rewired graphs (Degree-Preserving Randomization [22]) with similar N and L.

Figure. 2.7 illustrates the strong bipartite clustering characteristic in the Generalized Complementarity Random Hyperbolic Graph (GCRHG) model, as discussed in Chapter 5. It shows that the bipartite clustering coefficients C_4 are higher compared to those of the rewired (Degree-Preserving Randomization [22]) graphs, which have the same network size N and L. This indicates that the graph created by the GCRHG model with these parameters exhibits strong bipartite clustering.

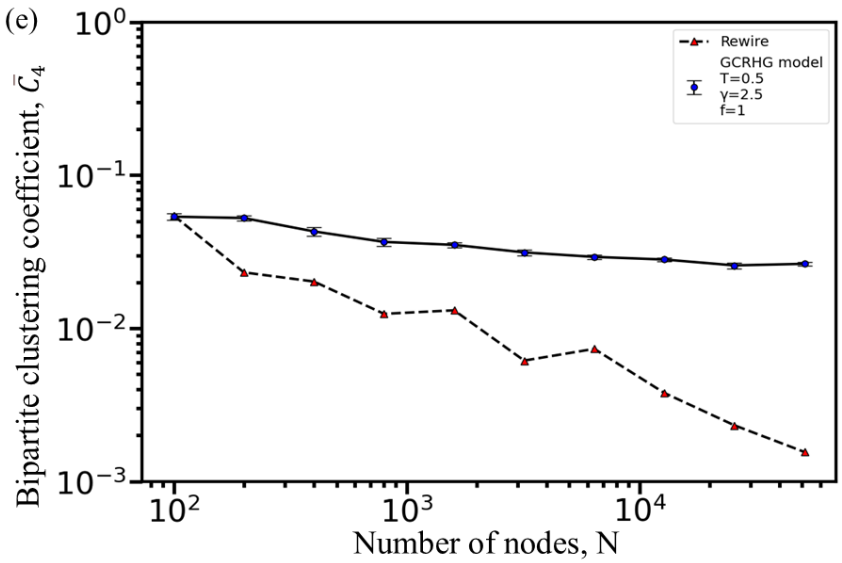

Figure 2.7: As shown in Fig. 5.7 (e), trends of C_4 for networks with strong bipartite clustering: The black line reflects the float of the C_4 in networks with strong bipartite clustering when the size N of the network increases. The black dotted line represents the C_4 values for the rewired graphs (Degree-Preserving Randomization [22]) with similar N and L.

2.4. Classical network models and their properties

The concepts of short path lengths, clustering, and scale-free degree distributions have sparked an interest in network modelling [6]. After understanding the topological properties of real-world networks, the idea of modelling is to propose a simple mechanism to capture or explain the property of interest. A natural step is to explore whether a mathematical model can be used to describe these properties. Indeed, several synthetic network models have been developed to facilitate the study and explanation of these properties. In the following, some classic synthetic models are briefly introduced to provide an overview of network modelling.

2.4.1. Erdos-Renyi (ER) model

The Erdős–Rényi (ER) random graph [23] is one famous random graph model. In [23], the ER model is defined as $G(N, L)$: N nodes are randomly connected with L links. Another form of the ER model is represented as $G(N, p)$ given by Gilbert et al.[24]. $G(N, p)$ consists of N nodes and each pair of nodes is independently connected to the probability p. This distinguishes two versions of the ER model. Then, we will work with $G(N, p)$ henceforth.

Most real-world networks appear disorganized and complex [25]. The ER model simplifies this by assuming that links between nodes are placed randomly, with the likelihood of a connection determined by a probability p. This randomness provides a basic framework for studying network structures.

Small-World Property: The ER model exhibits the small world property $\langle d \rangle \sim \ln N$ [10].

Scale-Free Property: The degree distribution of the ER model is the binomial distribution [26] following Eq.(2.16).

$$Pr[D = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2.16}$$

It is well approximated by a Poisson distribution in the $k \ll N$ limit.

Clustering: For the ER model, the average clustering coefficient depends on the size N as $\frac{1}{N}$ [10].

The Erdős–Rényi (ER) model exhibits the small-world property, ensuring short path lengths between nodes. However, it lacks the scale-free property, as its degree distribution follows a Poisson distribution. Additionally, the ER model does not exhibit strong clustering, as its average clustering coefficient decreases inversely with the network size.

2.4.2. Preferential Attachment (PA) model

Barabasi et al. observed that real-world networks follow a power-law degree distribution [19]. They argued that growth and preferential attachment coexist in real networks. To describe these two characteristics, they introduced the Preferential Attachment (PA) model which can generate networks with the scale-free property [19]. The size N graph of the PA model is defined as follows: (1) The PA model starts with m nodes without any link. (2) Add a new node with m links to m existing nodes each time. The probability that the new node connects to an existing node i is proportional to the current degree of node i following:

$$\Pi_i(t) = \frac{\mathrm{d}_i(t)}{\Sigma_j d_j(t)} \tag{2.17}$$

(3) Repeat (2) until size N is reached.

Generally, the PA model has two generic mechanisms. Growth: The network is continuously expanding by adding new nodes. Preferential attachment: Newly added nodes tend to preferentially connect to nodes that already have more connections. The PA model explains well the phenomenon of "the rich get richer" well, which is observed in many large complex networks.

Small-World Property: The observed relationship between $\langle d \rangle$ and N is [27, 28]:

$$\langle d \rangle \sim \frac{\ln(N)}{\ln(\ln(N))} \tag{2.18}$$

Scale-Free Property: The degree distribution of the PA model is given by Eq.(2.19)

$$Pr[D = k] \sim k^{-3} \tag{2.19}$$

The PA model has a power-law degree distribution with $\gamma = 3$.

Clustering: The average clustering coefficient decreases as a function of $\frac{(\ln(N))^2}{N}$ [29, 30].

The Preferential Attachment (PA) model captures several properties. It exhibits the Small-World property and displays the Scale-free property. However, its clustering is relatively weak for large networks.

2.4.3. Random Geometric Graph (RGG) model

The Random Geometric Graph (RGG) model has been introduced to model wireless networks, such as Wireless ad hoc networks [31]. In the RGG model, every node is a point in a *D*-dimensional hypercube $[0,1]^D$ [32]. Each coordinate of a point is drawn from a uniform distribution over the unit interval. Every node pair i and j is connected if $d_{ij} < R$, where R is the parameter and d_{ij} is the distance between node i and j.

The connection of nodes in the RGG model is closely related to their spatial distance. The spatial structure of the network based on the RGG model itself cannot be generated by ER model and PA model generation algorithms instead.

Small-World Property: For the short path property in the *D*-dimensional RGG model, it is polynomial $\langle d \rangle \sim N^{\frac{1}{D}}$.

Scale-Free Property: In the RGG model, a link exists only between nodes that are "close" to each other, where "close" means $d_{ij} < R$. Therefore, the degree distribution of a RGG model with average degree $\langle k \rangle$ is [32]:

$$Pr[D = k] \sim \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!} \tag{2.20}$$

It is close to the Poisson distribution [33].

Clustering: The clustering coefficient of the RGG model depends on the dimension D of the hypercube $[0,1]^D$. The detailed expression for C_D is [32]:

$$C_D = \begin{cases} 1 - H_D(1) & \text{even } D \\ \frac{3}{2} - H_D\left(\frac{1}{2}\right) & \text{odd } D \end{cases} \tag{2.21}$$

where

$$H_D(x) = \frac{1}{\sqrt{\pi}} \sum_{i=x}^{D/2} \frac{\Gamma(i)}{\Gamma\left(i + \frac{1}{2}\right)} \left(\frac{3}{4}\right)^{i+1/2} \tag{2.22}$$

For large D dimension, C_D is simplified to $C_D \sim 3\sqrt{\frac{2}{\pi D}}(\frac{3}{4})^{\frac{D+1}{2}}$.

In low-dimensional RGG models, such as *2*-dimensional model, the $C_D \approx 0.59$ is calculated by Eq.(2.21) [32]. For the high-dimensional RGG models, the $C_D \approx 0$.

In summary, the RGG model explains strong clustering but does not explain the Scale-Free and Small-World properties.

2.4.4. Random Hyperbolic Graph (RHG) model

The Random Hyperbolic Graph (RHG) model has been introduced to be a good statistical model that supports and explains real-world networks. It is a relatively simple model that explains all three properties: Small-World, Scale-Free and Clustering Coefficient. The applications of the RHG model include link prediction[34], routing and navigation [35, 36], and other things [37].

RHG model

Hyperbolic space is a type of non-Euclidean geometry characterized by a constant negative curvature $K = -\zeta^2$ [38]. A *d*-dimensional hyperbolic space with curvature is denoted as $\mathbb{H}^d_\zeta$. To represent the *2*-dimensional hyperbolic space, the Poincar´e unit disk model can be used by describing the entire infinite hyperbolic plane $\mathbb{H}^2$ as the interior of the Euclidean disk of radius one [39].

In this thesis, the RHG model is constructed in the latent space $\mathbb{H}^2$, with curvature $K = -1$. In this context, a Random Hyperbolic Graph (RHG) model in the *2*-dimensional hyperbolic space $\mathbb{H}^2$ on N nodes is defined as follows [39]: each of the N nodes is associated with a point **x** in the *two*-dimensional hyperbolic space. The polar coordinates of a point are denoted as (r_i, θ_i), where the coordinates are drawn uniformly at random in $\mathbb{H}^2$ as follows:

$$r_i \leftarrow \rho(r), r_i \in [0, R] \quad \theta_i \leftarrow \mathcal{U}(0, 2\pi), \tag{2.23}$$

where the probability density function $\rho(r)$ of r_i is given by Eq.(2.24) prescribed by the model. $R > 0$ is the radius which depends on the parameters: $< k >, N, T$andγ of the Poincar´e disk.

$$\rho(r) \sim \alpha e^{\alpha(r-R)} \tag{2.24}$$

where $\alpha = \frac{(\gamma-1)}{2}$, γ here is the power-law exponent for the degree distribution of the model. The links between pairs of nodes in the RHG model are generated independently, with their connection probabilities p_{ij} following Eq.(2.25). The connection probability p_{ij} is determined by a distribution in Fermi-Dirac shape.

$$p_{ij} = \frac{1}{1 + e^{\frac{(d_{ij}-R)}{2T}}} \tag{2.25}$$

where $T > 0$ is the temperature and d_{ij} is the hyperbolic distance. Due to Krioukov et al.[39] used the *native* representation of the hyperbolic space, which means all distance variables are their true hyperbolic values. Therefore, the radial coordinate $r \in [0, R]$ of a point is equal to its hyperbolic distance from the origin. And the hyperbolic distance d_{ij} between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ at polar coordinates (r_i, θ_i) and (r_j, θ_j) is

$$\cosh d_{ij} = \cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta\theta \quad \theta \in [0, 2\pi] \tag{2.26}$$

where $\Delta\theta = \pi - |\pi - |\theta_i - \theta_j||$, which is the angel between the points. Until now, the *2*-dimensional RHG model is completely specified by its latent space $\mathbb{H}^2$, the distribution of the coordinates of points, and the distance-based connection probability distribution. And, the RHG model in $\mathbb{H}^2$ is constructed in the following steps:

- Randomly select the coordinates of N points in $\mathbb{H}^2$
- Calculate the hyperbolic distances of all node pairs (i, j) by Eq.(2.26) in $\mathbb{H}^2$
- Randomly connect node pairs (i, j) independently with the connection probability p_{ij} related to the hyperbolic distance.

Overall, the RHG model is defined by these parameters: the properties of the disk, R which depends on the parameters: $\langle k \rangle$, N, Tandγ; the temperature $T \in (0, 1)$ and the radial component of the node distribution $\alpha = \frac{\gamma-1}{2}$, where γ typically follows the power-law exponent for the degree distribution of the real-world networks, as $\gamma > 2$. Meanwhile, the pseudocode algorithm of the RHG model is provided in Appendix A.

Small-World Property: The hubs and randomness contribute to the Small-World property of the RHG model. Nodes near the center of the Poincaré disk have an advantage, as almost all diameters ρ pass through these nodes. This results in the presence of many hubs in the RHG model. Additionally, the randomness arises from the statistical nature of the point distribution.

Scale-Free Property: The Scale-Free property of Random Hyperbolic Graphs (RHG) arises from the negative curvature of the underlying hyperbolic space [39]. The degree distribution of the RHG model is:

$$Pr[D = k] \sim k^{-\gamma} \tag{2.27}$$

where $\gamma = 2\alpha + 1$ when $\zeta = 1$.

Clustering: The strong clustering properties of the RHG model are a direct consequence of the hyperbolic geometry metric structure [39]. In hyperbolic space, the sum of the angles of a triangle is always less than π radians; thus, they become ”thinner. Figure. 2.8 illustrates points A and C are more likely to be connected due to the triangle inequality in hyperbolic geometry.

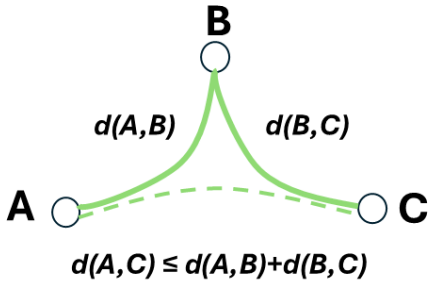

Figure 2.8: Points A and C will be more likely to be connected due to the triangle inequality. $d(A(B, C), C(A, B))$ indicates the hyperbolic distance between arbitrary two points

here, $d(A, C) \leq d(A, B) + d(B, C)$ indicates that points A and C are more likely to form a triangle with point B, which contributes to clustering. Moreover, the clustering coefficient in the model is strongly influenced by the temperature T. As $T \to 0$, the clustering coefficient reaches its maximum.

The RHG models reproduce **Strong Clustering**, **Scale-Free property** and **Small-World property**, which are important structural properties of real networks [39–42]. The RHG model owes these topological properties primarily to its intrinsic hyperbolic geometric nature. Compared to other classical synthetic models, the RHG model is a synthetic model that encompasses the most extensive range of topological properties so far.

3

Motivation: Similarity vs Complementarity in networks

In this chapter, we motivate our research on complementarity mechanisms by comparing similarity and complementarity mechanisms in network science. Firstly, we briefly introduce the concept that similarity is recognized as the leading principle in network science, positing that similar nodes are more likely to be connected. However, we observe that in certain networks, such as protein-protein interaction networks and production networks, connections are established based on the principle of complementarity. Secondly, we dive into the complementarity mechanism in networks by discussing the complementarity principle and introducing the minimal complementarity model, known as the Complementarity Random Hyperbolic Graph (CRHG) model [2]. Finally, we emphasize that the main objective of this thesis is to systematically document the topological properties of the CRHG model.

3.1. Similarity

In simple terms, the similarity is defined as follows: Similar objects i and j, have a high chance of connecting [43]. This concept has its roots in homophily [37, 44]. In sociology, homophily describes the tendency of an individual to socialize and bond with similar people [43]. Homophily is commonly found in many networks, where nodes tend to connect based on similarity [43, 45], and this phenomenon highlights similarity as the main principle in network science. Similarity is applied in various contexts, such as clustering [46, 47], node classification [48], and link prediction [49]. For instance, one common approach is to use common neighbours to predict missing links [49, 50], see Fig. 3.1. Nodes that share many common neighbours are considered similar and, consequently, are more likely to be connected.

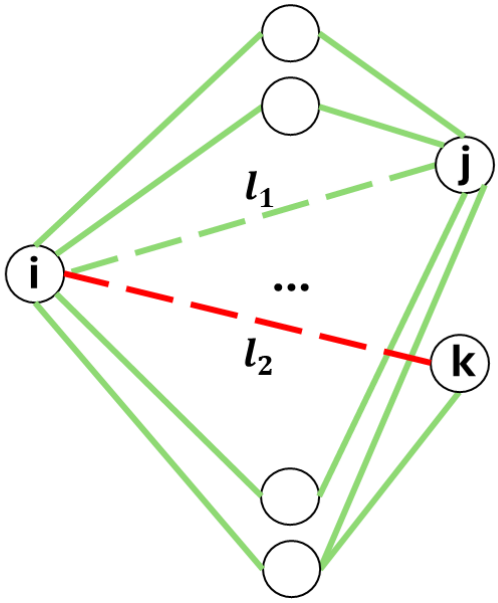

Figure 3.1: Nodes i and j share four common nodes, nodes i and k only share one common node. Therefore, the missing link l_1 (green dotted line) is more probable to exist than the missing link l_2 (red dotted line).

One property of similarity is transitity: If a node i is similar to node j and node j is similar to node k, then node i is similar to node k, see Fig. 3.2. In reality, it is a phenomenon that two people with a lot of common friends are always more likely to be friends [43], explaining the use of common neighbours in link prediction. Therefore, the transitivity of similarity implies the triangular closure mechanism [51], leading to a high density of triangles in social networks, where homophily is one of the key mechanisms. However, not all connectivity relationships between nodes can be explained by similarity; other connection mechanisms also exist.

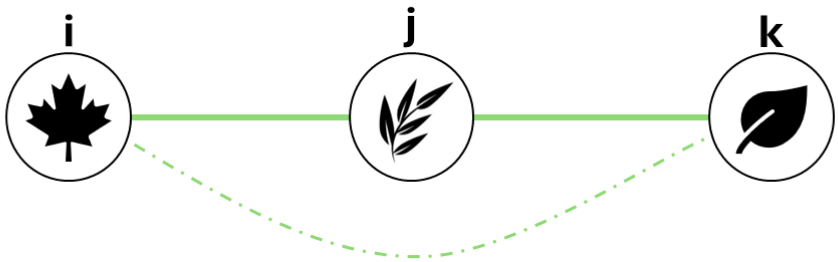

Figure 3.2: Similarity-based toy network with triangle closure mechanism. Nodes represent different kinds of leaves, they are similar since they are leaves. Node i is similar to node j and node j is similar to node k, then node i is similar to node k.

3.2. Complementarity

We now realize that in some networks, including the protein-protein interactions networks [52, 53] and production networks [54], connections are established following the principle of complementarity. For example, trading partners are often complementary in a firm production network. One firm may produce the right product to meet the needs of another firm. In network science, complementarity implies that two nodes provide features or attributes that are lacking in each other. Here, the features or attributes refer to functions, shapes, skills, and other related aspects. Numerous studies based on this concept have emerged since 2018 [1]. Unfortunately, theoretical foundations of complementarity are still in their infancy. A seminal work by G. Budel and M. Kitsak defined the complementary in complex networks in [2]. One of the notable outcomes is that they proposed a synthetic complementarity-based model known as the Complementarity Random Hyperbolic Graph (CRHG) model, which will be described in general terms here and rigorously defined in Chapter 4. CRHG model was conjectured to explain the topological properties of real complementarity-driven networks. The main goal of this thesis is to verify this hypothesis by documenting the topological properties of the CRHG model.

3.2.1. Complementarity principle

In simple terms, the complementary principle dictates that nodes with complementary properties are more likely to connect; see Fig. 3.3 [2]. Figure. 3.3 illustrates a complementarity-based toy network, consisting of four nodes. Each node's features are represented by different shapes. The connections between nodes i and j, j and k, as well as k and e, arise from the complementarity between their features. However, this does not imply that the nodes i and k are complementary and thus connected: Unlike similarity, complementarity is not transitive [2].

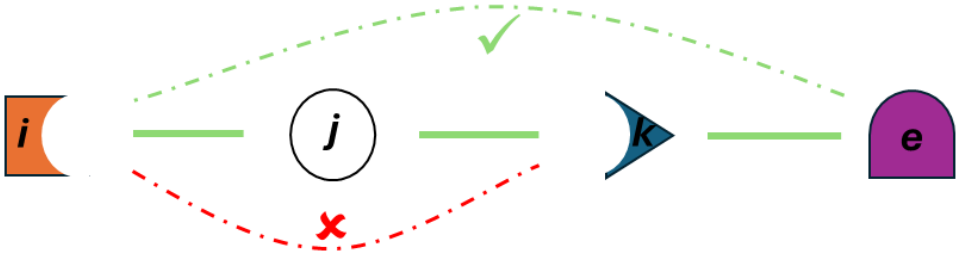

Figure 3.3: Complementarity-based toy network (node i, node j, node k, node e) and their latent feature connection patterns [2]: According to the complementarity of shapes, only shapes that match can be merged. Green solid lines indicate existing connections, while red and green dotted lines represent incompatible and potential connections, respectively. The colours correspond to different nodes. Triangles are not common but quadrilaterals are common.

From a structural perspective, Fig. 3.4 illustrates that if nodes i and k have common neighbors

and k is connected to e while i is not, then l_1 is a missing link. As a result, the triangle closure does not work in the complementary network. Instead, a quadrangle closure might work: If i and k share common neighbors and in addition k is connected to e but i is not, the l_1 could be a missing link. leading to the quadrangle closure rule.

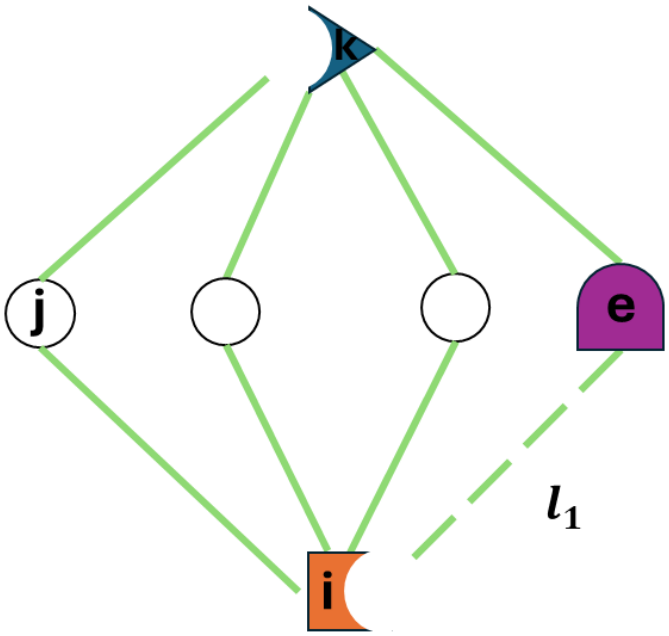

Figure 3.4: Nodes i and k share four common nodes and k is connected to e while i is not. Due to the complementarity between i and e, the missing link l_1 (green dotted line) is likely to exist.

3.2.2. Minimal complementarity model

G. Budel and M. Kitsak proposed a Complement Random Hyperbolic Graph model based on the complementarity principle [2].

The key assumption of the complementarity framework is that nodes in a network are represented not by one but several points in a latent space, see Fig. 3.5. These points can be regarded as distinct features or characteristics of an object. The complementarity of two nodes can then be quantified by measuring distances between points of different types. Figure. 3.5 is a simple geometric representation of complementarity between arbitrary two nodes. Node $i(j)$ is represented into points $i_1(j_1)$ and $i_2(j_2)$ in the latent space $\mathbb{M}$. These feature points correspond to the node's capabilities or attributes in two different domains by different colours. The complementarity of two nodes is defined on distances $d_{ij}^{12} \equiv d\,(i_1, j_2)$ and $d_{ij}^{21} \equiv d\,(i_2, j_1)$ between the complementarity feature points. The smaller one of these two distances, the higher the complementarity.

Different from similarity, it is sufficient for two nodes to complement each other either due to small d_{ij}^{12} or d_{ij}^{21}. For example, in a scientific collaboration network, the representation in Fig. 3.5 can describe the collaborative relationship between two researchers: the theoretical ability (i_1) of researcher i needs to be complementary to those of researcher j who has the

experimental ability (j_2), and vice versa.

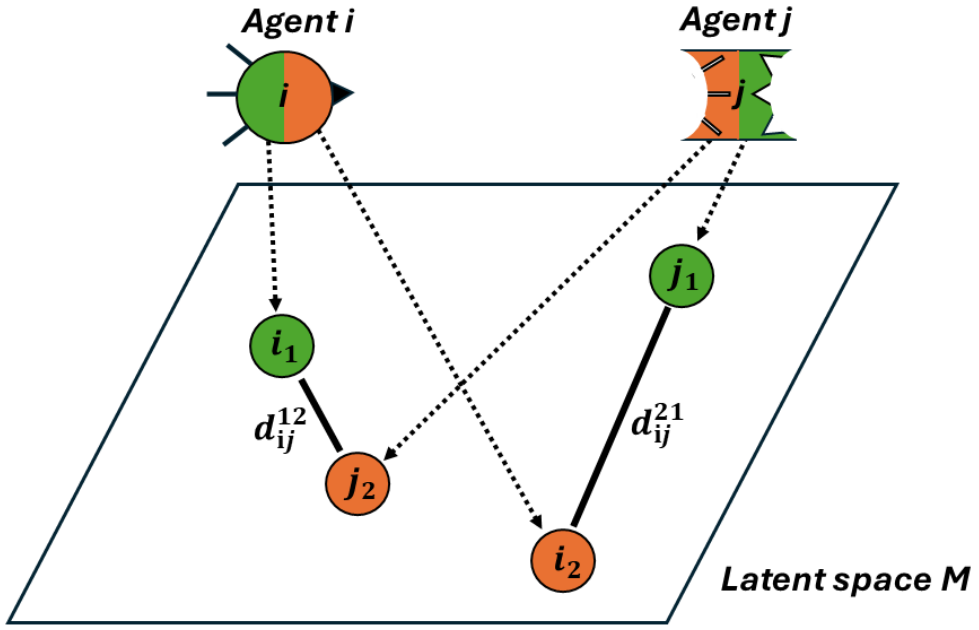

Figure 3.5: Simple geometric representation of complementarity between arbitrary two nodes in a network: When considering whether agent i (node i) is connected with agent j (node j). Each node($i(j)$) is decomposed into two feature points($i_1(j_1)$ and $i_2(j_2)$) according to its characteristic in two different domains by different colours. d_{ij}^{12} denotes the distance between feature point i_1 of node i and feature point j_2 of node j. This distance is small, reflecting a higher complementarity (i.e., the shapes fit perfectly). In contrast, the d_{ij}^{21} measures the distance between feature point i_2 of node i and feature point j_1 of node j; it is large, indicating lower complementarity (i.e., only a partial shape fit). The solid lines in the latent space $\mathbb{M}$ indicate cross-domain point relationships which are key to complementarity-driven networks.

3.3. Motivation

This minimal complementarity model has been called the Complementarity Random Hyperbolic Graph (CRHG) model. This model is using hyperbolic space as a latent space for two points per node. In general, any latent space $\mathbb{M}$ is feasible. However, hyperbolic space is specifically chosen because the CRHG model is inspired by the success of the RHG model which is constructed in the latent space $\mathbb{H}^2$ [2]. The CRHG model serves as a foundation for studying complementarity-driven mechanisms in networks. Unfortunately, the topological properties of the CRHG model have not been systematically documented. **Addressing this knowledge gap, is, therefore, the main objective of this thesis.**

The rest of this thesis is organized as follows. In Chapter 4, we will first define the complementarity random hyperbolic graph (CRHG) model. Then, investigate and discuss the Small-World property, Scale-Free property, Clustering Coefficient and Bipartite Clustering of the CRHG model in Sections 4.1 to 4.5.

In Chapter 5, an extended model of the CRHG model, the generalized complementarity random hyperbolic graph (GCRHG) model, will be introduced, and its properties will be analysed. I summarize findings and draw conclusions in Chapter 6.

4

Complementarity Random Hyperbolic Graph model and its properties

In this chapter, we dive into a specific synthetic complementarity-based model known as the Complementarity Random Hyperbolic Graph (CRHG) model. Firstly, we define it rigorously and explain the function of the key input parameters. In more detail, we analyze the effect of the input parameters within different ranges. Secondly, we systematically document the results of the measurement and the interpretation of the topological properties in the CRHG model. Finally, we summarize the place of the CRHG among other classic synthetic models.

4.1. Complementarity Random Hyperbolic Graph (CRHG) model

CRHG model

Following the minimal complementarity framework, G.Budel and M.Kitsak [2] propose the Complementarity Random Hyperbolic Graph (CRHG) model when the latent space is a hyperbolic disk $\mathbb{H}^2$, and the connection probability function follows a Fermi-Dirac distribution. It is derived and inspired by the RHG model which we described in Section 2.4.4. The Complementarity Random Hyperbolic Graph (CRHG) model in the *2*-dimensional hyperbolic space $\mathbb{H}^2$, with curvature $K = -1$ on N nodes is defined as follows [2]:

- Each of the N nodes is associated with two points in the *2*-dimensional hyperbolic space. Thus, each node i is characterized by points $\mathbf{x}_i \equiv \{r_i^1, \theta_i^1\}$ and $\mathbf{y}_i \equiv \{r_i^2, \theta_i^2\}$. The coordinates of the points are drawn independently at random in $\mathbb{H}^2$ as follows:

$$r_i^{12} \leftarrow \rho(r), r_i^{12} \in [0, R]; \quad \theta_i^{12} \leftarrow \mathcal{U}(0, 2\pi) \tag{4.1}$$

  where $\mathcal{U}$ is the uniform probability density function and the probability density function $\rho(r)$ of r_i^{12} is

$$\rho(r) = \alpha \frac{\sinh \alpha r}{\cosh \alpha R - 1} \tag{4.2}$$

  prescribed by the model. Here, $R > 0$ is the radius of the hyperbolic disk and $\alpha \in (0.5, 1)$ is the parameter controlling the density of the nodes in $\mathbb{H}^2$.

- For each node pair i and j, one computes two distances $d_{ij}^{12} = d_{\mathbb{H}^2}(\mathbf{x}_i, \mathbf{y}_j)$ and $d_{ij}^{21} = d_{\mathbb{H}^2}(\mathbf{x}_j, \mathbf{y}_i)$, where $d_{\mathbb{H}^2}(\mathbf{x}, \mathbf{y})$ is given by the hyperbolic law of cosines:

$$\cosh d_{\mathbb{H}^2}(\mathbf{x}, \mathbf{y}) = \cosh r^1 \cosh r^2 - \sinh r^1 \sinh r^2 \cos \Delta\theta^{12} \quad \theta \in [0, 2\pi] \tag{4.3}$$

  where $\Delta\theta^{12} = \pi - |\pi - |\theta^1 - \theta^2||$, which is the angle between the points.

- Each node pair ij is connected independently with probability:

$$p_{ij} = p\left(d_{ij}^{12}\right) + p\left(d_{ij}^{21}\right) - p\left(d_{ij}^{12}\right) * p\left(d_{ij}^{21}\right) \tag{4.4}$$

  where connection probability $p\,(d)$ has a Fermi-Dirac shape

$$p\,(d) = \frac{1}{1 + e^{\frac{(d-R)}{2T}}}, \tag{4.5}$$

  where $R > 0$ and $T \in [0, 1]$ are model parameters, as shown in Fig. 4.1, it is important to note that p_{ij} is nothing else but the union of two probabilities $p(d_{ij}^{12})$ and $p(d_{ij}^{21})$: In other word, either $\mathbf{x}_i$ complements $\mathbf{y}_j$ or $\mathbf{x}_j$ complements $\mathbf{y}_i$ or both.

The CRHG model has three parameters R, α and T. R is the radius of the hyperbolic disk. For a fixed number of nodes N, R is expected to control the node density, average degree of the resulting network.

Indeed, Budel and Kitsak establish that

$$\langle k \rangle = NI\frac{4}{\pi}\left(\frac{\gamma - 1}{\gamma - 2}\right)^2 e^{-R/2} \tag{4.6}$$

where

$$I = \left(\frac{\sin \pi T}{\pi T}\right)^{-1} \tag{4.7}$$

when $T \in (0, 1)$. Therefore, R depends on the parameters: $\langle k \rangle, N, T$ and γ. Parameter α controls the density of nodes in $\mathbb{H}^2$. If $\alpha = 1$, nodes distributed uniformly over $\mathbb{H}^2$. This is the case since the Poincar'e disk area is [39]

$$S = 2\pi\,(\cosh R - 1) \sim e^R \tag{4.8}$$

If $\alpha < 1$, more nodes are concentrated closer to $\mathbb{H}^2$ center.

Budel and Kitsak conjecture that the CRHG model is Scale-Free:

$$Pr[D = k] \sim k^{-\gamma} \tag{4.9}$$

and $\gamma = 2\alpha + 1$. We examine this relationship through simulations in Section 4.2. Temperature parameter T affects the range of connections. As $T \to 0$, the range of connections is short and $p\,(d)$ reduces to a step function: $p\,(d) \sim \Theta\,(d - R)$, see Fig. 4.1. According to Eq.(4.4)

and Eq.(4.5), for $p(d_{ij}^{12})$ (similar to $p(d_{ij}^{21})$), the term $e^{\frac{(d_{ij}^{12}-R)}{2T}}$ exhibits significant variations as $T \to 0$:

(1) When $(d_{ij}^{12} - R) < 0$, $p(d_{ij}^{12}) \approx 1$ leading to $p_{ij} \approx 1$.

(2) When $(d_{ij}^{12} - R) > 0$, $p(d_{ij}^{12}) \approx 0$, resulting in $p_{ij} \approx 0$.

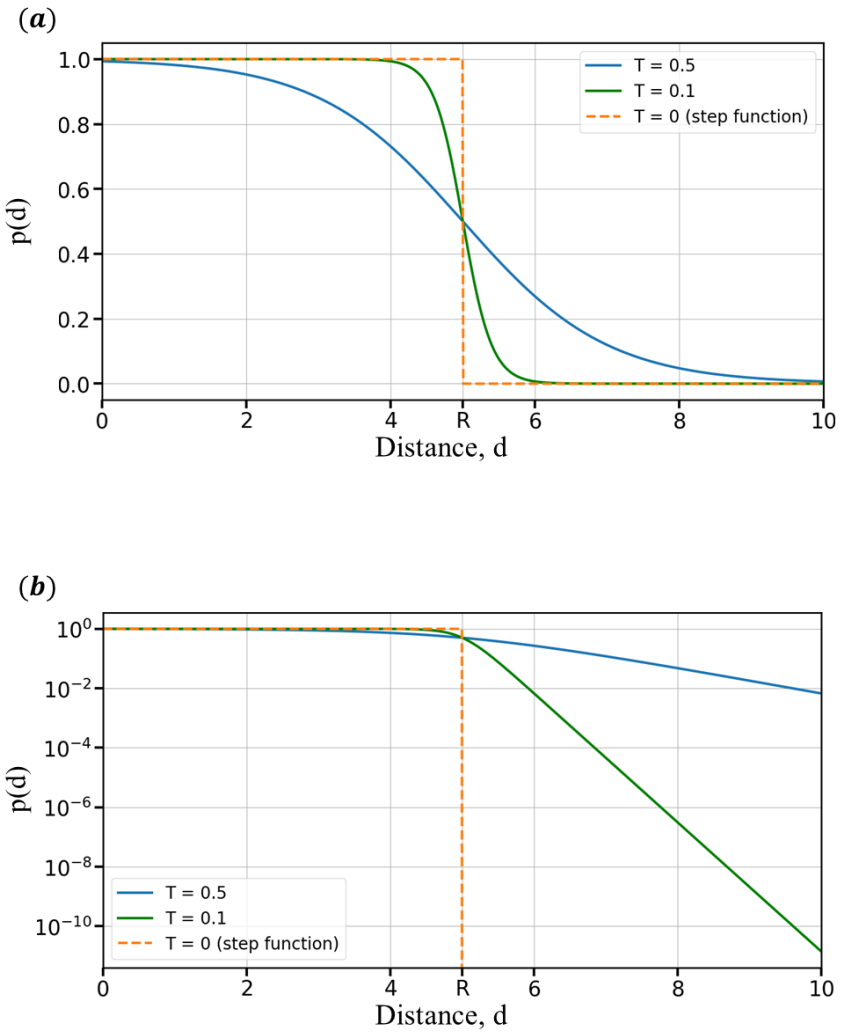

Figure 4.1: The relationship between distance d and $p\,(d)$ varies for different values of T. For instance, when $d \in [0, 10]$, $R = 5$, $T = \{0, 0.1, 0.5\}$. As $T \to 0$, the range of connections is short, and $p\,(d)$ approximates a step function, $p\,(d) \sim \Theta\,(d - R)$, as shown by the orange dotted line. In contrast, as $T \to 1$, the range of connections is long, and $p\,(d)$ varies more gradually, as illustrated by the green and blue lines. The panel (b) is the log-linear form.

In contrast, when $T \to 1$, the range of connections is long. Taking $p(d_{ij}^{12})$ (similar to $p(d_{ij}^{21})$) as an example, we observe that the term $e^{\frac{(d_{ij}^{12}-R)}{2T}}$ varies more gradually when $T > 0$, shown as the blue and green lines in Fig. 4.1.

In summary, the CRHG model with the target number of nodes N, average degree $\langle k \rangle$, degree distribution exponent γ and temperature T in $\mathbb{H}^2$ is constructed in the following steps:

- Randomly select the coordinates of $2N$ points in $\mathbb{H}^2$
- Calculate the hyperbolic distances of all point pairs $(\mathbf{x}, \mathbf{y})$ in $\mathbb{H}^2$
- Randomly connect node pairs (i, j) independently with the connection probability p_{ij} given by Eq.(4.4) related to the point-based connection probability.

When compared to the RHG, the CRHG model is defined by the same parameters as RHG. The main difference is that when we consider the connection probability of node pairs (i, j) being connected, two points are used to represent a node and decide whether the node pairs (i, j) are connected or not based on the connection probability of a two-by-two connection between these four points. The pseudocode algorithm of the CRHG model is shown in Appendix A.

4.2. Degree distribution of the CRHG model

In this section, with study degree distribution of the CRHG model. In [55], the authors conjectured that the degree distribution follows a power-law distribution in the CRHG model. For the CRHG model, one hypothesis is that the approximation of the degree distribution density follows a power law with exponent γ. To verify this hypothesis, we investigated the degree distribution of the CRHG model under different parameters. These parameters are the average degree $\langle k \rangle$, the temperature T and node density parameters α. The graph size of the CRHG model is $N = 1.024 * 10^5$. Figures. 4.2 and 4.3 illustrate that the CRHG model is characterized by a scale-free distribution $Pr[D = k] \sim k^{-\gamma}$ with $\gamma \in (2, 3)$. Further, the exponent γ seems to agree with the value $\gamma = 2\alpha + 1$, where α is the node density parameter, as conjectured in [55]. Table. 4.1 lists the parameters we used in the experiment where $\langle k \rangle = 5$ and $\langle k \rangle = 20$.

Average Degree $\langle k \rangle$	**Node density** α	**Theory** $\gamma = 2\alpha + 1$	**Temperature** T	$\hat{\gamma}$	$\lvert\hat{\gamma} - \gamma\rvert$
5	0.55	2.1	0.1	2.11	0.01
5	0.55	2.1	0.5	2.22	0.12
20	0.55	2.1	0.1	2.21	0.11
20	0.55	2.1	0.5	2.22	0.12
5	0.75	2.5	0.1	2.57	0.07
5	0.75	2.5	0.5	2.60	0.10
20	0.75	2.5	0.1	2.60	0.10
20	0.75	2.5	0.5	2.51	0.01
5	1.25	3.5	0.1	3.61	0.11
5	1.25	3.5	0.5	4.02	0.52
20	1.25	3.5	0.1	3.65	0.15
20	1.25	3.5	0.5	3.70	0.20

Table 4.1: Various combinations of average degree $\langle k \rangle = \{5, 20\}$, node density α (γ) and temperature T, with estimated $\hat{\gamma}$ and $|\hat{\gamma} - \gamma|$

Discussion and Observation

(1) The simulation results demonstrate that the CRHG model can be adapted to model scale-free networks. Notably, as the γ values increase while keeping other parameters constant, the Scale-Free property is preserved until a significant discrepancy emerges between the theoretical γ and the estimated $\hat{\gamma}$ values at $\gamma = 3.5$.

(2) γ is the same as in the RHG model, $\gamma = 2\alpha + 1$. The reason is that the CRHG is based on the RHG.

(3) The deviation of γ from the theoretical values is greater for larger α values.The reason is that large γ values lead to fewer hubs, making it harder to measure the power-law distribution. The maximum error reaches 0.52 under the conditions of $\langle k \rangle = 5$, $\gamma = 3.5$ and $T = 0.5$. This substantial gap indicates that the simulated values do not fit well with the theoretical values or the data are not sufficient, since large γ means fewer large k nodes reducing the range of k values for measuring γ. In other words, as shown in Fig. 4.3 (c), the approximated range of node degree $k \in [1, 10^4]$ for $\gamma = 2.1$. Figure. 4.3 (e) reflects that the range of node degree $k \in [1, 10^3]$ for $\gamma = 3.5$. To verify the scale-free property of the CRHG model with $\gamma = 3.5$, one assumption is that increasing the number of nodes beyond $N = 1.024 \times 10^5$ could improve the linear regression fit of the degree distribution, thereby better aligning the estimated $\hat{\gamma}$ value with its theoretical value.

Additionally, there is no significant relationship between the degree distribution and the temperature T, suggesting that changes in T have minimal impact on the degree distribution of the model, confirming the conjecture of Budel et al. in [55].

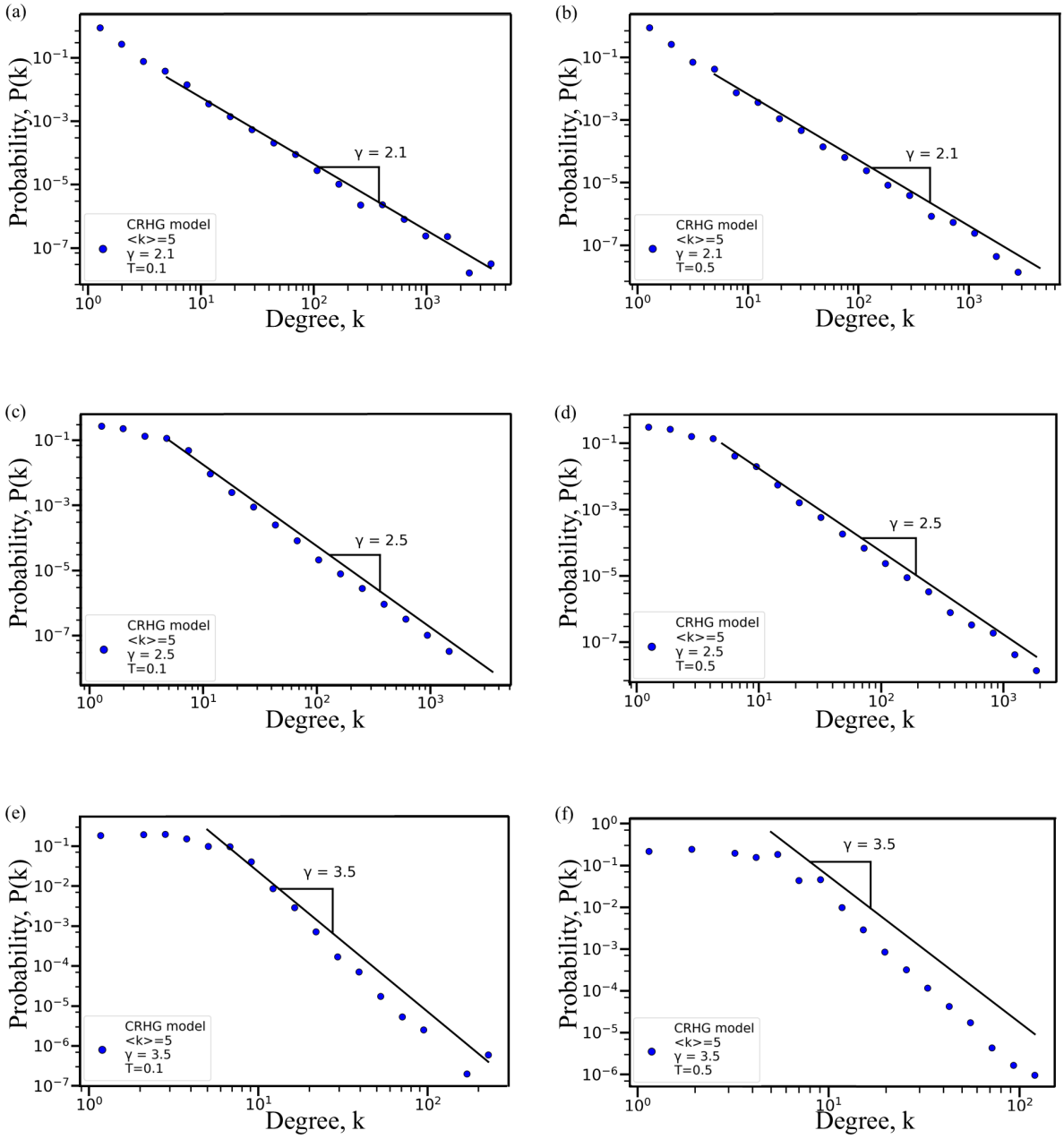

Figure 4.2: Degree distribution $Pr[D = k]$ of the CRHG model for $\langle k \rangle = 5$ and different input parameters: $N = 1.024 * 10^5, \alpha = \{0.55, 0.75, 1.25\}(\gamma = \{2.1, 2.5, 3.5\}), T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters. In each panel, we use the logarithmic binning [21] for the degree distribution with the number of binning 20. The line represents the theoretical conjectures γ value and the theoretical value is displayed as a slope above the straight line.

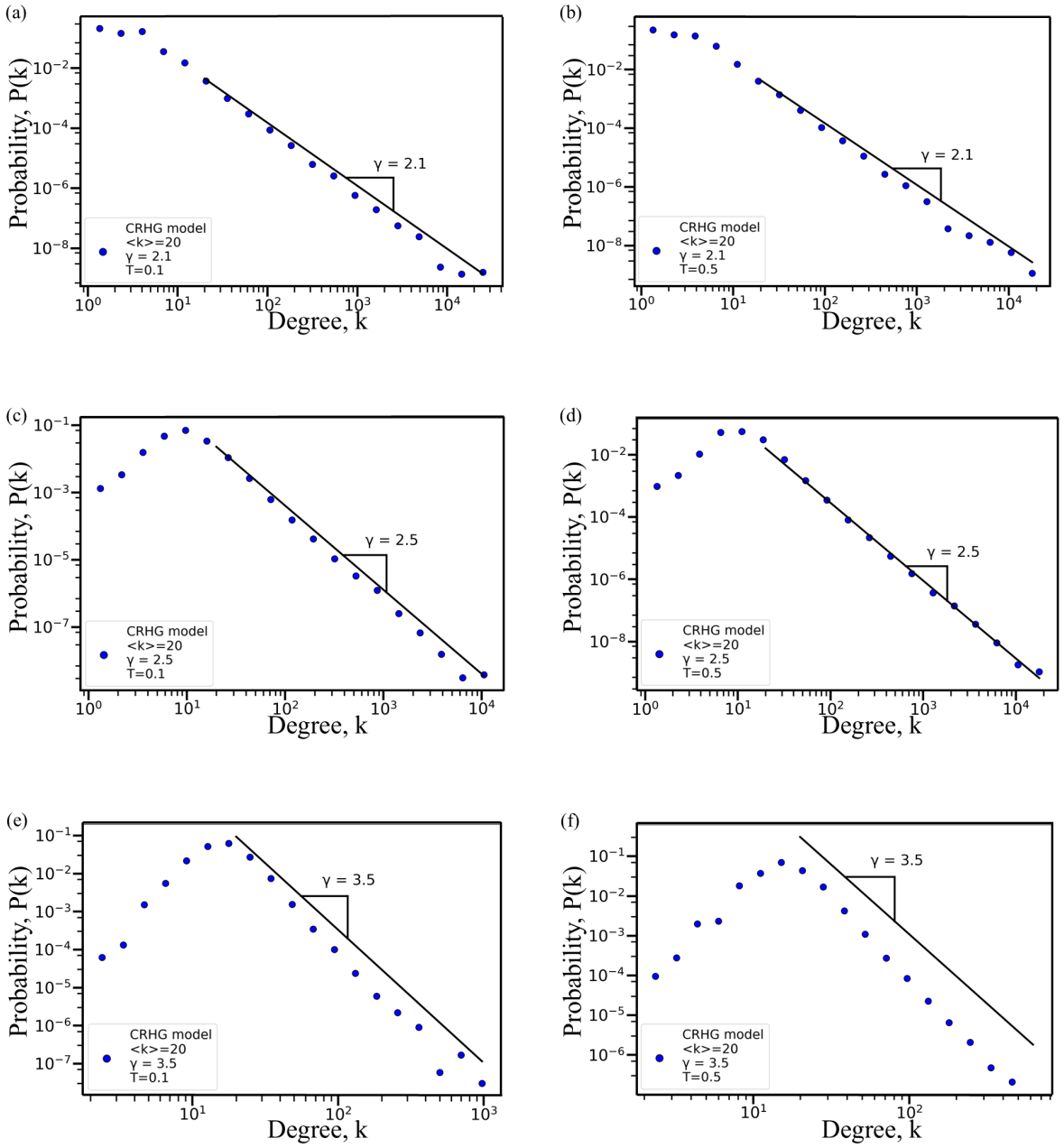

Figure 4.3: Degree distribution $Pr[D = k]$ of the CRHG model for $\langle k \rangle = 20$ and different input parameters: $N = 1.024 * 10^5, \alpha = \{0.55, 0.75, 1.25\}(\gamma = \{2.1, 2.5, 3.5\}), T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters. In each panel, we use the logarithmic binning [21] for the degree distribution with the number of binning 20. The line represents the theoretical conjectures γ value and the theoretical value is displayed as a slope above the straight line.

4.3. Small-World property of the model

The Small-World property is one of the basic properties in the Small-World networks. It is also a common topological property for synthetic models. One prominent characteristic is: The average shortest path length $\langle d \rangle$ typically scales as the logarithm of N [6], [9]:

$$\langle d \rangle \sim \ln N \tag{4.10}$$

Equation. (4.10) represents just one manifestation of the logarithmic scaling behaviour between the average shortest path length $\langle d \rangle$ and the network size N. This logarithmic relationship implies that the Small-World property is characterized not only in the fact that the absolute vale of $\langle d \rangle$ is small compared to the size of the network N, but also in the observation that $\langle d \rangle$ grows very slowly with N.

To investigate the Small-World property of the CRHG model, the most common method is to measure the average shortest path length $\langle d \rangle$ to reflect the characteristic of short paths in small worlds. We analyze multiple graph instances for each parameter combination in the CRHG model. In our simulations, we use the following parameters:

$$N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \quad \langle k \rangle = \{5, 20\}, \quad \gamma = \{2.1, 2.5, 3.5\}, \quad T = \{0.1, 0.5\}.$$

as input to produce the CRHG model graphs. Specifically, for each network size N, we generate 10 different CRHG graphs to reduce measurement errors. In each graph size N, we estimate the average shortest path length $\langle d \rangle$ over 10^4 random node pairs by computing the distance d_{ij} for 10^4 random node pairs. For each generated CRHG graph, we apply the "Degree Cutoffs" method. In this approach, all nodes with degree $k > \sqrt{N}$ are removed from the CRHG graph. We employ this method because we observed that the fluctuations in the average shortest path length $\langle d \rangle$ are significant due to the small number of node pairs used in estimating $\langle d \rangle$ or the k_{max} tends to change significantly in different N size graphs.

Our results are depicted separately in Fig. 4.5 and Fig. 4.6. Specifically, for $\gamma = 2.5$ and $\gamma = 3.5$, we observe logarithmic growth. When $\gamma = 2.1$, it is even sub-logarithmic (slower than logarithmic). For $\langle k \rangle = 20$, we see some "jumps" for small N, which is probably the small size effect. As a result, the CRHG model under various combinations of parameters has essentially Small-World properties.

Discussion and Observation

(1) In Fig. 4.5 and Fig. 4.6, the $\langle d \rangle$ values are small, $\langle d \rangle < 10$ for all model networks, which is consistent with the Small-World property. The Small-World property can arise from two sources: **Randomness**: Random links allow shortest to distant nodes, as seen in models like the ER model, and **Hubs**: Hubs bring a lot of nodes together, as observed in the BA model. Both of these sources are present in CRHG. The more hubs, the smaller the average distance $\langle d \rangle$, Specifically, **Randomness** is controlled by the parameter T; as T increases, there is more randomness in the CRHG model. In Section 4.1, we have analysed that when $T \to 1$, the range of connection is long. Comparing (c) and (d) of Fig. 4.5, as shown in Fig. 4.4 (a), we can see that as T grows, the value of $\langle d \rangle$ in Fig. 4.5 (d) decreases as a whole in comparison with Fig. 4.5 (c) while accompanying the slow growth of N. Meanwhile, **Hubs** are influenced by the parameter γ; a lower γ leads to more hubs in the CRHG model. As shown in Fig. 4.5 (a) and (c), $\langle d \rangle$ grows slower with lower γ in Fig. 4.4 (b).

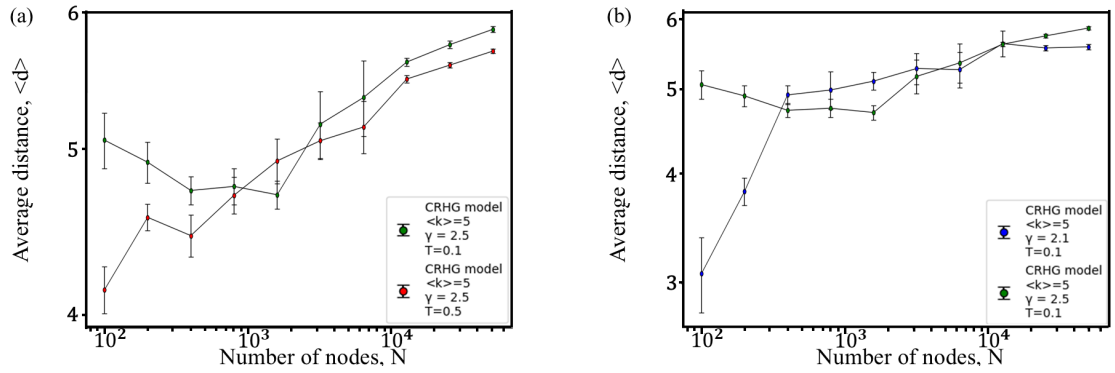

Figure 4.4: Comparison of $\langle d \rangle$ curves for different input T and γ values

(2) $\langle d \rangle$ is growing slower than logarithmically for $\gamma = 2.1$, as shown in Fig. 4.5 (a) (b) and Fig. 4.6 (a)-(d). In particular, when $N > 2^2 \times 10^2$, $\langle d \rangle$ tends to stabilize as N increases. In subplots (c)-(f) of Fig. 4.5 and subplots (e)-(f) of Fig. 4.6 , $\langle d \rangle$ increases slowly with the size of the network N for $\gamma = \{2.5, 3.5\}$ too. This means $\langle d \rangle$ growing slower than logarithmically. A reasonable explanation for the vanished constant value can be attributed to the limited range of N.

(3) From the perspective of the parameters, we find that when $\langle k \rangle$ and γ are held constant, variations in temperature T have little effect on the results. In contrast, when $\langle k \rangle$ and T remain fixed, an increase in γ and N leads to an accelerating growth of $\langle d \rangle$. This may be due to the fact that lower γ values correspond to a higher probability of generating nodes with high k(hubs); as a result, $\langle d \rangle$ tends to stabilize.

In Fig. 4.6, the main difference compared to Fig. 4.5 lies in the application of a higher average degree, $\langle k \rangle = 20$. Here, we observe that the 'Degree Cutoffs' method has a negative effect on the small-size model networks. Specifically, the nodes removed by this method result in disconnected graphs. Therefore, after considering the trade-offs, we decided not to apply the "Degree Cutoffs" method for graphs with $N = [10^2, 2 \times 10^2]$ in Fig. 4.6 to avoid excessively impacting $\langle d \rangle$ in these cases. Consequently, the gap in the $\langle d \rangle$ values between $N = 2 \times 10^2$ and $N = 4 \times 10^2$ in all subplots of Fig. 4.6 maybe caused by this processing or the small size effect. In comparison, as N increases, most subplots in Fig. 4.6 exhibit a nearly constant $\langle d \rangle$, which could be explained by the presence of more nodes with high k (hubs) that increases the likelihood of connecting any two arbitrary nodes.

Overall, the graphs of the CRHG model exhibit the Small-World property, as demonstrated in Fig. 4.5 and Fig. 4.6.

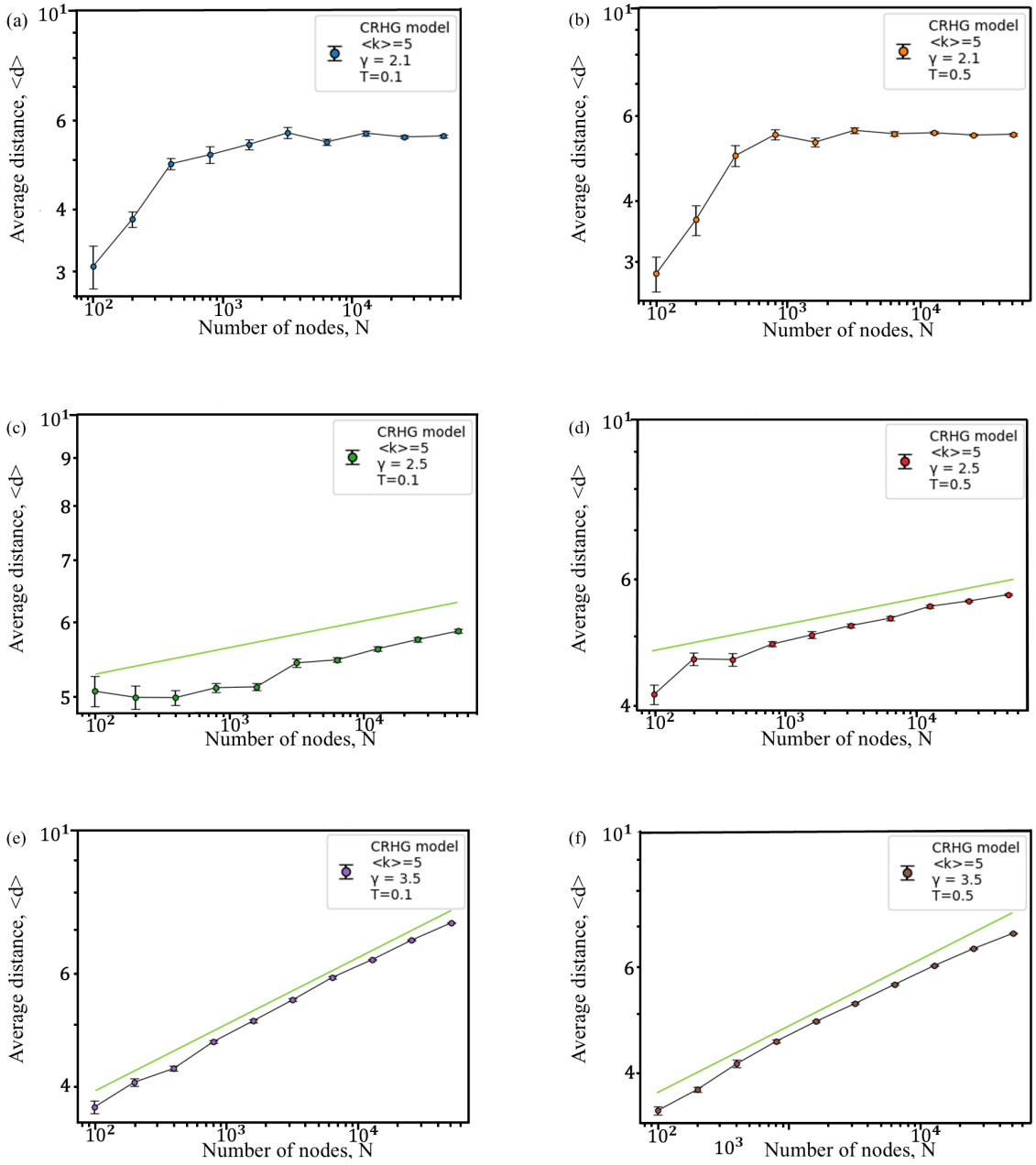

Figure 4.5: The average shortest path length $\langle d \rangle$ of the CRHG model ($\langle k \rangle = 5$) in different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \langle k \rangle = 5, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T.

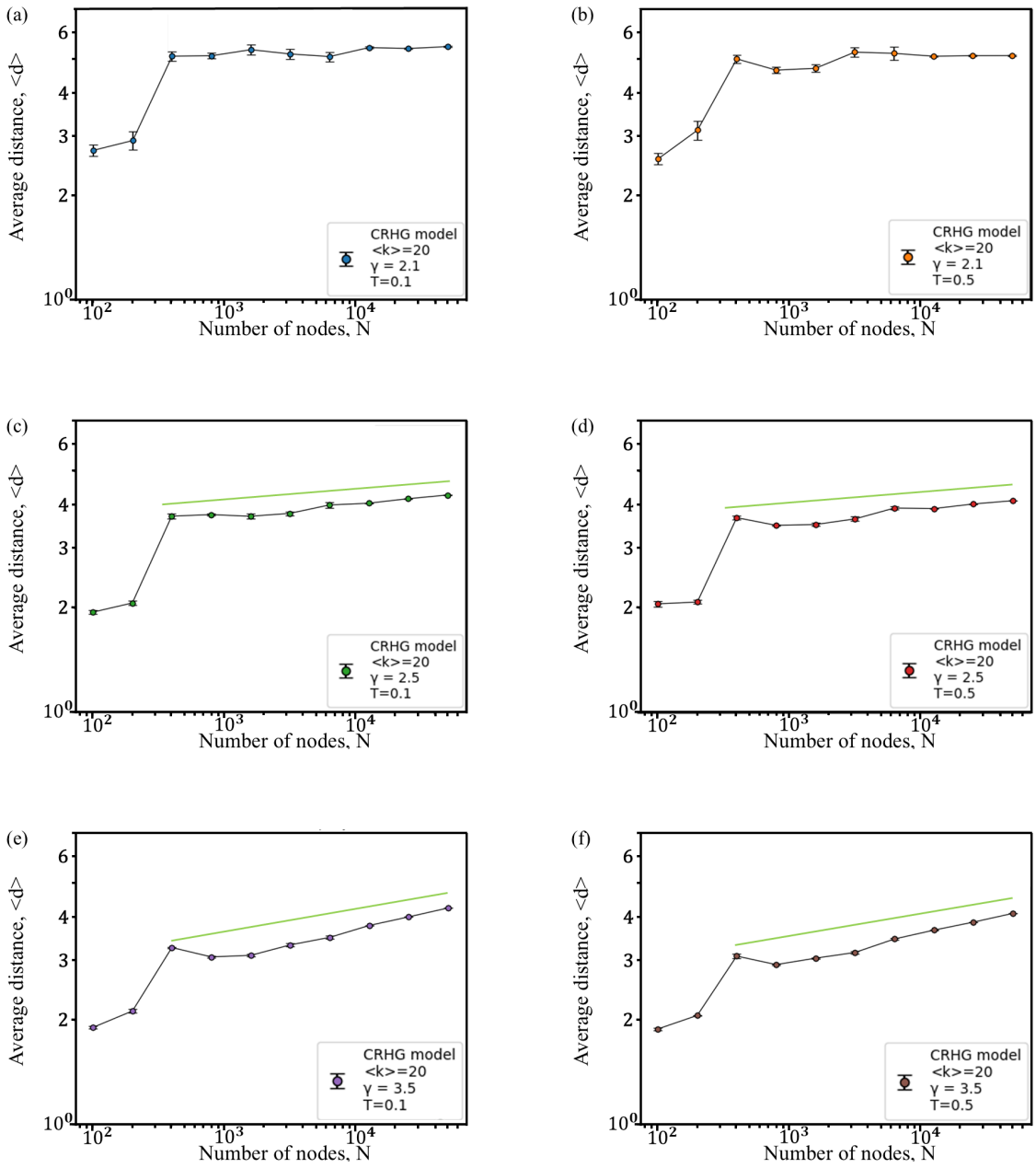

Figure 4.6: The average shortest path length $\langle d \rangle$ of the CRHG model ($\langle k \rangle = 20$) in different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \langle k \rangle = 20, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T.

4.4. Clustering coefficient, C_3

As mentioned in Section 2.2.3 and 2.3.3, the clustering coefficients describe the tendency of nodes in a graph to form closely connected groups which are represented as closed triangles. We have known that the RHG model exhibits strong clustering. Then, what is the clustering in the complementarity-based CRHG model? To investigate clustering in the CRHG model, we measure the clustering coefficient $\bar{C}_3$ as a function of network size N. Meanwhile, for every graph, we also measure the $\bar{C}_3$ of one degree-preserving rewired graph for comparison and reference. In our simulations, we use the following parameters:

$$N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \quad \langle k \rangle = \{5, 20\}, \quad \gamma = \{2.1, 2.5, 3.5\}, \quad T = \{0.1, 0.5\}.$$

as input to produce the CRHG model graphs.

After that, we calculate $\bar{C}_3(k)$ of nodes with k for large N, based on the 10 sample graphs of the CRHG model with $N = 1.024 * 10^5$. As seen in Fig. 4.7 to Fig. 4.10, the CRHG model under various combinations of parameters has vanishing clustering.

Discussion and Observation: $\bar{C}_3$ vs N

(1) Figures 4.7 and 4.8 depict $\bar{C}_3$ as a function of network size N. The values of the clustering coefficients $\bar{C}_3$ are generally very low in the CRHG model graphs, as shown in the figures, the clustering coefficients $\bar{C}_3 \in [10^{-4}, 10^{-1}]$, thus the C_3 is small. To establish this result we compare CRHG to randomize networks, with preserved node degrees. The clustering coefficient $\bar{C}_3$ is comparable to degree-preserving randomizations. Randomized graphs are nothing else but random configurational models and are known for their small clustering. Since C_3 in CRHG is very close to random rewired counterparts, we conclude that C_3 is weak.

(2) With increasing graph size N, the clustering coefficients $\bar{C}_3$ decrease as the variation trend of the clustering coefficients in rewired random graphs. This is due to the triangle closure is not enforced in complementarity. $\bar{C}_3$ is almost the same as that of randomized graphs. All these reflect that the CRHG model lacks strong clustering.

Additionally, the change in temperature T does not affect the tendency of $\bar{C}_3$ to decrease, as shown in the panels (a) and (b) of Fig. 4.7 and 4.8. This is in contrast to the RHG model where temperature T controls clustering. Instead, an increase in the value of the γ causes the decrease in $\bar{C}_3$ to be faster, as shown in Fig. 4.7 (a) and Fig. 4.7 (c).

Discussion and Observation: $\bar{C}_3(k)$ vs k:

We calculate $\bar{C}_3(k)$ of the nodes with degree k based on the sample graphs 10 of the CRHG model that set the same input parameters. When plotting the figures, we applied linear binning [21] with 20 bins for the $\bar{C}_3(k)$ curves to capture the overall variation trend and reduce noise. Figure 4.9 and Figure 4.10 illustrate that the reduction of $\bar{C}_3(k)$ value is significant when the $\langle k \rangle$ is increasing.

(1) In general, the value of $\bar{C}_3(k)$ decreases with increasing k, except in the case of $\gamma = 3.5$. This is common in most networks. Large degree k need $\frac{k(k-1)}{2}$ links to connect all the neighbours, which is very hard to achieve. In that case $\gamma = 3.5$, as shown in Fig. 4.9 (e), (f) and Fig. 4.10 (e), (f), the value of $\bar{C}_3(k)$ is a constant that tends to 0 with small upward and downward

fluctuations. This reflects the clustering is nearly vanishing. In conclusion, $\bar{C}_3(k)$ decreases as a function of $\langle k \rangle$. This is expected and seen in other Scale-Free models because for large k lots of links are required to connect neighbors. The results in Fig. 4.9 and Fig. 4.10 also reflect the vanishing clustering in the CRHG model.

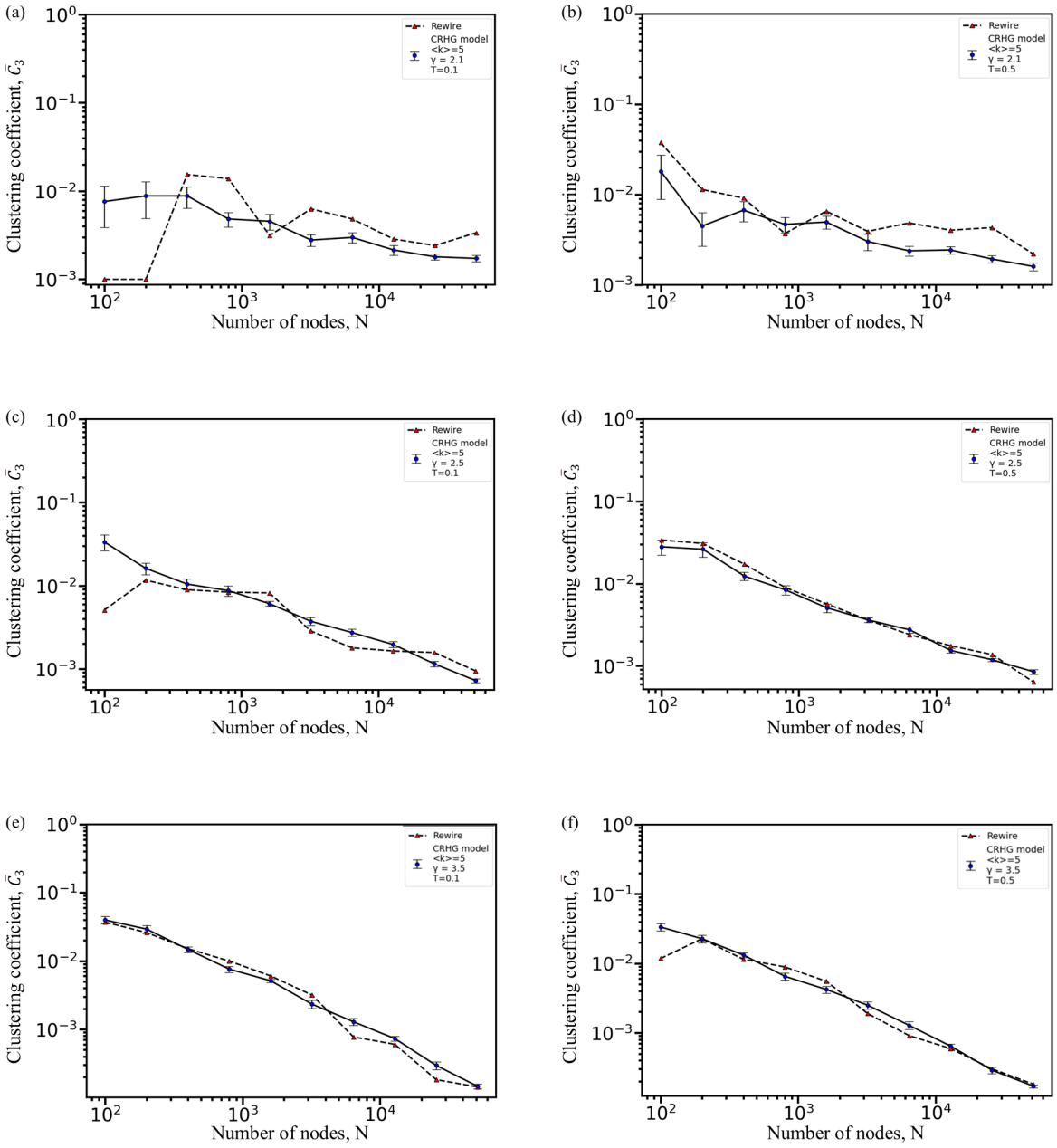

Figure 4.7: The clustering coefficient $\bar{C}_3$ of the CRHG model for $\langle k \rangle = 5$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, each red triangle represents the $\bar{C}_3$ of CRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_3$ with the growth of N in rewired random graphs.

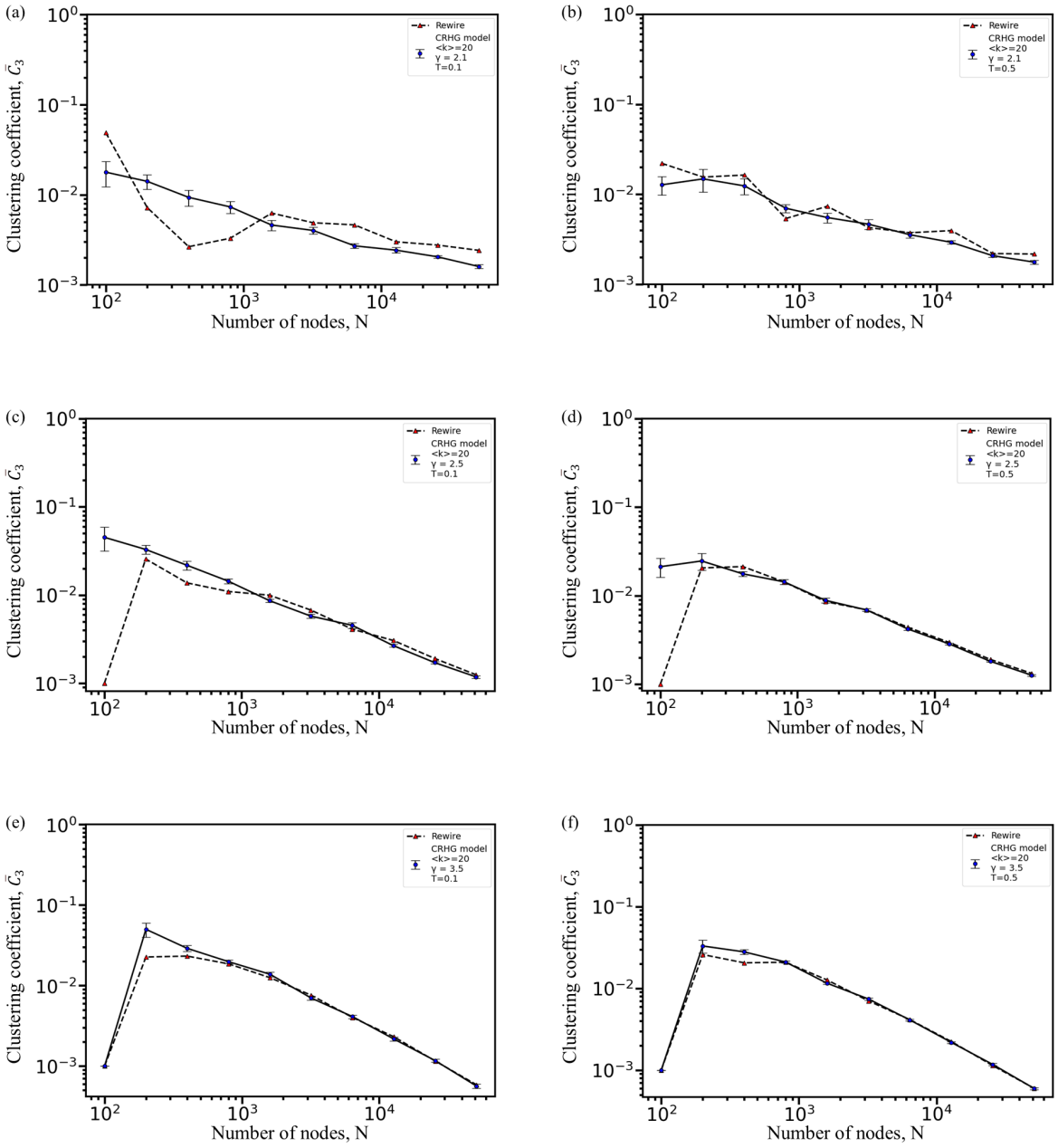

Figure 4.8: The clustering coefficient $\bar{C}_3$ of the CRHG model for $\langle k \rangle = 20$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, each red triangle represents the $\bar{C}_3$ of CRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_3$ with the growth of N in rewired random graphs.

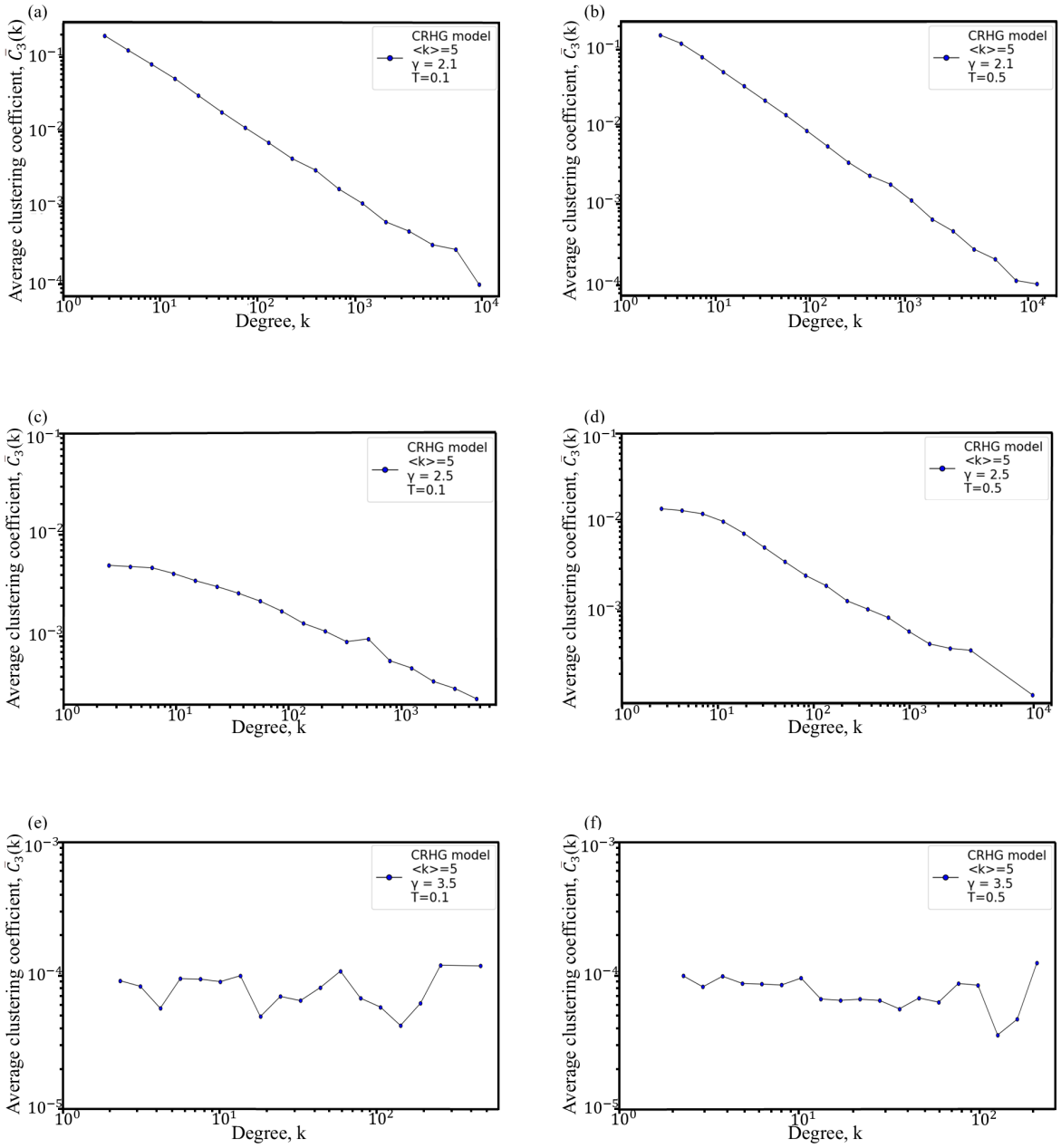

Figure 4.9: The average clustering coefficient $\bar{C}_3$ of the nodes with degree k of the CRHG model for $\langle k \rangle = 5$ and different input parameters: $N = 1.024 * 10^5, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, the linear binning [21] is applied with 20 bins for the $\bar{C}_3(k)$ curves to capture the overall variation trend.

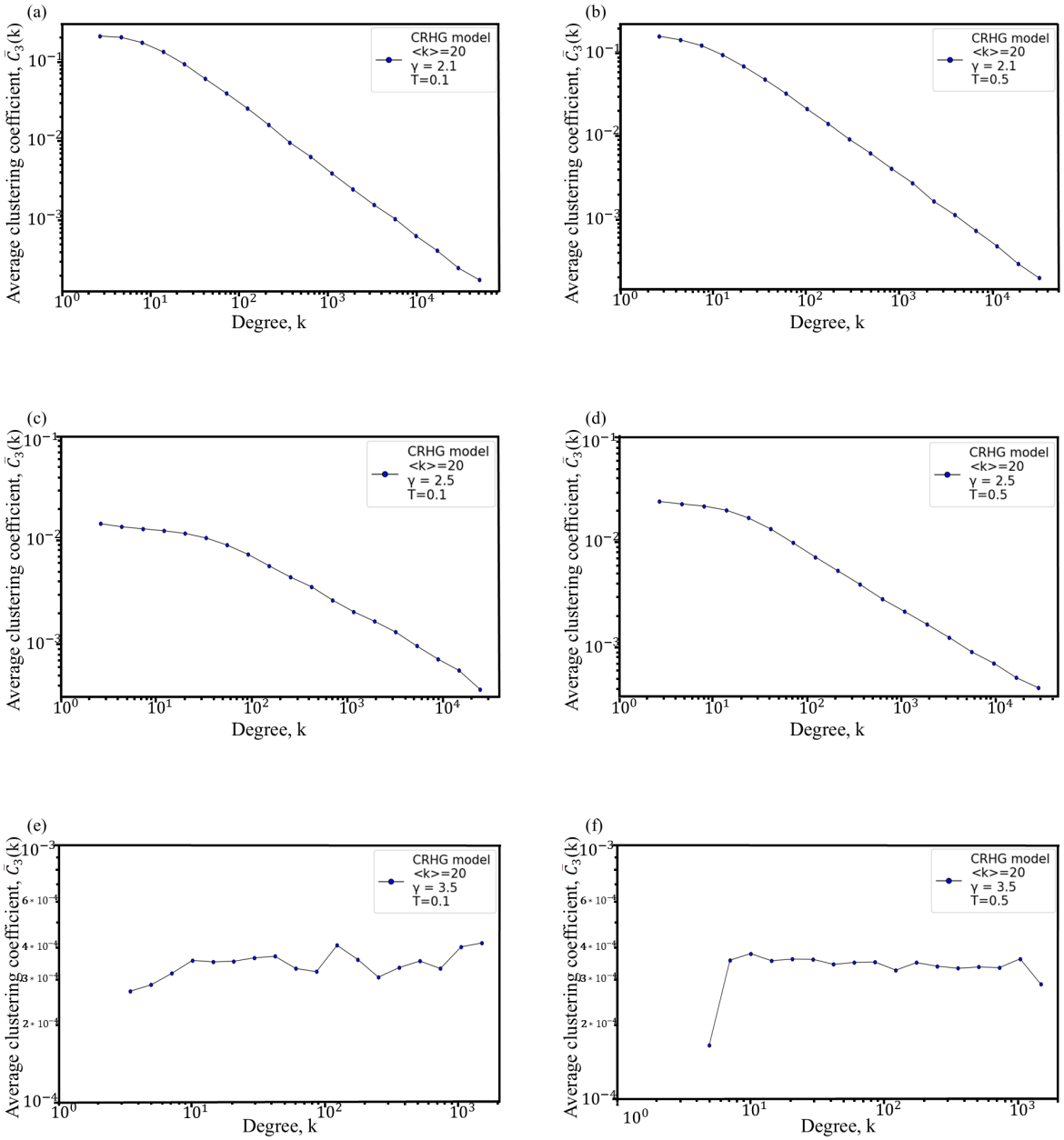

Figure 4.10: The average clustering coefficient $\bar{C}_3$ of the nodes with degree k of the CRHG model for $\langle k \rangle = 20$ and different input parameters: $N = 1.024 * 10^5, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, the linear binning [21] is applied with 20 bins for the $\bar{C}_3(k)$ curves to capture the overall variation trend.

4.5. Bipartite clustering, C_4

Observing the vanishing clustering in the CRHG model raises an important question: What other topological properties of interest might be exhibited in the CRHG model? Starting with the definition and construction of the CRHG model, one of its key features lies in its mechanism of connecting nodes based on complementarity, as we discussed in Section 3.2.1. This naturally leads us to hypothesize there are more rectangular closures in the CRHG model, thus bipartite clustering may play a dominant role in CRHG graphs. To investigate this, we measure the bipartite clustering coefficient $\bar{C}_4$ as a function of network size N. For comparison and reference, we also compute the $\bar{C}_4$ for degree-preserving rewired versions of each graph. These rewired graphs serve as a baseline to help isolate the effects of the complementarity-driven mechanism. In our simulations, the following parameters are used:

$$N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \quad \langle k \rangle = \{5, 20\}, \quad \gamma = \{2.1, 2.5, 3.5\}, \quad T = \{0.1, 0.5\}.$$

as input to produce the CRHG model graphs.

Then, we calculate $\bar{C}_4(k)$ of nodes with k based on the 10 sample graphs of the CRHG model with $N = 1.024 * 10^5$. As shown in Figs. 4.11 to 4.14, the CRHG model under various combinations of parameters has non-vanishing bipartite clustering.

Discussion and Observation: $\bar{C}_4$ vs N:

(1) In Figs. 4.11 and 4.12, the value of $\bar{C}_4$ is small, but at the same time is much larger than that of the randomized counterparts, by an order of magnitude or more.

(2) $\bar{C}_4$ only weakly changes as N increases, almost saturation at a constant value. The reason is that rectangular is expected for complementarity networks. As shown in Figs. 4.11 to 4.12, these figures illustrate that as the size of the graph N increases, the value $\bar{C}_4$ fluctuates around a constant value and does not disappear. This phenomenon is most notable in Fig. 4.11 (e), (f) and Fig. 4.12 (e), (f). $\bar{C}_4$ values of the CRHG model may exceed those of the randomized rewired graphs, by an order of magnitude.

In conclusion, we can find that the CRHG model has excellent retention of bipartite clustering compared to its degree-preserving rewired graph.

Discussion and Observation: $\bar{C}_4(k)$ vs k:

We calculate $\bar{C}_4(k)$ of the nodes with k based on the sample graphs 10 of the CRHG model that set the same input parameters. When plotting the figures, we applied linear binning [21] with 20 bins for the $\bar{C}_4(k)$ curves to capture the overall variation trend.

(1) As shown in Fig. 4.13 and Fig. 4.14, they illustrate that as $\langle k \rangle$ increases, the value of $\bar{C}_4(k)$ decreases slowly as a function of k. As the same explanation as for $\bar{C}_3(k)$, this is common in most networks. Large degrees k need $\frac{k(k-1)}{2}$ links to connect all the neighbours, which is very hard to achieve. In Fig. 4.13 (a), (b) and Fig. 4.14 (a), (b), the bipartite clustering coefficient $\bar{C}_4$ is stabilized at a constant value. In conclusion, the results in Fig. 4.13 and Fig. 4.14 also reflect the non-vanishing bipartite clustering in the CRHG model.

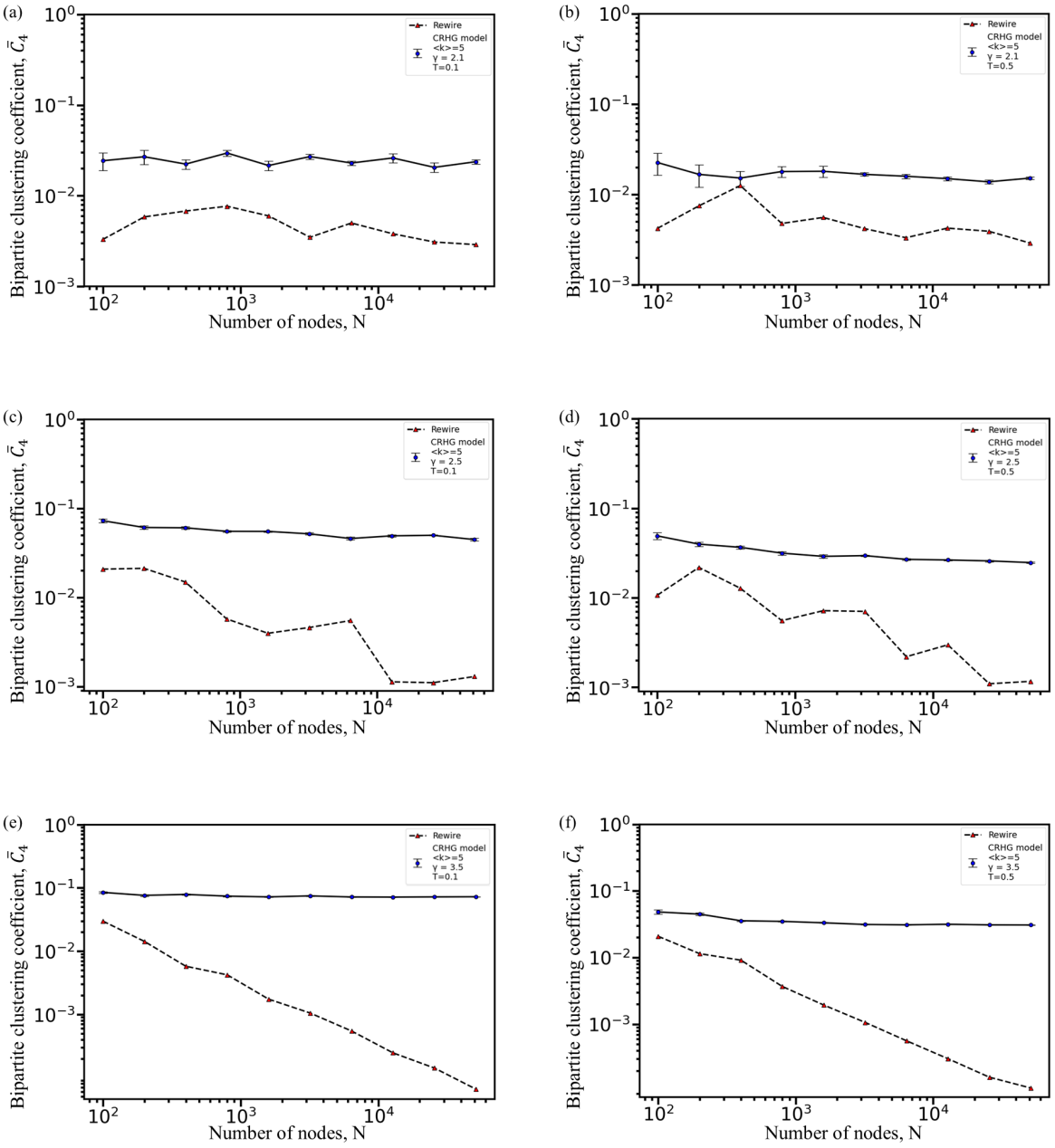

Figure 4.11: The bipartite clustering coefficient $\bar{C}_4$ of the CRHG model for $\langle k \rangle = 5$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, each red triangle represents the $\bar{C}_4$ of CRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_4$ with the growth of N in rewired random graphs.

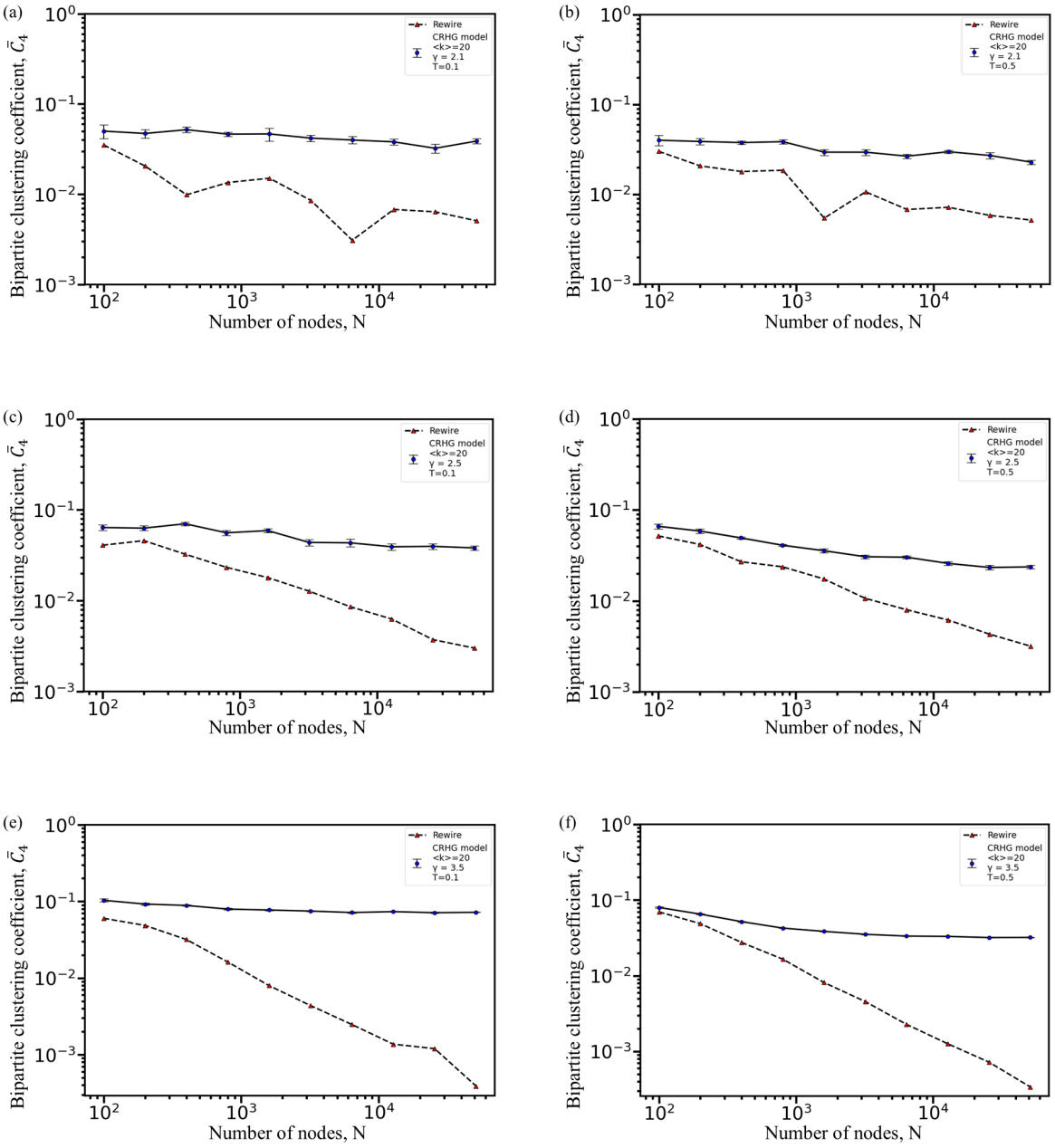

Figure 4.12: The bipartite clustering coefficient $\bar{C}_4$ of the CRHG model for $\langle k \rangle = 20$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, each red triangle represents the $\bar{C}_4$ of CRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_4$ with the growth of N in rewired random graphs.

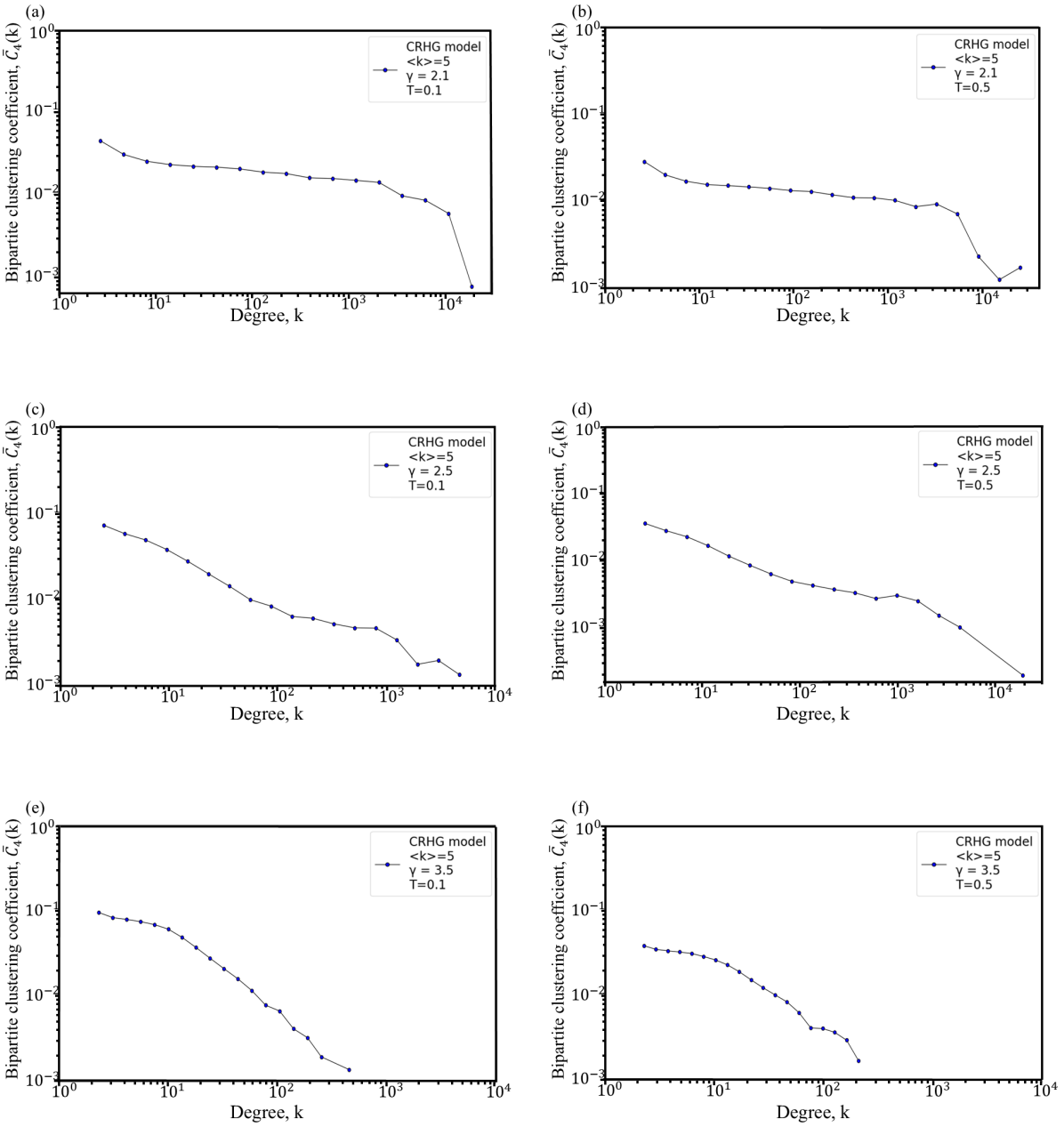

Figure 4.13: The average clustering coefficient $\bar{C}_4$ of the nodes with degree k of the CRHG model for $\langle k \rangle = 5$ and different input parameters: $N = 1.024 * 10^5, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, the linear binning [21] is applied with 20 bins for the $\bar{C}_4(k)$ curves to capture the overall variation trend.

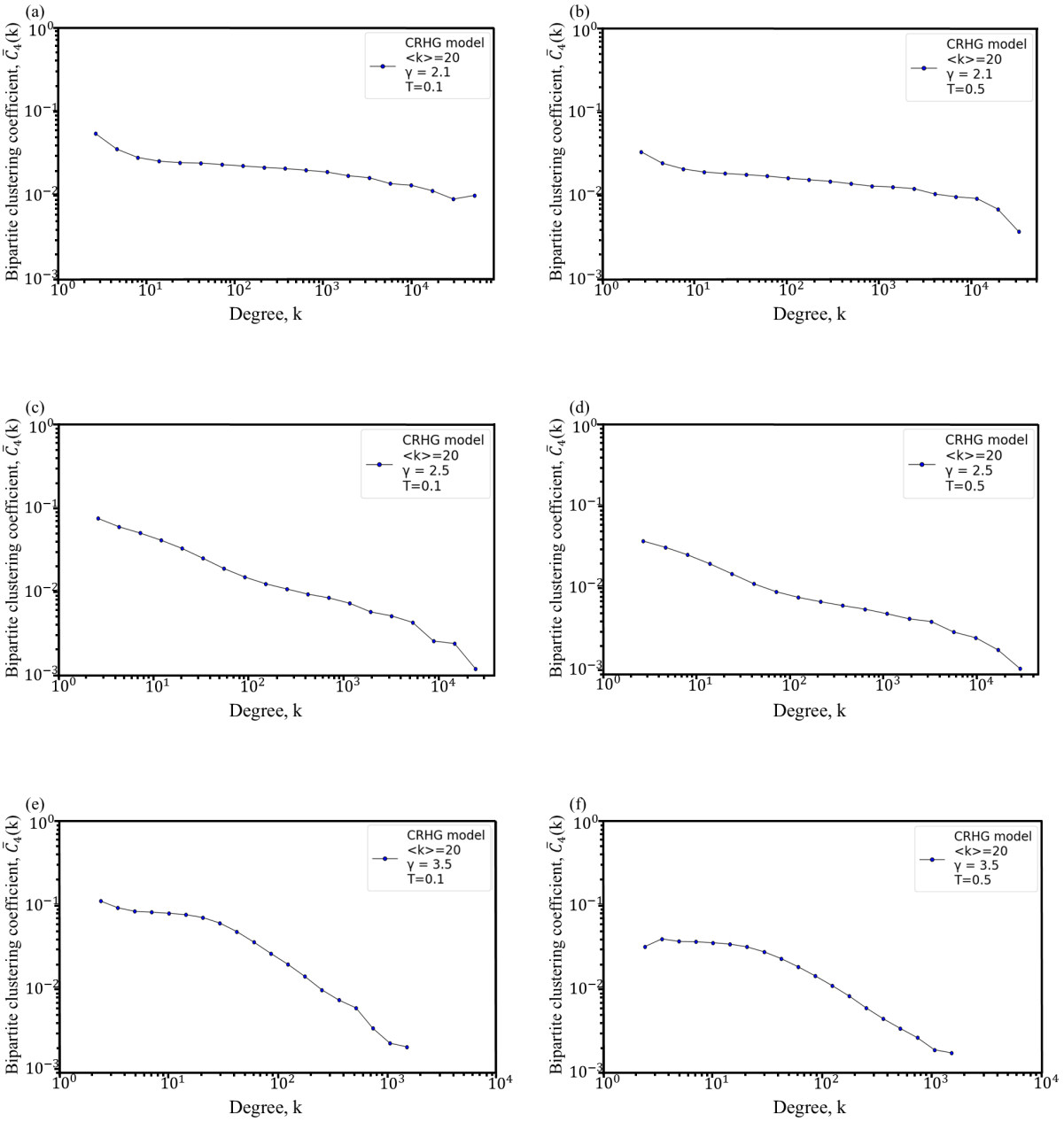

Figure 4.14: The average clustering coefficient $\bar{C}_4$ of the nodes with degree k of the CRHG model for $\langle k \rangle = 20$ and different input parameters: $N = 1.024 * 10^5, \gamma = \{2.1, 2.5, 3.5\}, T = \{0.1, 0.5\}$. Panels (a)-(f) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T. In each subplot, the linear binning [21] is applied with 20 bins for the $\bar{C}_4(k)$ curves to capture the overall variation trend.

4.6. Intermediate summary: The place of the CRHG among other models

The Complementarity Random Hyperbolic Graph (CRHG) model exhibits three fundamental properties: the Scale-Free property, the Small-World property, and Non-vanishing bipartite clustering at the same time the clustering coefficient C_3 is vanishing. These features position the CRHG model as a good synthetic network model with unique advantages compared to other classic models. A brief overview of properties in different synthetic models is shown in Table. 4.2.

Properties / Models	Degree distribution ($Pr[D=k]$)	Small-World ($\langle d \rangle$)	Clustering coefficient (C_3)	Bipartite clustering (C_4)
ER	Poisson	Yes, $\langle d \rangle \sim \ln N$	No, $C_3 \sim \frac{1}{N}$	
PA	$Pr[D=k] \sim k^{-\gamma}$ $(\gamma = 3)$	Yes, $\langle d \rangle \sim \frac{\ln(N)}{\ln(\ln(N))}$	No, $C_3 \sim \frac{(\ln(N))^2}{N}$	
RGG	Poisson	No, $\langle d \rangle \sim N^{\frac{1}{D}}$	Strong, $C_3 = \mathrm{O}(1)$	
RHG	$Pr[D=k] \sim k^{-\gamma}$ $\gamma \in (2,3)$	Yes, $\langle d \rangle \sim \ln N$	Strong, $C_3 = \mathrm{O}(1)$	Strong, $C_4 = \mathrm{O}(1)$
CRHG	$Pr[D=k] \sim k^{-\gamma}$ $\gamma \in (2,3)$	Yes	Weak, $C_3 = \mathrm{O}(N^{-a})$ $a > 0$	Strong, $C_4 = \mathrm{O}(1)$

Table 4.2: The brief overview of the topological properties among different synthetic models: The cross means we did not measure the values of that part; "Yes" means the synthetic model has the property; "No" means the synthetic model does not have the property

Comparison with ER and PA Models: The CRHG model has Scale-Free and Small-World properties with the Erdős–Rényi (ER) and Preferential Attachment (PA) models. These properties reflect the fact that the CRHG model can exhibit the general topological properties of real-world networks.

Comparison with the RGG Model: Unlike the Random Geometric Graph (RGG) model, the CRHG model demonstrates Non-vanishing bipartite clustering. This kind of clustering usually reflects the complementary characteristics of the real-world networks. Therefore, the CRHG model may be more advantageous in fitting real-world networks following the principle of complementarity, such as protein-protein interactions networks, and production networks.

Comparison with the RHG Model: The primary difference between the CRHG model and the Random Hyperbolic Graph (RHG) model depends on their node connection mechanisms. While the RHG model relies on a similarity mechanism to form connections between nodes, the CRHG model employs a complementarity mechanism. This means that the CRHG model not only exhibits the characteristics of its complementarity mechanism but also possesses the basic topological properties of real-world networks. This further illustrates that the CRHG model provides a significant reference value for explaining complementary mechanisms in real-world networks.

In summary, the CRHG model stands out by combining Scale-Free and Small-World properties with Non-vanishing bipartite clustering. It is a novel combination among other synthetic models. Its unique complementary connectivity mechanism makes it an innovative synthetic model. This model is particularly effective for modelling complex systems formed by complementarity-driven networks.

5

The Generalized Complementarity Random Hyperbolic Graph model and its properties

Motivation: In Section 2.4.4, we have already introduced the RHG model which can perform the similarity-driven mechanism in networks well. Then, we define the CRHG model which serves as a foundation for studying networks with complementarity-driven mechanisms and discuss the topological properties of CRHG. However, RHG or CRHG has only a single mechanism characteristic either similarity (RHG) or complementarity (CRHG). Can we define a synthetic network model that can be smoothly transition between similarity and complementarity?

5.1. Generalized Complementarity Random Hyperbolic Graph (GCRHG) model

GCRHG model

The generalized CRHG model is based on the CRHG model we defined before, but it has one more parameter $f \in [0, 1]$ which defines the relations between coordinates of the same node. Compared with the RHG in Section 4.1, in GCRHG, with probability f, coordinates $\mathbf{x}_i$ and $\mathbf{y}_i$ are selected independently, and with probability $1 - f$, $\mathbf{x}_i = \mathbf{y}_i$ In more precise terms, the GCRHG model is defined as follows [2]:

- Each of the N nodes is associated with two points in the *2*-dimensional hyperbolic space. Thus, each node i is characterized by $\mathbf{x}_i \equiv \{r_i^1, \theta_i^1\}$. With probability f, the $\mathbf{y}_i \equiv \{r_i^2, \theta_i^2\}$. With probability $1 - f$, radial and angular positions of the second point are equal to those of the first point, meaning $\mathbf{y}_i \equiv \{r_i^1, \theta_i^1\}$. The coordinates of points are drawn uniformly at random in $\mathbb{H}^2$ as follows:

$$r_i^{12} \leftarrow \rho(r), r_i^{12} \in [0, R] \quad \theta_i^{12} \leftarrow \mathcal{U}(0, 2\pi), \tag{5.1}$$

  where the probability density function $\rho(r)$ of r_i^{12} is given by Eq.(5.2)

$$\rho(r) = \alpha \frac{\sinh \alpha r}{\cosh \alpha R - 1} \tag{5.2}$$

prescribed by the model. Here $R > 0$ is the radius of the hyperbolic disk and $\alpha \in (0.5, 1)$ is the parameter controlling the density of the nodes in $\mathbb{H}^2$.

- For each node pair i and j, one computes one distance as Eq.(2.26) when $f = 0$. For $f = 1$, one computes two distances as Eq.(4.3).
- The connection probability p_{ij} is determined after $\mathbf{x}_i$ and $\mathbf{y}_j$ are selected. This part is identical to CRHG.

in $f = 0$ the GCRHG model is equivalent to the RHG model. Otherwise, when $f = 1$ the GCRHG model is the CRHG model. When $f \in (0, 1)$, the GCRHG model is in between, which means that the nodes in the target network perform the mapping and connecting of the points using the RHG and CRHG mechanisms at the same time. In other words, GCRHG allows for smooth turning between similarity and complementarity. The *2*-dimensional GCRHG model has four parameters f, R, α and T. The only difference is the radio f, which controls the percentage of RHG-typed points versus CRHG-typed points.

The GCRHG model with the target number of nodes N, average degree $\langle k \rangle$, degree distribution exponent γ, radio f and temperature T in $\mathbb{H}^2$ is constructed in the following steps:

- Randomly select the coordinates of $2N$ points in $\mathbb{H}^2$: Each node, with probability $f \in [0, 1]$ has two independent points $\{r_i^1, \theta_i^1\}$ and $\{r_i^2, \theta_i^2\}$; with probability $1 - f$ has identical points $r_i^1 = r_i^2$ and $\theta_i^1 = \theta_i^2$
- Calculate the hyperbolic distances of all point pairs $(\mathbf{x}, \mathbf{y})$ in $\mathbb{H}^2$
- Randomly connect node pairs (i, j) independently with the connection probability p_{ij} related to the point-based connection probability.

Overall, the GCRHG model is defined by the same parameters as the CRHG model, except for the radio f. The generalized CRHG (GCRHG) for $f = 1$ is nothing else but the CRHG model. However, for $f = 0$, the GCRHG model becomes analogous to the RHG because every node has effectively one point. The pseudocode algorithm of the GCRHG model is shown in Appendix A.

5.2. Degree distribution of the GCRHG model

We established that the graphs produced by the CRHG model are Scale-Free, $Pr[D = k] \sim k^{-\gamma}$ with $\gamma = 2\alpha+1$, and we are interested in whether the GCRHG model has the Scale-Free property with floating f values under different γ values. In the simulation, we set different γ value along with $N = 10^5, \langle k \rangle = 10, T = 0.5$ as input parameters. Figure. 5.1 shows that the GCRHG model performs as a Scale-Free network when different $\gamma \in (2, 3)$ are chosen with floating f values. Table. 5.1 lists the combination with the estimated $\hat{\gamma}$.

Ratio f	**Node density** α	**Theory** $\gamma = 2\alpha + 1$	**Average Degree** $\langle k \rangle$	**Temperature** T	$\hat{\gamma}$	$\lvert\hat{\gamma} - \gamma\rvert$
0	0.55	2.1	10	0.5	2.17	0.07
0.1	0.55	2.1	10	0.5	2.18	0.08
0.5	0.55	2.1	10	0.5	2.22	0.12
0.9	0.55	2.1	10	0.5	2.17	0.07
1	0.55	2.1	10	0.5	2.15	0.05
0	0.75	2.5	10	0.5	2.40	0.1
0.1	0.75	2.5	10	0.5	2.51	0.01
0.5	0.75	2.5	10	0.5	2.56	0.06
0.9	0.75	2.5	10	0.5	2.60	0.1
1	0.75	2.5	10	0.5	2.58	0.08
0	1	3	10	0.5	3.07	0.07
0.1	1	3	10	0.5	3.10	0.1
0.5	1	3	10	0.5	3.10	0.1
0.9	1	3	10	0.5	3.32	0.32
1	1	3	10	0.5	3.30	0.3

Table 5.1: Various combinations of ratios f, node density $\alpha(\gamma)$, average degree $\langle k \rangle$ and temperature T, with estimated $\hat{\gamma}$ and $|\hat{\gamma} - \gamma|$

Discussion and Observation

As f increases monotonically, the designed simulation represents an evolutionary process transitioning from a pure RHG model to a pure CRHG model. Since RHG and CRHG exhibit Scale-Free property, we have reason to assume that the GCRHG model is theoretically expected to possess Scale-Free property with $\gamma \in (2, 3)$.

With floating f value, we find that the estimated $\hat{\gamma}$ values are generally close to the theoretical critical values $\gamma \in \{2.1, 2.5, 3\}$. This indicates that the GCRHG model is capable of generating scale-free networks. However, some discrepancies exist between the theoretical γ and the estimated $\hat{\gamma}$. As shown in Table. 5.1, the maximum error is observed to be 0.32 when f equals 0.9. In other words, the $|\hat{\gamma} - \gamma|$ gets worse when $\alpha \to 1$ and $f \to 1$. We hypothesize that when $\alpha \to 1$ we have fewer existing hubs in the model and it is harder to measure γ. When $f \to 1$ we have a pure CRHG model. In this case, $\gamma = 2\alpha + 1$ is less accurate than in the $f = 0$ case, when we have RHG.

In conclusion, we observe that the degree distributions follow a power-law distribution, characterized by a fat tail as the degree increases. Although some discrepancies are noted between the estimated $\hat{\gamma}$ and theoretical γ values, the degree distribution remains largely unchanged. Therefore, it can be objectively stated that the GCRHG model exhibits scale-free properties under the parameter settings used in this thesis.

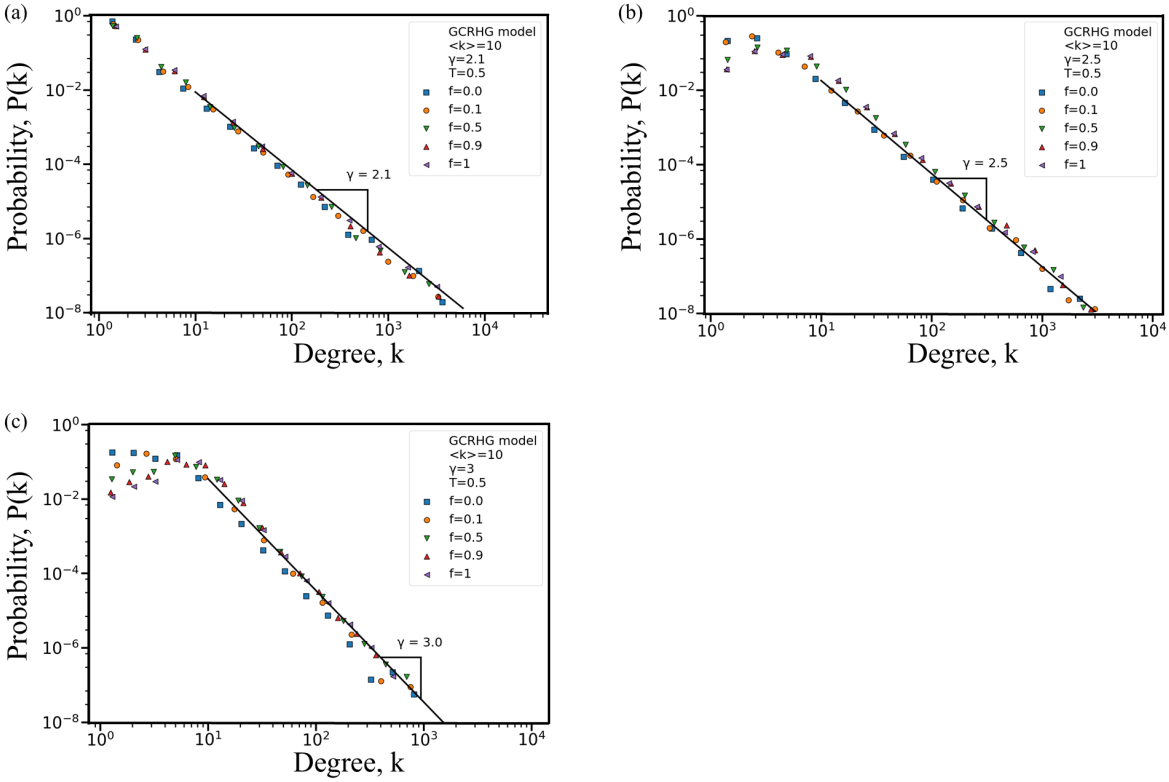

Figure 5.1: Degree distribution $Pr[D = k]$ of the GCRHG model for $f = \{0, 0.1, 0.5, 0.9, 1\}$ and different input parameters: $N = 10^5, \langle k \rangle = 10, \gamma = \{2.1, 2.5, 3\}, T = 0.5$. We use the logarithmic binning [21] for the degree distribution with the number of binning 15. The line represents the theory linear regression of γ value, and the theoretical value is displayed as a slope above the straight line. It does not start at the beginning because the linear part at large values of the degree k is the focus of our attention. Here, different colours as well as shapes represent different values of f.

5.3. Clustering coefficient, C_3

In Chapter 4, we established that the CRHG model has vanishing clustering C_3. And we have known that the RHG model demonstrates strong clustering. As a generalized model, it is natural to ask: what kinds of clustering behaviours exist in the GCRHG model? In other words, will the clustering coefficient in the GCRHG model be the same as those in the original CRHG and RHG models? To explore this, we measure the average clustering coefficient $\bar{C}_3$ as a function of network size N. Additionally, for every graph, we compute the $\bar{C}_3$ of its degree-preserving rewired graph for reference. In our simulation, as shown in Fig. 5.2, Fig. 5.3 and Fig. 5.4, we use the following parameters:

$$N = \{10^2, 2\times10^2, ..., 2^9\times10^2\}, f = \{0, 0.1, 0.5, 0.9, 1\}, \langle k\rangle = 10, \gamma = \{2.1, 2.5, 3\}, T = 0.5$$

as input to produce the GCRHG model graphs. We also plot the $\bar{C}_3$ as a function of increasing ratio f in Fig. 5.2 (f), Fig. 5.3 (f) and Fig. 5.4 (f).

Then, we calculate $\bar{C}_3(k)$ of nodes with degrees k based on the 10 sample graphs of the GCRHG model with $N = 10^5$, see Fig. 5.5.

Observation and Discussion: $\bar{C}_3$ vs N

(1) $\bar{C}_3$ does not depend on N for all f values except $f = 1$. In other words, C_3 is nearly constant as a function of N, except for $f = 1$.

(2) $\bar{C}_3$ decreases as f increases. Notably, the $\bar{C}_3$ drops sharply as f approaches $f = 1$. We conjecture that once $f \neq 1$, similarity appears. In other words, even if only a few nodes display similar behaviour, the clustering coefficient is no vanishing.

Overall, except for the case when $f = 1$ (representing the pure CRHG model), the clustering of the GCRHG model is non-vanishing. Furthermore, when $f = \{0, 0.1, 0.5\}$, the clustering is particularly strong. The trend means that the model turns into a pure CRHG model with increasing f, especially, with $f \in (0.9, 1)$ moment, the change is notable, see panel (e) in Fig. 5.2, Fig. 5.3 and Fig. 5.4.

Observation and Discussion: $\bar{C}_3(k)$ vs k:

We calculate $\bar{C}_3(k)$ of the nodes with k based on the sample of 10 graphs of the GCRHG model that set the same input parameters. When plotting the figures, we applied linear binning [21] with 20 bins for $\bar{C}_3(k)$ to capture the overall variation trend.

In Fig. 5.5, we observe that when $f = \{0, 0.1, 0.5\}$, the value of $\bar{C}_3(k)$ for nodes with low k is very high. This indicates the presence of strong clustering in these cases, particularly for smaller values of f. As shown in Fig. 5.5 (b) and (c), the value of $\bar{C}_3(k)$ decreases significantly as f increases from 0.9 to 1 when $\gamma = 2.5$. Furthermore, as $\gamma = 3$, the value of $\bar{C}_3(k)$ approaches zero, indicating the absence of clustering under these conditions.

Overall, the clustering in the GCRHG model is closely related to the parameter f. Specifically, the adjustment of f makes the model dynamically reflect the disappearance of clustering from the RHG model to the CRHG model. In other words, the GCRHG model shows strong clustering coefficient when it is close to the pure RHG model. Notably, when f changes from 0.9 to 1, the C_3 decreases rapidly. We conjecture that this is because, at $f = 0.9$, triangle clustering arising from the RHG model component within the CRHG graph still dominates. While for $f = 1$, the triangle clustering is vanishing.

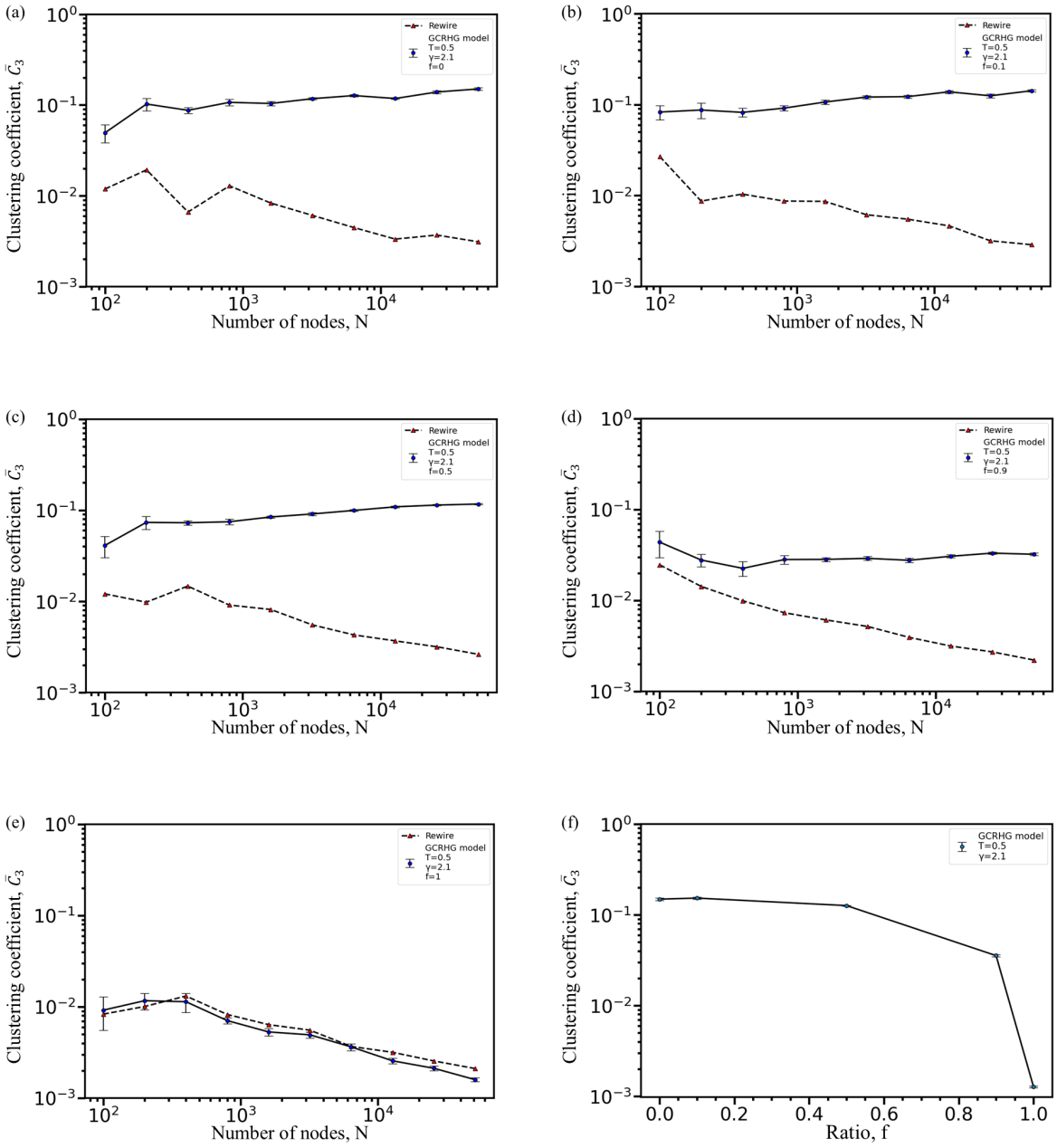

Figure 5.2: The clustering coefficient $\bar{C}_3$ of the GCRHG model for $\gamma = 2.1$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}$, $\langle k \rangle = 10, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. Panels (a)-(e) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T, f. In each panel, each red triangle represents the $\bar{C}_3$ of GCRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_3$ with the growth of N in rewired random graphs. Panels (f) shows the $\bar{C}_3$ as a function of increasing ratio f with $N = 10^5, \langle k \rangle = 10, T = 0.5, \gamma = 2.1$.

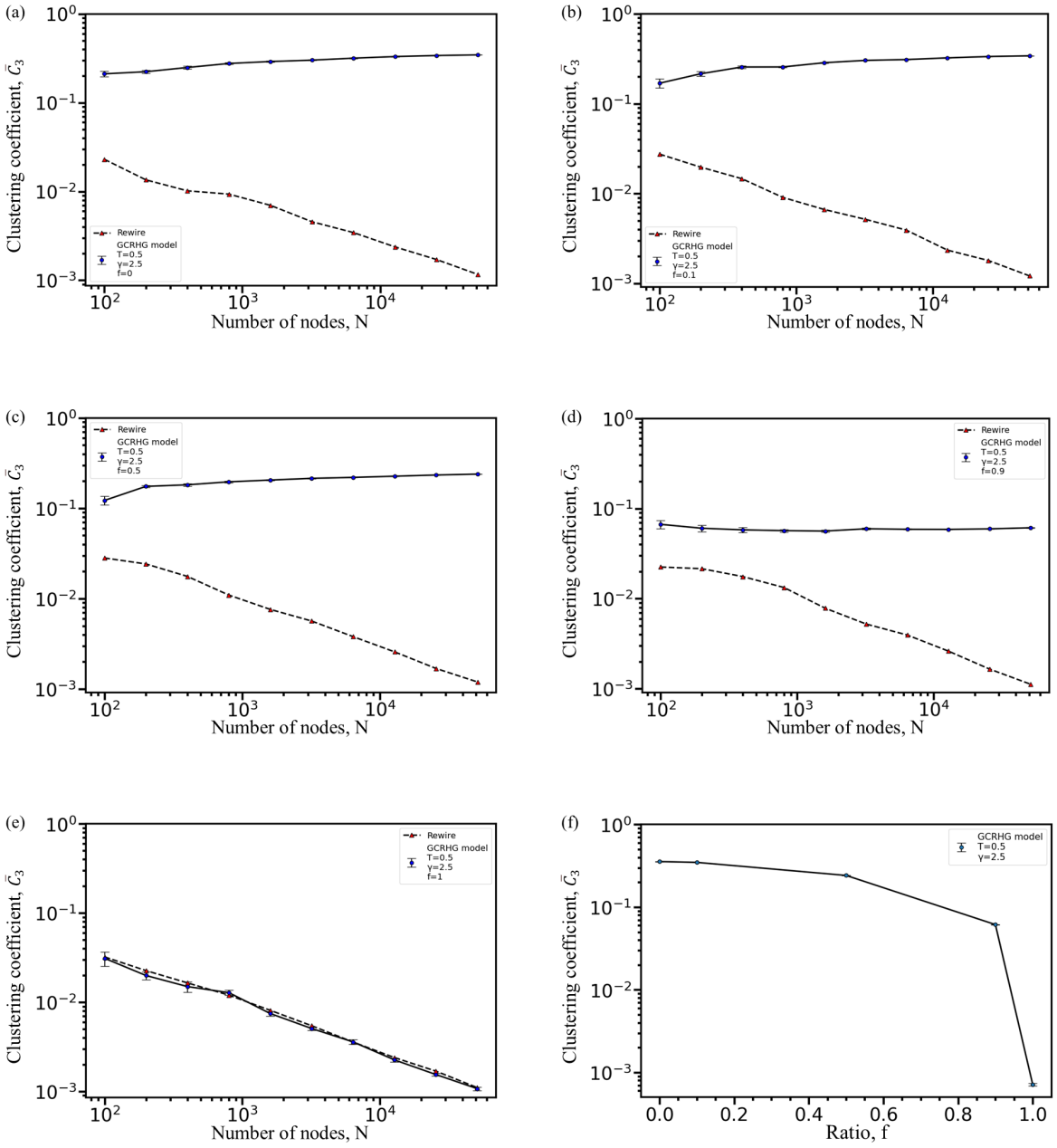

Figure 5.3: The clustering coefficient $\bar{C}_3$ of the GCRHG model for $\gamma = 2.5$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}$, $\langle k \rangle = 10, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. Panels (a)-(e) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T, f. In each panel, each red triangle represents the $\bar{C}_3$ of GCRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_3$ with the growth of N in rewired random graphs. Panels (f) shows the $\bar{C}_3$ as a function of increasing ratio f with $N = 10^5, \langle k \rangle = 10, T = 0.5, \gamma = 2.5$.

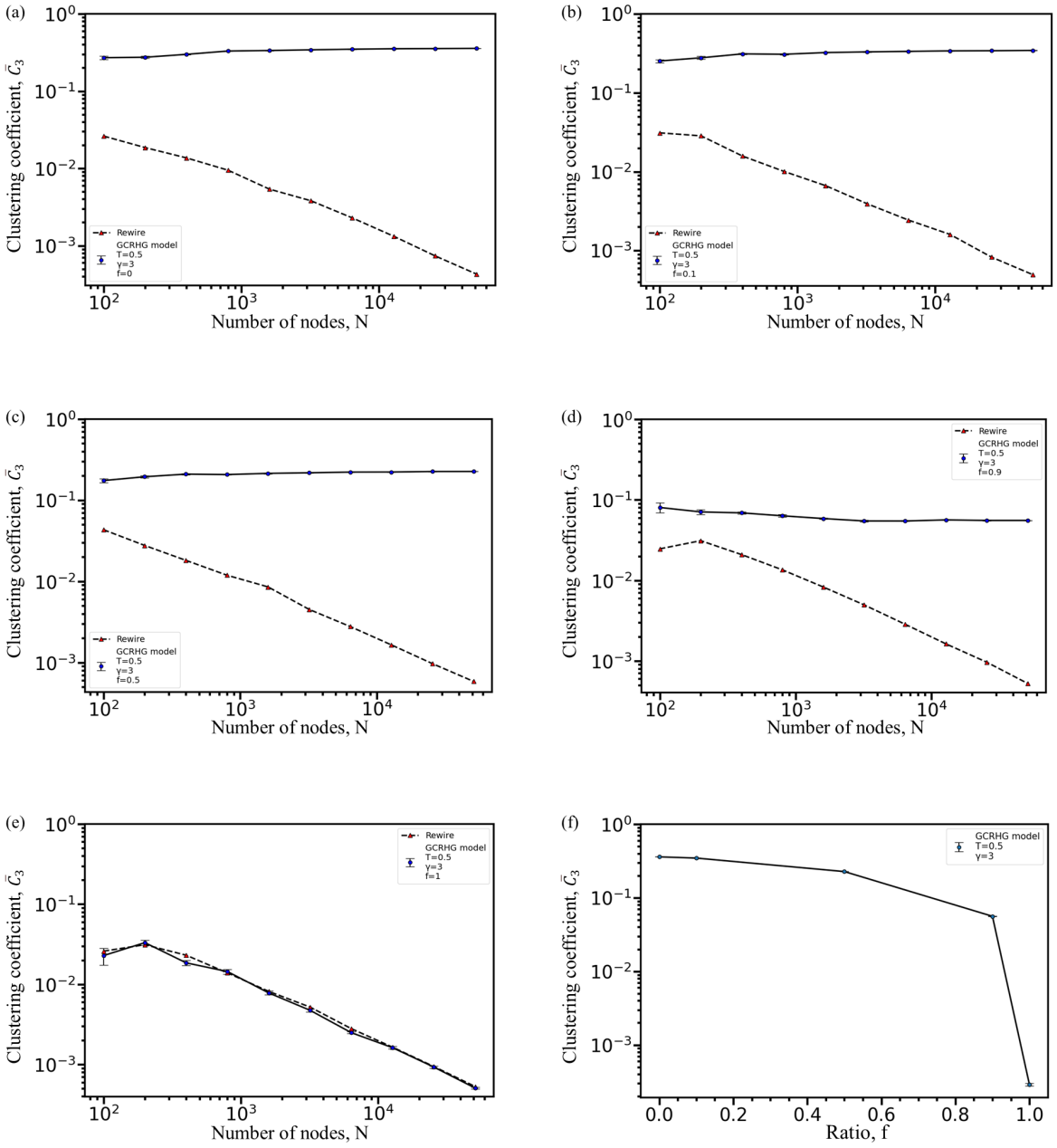

Figure 5.4: The clustering coefficient $\bar{C}_3$ of the GCRHG model for $\gamma = 3$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}$, $\langle k \rangle = 10, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. Panels (a)-(e) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T, f. In each panel, each red triangle represents the $\bar{C}_3$ of GCRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_3$ with the growth of N in rewired random graphs. Panels (f) shows the $\bar{C}_3$ as a function of increasing ratio f with $N = 10^5$, $\langle k \rangle = 10, T = 0.5, \gamma = 3$.

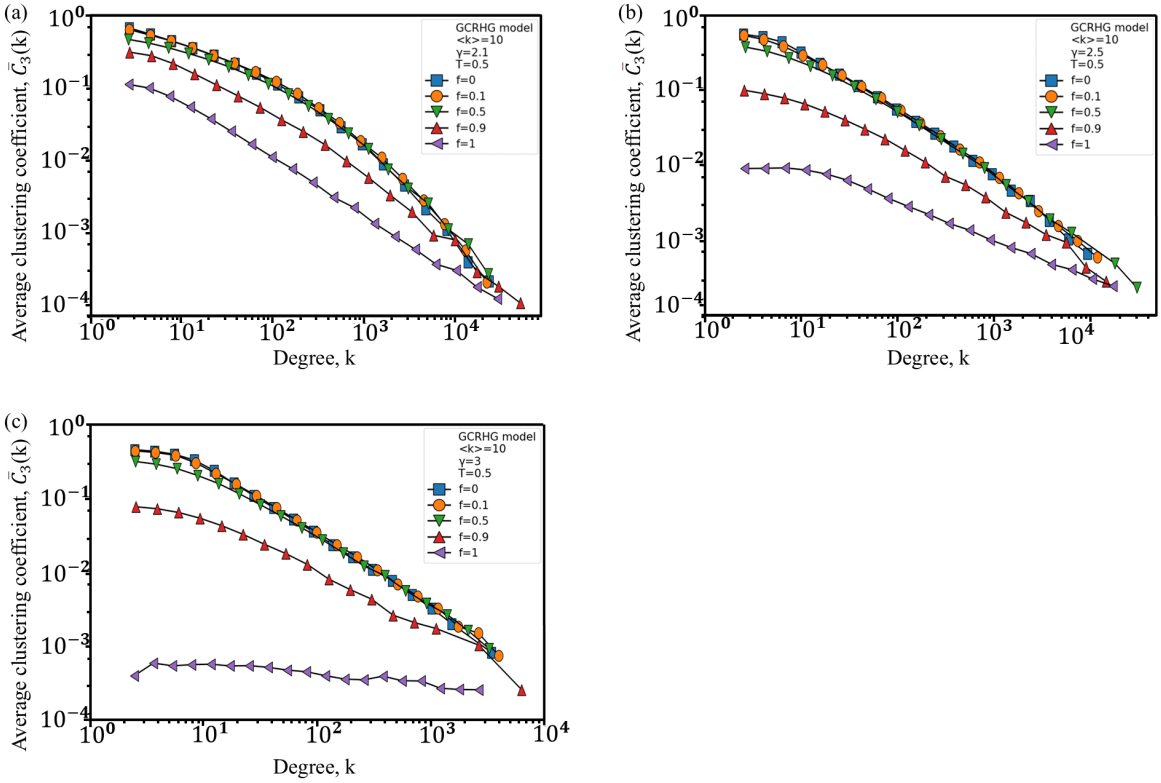

Figure 5.5: The average clustering coefficient $\bar{C}_3$ of the nodes with degree k of the GCRHG model with different input parameters: $N = 10^5, \langle k \rangle = 10, \gamma = \{2.1, 2.5, 3\}, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. The linear binning [21] is applied with 20 bins for the $\bar{C}_3(k)$ curves to capture the overall variation trend. Here, different colours as well as shapes represent different values of f.

5.4. Bipartite clustering, C_4

We also measure the average bipartite clustering coefficient $\bar{C}_4$ as the network size N increases given by Eq.(5.3)

$$\bar{C}_4 = \frac{1}{N}\sum_{i=1}^{N} C_4(i) \tag{5.3}$$

Additionally, for every graph, we compute the $\bar{C}_4$ of its degree-preserving rewired graph for reference. In our simulation as shown in Figs. 5.6 to 5.8 , we use the following parameters:

$$N = \{10^2, 2\times 10^2, ..., 2^9\times 10^2\}, f = \{0, 0.1, 0.5, 0.9, 1\}, \langle k\rangle = 10, \gamma = \{2.1, 2.5, 3\}, T = 0.5$$

as input to produce the GCRHG model graphs. The $\bar{C}_4$ as a function of increasing ratio f in Fig. 5.6 (e), Fig. 5.7 (e) and Fig. 5.8 (e).

In Fig. 5.9, the $\bar{C}_4(k)$ of nodes with k based on the 10 sample graphs of the GCRHG model with $N = 10^5$ is shown.

Observation and Discussion: $\bar{C}_4$ **vs** N

(1) $\bar{C}_4$ is close to random for similarity-based networks. Especially, as shown in Fig. 5.6 (a) and (b).

(2) $\bar{C}_4$ increases as f increases. Therefore, $\bar{C}_4$ is larger for purely complementarity-based networks. As shown in panels (c), (d), and (e) of Figs. 5.6 to 5.8, when $f = \{0.5, 1\}$, the $\bar{C}_4$ coefficient stabilizes as N increases and remains non-zero. This behaviour occurs because the GCRHG model increasingly resembles a pure CRHG model within this parameter range.

Observation and Discussion: $\bar{C}_4(k)$ **vs** k

We calculate $\bar{C}_4(k)$ of the nodes with k based on the sample of 10 graphs of the GCRHG model that set the same input parameters. When plotting the figures, we applied linear binning [21]with 20 bins for the $\bar{C}_4(k)$ to capture the overall variation trend.

In Fig. 5.9, we observe that when $f = \{0.9, 1\}$, the $\bar{C}_4(k)$ values in the GCRHG model are significantly higher than those of other curves. This indicates a pronounced bipartite clustering effect for the GCRHG model when it tends to be a pure CRHG model.

Comparing Fig. 5.9. (a) and (b), specifically the two curves with $f = 0.5$, we observe that as the value of γ increases, the green curve's $\bar{C}_4$ behaviour gradually converges toward that of the pure RHG model. This suggests that for the GCRHG model with $f = 0.5$, the bipartite clustering effect becomes weaker with increasing γ.

Overall, the bipartite clustering effect is most pronounced in the GCRHG model when $f = \{0.9, 1\}$. This observation implies that as the GCRHG model transitions toward a pure CRHG model, bipartite clustering dominates. In other words, this phenomenon results from the non-vanishing bipartite clustering inherent to the CRHG model.

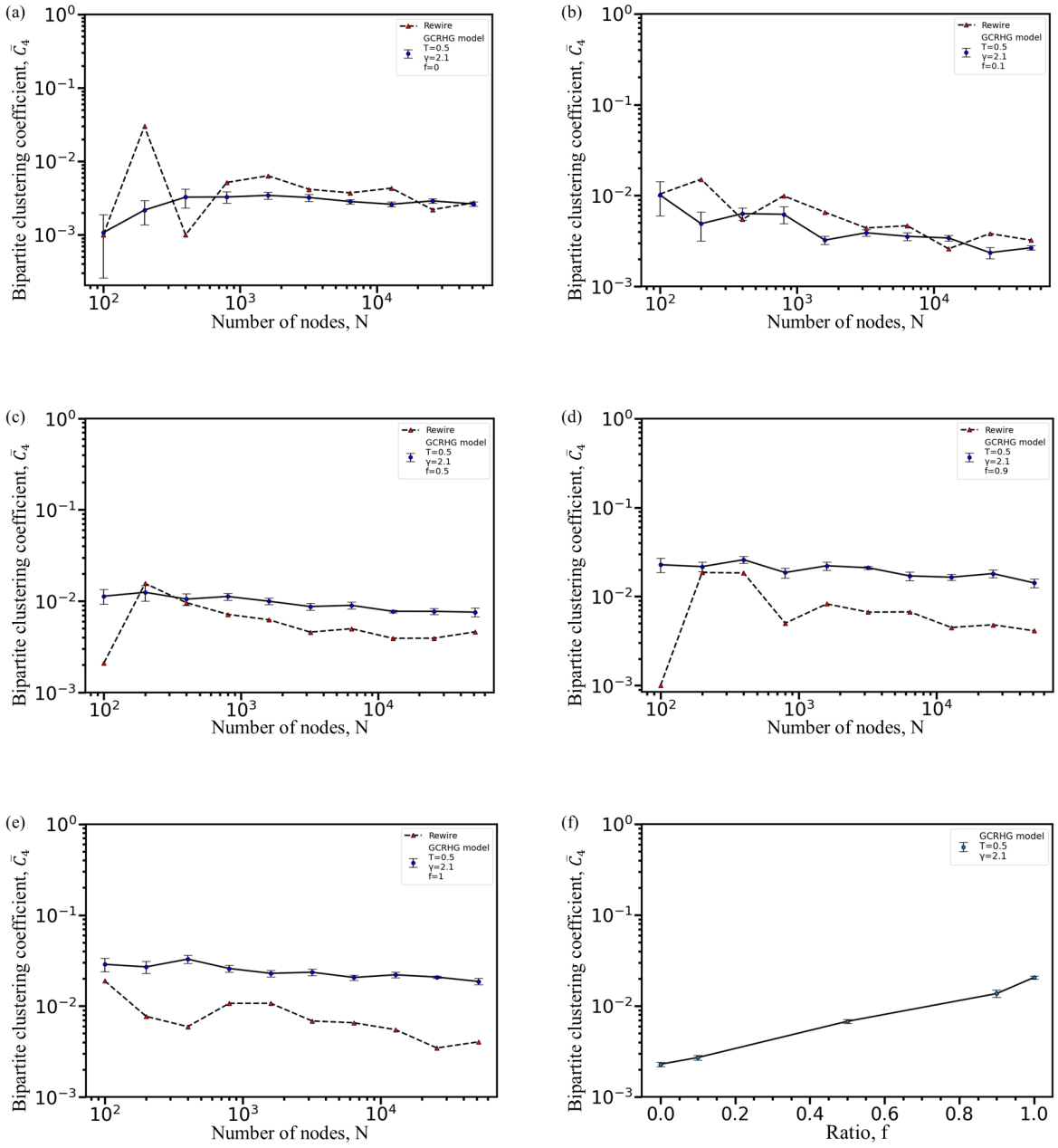

Figure 5.6: The bipartite clustering coefficient $\bar{C}_4$ of the GCRHG model for $\gamma = 2.1$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}$, $\langle k \rangle = 10, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. Panels (a)-(e) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T, f. In each panel, each red triangle represents the $\bar{C}_4$ of GCRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_4$ with the growth of N in rewired random graphs. Panels (f) shows the $\bar{C}_4$ as a function of increasing ratio f with $N = 10^5, \langle k \rangle = 10, T = 0.5, \gamma = 2.1$.

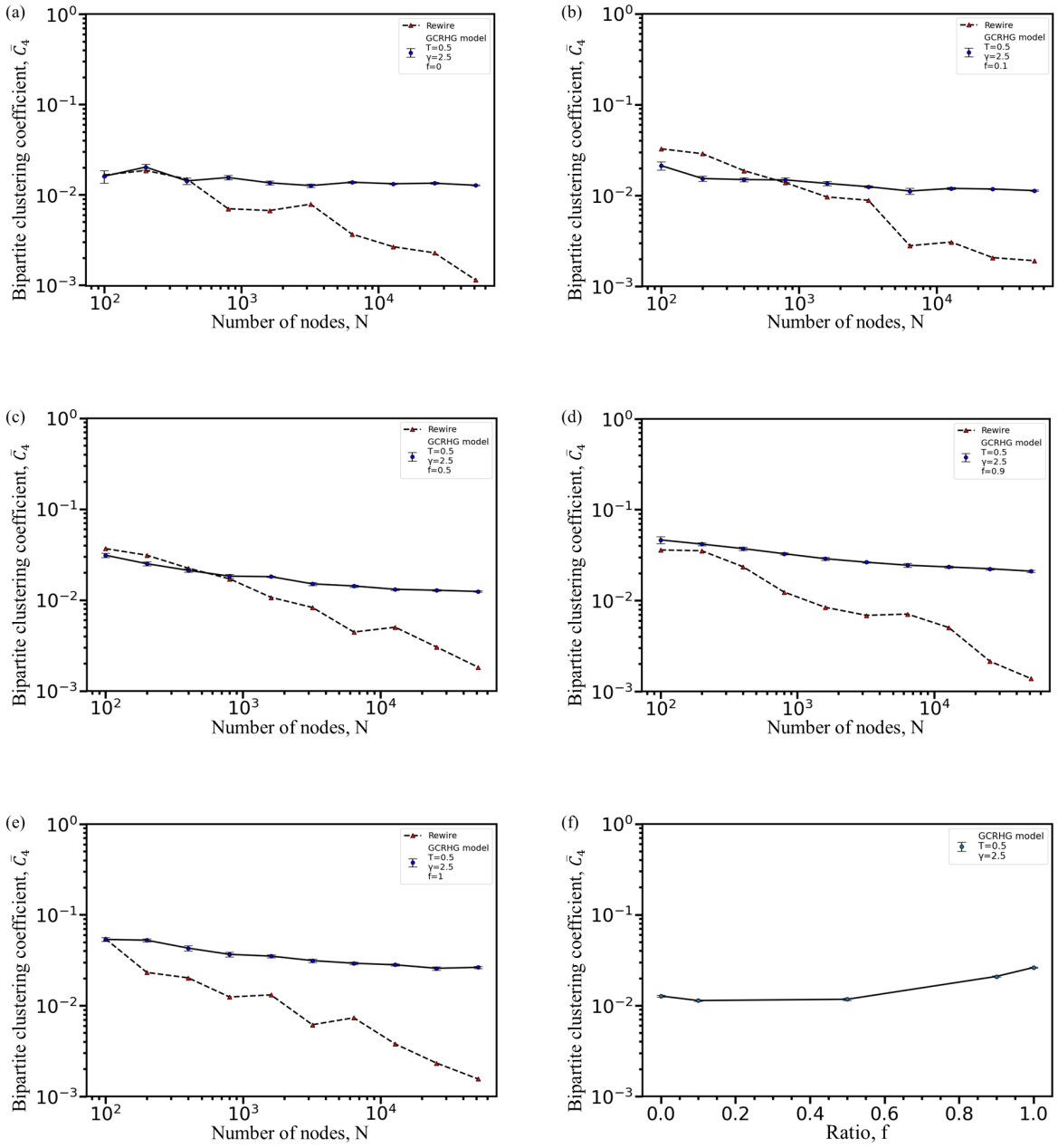

Figure 5.7: The bipartite clustering coefficient $\bar{C}_4$ of the GCRHG model for $\gamma = 2.5$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}$, $\langle k \rangle = 10, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. Panels (a)-(e) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T, f. In each panel, each red triangle represents the $\bar{C}_4$ of GCRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_4$ with the growth of N in rewired random graphs. Panels (f) shows the $\bar{C}_4$ as a function of increasing ratio f with $N = 10^5, \langle k \rangle = 10, T = 0.5, \gamma = 2.5$.

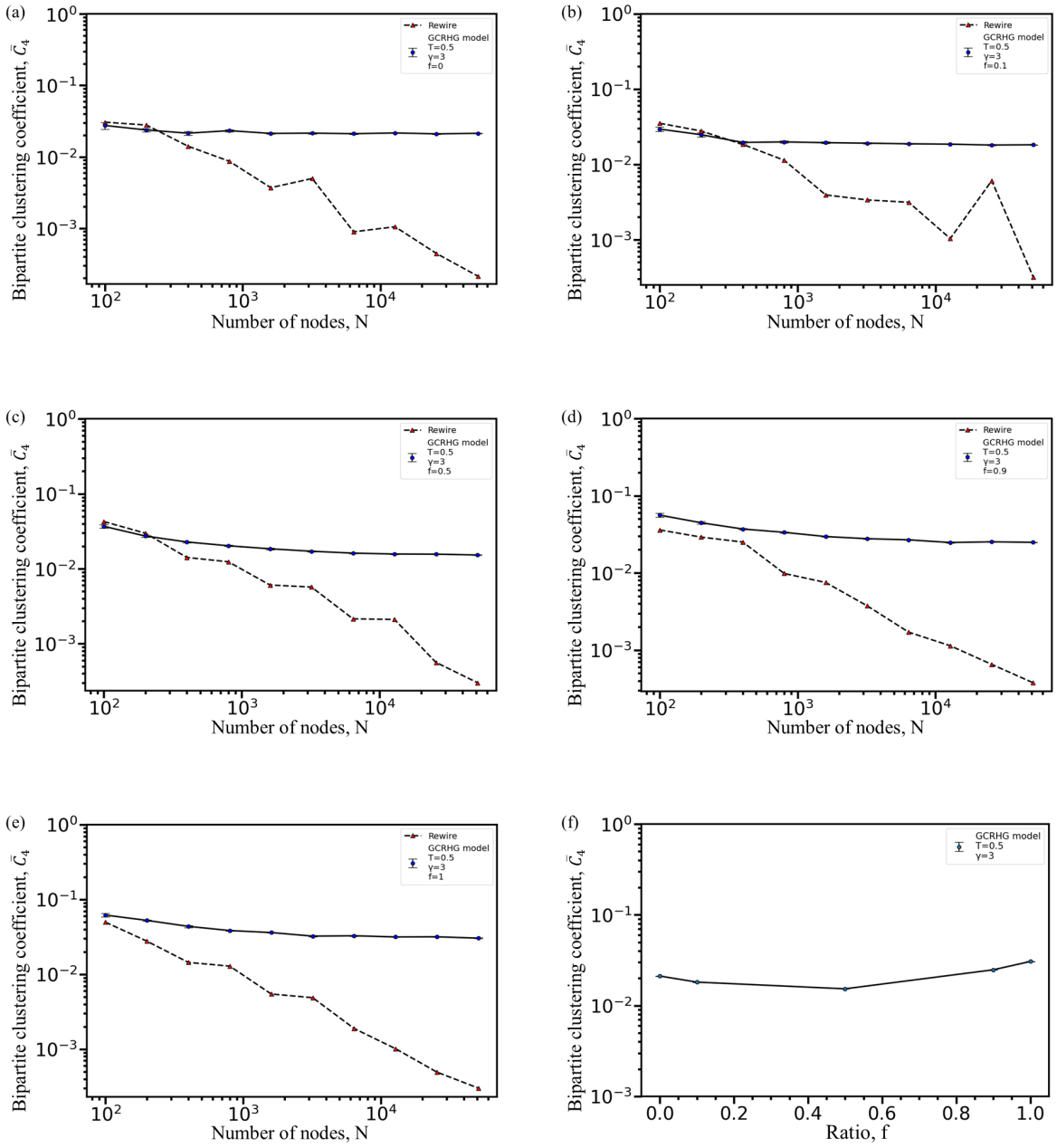

Figure 5.8: The bipartite clustering coefficient $\bar{C}_4$ of the GCRHG model for $\gamma = 3$ and different input parameters: $N = \{10^2, 2 \times 10^2, ..., 2^9 \times 10^2\}$, $\langle k \rangle = 10, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. Panels (a)-(e) illustrate the combined effects of varying input model parameters $\langle k \rangle$, γ, T, f. In each panel, each red triangle represents the $\bar{C}_4$ of GCRHG's degree-preserving rewired random graph. The dotted line formed by the red triangles represents the trend of $\bar{C}_4$ with the growth of N in rewired random graphs. Panels (f) shows the $\bar{C}_4$ as a function of increasing ratio f with $N = 10^5$, $\langle k \rangle = 10, T = 0.5, \gamma = 3$.

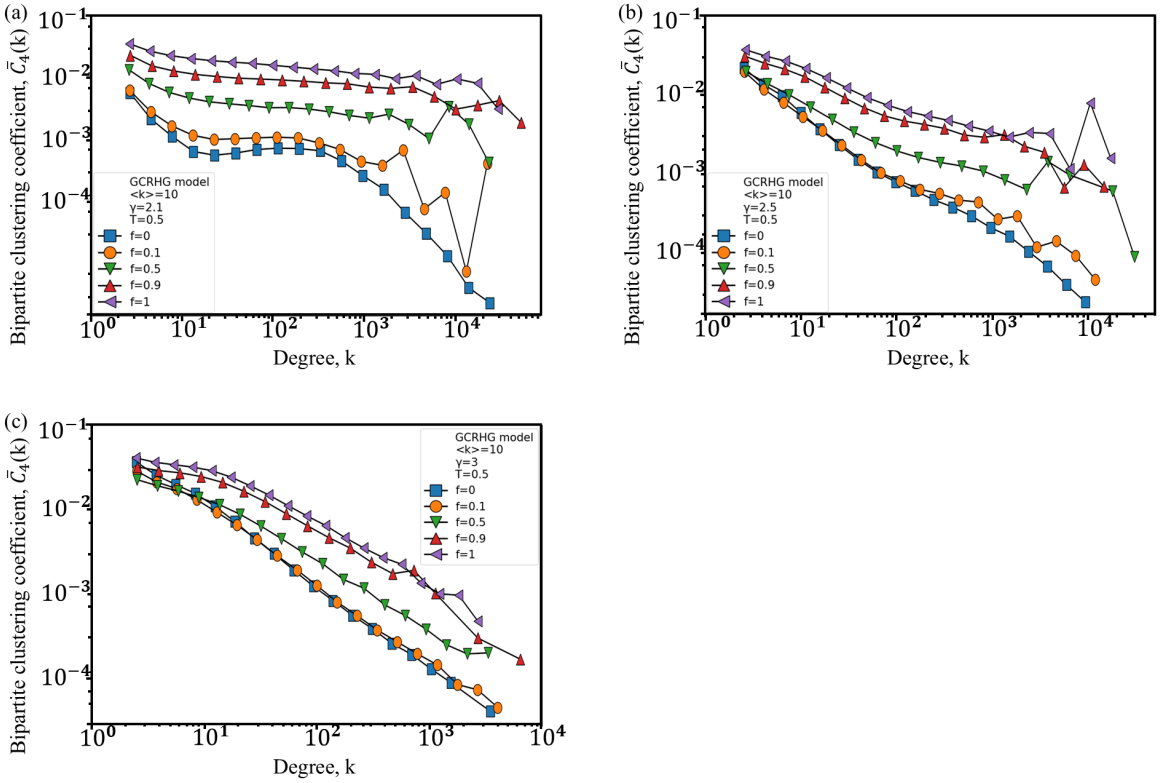

Figure 5.9: The average clustering coefficient $\bar{C}_4$ of the nodes with degree k of the GCRHG model with different input parameters: $N = 10^5, \langle k \rangle = 10, \gamma = \{2.1, 2.5, 3\}, T = 0.5, f = \{0, 0.1, 0.5, 0.9, 1\}$. The linear binning [21] is applied with 20 bins for the $\bar{C}_4(k)$ curves to capture the overall variation trend. Here, different colours as well as shapes represent different values of f.

Summary: The place of the GCRHG among other models

The Generalized Complementarity Random Hyperbolic Graph (GCRHG) model exhibits four fundamental properties: Scale-Free property, Small-World property, clustering coefficient and bipartite clustering. A brief overview of properties in different synthetic models is shown in Table. 5.2.

Models \ Properties	Degree distribution ($Pr[D=k]$)	Small-World ($\langle d \rangle$)	Clustering coefficient (C_3)	Bipartite clustering (C_4)
ER	Poisson	Yes, $\langle d \rangle \sim \ln N$	No, $C_3 \sim \frac{1}{N}$	
PA	$Pr[D=k] \sim k^{-\gamma}$ $(\gamma = 3)$	Yes, $\langle d \rangle \sim \frac{\ln(N)}{\ln(\ln(N))}$	No, $C_3 \sim \frac{(\ln(N))^2}{N}$	
RGG	Poisson	No, $\langle d \rangle \sim N^{\frac{1}{D}}$	Strong, $C_3 = \mathrm{O}(1)$	
RHG	$Pr[D=k] \sim k^{-\gamma}$ $\gamma \in (2,3)$	Yes, $\langle d \rangle \sim \ln N$	Strong, $C_3 = \mathrm{O}(1)$	Strong, $C_4 = \mathrm{O}(1)$
CRHG	$Pr[D=k] \sim k^{-\gamma}$ $\gamma \in (2,3)$	Yes	Weak, $C_3 = \mathrm{O}(N^{-a})$ $a > 0$	Strong, $C_4 = \mathrm{O}(1)$
GCRHG	$Pr[D=k] \sim k^{-\gamma}$ $\gamma \in (2,3)$		Strong, $C_3 = \mathrm{O}(1))$	Strong, $C_4 = \mathrm{O}(1)$

Table 5.2: The brief overview of the topological properties among different synthetic models including GCRHG: The cross means we did not measure the values of that part; "Yes" means the synthetic model has the property; "No" means the synthetic model does not have the property

In this chapter, GCRHG allows for fine-tuning between similarity and complementarity mechanisms by tuning the correlations between node coordinates. It represents strong clustering coefficient when $f \to 0$ and strong bipartite clustering as $f \to 1$. The results in Section 5.3 (2), Section 5.4 (1) and Section 5.4 (2) are not expected and require further investigation. Since similarity networks enforce triangle closure, one would intuitively expect that triangle closure would result in the closure of large loops, like rectangular loops. However, it is seen that the GCRHG models nevertheless, have a stronger bipartite clustering coefficient.

6

Conclusion

In this chapter, we first mention the objective of this thesis and review the main results in each chapter. Later, we offer suggestions and feedback for follow-up research based on the existing observations.

6.1. Conclusion

The main objective of this thesis is to address the knowledge gap: although a complementarity-based network model has been proposed in previous studies, its topological properties have not been systematically examined. To fill in this gap, this work systematically studies complementarity network models (CRHG, GCRHG) and documents their topological properties.

Overall, we have studied and interpreted the topological properties of the CRHG and GCRHG models. The main results of this research are: 1) The CRHG model can explain the properties of real networks. 2) The GCRHG model allows for fine-tuning between the similarity and complementarity mechanisms. 3) Some topological results were expected and previously assumed in CRHG: Scale-Free property, Small-World property, Weak C_3 and Strong C_4. We establish that this is indeed correct. The observations in Section 5.3 (2), Section 5.4 (1), and Section 5.4 (2) are not much expected and require further investigations.

In Chapter 3, we compare the similarity and complementarity principles in complex networks and emphasise the importance of complementarity. After, the minimal complementarity model is introduced, setting the stage for a complementarity-based model known as the Complementarity Random Hyperbolic Graph (CRHG) in Chapter 4.

In Chapter 4, we first define the Complementarity Random Hyperbolic Graph (CRHG) model. Then, we do the simulations to establish the Small-World property, Scale-Free property, vanishing Clustering Coefficient and non-vanishing Bipartite Clustering of the CRHG model. We find that the CRHG model stands out by combining Scale-Free and Small-World properties with Non-vanishing biparting clustering. This model is particularly effective for modelling complex networks formed by the complementarity mechanism.

To transition smoothly between similarity and complementarity, in Chapter 5, the Generalized Complementarity Random Hyperbolic Graph (GCRHG) model has been defined. The

model exhibits the Scale-Free property. With different input parameters f, the GCRHG model has clustering coefficient and non-vanishing bipartite clustering. This demonstrates that the GCRHG model can be used flexibly to capture both similarity and complementarity features in networks. When $f = 0$ the GCRHG model is equivalent to the RHG model, GCRHG model is the CRHG model as $f = 1$. When $f \in (0, 1)$, the nodes in the target network are mapped using the RHG and CRHG mechanisms at the same time.

Overall, we document and interpret the topological properties of complementarity-based spatial graph models. Additionally, we conduct partial simulation verification of the theoretical topological properties of synthetic complementarity-based models, providing a reference for their future development and applications.

6.2. Discussion

We have already performed simulation tests on the Scale-Free property, Small-World property, clustering coefficient and bipartite clustering of complementarity-based network models. Most of the measurement results are in line with our expectations. However, these are only verifications from the simulation perspective. The theoretical proof of these topological properties for the CRHG model has not yet been made. For future research, the results of this thesis may provide a guideline for theoretical proofs. One of these perspectives is to explain the topological properties exhibited by the CHRG model from the perspective of hidden variable formalism. The study of these topological properties helps us understand and analyze the characteristics of real networks. Especially, when f changes between $\{0.9, 1\}$, the clustering coefficient varies greatly in the Generalized Complementarity Random Hyperbolic Graph (GCRHG) model. As shown in panels (d) and (e) of Fig. 5.2, Fig. 5.3 and Fig. 5.4, the clustering coefficient C_3 trends from stable to decreasing. This phenomenon may be related to the density of triangular subgraphs and quadrilateral subgraphs in the GCRHG when $f \in (0.9, 1)$. To investigate the reason for this phenomenon, one possible way is that we can figure out the percentage of points mapped as the RHG or CRHG model. In other words, we can label all the corresponding points of a node in the process of generating the GCRHG.

After we fully understand the complementarity-based network models, another research direction could be exploring the relationship between these models and their real-world applications and technological developments. As we mention in Section 2.4.4, we know that the RHG model has some real applications, including link prediction, routing and navigation. Since complementarity-based network models are inspired by the success of the RHG model, they may also be relevant to some real-world applications.

Bibliography

[1] István A Kovács et al. “Network-based prediction of protein interactions”. In: *Nature communications* 10.1 (2019), p. 1240.

[2] Gabriel Budel and Maksim Kitsak. “Complementarity in complex networks”. In: *arXiv preprint arXiv:2003.06665* (2020).

[3] Timon Elmer, Kieran Mepham, and Christoph Stadtfeld. “Students under lockdown: Comparisons of students’ social networks and mental health before and during the COVID-19 crisis in Switzerland”. In: *Plos one* 15.7 (2020), e0236337.

[4] Weiwei Jiang. “Graph-based deep learning for communication networks: A survey”. In: *Computer Communications* 185 (2022), pp. 40–54.

[5] Mikaela Koutrouli et al. “A guide to conquer the biological network era using graph theory”. In: *Frontiers in bioengineering and biotechnology* 8 (2020), p. 34.

[6] Réka Albert and Albert-László Barabási. “Statistical mechanics of complex networks”. In: *Reviews of modern physics* 74.1 (2002), p. 47.

[7] Piet Van Mieghem. *Performance analysis of complex networks and systems*. Cambridge University Press, 2014.

[8] Krishnaiyan Thulasiraman and Madisetti NS Swamy. *Graphs: theory and algorithms*. John Wiley & Sons, 2011.

[9] Duncan J Watts and Steven H Strogatz. “Collective dynamics of ‘small-world’networks”. In: *nature* 393.6684 (1998), pp. 440–442.

[10] Márton Pósfai and Albert-László Barabási. *Network science*. Citeseer, 2016.

[11] Maksim Kitsak, Fragkiskos Papadopoulos, and Dmitri Krioukov. “Latent geometry of bipartite networks”. In: *Physical Review E* 95.3 (2017), p. 032309.

[12] Peng Zhang et al. “Clustering coefficient and community structure of bipartite networks”. In: *Physica A: Statistical Mechanics and its Applications* 387.27 (2008), pp. 6869–6875.

[13] Stanley Milgram. “The small world problem”. In: *Psychology today* 2.1 (1967), pp. 60–67.

[14] Ithiel de Sola Pool and Manfred Kochen. “Contacts and influence”. In: *Social networks* 1.1 (1978), pp. 5–51.

[15] Jeffrey Travers and Stanley Milgram. “An experimental study of the small world problem”. In: *Social networks*. Elsevier, 1977, pp. 179–197.

[16] Paul Erdos, Alfréd Rényi, et al. “On the evolution of random graphs”. In: *Publ. math. inst. hung. acad. sci* 5.1 (1960), pp. 17–60.

[17] Erzsébet Ravasz Regan. “Networks: Structure and dynamics”. In: *Statistical and Nonlinear Physics*. Springer, 2022, pp. 575–597.

[18] Alain Barrat and Martin Weigt. “On the properties of small-world network models”. In: *The European Physical Journal B-Condensed Matter and Complex Systems* 13 (2000), pp. 547–560.

[19] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *science* 286.5439 (1999), pp. 509–512.

[20] Stefan Bornholdt and Holger Ebel. “World Wide Web scaling exponent from Simon’s 1955 model”. In: *Physical Review E* 64.3 (2001), p. 035104.

[21] Ethan P White, Brian J Enquist, and Jessica L Green. “On estimating the exponent of power-law frequency distributions”. In: *Ecology* 89.4 (2008), pp. 905–912.

[22] Sergei Maslov and Kim Sneppen. “Specificity and stability in topology of protein networks”. In: *Science* 296.5569 (2002), pp. 910–913.

[23] P ERDdS and A R&wi. “On random graphs I”. In: *Publ. math. debrecen* 6.290-297 (1959), p. 18.

[24] Edgar N Gilbert. “Random graphs”. In: *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[25] Mark EJ Newman. “The structure and function of networks”. In: *Computer Physics Communications* 147.1-2 (2002), pp. 40–45.

[26] David Chandler. “Introduction to modern statistical”. In: *Mechanics. Oxford University Press, Oxford, UK* 5.449 (1987), p. 11.

[27] Reuven Cohen and Shlomo Havlin. “Scale-free networks are ultrasmall”. In: *Physical review letters* 90.5 (2003), p. 058701.

[28] Béla Bollobás* and Oliver Riordan. “The diameter of a scale-free random graph”. In: *Combinatorica* 24.1 (2004), pp. 5–34.

[29] Konstantin Klemm and Victor M Eguiluz. “Growing scale-free networks with small-world behavior”. In: *Physical Review E* 65.5 (2002), p. 057102.

[30] Béla Bollobás and Oliver M Riordan. “Mathematical results on scale-free random graphs”. In: *Handbook of graphs and networks: from the genome to the internet* (2003), pp. 1–34.

[31] Chai-Keong Toh. *Wireless ATM and ad-hoc networks: Protocols and architectures*. Springer Science & Business Media, 2012.

[32] Jesper Dall and Michael Christensen. “Random geometric graphs”. In: *Physical review E* 66.1 (2002), p. 016121.

[33] Frank A Haight. “Handbook of the Poisson distribution”. In: *(No Title)* (1967).

[34] Maksim Kitsak, Ivan Voitalov, and Dmitri Krioukov. “Link prediction with hyperbolic geometry”. In: *Physical Review Research* 2.4 (2020), p. 043113.

[35] Ivan Voitalov et al. “Geohyperbolic routing and addressing schemes”. In: *ACM SIGCOMM Computer Communication Review* 47.3 (2017), pp. 11–18.

[36] Marian Boguna, Dmitri Krioukov, and Kimberly C Claffy. “Navigability of complex networks”. In: *Nature Physics* 5.1 (2009), pp. 74–80.

[37] Marian Boguna et al. “Network geometry”. In: *Nature Reviews Physics* 3.2 (2021), pp. 114–135.

[38] James W Cannon et al. “Hyperbolic geometry”. In: *Flavors of geometry* 31.59-115 (1997), p. 2.

[39] Dmitri Krioukov et al. “Hyperbolic geometry of complex networks”. In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 82.3 (2010), p. 036106.

[40] M Ángeles Serrano, Dmitri Krioukov, and Marián Boguná. “Self-similarity of complex networks and hidden metric spaces”. In: *Physical review letters* 100.7 (2008), p. 078701.

[41] Dmitri Krioukov et al. “Curvature and temperature of complex networks”. In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 80.3 (2009), p. 035101.

[42] Fragkiskos Papadopoulos et al. “Popularity versus similarity in growing networks”. In: *Nature* 489.7417 (2012), pp. 537–540.

[43] Miller McPherson, Lynn Smith-Lovin, and James M Cook. “Birds of a feather: Homophily in social networks”. In: *Annual review of sociology* 27.1 (2001), pp. 415–444.

[44] Pitirim A Sorokin. *Social mobility*. Harper & Brothers, 1927.

[45] Linton C Freeman. “Some antecedents of social network analysis”. In: *Connections* 19.1 (1996), pp. 39–42.

[46] Mark EJ Newman. “Modularity and community structure in networks”. In: *Proceedings of the national academy of sciences* 103.23 (2006), pp. 8577–8582.

[47] Erzsébet Ravasz et al. “Hierarchical organization of modularity in metabolic networks”. In: *science* 297.5586 (2002), pp. 1551–1555.

[48] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. “Semi-supervised learning using gaussian fields and harmonic functions”. In: *Proceedings of the 20th International conference on Machine learning (ICML-03)*. 2003, pp. 912–919.

[49] David Liben-Nowell and Jon Kleinberg. “The link prediction problem for social networks”. In: *Proceedings of the twelfth international conference on Information and knowledge management*. 2003, pp. 556–559.

[50] Tao Zhou, Linyuan Lü, and Yi-Cheng Zhang. “Predicting missing links via local information”. In: *The European Physical Journal B* 71 (2009), pp. 623–630.

[51] Aili Asikainen et al. “Cumulative effects of triadic closure and homophily in social networks”. In: *Science Advances* 6.19 (2020), eaax7310.

[52] Katja Luck et al. “A reference map of the human binary protein interactome”. In: *Nature* 580.7803 (2020), pp. 402–408.

[53] Chris Stark et al. “BioGRID: a general repository for interaction datasets”. In: *Nucleic acids research* 34.suppl_1 (2006), pp. D535–D539.

[54] Sven-Volker Rehm and Lakshmi Goel. “Using information systems to achieve complementarity in SME innovation networks”. In: *Information & Management* 54.4 (2017), pp. 438–451.

[55] GJA Budel. “Complementarity and Similarity in Complex Networks”. In: (2024).

A

Appendix

In this Appendix, we provide the pseudo-code for generating the Random Hyperbolic Graph (RHG) model, Complementarity Random Hyperbolic Graph (CRHG) model, and Generalized Complementarity Random Hyperbolic Graph (GCRHG) model.

The RHG model algorithm:

Algorithm 1 Random Hyperbolic Graph (RHG) model

Require: N, $Tor\left(\frac{1}{\beta}\right), \gamma, \zeta = 1, <$ k $>$

$D_{i,j} \leftarrow$ **def:**(node i(r_i, θ_i), node j(r_j, θ_j), $\zeta = 1$) {Calculate Hyperbolic Distance between node i and node j}

$R \leftarrow$ **def:**(N, $<$ k $>, \beta, \zeta = 1, \gamma$) {Construct the H^2 disk of reference R based on different T conditions}

Calculate R based on N, $<$ k $>, \beta, \zeta = 1, \gamma$
Initialize empty H^2_node node list
for $i = 0$ to N **do**
  Generate $r[i]$ and $\theta[i]$ for each node
  Append $(r[i], \theta[i])$ to H^2_node
end for
for each pair (i, j) with $j > i$ **do**
  Compute $D_{i,j}$ between nodes i and j
  {To place N nodes i.i.d.ly in $H2$ disk}
  if $T \neq 0$ **then**
    Compute edge connecting probability P
    if $random.uniform(0, 1) \leq P$ **then**
      Add edge (i, j) to file(.dat)
    end if
  else
    if $D_{i,j} \leq R$ **then**
      Add edge (i, j) to file(.dat)
    end if
  end if
end for

The CRHG model algorithm:

Algorithm 2 Complementarity Random Hyperbolic Graph (CRHG) model

Require: N, $T \neq 0 or \left(\frac{1}{\beta}\right) \neq 0, \gamma, \zeta = 1 <$ k $>$

$D_{i,j} \leftarrow$ **def:**(node i(r_i, θ_i), node j(r_j, θ_j), $\zeta = 1$) {Calculate Hyperbolic Distance between node i and node j}

$R \leftarrow$ **def:**(N, $<$ k $>, \beta, \zeta = 1, \gamma$) {Construct the H^2 disk of reference R based on different T conditions}

Calculate R based on N, $<$ k $>, \beta, \zeta = 1, \gamma$

Initialize empty H^2_node node list

for $i = 0$ to N **do**

Generate ($r[i_1]$,$\theta[i_1]$) and ($r[i_2]$,$\theta[i_2]$) for each node {Set two-point position for one node}

Append ($r[i_1]$,$\theta[i_1]$) and ($r[i_2]$,$\theta[i_2]$) to H^2_node

end for

for each pair (i, j) with $j > i$ **do**

Compute D_{i_1,j_2} between point i_1 and j_2

Compute D_{i_2,j_1} between point i_2 and j_1

{To place N nodes i.i.d.ly in H^2 disk}

Compute edge connecting probability P {Depend on D_{i_1,j_2} and D_{i_2,j_1}}

if $random() \leq P$ **then**

Add edge (i, j) to file(.dat)

end if

end for

The GCRHG model algorithm:

Algorithm 3 Generalized Complementarity Random Hyperbolic Graph (GCRHG) model

Require: N, $T \neq 0 or \left(\frac{1}{\beta}\right) \neq 0, \gamma, \zeta = 1 <$ k $>, f$

$D_{i,j} \leftarrow$ **def:**(node i(r_i, θ_i), node j(r_j, θ_j), $\zeta = 1$) {Calculate Hyperbolic Distance between node i and node j}

$R \leftarrow$ **def:**(N, $<$ k $>, \beta, \zeta = 1, \gamma$) {Construct the H^2 disk of reference R based on different T conditions}

Calculate R based on N, $<$ k $>, \beta, \zeta = 1, \gamma$

Initialize empty H^2_node node list

for $i = 0$ to N **do**

  Generate ($r[i_1]$,$\theta[i_1]$) and ($r[i_2]$,$\theta[i_2]$) for each node {Set two-point position for one node}

  if $random.uniform(0, 1) \leq f$ **then**

    Append ($r[i_1]$,$\theta[i_1]$) and ($r[i_2]$,$\theta[i_2]$) to H^2_node

  else

    Append ($r[i_1]$,$\theta[i_1]$) and ($r[i_1]$,$\theta[i_1]$) to H^2_node

  end if

end for

for each pair (i, j) with $j > i$ **do**

  if $i_1 == i_2$ and $j_1 == j_2$ **then**

    Compute $D_{i,j}$ between nodes i_1 and j_1

    Compute edge connecting probability P

    if $random.uniform(0, 1) \leq P$ **then**

      Add edge (i, j) to file(.dat)

    end if

  else

    Compute D_{i_1,j_2} between point i_1 and j_2

    Compute D_{i_2,j_1} between point i_2 and j_1

    {To place N nodes i.i.d.ly in H^2 disk}

    Compute edge connecting probability P {Depend on D_{i_1,j_2} and D_{i_2,j_1}}

    if $random.uniform(0, 1) \leq P$ **then**

      Add edge (i, j) to file(.dat)

    end if

  end if

end for