博士学位論文

Network analysis of medical knowledge to investigate the structure of medical service: Toward a mathematical approach to medical service

(医療構造を調べるための医療知識のネットワーク分析: 医療に対する数学的アプローチを目指して)

日月　　裕

北海道大学大学院理学院
数学専攻

2015 年 6 月

Summary

Medical knowledge is commonly described by natural language. Since 2001, a complex network analysis has been applied to natural language. This analysis found that the language network has small-world and scale-free properties. Here, we applied this natural language analysis to medical knowledge and found similarities between the structure of medical knowledge and that of clinical practice.

We analyzed “Harrison's Principles of Internal Medicine, 15th Edition”, which is a major textbook in internal medicine. To determine the structure of medical knowledge rather than language, we confined the subjects of study only to medical terms. We then constructed a medical knowledge network (MKN) as follows. First, we defined medical terms as the nodes of the network. Then, we defined edges that mutually connect a pair of terms in one sentence.

We calculated the average path length and the average clustering coefficient of the MKN. The average path length was 4.317, and the average clustering coefficient was 0.86, the implication being that the MKN had nearly the same average path length and a far larger average clustering coefficient compared with the corresponding random graph. These findings suggest that the MKN has the small-world property.

We also found that the degree distribution of the MKN exhibited a power law with a fast decaying tail. This finding indicates that the MKN is a truncated scale-free network. The exponent was 2.045, which is consistent with many other complex networks.

We also investigated the hierarchical structure of the MKN. A network model that produces a network with a hierarchical structure was recently proposed. According to this model, a network having the hierarchical structure has the following two features: the average clustering coefficient is independent of the size of the network, and the average of the clustering coefficients of nodes with k edges follows the scaling law $C(k) \sim k^{-1}$. Our analysis of the MKN determined that it has these two features.

As described above, the MKN displays small-world, scale-free, and hierarchical properties. The

small-world property may help a clinician make the most appropriate diagnosis. In contrast, scale-free and hierarchical properties are related to the mechanism generating the MKN. The MKN would develop and evolve under the following two principles: preferential attachment and preservation of the preexisting hierarchical structure. These two principles may give the MKN its precise structure.

We next considered whether clinical practice has properties similar to the MKN. To answer this question, we applied network analysis to a database of "disease" in a hospital information system, which should reflect the clinical behavior of doctors. This database record contained several items in addition to "diagnosis", such as "patient ID", "department code", and "doctor ID". Therefore, by assigning these items to nodes and mutually connecting all nodes in each record by edges, we constructed the diagnosis database network (DDN). We then applied network analysis to the DDN, and we found that the DDN also has small-world, scale-free, and hierarchical features.

Moreover, we found that both the diagnosis frequency distribution in the hospital and the diagnosis degree distribution of the medical knowledge network obey a similar power law. It used to be thought that diagnosis frequency is an objective index existing in the world; however, given the similarity between the diagnosis frequency of the hospital and the diagnosis degree of the MKN, it could be that diagnosis frequency is influenced by medical knowledge. Medical knowledge influences clinical practice, and this practice influences the frequency of diagnosis. Because doctors diagnose through clinical practice, clinical practice is thought to be the observation of disease. Therefore, the fact that clinical practice influences diagnosis frequency implies that diagnosis frequency is not an objective index but, to a certain degree, a subjective index, the value of which varies somewhat with the observation of disease.

These findings also have several mathematical implications:

1. The network derived from medical knowledge (MKN) cannot be expressed by Erdős-Rényi random graphs.

2. We require the concept of a complex network to investigate this network.

3. To express this network, we require a new mathematical model.

We consider that medical service is a complex system composed of multiple interconnected elements. Examples of these elements are patients, medical staff (e.g., doctors), medical knowledge, and the results of clinical practice (medical results). This study shows that we can mathematically express medical knowledge (MKN) and medical results (DDN) by constructing networks. Mathematically expressing medical knowledge and medical results leads to the possibility of mathematically analyzing the complex system of medical service. These mathematical analyses can be called “medical complex systemology”.

Keywords: small world; scale-free; complex network; natural language; medical knowledge

CONTENTS

1. Introduction[1]

Medical knowledge is extremely complicated. The number of diagnoses alone amounts to tens of thousands. Other than diagnoses, an enormous quantity of knowledge is involved, such as symptoms, results of clinical tests, pathological knowledge, and anatomical knowledge. Clinicians must make appropriate diagnoses on the basis of such complicated knowledge. How does a clinician select the appropriate diagnosis? One hypothesis is that the structure of medical knowledge itself helps a clinician diagnose. An approach to understanding the structure of medical knowledge is to classify that knowledge by using some predefined criteria, such as the International Classification of Diseases (ICD) [2]. If the purpose of using the classification is appropriate for the criteria, this approach is very useful. However, the criteria do not always fit actual clinical practice.

Complex networks have been studied extensively to determine the structure of many real systems such as the World Wide Web (WWW), the Internet, and biological and social networks. These complex networks are frequently small-world and scale-free [3–7]. Small-world networks have a large clustering coefficient and a small average path length. Scale-free networks are characterized by a power-law decay of the degree distribution $p(k) \sim k^{-\alpha}$ (see Sec. 2.5). The hierarchical organization of these complex networks has also recently been investigated [8, 9].

Medical knowledge is commonly described by natural language. Unlike classification by criteria, many properties of diseases are freely described. In description by natural language, various components of this medical knowledge are mutually connected in context, and a network is produced. This network may, in turn, influence clinical practice. However, the structure of this network and its influence on clinical practice are virtually unknown. Since 2001, a complex network analysis has been applied to natural language [10, 11]. This analysis found that the language network has small-world and scale-free properties [12, 13]. Here, we applied this natural language analysis to medical knowledge and constructed a medical knowledge network (MKN).

[1] This work is based on the article [1] and the main parts of several sections (Secs.1, 4, 5, and 6) are quoted from [1].

Next, we determined that the MKN has small-world and scale-free properties. Moreover, we found similarities between the structure of medical knowledge and that of clinical practice.

These findings also have several mathematical implications. The network structure of medical knowledge found in this paper cannot be expressed by Erdős-Rényi random graphs [14–16]. Although the random graph has been the most studied model in graph theory, it is insufficient for analyzing the medical knowledge network. The MKN would be classified as a complex network, but the concept of the complex network has not been established in mathematics. We consider that it is necessary to establish the mathematical concept of the complex network and construct a new mathematical framework to investigate the medical knowledge network.

2. Review of network theory and complex network

In this paper, medical knowledge is investigated by using network theory. Thus, we review the terminology of network theory and some of its significant results necessary for understanding the paper [17–19].

Unless otherwise noted, the following symbols are used in the paper.

n: the order of a graph (i.e., the total number of nodes).

M: the size of a graph (i.e., the total number of edges).

k_i: the degree of node i.

$\langle k \rangle$: the average degree of a graph.

$p(k)$: the degree distribution of a graph.

P_{ik}: a path between nodes i and j.

d_G: the diameter of graph G.

l_{ij}: the shortest path length between nodes i and j.

$\langle l \rangle$: the average path length of a graph.

$C(i)$: the cluster coefficient of node i.

$\langle \mathrm{C} \rangle$: the average clustering coefficient of a graph.

2.1 Graph (Network)[2]

Definition (Undirected Graph) A graph (or a network) is defined by a pair of set $G \equiv (V, E)$, where $V \equiv \{v_1, v, \dots, v_n\}$ is a set of nodes (or vertices), and $E \equiv \{e_1, e_2, \dots, e_M\}$ is a set of edges (or links). Edges are unordered pairs of nodes (v_i, v_j).

The number of nodes n is called the order of the graph, and the number of edges M is called the size of the graph. A node has a unique label. Usually, we label the nodes with integer

[2] In this paper, we use both "network" and "graph" in exactly the same sense. The word "graph" is used mainly in mathematical fields, while "network" is used mainly in reference to real world networks and their applications.

labels $1, 2, \ldots, n$. When there is no misunderstanding, we identify the node with its label, and we can simply call a node i instead of v_i. The edge $e_k = (i, j)$ connects nodes i and j. In an undirected graph, the edges (i, j) and (j, i) are the same.

Figure 2-1 is an illustration of a graph in which the nodes and the edges are respectively expressed by circles and lines. In practical applications, a node indicates a concrete subject such as a person, the World Wide Web, or a word, and an edge indicates a concrete relationship between two nodes such as a friendship, a link, or a context, respectively.

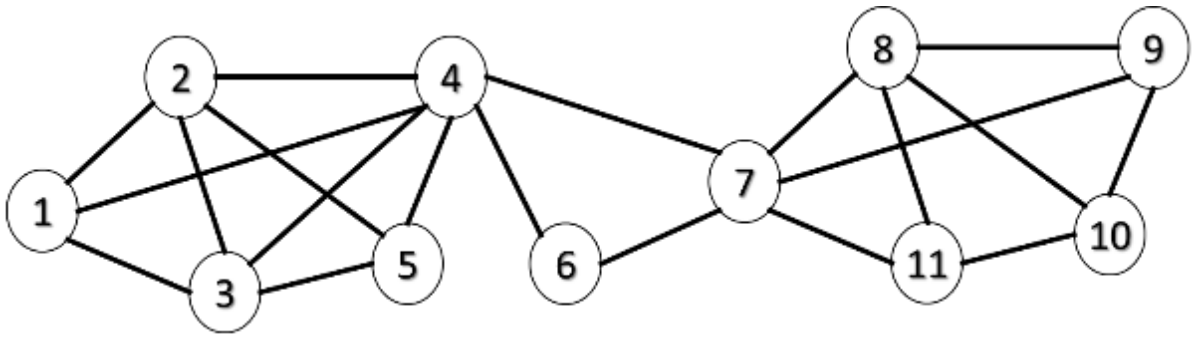


Figure 2-1. Illustration of a graph.

Definition (Subgraph) A graph $G' \equiv (V', E')$ is said to be a subgraph of a graph $G = (V, E)$ if all the nodes in V' belong to V and all the edges in E' belong to E, i.e., $E' \subset E$ and $V' \subset V$.

Definition (Path) A path $P_{i_0\, i_l}$ in a graph $G = (V, E)$ is an ordered collection of $l + 1$ nodes $V_P \equiv \{i_0, i_1, \ldots, i_l\}$ and l edges $E_P \equiv \{(i_0, i_1), (i_1, i_2), \ldots, (i_{l-1}, i_l)\}$, where $i_k \in V$ and $(i_{k-1}, i_k) \in E$, for all k. The path $P_{i_0\, i_l}$ connects nodes i_0 and i_l.

A closed path is a path $P_{i_0\, i_l}$ where $i_0 = i_l$ and $i_k \neq i_p$ for $k \neq p, 0 < k, p < l$.

Definition (Connectivity) A graph is said to be connected if there exists a path connecting any two nodes in the graph.

Examples of graphs (Fig. 2-2)

A *complete graph* is a simple undirected graph in which every pair of distinct nodes is connected by a unique edge. A complete graph therefore has $n(n-1)/2$ edges.

A *lattice graph* is a graph whose drawing is embedded in some Euclidean space and forms a regular tiling.

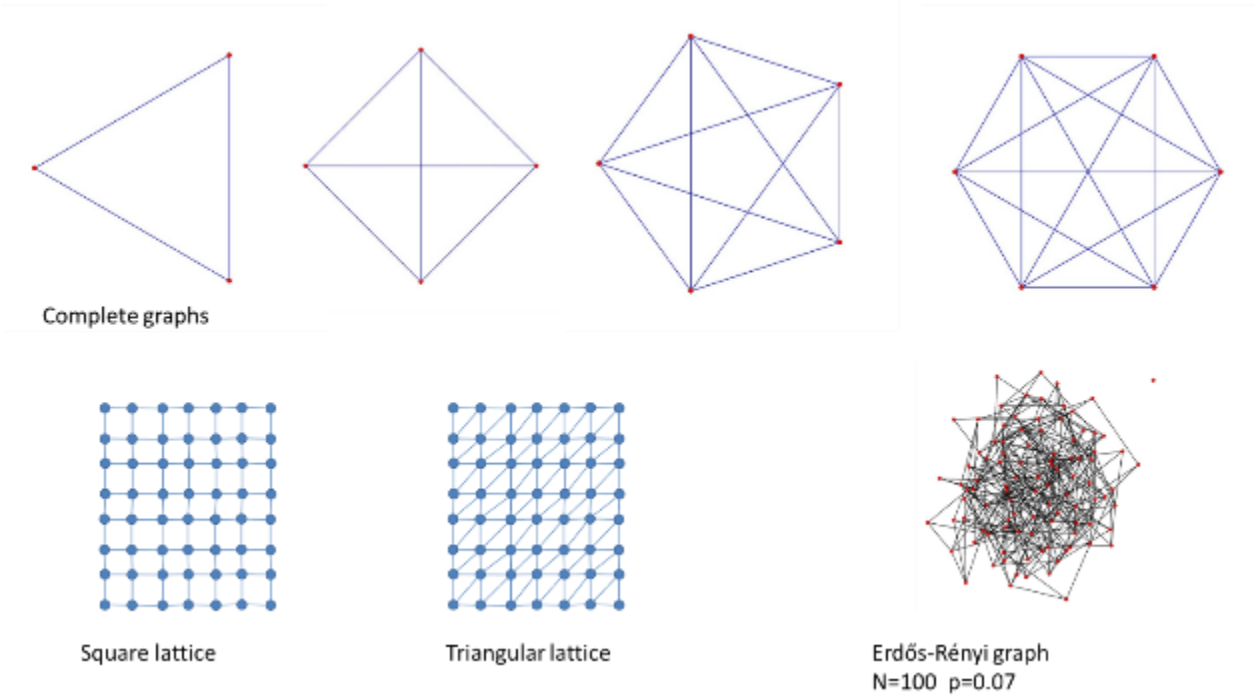


Figure 2-2. Examples of graphs drawn by the network analysis software program Pajek [21].

A *tree* is a connected graph that has no closed path.

A *random network* is a network generated by a random process. Although there exist many random networks, the Erdős-Rényi graph $G_{n,p}$ is the most common [14–16, 20]. The Erdős-Rényi (ER) graph (or the Edgar-Gilbert graph) is a graph with n nodes in which each of the $n(n-1)/2$ possible edges is present with probability p (the connection probability**)** and absent with probability $1-p$. A *random graph* refers almost exclusively to an Erdős-Rényi graph.

Examples of real-world networks

An *actor network* is a network whose nodes are actors. Two actors are connected by an edge if they played together in a film.

A *language network* is a network whose nodes are words. Two words can be connected in the following ways [10–13, 22].

1) Co-occurrence networks: Two words are connected if they appear together within at least one sentence.

2) Syntactic networks: These are based on constituent structures that depend on the language grammar.

3) Semantic networks: Starting from individual words that lexicalize concepts, these are built by mapping out basic semantic relations such as hypernym-hyponym relations ("flower" is a hypernym of "rose"), part-whole, or binary opposition (antonym).

Although the structures of these three networks are considered to be very different, each of them exhibits *scale-free* and *small-world* properties (see Sections 3.3 and 3.4). In this paper, we construct a medical knowledge network by a method similar to a co-occurrence network.

World Wide Web (WWW) network: There are many Internet websites around the world, and each website has any number of documents and links to other websites. The World Wide Web network consists of these websites (nodes) and their links (edges) [23].

The *Internet at the router level*: The Internet consists of a huge number of routers and communication cables, each of which connects several other routers. This is a physical network.

A *power grid* is a network that consists of electric power plants and power lines. This is a physical network.

2.2 Indexes characterizing graphs

Definition (Path length) The path length is the number of edges in the path. The shortest path length between nodes *i* and *j* (expressed as l_{ij}) is called the distance between *i* and *j*.

The diameter d_G of a graph *G* is defined as

$$d_G \equiv \max_{i,j} l_{ij} \quad (G \text{ is connected})$$

$$\equiv \infty \quad (G \text{ is not connected}) . \qquad (2-1)$$

The average shortest path length $\langle l \rangle$ is defined as the average value of l_{ij} over all possible pairs of nodes in a graph:

$$\langle l \rangle \equiv \frac{1}{n(n-1)} \sum_{i,j} l_{ij} . \qquad (2-2)$$

By definition $\langle l \rangle \leq d_G$, and in the case of a well-behaved and bounded shortest path length distribution, it is possible to show heuristically that in many cases $\langle l \rangle$ and d_G behave in the same way with the network order *n* [18]. In general, $\langle l \rangle$ increases with the graph order. However, $\langle l \rangle$ is not so large compared with the graph order *n* in many real-world networks. When $\langle l \rangle$ grows only logarithmically with *n*, the network is said to be *small*.

Definition (Degree) The degree k_i of node *i* is defined as the number of edges in the graph leaving from *i*. The k_i nodes connected to node *i* is said to be the neighbors of *i*.

Definition (Degree distribution) The degree distribution $p(k)$ of a graph is defined as the probability that a node has degree *k*. That is,

$$p(k) \equiv \frac{\text{the number of nodes with degree } k}{n} . \qquad (2-3)$$

The average degree of a graph is defined as the average value of *k* over all the nodes in the graph:

$$\langle k \rangle \equiv \frac{1}{n} \sum_{i=1}^{n} k_i = \sum_{k=0}^{\infty} k p(k) . \qquad (2-4)$$

For a random graph, $\langle k \rangle = (n-1)p \sim np$. In real-world significant networks, $\langle k \rangle$ is not very

large, but not so small that the network is disconnected. Hence, we usually suppose $n \gg \langle k \rangle \gg \log n \gg 1$, where $\langle k \rangle \gg \log n$ guarantees that a random graph will be connected (see theorem 2.1) [3, 14].

Definition (Clustering coefficient) The clustering coefficient $C(i)$ of node i is defined as the ratio of the number of edges within the neighbors of i to the possible number of such edges. If the degree of node i is k_i and the neighbors of i have e_i edges among them, we have

$$C(i) \equiv \frac{e_i}{k_i(k_i-1)/2} \qquad (k_i > 1)$$
$$\equiv 0 \qquad (k_i \leq 1). \qquad (2-5)$$

Clustering implies the property that if node i is connected to node j and at the same time i is connected to l, then j is also connected to l with a high probability. Moreover, a large $C(i)$ implies that the periphery of node i is dense with edges.

The average clustering coefficient of a graph is simply given by

$$\langle C \rangle \equiv \frac{1}{n}\sum_{i=1}^{n} C(i). \qquad (2-6)$$

Since $0 \leq C(\mathrm{i}) \leq 1$, then $0 \leq \langle C \rangle \leq 1$.

An Erdős-Rényi graph has a small $\langle C \rangle$, or $\langle C \rangle \to 0$ when $n \to \infty$. From this, a large $\langle C \rangle$ implies that there exists deviation of the edge density in a network. Moreover, the average clustering coefficient is a measure of the intrinsic potential modularity of the network, for example, of a metabolic network [8].

Properties of various networks

Table 2-1 shows the properties of several graphs [18, 19]. In summary, lattices have a large average shortest path length (or a large diameter), and some lattices have a finite positive clustering coefficient, regardless of the order of the graph. On the contrary, a random graph (Erdős-Rényi

graph) has a small average path length, and its clustering coefficient approaches zero with increasing n. Compared with these properties, a real-world network has a small average path length and a large clustering coefficient. Therefore, we cannot completely investigate the real-world network by using these simple network models.

Table 2-1. Properties of some graphs.

	Complete graph	Lattice		Erdős-Rényi graph $G_{n,p}$
		square lattice	triangular lattice	(when n is large)
k_i or $\langle k \rangle$	$n-1$	4	6	np
$\langle l \rangle$	1	$\approx \sqrt{n}$		$\approx \frac{\log n}{\log\langle k \rangle}$ *
$\langle C \rangle$	1	0	2/5	$\approx \frac{\langle k \rangle}{n}$ *
$p(k)$	$\delta_{k,N-1}$	$\delta_{k,4}$	$\delta_{k,6}$	$\approx \frac{e^{-\lambda}\lambda^k}{k!}$ $(\lambda = np)$**

* see Sec. 2.3

Given two functions $f(n)$ and $g(n)$, when $\lim_{n\to\infty} \frac{f(n)}{g(n)} = A$, where A is a finite positive value, we denote $f(n) \approx g(n)$. When $\lim_{N\to\infty} \frac{f(n)}{g(n)} = 0$, we denote $f(n) \preccurlyeq g(n)$.

**For a random graph $G_{n,p}$, $p(k) = \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k}$. When n is large and p is small, whereas the product $\lambda = np$ is of moderate magnitude, $p(k)$ is approximated by a Poisson distribution, i.e., $p(k) \approx \frac{e^{-\lambda}\lambda^k}{k!}$.

2.3 Random graph model

The random graph model is a concept introduced by Erdős and Rényi [14–16, 20]. The model $\mathcal{G}(n,p)$ is the set of all random graphs $G_{n,p}$ with node set $V = \{1, 2, \ldots, n\}$ in which the edges are chosen independently and with probability p. Therefore, for any graph $G_{n,p}$ in $\mathcal{G}(n,p)$, we can consider the probability $P(\{G_{n,p}\})$: if $G_{n,p}$ has m edges ($0 \le m \le M$), then

$$P(\{G_{n,p}\}) = p^m q^{M-m} \quad \text{where } M = \binom{n}{2} \text{ and } q = p - 1 .$$

Therefore, if A is a subset of $\mathcal{G}(n,p)$, then $P(A) = \sum_{G_{n,p} \in A} P(\{G_{n,p}\})$.

If Q is a property that a graph may or may not possess, the probability $P(Q)$ that the random graph possesses the property Q is defined as $P(Q) = \mathrm{P}(\{G_{n,p} \in \mathcal{G}(n,p): G_{n,p}$ possesses the property $Q\})$. For instance, $P\{G_{n,p}$ is connected$\} = P(\{G_{n,p}: G_{n,p}$ is connected$\})$. This probability varies depending on the function $p(n)$. Erdős and Rényi studied the evolution of random graphs [15]. They made an issue of "typical" properties, i.e., such properties for which the probability tends to 1 if $n \to \infty$ when $p = p(n)$. In addition, almost every (a.e.) graph in $\mathcal{G}(n,p)$ is said to have the property Q if $\lim_{n\to\infty} P(Q) = 1$ when $p = p(n)$.

There are many properties of random graphs. The following theorems are important in this study.

Theorem 2.1[3] (Erdös and Rényi, 1959 [14] and Bollobás, 2001 [20])

Let c be a fixed real number and let $p = \frac{\{\log n + c + o(1)\}}{n}$, *then*

$$P\{G_{n,p} \text{ is connected}\} \xrightarrow{n\to\infty} e^{-e^{-c}}.$$

If $p_1 > p_2$, then $P\{G_{n,p_1}$ is connected$\} \geq P\{G_{n,p_2}$ is connected$\}$. Let $p' = (1+\varepsilon)\log n/n$ and let $c = \varepsilon \log n_c$ for an arbitrary constant c. Since $p' = \frac{(1+\varepsilon)\log n}{n} > \frac{\log n + c + o(1)}{n} = p$ for $n > n_c$, then

$$P\{G_{n,p'} \text{ is connected}\} \geq P\{G_{n,p} \text{ is connected}\} \xrightarrow{n\to\infty} e^{-e^{-c}}.$$

Therefore, $P\{G_{n,p'}$ is connected$\} \xrightarrow{n\to\infty} 1$. Namely, when $p = (1+\varepsilon)\log n/n$, almost every graph is connected. From these considerations, we can assume that the network is connected for large n if $p > \frac{\log n}{n}$. Using the average degree $\langle k \rangle$, if $\langle k \rangle > \log n$, then the network is assumed to be connected.

The diameter is an important index in various applications of network theory. There are many papers investigating the diameter of random graphs [24–27]. Bollobás [24] showed that the

[3] Given two functions $f(n)$ and $g(n)$, when $\lim_{n\to\infty} \frac{f(n)}{g(n)} = A$, where A is a finite positive value, we denote $f(n) = O(g(n))$ or $f(n) \approx g(n)$, and when A = 0, we denote $f(n) = o(g(n))$.

diameter of a.e. graph in $\mathcal{G}(n,p)$ is concentrated on at most two values if $\frac{np}{\log n} \to \infty$:

$$\left\lfloor \frac{\log n + \log\log n}{\log\langle k\rangle} \right\rfloor \leq d_{G_{n,p}} \leq \left\lceil \frac{\log n + \log\log n + 1}{\log\langle k\rangle} \right\rceil$$ [4].

However, the condition that $\frac{np}{\log n} \to \infty$ is seldom satisfied in real world networks. On the contrary, a sparse graph ($np \leq \mathrm{c}\log n$ for some constant $c > 1$) is common in the real world. Theorem 2.1 implies that if $np > \log n$, then the network can be assumed to be connected. For a network that is sparse and connected ($\log n < np < c\log n$), Chung et al. proved the following three theorems [27].

Theorem 2.2 (Chung and Lu, 2001 [27])

If $np \geq c\log n$ *for some constant* $c > 8$, *the diameter of almost every random graph* $G_{n,p}$ *is concentrated on at most two values at* $\frac{\log n}{\log np} = \frac{\log n}{\log\langle k\rangle}$.

Theorem 2.3 (Chung and Lu, 2001 [27])

If $np \geq c\log n$ *for some constant* $c > 2$, *the diameter of almost every random graph* $G_{n,p}$ *is concentrated on at most three values at* $\frac{\log n}{\log np} = \frac{\log n}{\log\langle k\rangle}$.

Theorem 2.4 (Chung and Lu, 2001 [27])

If $np \geq c\log n$ *for some positive constant c,*

$$\left\lfloor \frac{\log\left(\frac{cn}{11}\right)}{\log np} \right\rfloor \leq d_{G_{n,p}} \leq \left\lceil \frac{\log\left(\frac{33c^2}{400} n\log n\right)}{\log np} \right\rceil + 2\left\lfloor \frac{1}{c} \right\rfloor + 2 .$$

The diameter of almost every random graph $G_{n,p}$ *is concentrated on at most* $2\left\lfloor \frac{1}{c} \right\rfloor + 2$ *values.*

---

[4] $\lfloor x \rfloor$: the floor of $x = \max\{m \in \mathbb{Z} \mid m \leq x\}$, $\lceil x \rceil$: the ceiling of $x = \min\{n \in \mathbb{Z} \mid n \geq x\}$.

Ultimately, when $np \geq \log n$, the diameter of $G_{n,p}$ is near $\frac{\log n}{\log np} = \frac{\log n}{\log\langle k\rangle}$.

Average path lengths are more frequently used than diameters in many applications. Newman [28] described that the average path length is approximately equal to the diameter. Therefore, the average path length $\langle l\rangle$ is also near $\frac{\log n}{\log np} = \frac{\log n}{\log\langle k\rangle}$. All the networks constructed in this study satisfy the condition that $\log n \leq np \leq 2\log n$. Therefore, if these networks can be expressed by random graphs, then their average path lengths should be near $\frac{\log n}{\log\langle k\rangle}$.

The clustering coefficient $C(i)$ and the average clustering coefficient $\langle C\rangle$ of the random graph are random variables in the probability space $\mathcal{G}(n,p)$. Therefore, we can define their expectations, i.e., $E(C(i))$ and $E(\langle C\rangle)$, in the probability space $\mathcal{G}(n,p)$. We use the following theorem in this paper [17].

Theorem 2.5 (Konnno and Ide, 2008 [17])

Let $C(i)$ be the cluster coefficient of any node i of a random graph $G_{n,p}$ and let $E(C(i))$ be the expectation of $C(i)$, then

$$E\big(C(i)\big) = p[1-(1-p)^{n-1}\{1+(n-2)p\}]\,.$$

Because $E(C(i)) \xrightarrow{n\to\infty} p$, the expectation of the average clustering coefficient $E(\langle C\rangle) \xrightarrow{n\to\infty} p$. If a real network having n nodes and M edges is expressed by a random graph model, then its average clustering coefficient $E(\langle C\rangle) \sim p = \langle k\rangle/n = 2M/n(n-1)$.

2.4 Small-world

When a network has a large average clustering coefficient $\langle C\rangle$ and a small average shortest path length $\langle l\rangle$, it is called a "small-world" network. The large $\langle C\rangle$ means that $\langle C\rangle \to A(>0)$, when $n \to \infty$, and the small $\langle l\rangle$ means that $\langle l\rangle$ grows only logarithmically with n. Given the impossibility of changing n in a real-world network, we compare $\langle C\rangle$ and $\langle l\rangle$ to those of the ER graph with the same

numbers of nodes n and edges M, i.e., the connection probability $p = 2M/n(n-1)$. When $\langle l \rangle \leq \langle l \rangle_{ER}$ and $\langle C \rangle \gg \langle C \rangle_{ER}$, we determine that the network is a small-world network. Many real-world networks are small-world networks such as the WWW, actor networks, and language networks.

In a small-world network, every node can be reached within a few steps. When a network is constructed for communication, this property makes information rapidly pass throughout the network. If a sexual-contact network has the small-world property, a virus like HIV that causes an infectious disease can spread quickly all over the world [29].

Watt and Strogatz found that networks that are not small-world networks, such as lattice networks, can be made to have the small-world property by adding a few shortcut edges connecting two randomly selected nodes to the network (Fig. 2-3) [3]. Previously, disease infections were usually limited to a geographical neighborhood; nowadays they can quickly spread to distant people all over the world owing to express transportation systems such as airplanes, cars, and high-speed railways. These express transportation systems correspond to shortcuts in the network.

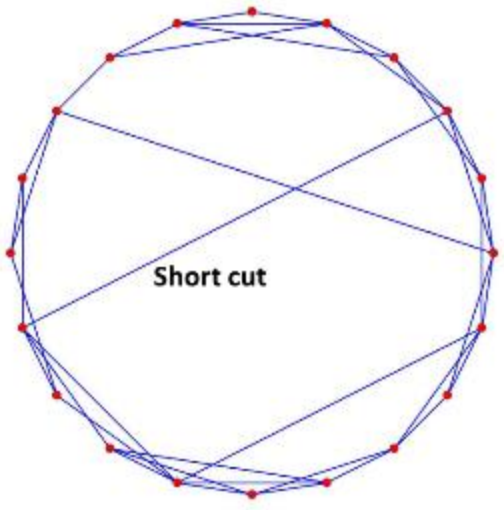


Figure 2-3. The small-world network by Watts and Strogatz drawn by the network analysis software Pajek [20].

Another property of small-world networks is a large average clustering coefficient. The concept of

clustering in a network refers to the tendency observed in many natural networks to form cliques in the neighborhood of any given node [18, p10]. Nodes in the cliques are densely related to each other, i.e., they connect each other, implying that the network is difficult to break.

2.5 Scale-free networks

A scale-free network is a network whose degree distribution follows a power law. That is, the degree distribution $p(k)$ is expressed as follows:

$$p(k) = Ak^{-\alpha} \qquad \text{for} \quad k > k_{min}\,. \tag{2-7}$$

There must be a lowest value k_{min} above which the power law is obeyed. The constant α is called the exponent of the power law (or the power exponent). Graphs of $p(k)$ appear as straight lines in logarithmic scales on both axes on a so-called log-log plot, and we can calculate α from the slopes of the lines.

The power law is frequently observed in many real-world networks, e.g., the WWW, language networks, and actor networks. However, networks constructed from simple models, e.g., Erdős-Rényi graphs, lattices, and trees, are not scale-free networks. Scale-free networks have nodes that connect a huge number of nodes, i.e., the nodes have a very large degree. These nodes are called hubs. Because hubs can connect a huge number of nodes, the diameter of the network is subject to becoming smaller.

The properties of the network change according to the size of the power exponent. If a power law is approximated by a continuous distribution, we obtain the following equations:

$$\int_{k_{min}}^{\infty} Ak^{-\alpha}dk = A[k^{-\alpha+1}/(-\alpha+1)]_{k_{min}}^{\infty} = \frac{A}{\alpha-1}k_{min}^{-\alpha+1}\ (\alpha > 1), \tag{2-8}$$

$$\int_{k_{min}}^{\infty} kAk^{-\alpha}dk = \frac{A}{\alpha-2}k_{min}^{-\alpha+2} \qquad (\alpha > 2), \tag{2-9}$$

$$\int_{k_{min}}^{\infty} k^2Ak^{-\alpha}dk = \frac{A}{\alpha-3}k_{min}^{-\alpha+3} \qquad (\alpha > 3). \tag{2-10}$$

From these equations, we obtain the following properties:

1) $\alpha > 1$.

2) For $1 < \alpha \leq 2$, the mean of the degree values becomes infinite, i.e., the degree distribution has no finite mean.

3) For $2 < \alpha \leq 3$, the mean square of the degree values become infinite, i.e., the degree distribution has no finite variance.

"No finite mean" means that in real data, the mean value changes according to the size of the sampled data. In addition, "no finite variance" means that variations in the means are so large that the mean value becomes meaningless. Accordingly, the properties of the network discontinuously change at values of the power exponent of 2 and 3. Thus, it is thought that scale-free networks can be classified by the value of the power exponent.

Several models that produce scale-free networks have been proposed. Among them, the "preferential attachment" model (BA model) by Barabási and Albert is the most important [4]. This model is a "rich-gets-richer" model (Yule process) [30–31]. The BA model is as follows:

> *Starting with a connected network with a small number (m_0) of nodes, at every time step a new node with m ($\leq m_0$) edges is added to the network. These edges connect the new node to m different nodes already present in the system. To incorporate preferential attachment, it is assumed that the probability Π_i that a new node will be connected to node i depends on the degree k_i of that node, such that* $\Pi_i = \frac{k_i}{\sum_{j=1}^{n} k_j} \quad (1 \leq i \leq n)$.

After *t* time steps the model leads to a network with $t + m_0$ nodes and *mt* edges. A numerical simulation demonstrates that this network evolves into a scale-free network with the probability that a node has *k* edges following a power law with power exponent $\alpha = 2.9 \pm 0.1$. Barabási et al. have also analytically calculated the exponent to be $\alpha = 3.0$.

Although the BA model exhibits $\alpha = 3.0$, many real-world networks exhibit scale-free networks with exponents ranging from 2 to 3. However, this is not crucial because the exponent can be

changed by slightly changing the BA model. Moreover, several variations of the BA model have been proposed [5–9].

The development of power-law scaling in the model indicates that growth and preferential attachment play important roles in network development. To verify that both ingredients are necessary, Barabási et al. investigated two variants of the model. The growing character of the network was retained in Model *A*, but preferential attachment was eliminated. Model *B* did not have the growing character but kept preferential attachment. As a result, neither model exhibited the scale-free feature [4, 32].

2.6 Truncation of the power law

The BA model produces a network with a degree distribution that obeys a complete power law. However, most real-world networks alleged to have a power-law distribution actually have degree distributions with a power-law regime followed by a sharp cutoff, such as an exponential or Gaussian decay of the tail [33]. This type of distribution is called a truncated power law (Fig. 2-4).

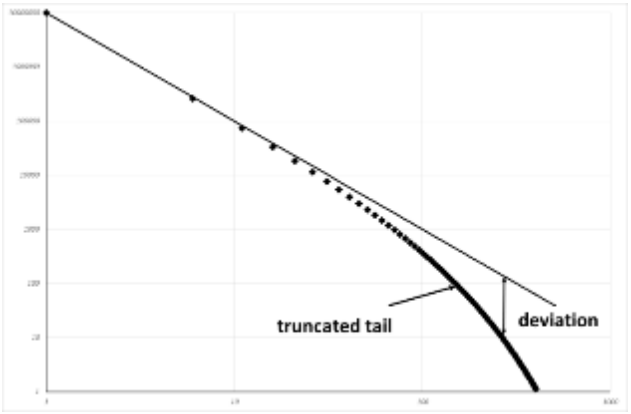


Figure 2-4. Log-log plot of a truncated power law.

A scale-free network emerges in the context of a growing network in which new nodes connect preferentially to the more highly connected nodes in the network (preferential attachment), i.e., the probability Π_i that a new node will be connected to node i is described by the formula $\Pi_i =$

$\frac{k_i}{\sum_{j=1}^{n} k_j}$ $(1 \leq i \leq n)$. However, preferential attachment is not satisfied without preconditions. Several constraints on preferential attachment are suggested as follows.

1) Information filtering [34]

Although an added new node must know the degrees of all existing nodes to fulfill the preferential attachment, this situation is not plausible in a large, growing network. It is likely that a new node of the large, growing network will know only information concerning a subset of existing nodes (filtered information). Then the new node will make decisions on which node to connect with, based on the filtered information.

Specifically, preferential attachment of the BA model is modified as follows. Let *f* be the constant fraction of "filtered nodes" in the network. The probability *Π(i, t)* that a new node will be connected to node *i* at time *t* is described by the following formula:

$$\Pi(i,t) = \frac{k_i}{\sum_{j \in C} k_j} \ , \qquad (2-11)$$

where C is a randomly selected subset of nodes containing $n(t) = (t + m_0)f$ nodes (m_0 is the number of nodes at the start).

2) Aging of the nodes [33]

It is unlikely that a node can permanently receive new edges. Even a highly connected node will, eventually, stop receiving new edges. The node is still part of the network and contributes to network statistics, but it no longer receives edges.

This idea is achieved by modifying the BA model. First, nodes are classified into one of two groups: active or inactive. Inactive nodes cannot receive new edges. All new nodes are created active, but they may become inactive after each time step with a constant probability. Ultimately, the probability *Π(i, t)* that a new node will be connected to node *i* at time *t* is described by

$$\Pi(i,t) = \frac{k_i}{\sum_{j \in A} k_j} \ , \qquad (2-12)$$

where A is a subset of nodes containing only active nodes at time t.

3) Limited capacity of a node to receive new edges [33, 35]

This is based on the idea that there are physical costs involved in adding new edges. In the real world, an edge has physical entities such as electric cables or airline routes, which have related costs. Thus, there is a limitation on a node receiving edges.

In this constraint, nodes are classified into two groups: active or inactive, as is the case with "Aging". Every node is created active, but it becomes inactive when its degree reaches a maximum number of edges k_{max}. The probability $\Pi(i, t)$ is described by the same equation as the "Aging" constraint (eq. (2-12)).

Introducing these constraints to any of the models leads to cutoffs on the power-law decay of the tail of the degree distribution.

2.7 Scale-free and small-world properties

It is known that the power exponent of a scale-free network is associated with the average shortest path length $\langle l \rangle$. Cohen and Havlin show the following relations [36].

Let α be the power exponent of the degree distribution, then

1) for $2 < \alpha < 3$

$$\langle l \rangle \approx \log\log n, \qquad (2-14)$$

2) for $\alpha = 3$

$$\langle l \rangle \approx \frac{\log n}{\log\log n}, \qquad (2-15)$$

3) and for $\alpha > 3$

$$\langle l \rangle \approx \log n. \qquad (2-16)$$

Thus, a scale-free network has an average shortest path length that grows only logarithmically with n. This is also one of the conditions for a small-world network. The other condition, "large average clustering coefficient", is not always satisfied in a scale-free network. For example, the BA model

has an average clustering coefficient that approaches zero with increasing *n*. Overall, whether a scale-free network is small-world depends on its cluster coefficient.

2.8 Hierarchical structure

The BA model is a scale-free network but not small-world because the average clustering coefficient approaches zero with increasing *n*. Several scale-free and small-world networks have been proposed. Some of them are hierarchical models constructed by using the hierarchical rule (or recursive rule), which is the process of repeating items in a self-similar way as used in deterministic fractals. A model based on this hierarchical rule is called a network that has a hierarchical structure.

Among hierarchical models, Ravasz's model is the most important [8, 9]. This model is constructed as follows (Fig. 3-4).

Step 0) *Start from a complete network consisting of n_0 nodes. Make one of the nodes the central node.*

Step 1) *Create n_0-1 identical replicas. Connect the peripheral nodes of each replica to the central node of the original network.*

Step 2) *Create n_0-1 replicas of the obtained network. Connect the peripheral nodes to the central node of the original network.*

Step 3) *Repeat step 2.*

Although Ravesz's model itself is a deterministic model, it can be easily converted to a random model involving preferential attachment similar to the BA model. These hierarchical models have two significant features:

1) A system-size independent clustering coefficient.

2) The average of the clustering coefficients of nodes with *k* edges follows the power law

$$C(k) \approx \frac{1}{k}\,. \qquad (2-17)$$

These two features are used to verify whether a network is hierarchical. By examining these features, it is verified that actor, language, and WWW networks have hierarchical features, whereas the Internet at the router level and power grid networks do not. However, it should be noted that hierarchical structure is not precisely defined.

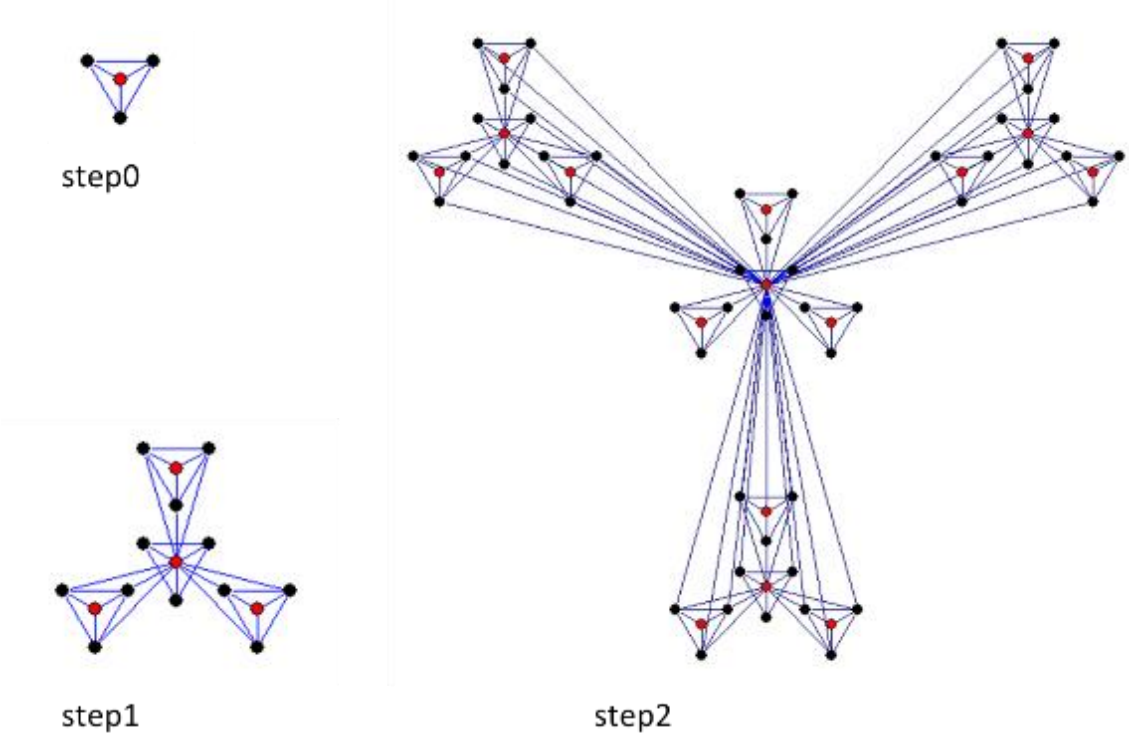


Figure 3-4. Rabasz's model drawn by the network analysis software Pajek [21].

2.9 Complex network

The definition of a complex network is controversial. Real-world networks have features different from simple network models such as a lattice network and the Erdős-Rényi network. In particular, many real-world networks have small-world and scale-free features. Some researchers call a network that has these features a complex network. Others consider that the concept of a complex network relates to complex systems. While the precise definition of complex systems is also subjective, it is plausible that the following few basic features characterize complex systems [18]:

1) Complex systems have emergent phenomena in the sense that these phenomena are the spontaneous outcome of the interactions among the many constituent units (self-organization

phenomena).

2) Decomposing the system and studying each subpart in isolation do not allow an understanding of the whole system and its dynamics.

3) Complex systems have complications at all scales (possibly under physical constraints) of the system. Those complications may produce a long-tail distribution like a power-law distribution.

The relationship between scale-free and self-similar properties has been investigated in several papers [5, 6, 37]. Barabási et al. [5] pointed out that although their model, which has the scale-free property, seems to display self-similarity, this self-similarity is not complete. This is because a model that has the scale-free property needs global information about the network. For example, the central node in Rabasz's model (Fig. 3-4, step 0) keeps a detailed record of the system size through the number of edges it has. Such global information is never present in a local element of a fractal. However, Song et al. [37] have argued that a variety of real networks that have a power-law distribution possess the self-similar property.

Classification of complex networks

As described above, real-world networks and complex networks can be classified by features such as those summarized in Table 2-2, which are not always independent. For example, in a network whose degree distribution follows a power-law, $\langle l \rangle$ grows only logarithmically with n. However, we can classify networks to some degree, and two networks that have the same features can be said to be similar networks.

Table 2-2. Properties of complex networks.

Network	$\langle l \rangle$	$\langle C \rangle$	$p(k)$	Other features
Small-world network	$\preccurlyeq \log n$ or $\approx \log n$	$\approx C_0 (> 0)$	various	
Scale-free network	$\preccurlyeq \log n$ or $\approx \log n$	various	power law	
Hierarchical model	$\preccurlyeq \log n$ or $\approx \log n$	constant	power law	$C(k) \approx \frac{1}{k}$
Self-similar network	$\preccurlyeq \log n$ or $\approx \log n$	various	power law	self-similarity

$\langle l \rangle$ is the average shortest path length, $\langle C \rangle$ the average clustering coefficient, n the order of the network, $p(k)$ the degree distribution of the network, and $C(k)$ the average of the clustering coefficients of nodes with k edges. Given two functions $f(n)$ and $g(n)$, when $\lim_{n\to\infty} \frac{f(n)}{g(n)} = A$, where A is a finite positive value, we denote $f(n) \approx g(n)$. When $\lim_{N\to\infty} \frac{f(n)}{g(n)} = 0$, we denote $f(n) \preccurlyeq g(n)$.

3. Background of the study

Long-tail distributions like the power-law distribution have been studied in many fields. For example, the word frequency of natural language, the relationship between the magnitude and the total number of earthquakes (Gutenberg-Richter law), and income distribution in economics are expressed by power-law distributions [38–43]. The power-law distribution has also been studied in the field of medical service: it is known that the frequency of diagnoses and the length of stay of hospital patients exhibit power-law distributions [44–47].

The power law has also been highlighted in relation to nonlinear dynamics that are formulated in fractal geometry and chaos theory [41, 48]. In systems consisting of many elements, an inter-element correlation (interaction) over a long period is undoubtedly related to the origin of a power law [49]. The probability of word appearance in natural language is affected by the context. In other words, the choice of words is context-dependent. This is considered to be a necessary condition to create a power-law distribution.

It is said that a quantity x obeys a power law if its probability density $p(x)$ is described by

$$p(x) \propto x^{-\alpha} \qquad (x \geq x_{min} > 0), \tag{3-1}$$

where α is a constant parameter of the distribution known as the (power) exponent. The magnitude of the exponent α determines the tail behavior. A power-law distribution with an exponent less than 2 has no finite mean, whereas one with an exponent less than 3 has no finite variance.

In many cases, it is useful to consider the complementary cumulative distribution (CCD) of the distribution $p(x)$:

$$P(x) \equiv \int_x^{\infty} p(x')dx' = \left(\frac{x}{x_{min}}\right)^{-\alpha+1}. \tag{3-2}$$

For discrete cases, the power-law distribution is sometimes called “Zipf’s law”. The frequency F_n of the quantity of the rank n is expressed as

$$F_n = \frac{A}{n^{\zeta}}, \tag{3-3}$$

where A is a constant, and ζ is referred to as the Zipf exponent. The power exponent α is expressed by ζ as follows (Appendix 1):

$$\alpha = \frac{\zeta + 1}{\zeta} . \qquad (3-4)$$

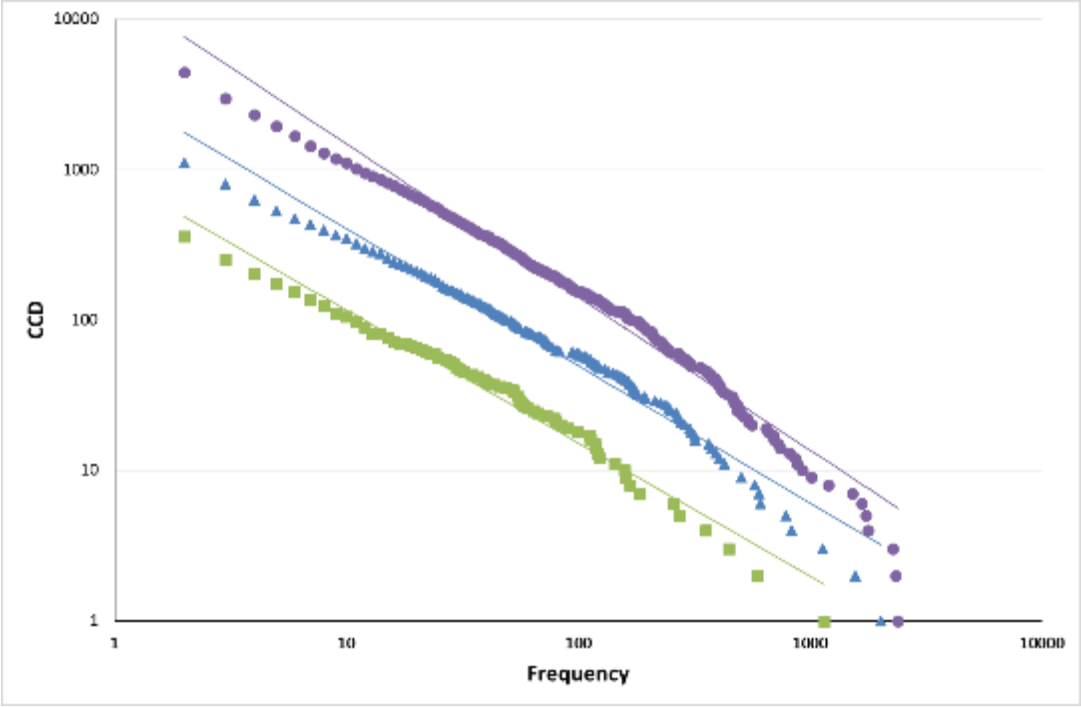
The graphs of the functions (3-1) to (3-3) appear as straight lines in a log-log plot, and we can calculate α and ζ by their slopes. Although all three functions may lead to the same value of α, the CCD usually leads to the most accurate value (Appendix 2). Therefore, we typically express a power-law distribution by its CCD.

Figure 3-1 shows the CCDs of the diagnosis frequency for several departments of a hospital [44]. In these plots, the CCDs are plotted on a logarithmic scale. The regression lines for each dataset are also shown. The good fits by the regression lines mean that for all departments, the data follow a power law. The power exponents calculated from the slopes of the regression lines show that all the power exponents are close to 2 (Table 3-1).

Table 3-1. Results of regression analyses.

	free written group		
	α (mean ± SE)	R^2	N
total	2.01±0.011	0.97	219
internal medicine	1.91±0.014	0.97	127
pediatrics	1.83±0.016	0.97	70
surgery	1.96±0.022	0.97	68
neurosurgery	2.05±0.017	0.99	34
ophthalmology	2.04±0.010	0.99	52
orthopedics	2.24±0.013	0.99	51
cardiac surgery	1.97±0.022	0.97	68
dermatology	2.07±0.011	0.99	41
urology	1.87±0.018	0.98	49

α is the power exponent, R^2 the coefficient of determination in linear regression analysis, N the number of data, and SE the standard error. The coefficients of determination and the power exponents are obtained by linear regression analyses for the log-log plots of the CCDs of diagnosis frequencies.



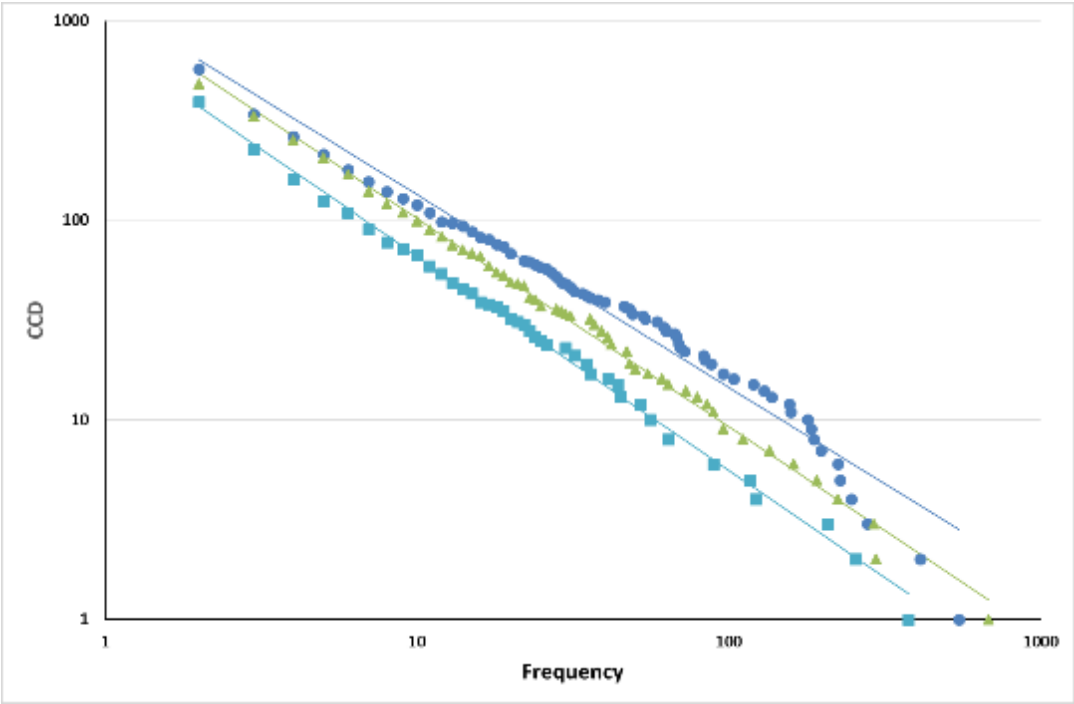


Figure 3-1. The CCDs of diagnosis frequencies for several departments of a hospital.

These results imply that a patient diagnosis is affected by the diagnoses of previous patients. This is surprising, because it strongly differs from the conventional understanding that a diagnosis is dependent only on the individual patient's condition.

We further review the process involved when making a diagnosis. Not all clinical diagnoses are independent. In fact, some diagnoses are strongly related. For example, a patient who shows symptoms of a common cold may be diagnosed with one of the following diseases: common cold, upper respiratory infection, acute rhinitis, acute pharyngitis, or acute bronchitis. These names are easily used interchangeably (or diagnosed in error). In this sense, these terms are closely related; in other words, the "distance" between them is short. In contrast, because a femoral fracture and an upper respiratory infection would not be confused, these two diagnoses will be unrelated, and the distance between them will be great.

When diagnosing a patient, a doctor chooses the diagnosis that accommodates symptoms most similar to those of the patient. Let us consider a situation where several close diagnoses fit the symptoms of the patient, but there is no absolute standard for selecting one of them because of the vagueness of each diagnosis. In such a case, the selection of a diagnosis depends on the doctor's subjective opinion. This situation is similar to the selection from many plausible words when writing a sentence. In natural language, a word is selected based on context. Similarly, the selection of a diagnosis is affected by many factors such as the diagnoses of previous patients with similar symptoms, the diagnosis frequency distribution of that hospital, and the doctor's skill. Consequently, the selection of a diagnosis is not only dependent on the patient's condition, but also on the condition of the doctor, previous patients**,** and the hospital where the doctor works. These factors would create an inter-diagnostic correlation, and this correlation may result in a power-law distribution.

As mentioned above, the frequency of diagnoses is influenced by the doctor's decision-making process when diagnosing. For this reason, an intimate evaluation of this process is required to

investigate the mechanism that produces the frequency distribution of diagnoses. It is, however, difficult to evaluate this by a simple model because a doctor's decision-making is far from being simple. More than anything, a doctor's decision-making is based on a large amount of medical knowledge. In other words, diagnoses by doctors are influenced by medical knowledge. Investigation of the structure of medical knowledge is, therefore, indispensable to evaluate the decision-making of doctors. For these reasons, our aim in this study is to clarify the structure of medical knowledge.

4. Methods

4.1 Construction of the medical knowledge network (MKN)

We analyzed “Harrison's Principles of Internal Medicine, 15th Edition”, which is one of the most important textbooks on internal medicine. We analyzed three chapters: “Disorders of the Cardiovascular System”, “Disorders of the Respiratory System”, and “Neurologic Disorders”. All 38 353 sentences that comprise these chapters were analyzed. Because the purpose of our analysis was to determine the structure of medical knowledge rather than that of language, we confined the objects of study only to medical terms consisting of several words related to medical knowledge. We then classified these terms into four categories: “diagnosis”, “subjective symptom”, “objective symptom”, and “other medical terms”. For instance, “stomach cancer” is a term consisting of two words and classified as “diagnosis”. These classifications were made by two medical doctors and a healthcare information technologist. However, in this study we used only the difference between “diagnosis” and the other three categories. All categories but “diagnosis” remain for future analyses. We then constructed the medical knowledge network by first defining the medical terms as nodes of the network, then defining the edges that mutually connected a pair of terms in a sentence (Figs. 4-1 and 4-2).

4.2 Construction of the diagnosis database network

We also applied the network analyses to a database of “disease” in the hospital information system of Toyonaka Municipal Hospital in Japan. This database consisted of 218 063 records, which represented 2 years of data. Each database record contained several items in addition to “diagnosis”, such as “patient ID”, “department code”, and “doctor ID”. Therefore, by assigning these items to nodes and mutually connecting all nodes in each record by edges, we constructed the diagnosis database network (DDN, Fig. 4-2). Because this network was too large to analyze, we analyzed only a partial network consisting of randomly selected nodes.

Our use of these data was approved by the Toyonaka Municipal Hospital. All of the patient data were de-identified, and because the analysis was purely statistical, there were no ethical issues.

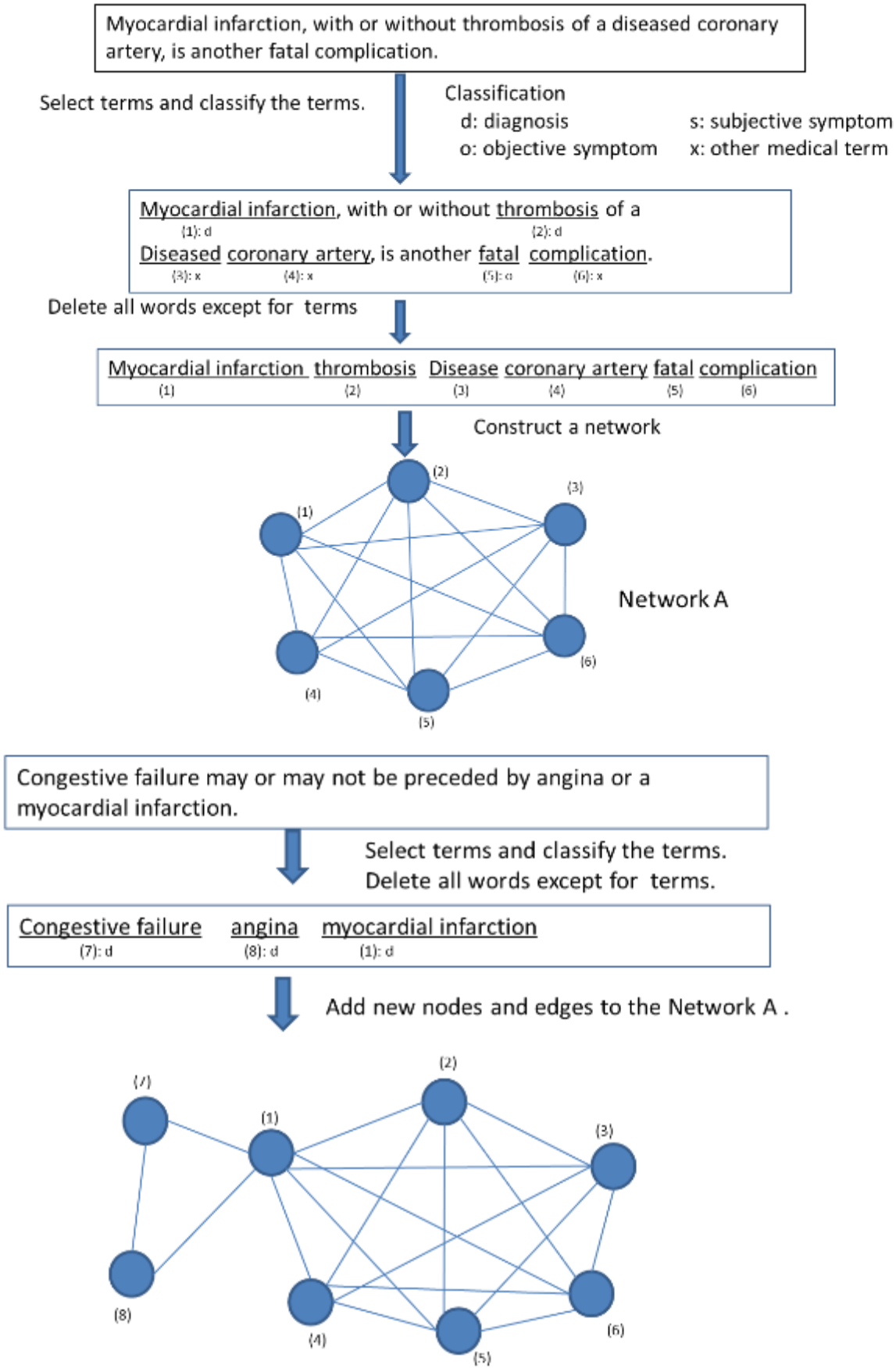


Figure 4-1. Construction of the MKN.

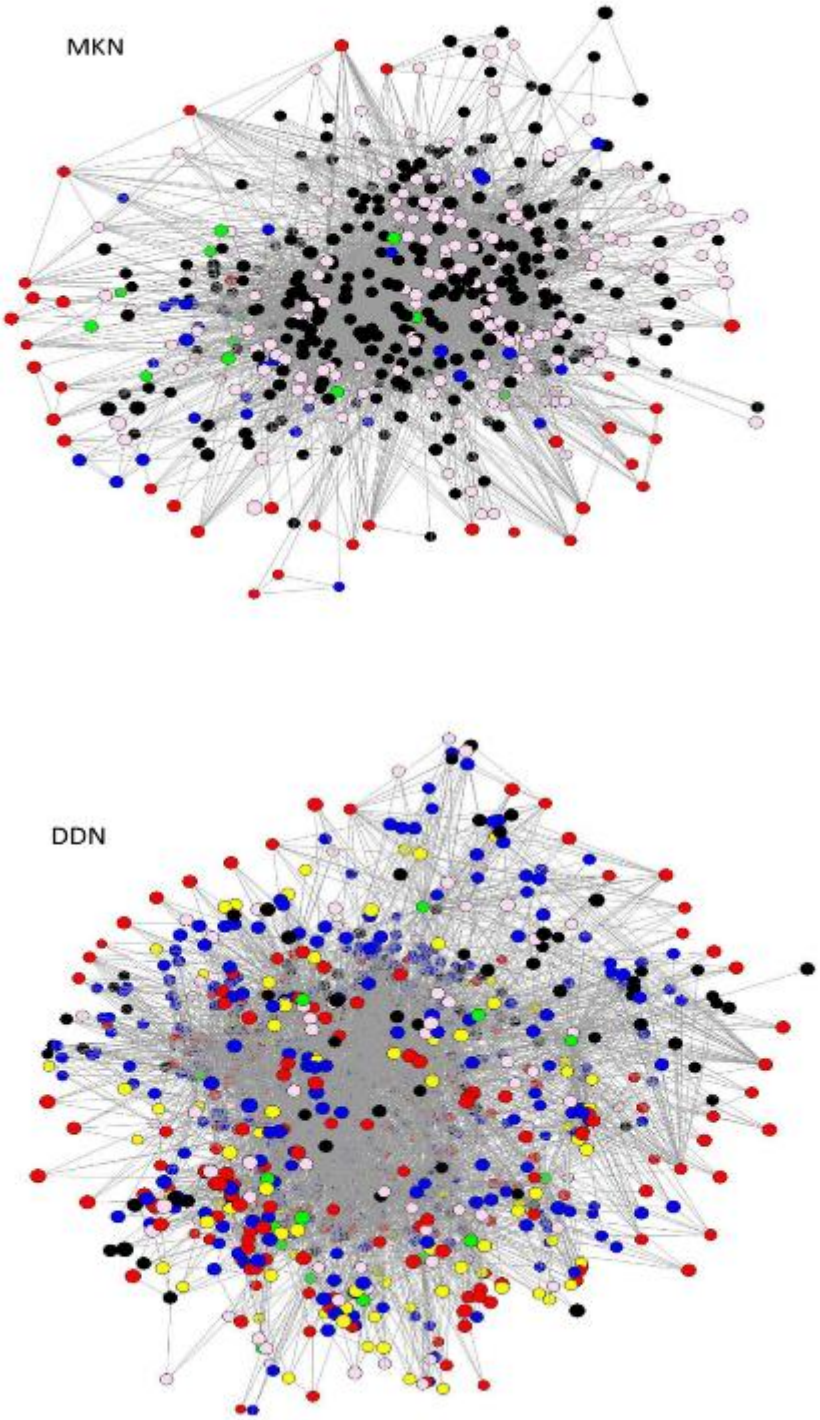


Figure 4-2. MKN and DDN. The original networks have so many nodes that they cannot be drawn. Thus, here only the subnetworks are depicted, with about 1000 nodes.

MKN: ●, diagnosis; ●, sub. sym.; ●, obj. sym.; ●, other med. term; ●, others

DDN: ●, diagnosis; ●, pt. ID; ●, dep.; ●, Dr.; ●, Pc. ID; ●, data

5. Numerical analysis and conclusions

We calculated the average path length and the average clustering coefficient of the MKN (see definitions in Sec. 2) to be 4.317 and 0.86, respectively. The implication is that the MKN has nearly the same average path length and a far larger average clustering coefficient compared with the corresponding random network with the same number of nodes and edges as the original network. These findings suggest that the MKN has the small-world property (Table 5-1), which may be useful for a clinician to quickly find the correct diagnosis among a large number of diseases. In a small-world network, every node (i.e., every diagnosis) can be reached within a few steps [10, 22].

Table 5-1. Profiles of the MKN and the DDN.

network	*nodes*	*edges*	$\langle k \rangle$	$\langle C \rangle$	$\langle C \rangle_{random}$	$\langle l \rangle$	$\langle l \rangle_{random}$
MKN	47769	884613	37.0	0.86	7.75×10^{-4}	4.317	3.07
DDN	44997	505565	22.5	0.83	4.99×10^{-4}	3.894	3.44

$\langle k \rangle$ is the average degree, $\langle C \rangle$ the average clustering coefficient, $\langle C \rangle_{random}$ the average clustering coefficient of the corresponding random network estimated from Table 2.1 (Sec. 2.2), $\langle l \rangle$ the average path length, and $\langle l \rangle_{random}$ the average path length of the corresponding random network estimated from Table 2.1 (Sec. 2.2). For both the MKN and DDN, the average path length is close to that expected for a random graph and $C \gg C_{random}$, meaning that both the MKN and DDN have small-world features. For the DDN, we constructed a partial network by randomly selecting about 45 000 nodes from the DDN and analyzed this partial network. (Modified from Table 1 in Ref. [1].)

We also found that the degree distribution of the MKN exhibits a power law with a rapidly decaying tail (Fig. 5-1a) [8, 33, 50]. This finding indicates that the MKN is a truncated, scale-free network. The exponent is 2.045, which is consistent with many other complex networks whose exponents range from 2 to 3 [4, 19]. Scale-free behavior is a consequence of two generating mechanisms: networks expanding continuously by the addition of new nodes, and new nodes attaching preferentially to already well-connected sites (preferential attachment) [4, 51]. If preferential attachment is completely fulfilled, the network exhibits entirely scale-free behavior; however, if preferential attachment is not completely fulfilled because of, for example, limitations on the available information, scale-free behavior is truncated [31, 34, 52, 53]. For the MKN, our

results suggest that preferential attachment may be somewhat restricted.

We also investigated the hierarchical structure of the MKN. The preferential attachment model is a simple network growth model with no hierarchical structure [4]. In contrast, many networks in nature and society have some form of hierarchical structure. Recently, a network model that produces a network with a hierarchical structure was proposed [5, 7, 9]. According to this model, a network that has a hierarchical structure has the following two features: the average clustering coefficient is independent of the size of the network, and the average of the clustering coefficients of nodes with k edges follows the scaling law $C(k) \sim k^{-1}$. Our analysis of the MKN determined that it has these two features (Fig. 5-1b, c).

We next considered whether clinical practice also has small-world and scale-free properties. If the structure of the MKN reflects only that of language itself, its application to real medical services would be limited. However, if its structure affects the clinical behavior of medical professionals, its meaning would be significant. To answer this question, we applied network analysis to a database of "disease" in a hospital information system, which should reflect the clinical behavior of doctors. This database was comprised of clinical diagnoses that the doctors had entered during daily medical examinations. We constructed a network derived from this database (DDN, see Methods section).

Table 5-1 shows that the DDN has the small-world property. Figure 5-2 shows that the DDN is a truncated, scale-free, hierarchical network. Furthermore, the average clustering coefficient and the power exponent of the DDN were 0.83 and 2.084, respectively, which are values similar to those of the MKN. Thus, both the DDN and the MKN have similar network structures.

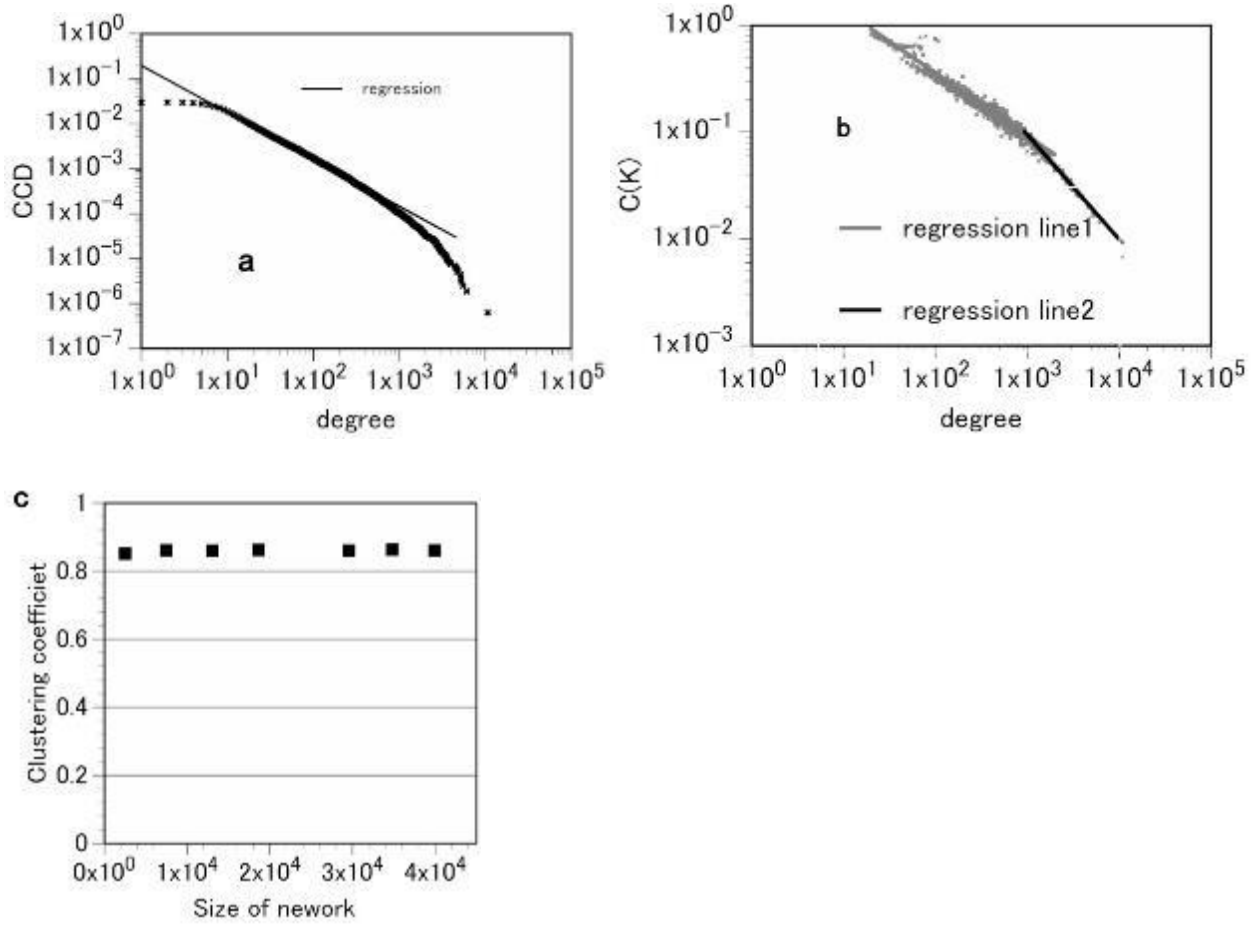


Figure 5-1. Attributes of the MKN. (a) The complementary cumulative distribution (CCD) of degrees for the MKN, which is defined as $CCD(k) = \sum_{j>k} p(j)$, where *p(j)* is the probability of a node having degree *j* (i.e., the degree distribution). The exponent of the degree distribution is 2.045. This figure indicates that this degree distribution follows a truncated power law (scale-free). A truncated, scale-free distribution is a distribution that decays according to the power law $p(k) \sim k^{-\alpha}$ followed by a sharp cutoff. (b) The average of the clustering coefficients of nodes with *k* edges, *C(k)*, which follows the power law $C(k) \sim k^{-\nu}$, where ν = 0.59 and 0.98 for small and large values of *k*, respectively. This means that, at least for large *k*, the clustering coefficient follows the scaling law $C(k) \sim k^{-1}$. (c) The average clustering coefficient vs. the size of the network. We calculated the average clustering coefficients of partial networks we constructed by randomly eliminating several nodes from the MKN. This figure shows that the average clustering coefficient of the MKN is independent of the network size. These data demonstrate that the MKN has both properties of a hierarchical network. (Originally Fig. 1 in Ref. [1].)

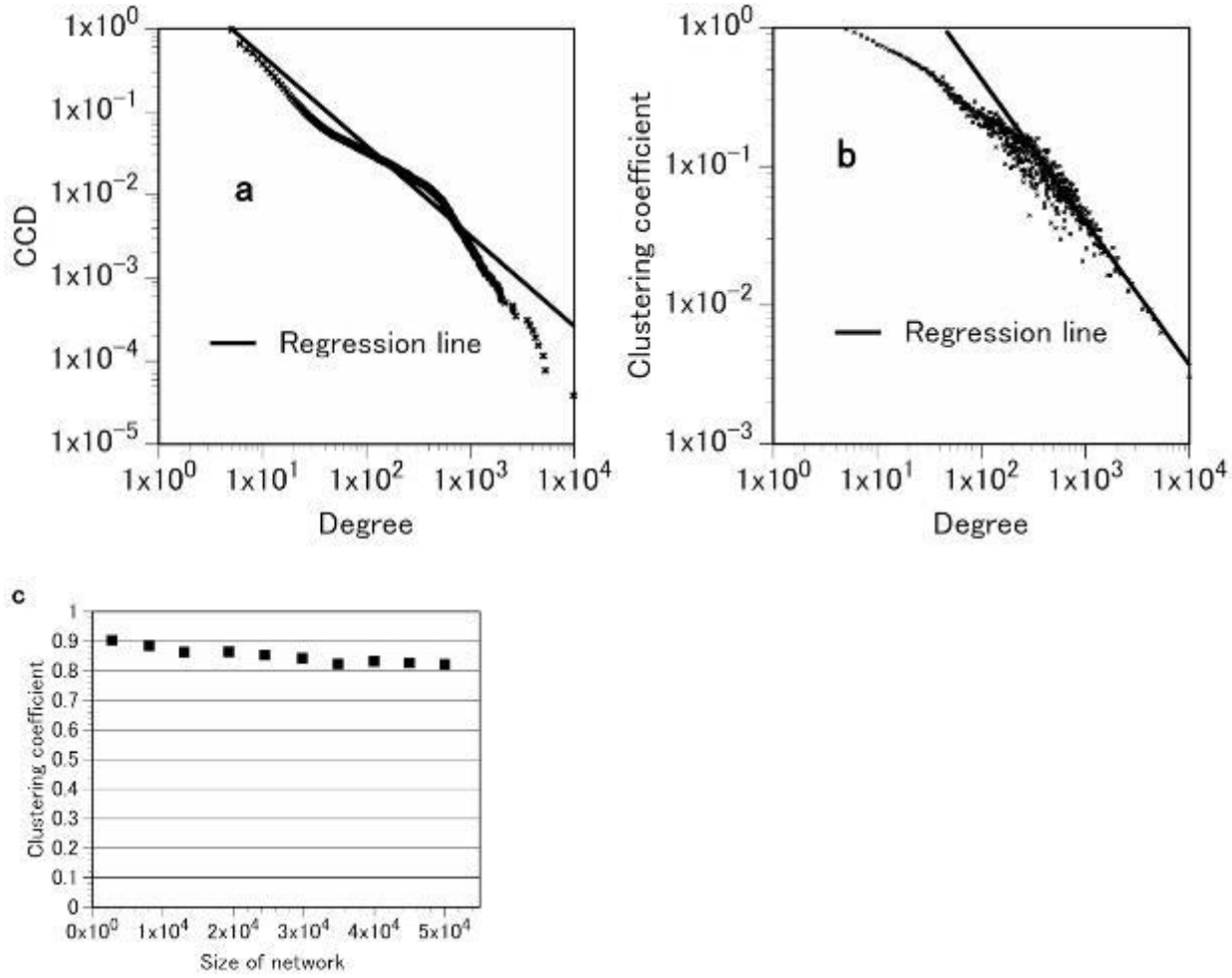


Figure 5-2. Attributes of the DDN of Toyonaka Municipal Hospital. (a) The CCD distribution of the DDN. These data show that the DDN degree distribution follows a truncated power law; the exponent is 2.084, which is similar to that of the MKN. (b) The average of the clustering coefficients of nodes with k edges, $C(k)$. As with the MKN, $C(k)$ of the DDN follows the scaling law $C(k) \sim k^{-\nu}$, where $\nu = 1.03$ for large k. (c) The average clustering coefficient vs. the size of network. The average clustering coefficient of the DDN is independent of the system size. These data indicate that the DDN has both properties of a hierarchical network, similar to the MKN. (Originally Fig. 2 in Ref. [1].)

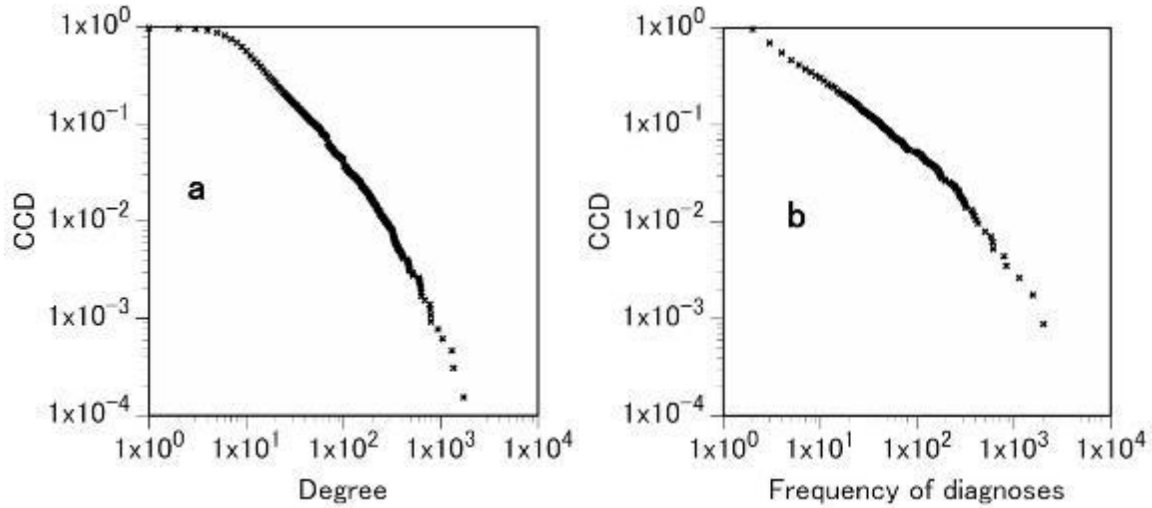


Figure 5-3. Degree and frequency of diagnoses. (a) The CCD degree distribution of the partial network of the MKN, consisting of only those nodes that were classified as "diagnoses" and their edges. The diagnosis degree distribution follows a power law with an exponent of 2.10. (b) The CCD frequency distribution of internal medicine at Toyonaka Municipal Hospital. The diagnosis frequency distribution follows a power law with an exponent of 1.84. (Originally Fig. 3 in Ref. [1].)

For the final analysis in this study, we asked whether the degree distribution of diagnoses (diagnosis degree distribution) of the MKN obeys a power law similar to the frequency distribution of clinical diagnoses (diagnosis frequency distribution). We extracted the nodes classified as "diagnoses" from the MKN and analyzed their degree distribution. We also analyzed the frequency distribution of internal medicine diagnoses from the hospital data. Both the diagnosis degree distribution of the MKN and the diagnosis frequency distribution of internal medicine followed truncated power laws with similar exponents of 2.10 and 1.84, respectively (Fig. 5-3). In fact, the diagnosis frequencies of not only internal medicine but also other departments followed truncated power laws with exponents ranging from 1.76 to 2.20 [44, 45]. These findings may have important implications. In a knowledge network, a node is a term. Consequently, a term with a large degree is connected to many other terms, the result being that it frequently appears in many sentences. In fact, according to an investigation of a general text, the frequency of word appearance in a text and the degree of the network constructed from that text are positively correlated [55]. However, in this study the frequency is not the frequency of diagnosis appearance in the text but that of diagnosis appearance in the hospital. This distinction between our study and the preceding investigation [55]

is important.

Before presenting our conclusions, we describe the mathematical implications of the numerical analysis. First, having the small-world property implies that a network cannot be expressed by an Erdős-Rényi random graph. Table 5-1 shows that both the MKN and DDN have average degrees that are greater than $\log n$. Therefore, if these networks are expressed by random graphs, then the random graphs are connected. The connected random graph has an average path length near $\frac{\log n}{\log np} = \frac{\log n}{\log\langle k\rangle}$ (Sec. 2.3). Table 5-1 shows that the average path lengths of the MKN and the DDN are both near $\frac{\log n}{\log np}$. These findings are compatible with the hypothesis that these networks are expressed by random graphs. However, these networks have far larger average clustering coefficients compared with the corresponding random networks. These findings are inconsistent with Theorem 2.5 and imply that these networks cannot be expressed by random graphs. Moreover, these networks exhibit scale-free behavior; their degree distributions exhibit a power law. However, the degree distribution of the random graph follows a Poisson distribution. Therefore, these findings are also inconsistent with the random graph. Thus, the numerical analysis shows that both the networks investigated in this study cannot be expressed by Erdős-Rényi random graphs.

In addition to these facts, the networks investigated in this study have a hierarchical structure. The hierarchical structure has only an empirical definition and no mathematical definition. We therefore require mathematical definitions of hierarchical structure and the construction of a new mathematical model to investigate the MKN and DDN in detail.

The conclusions of the numerical analysis are summarized as follows:

1) The networks derived from medical knowledge and medical practice have multiple common features. The medical significance of this fact is described in the next section.
2) These networks cannot be expressed by Erdős-Rényi random graphs.
3) We require the concept of a complex network to investigate these networks.
4) To express these networks, we require a new mathematical model.

6. Discussion

As described above, the MKN displays small-world, scale-free, and hierarchical properties. The small-world property may help a clinician make the most appropriate diagnosis. In contrast, scale-free and hierarchical properties are related to the mechanism generating the MKN [4, 9]. Medical knowledge is always changing as it adapts to change or progress in social situations and medical science. New medical terms, such as new diagnoses, are constantly being added to medical knowledge, and definitions of diagnoses frequently vary. The addition of a new term corresponds to the addition of a new node in the MKN. This new node is attached to already existing nodes. In this example, the possibility of attachment to frequently used terms (nodes) is thought to be large. This means that preferential attachment is applied to the MKN. Medical knowledge intrinsically involves a complex hierarchical structure. Therefore, new nodes need to be added to the MKN without damaging the preexisting hierarchical structure. Thus, the MKN would develop and evolve under the following two principles: preferential attachment and preservation of the preexisting hierarchical structure. These two principles may give the MKN its precise structure. These results suggest that network analysis of medical texts may provide new insights into the genesis of medical knowledge.

We showed that the diagnosis degree distribution of the MKN obeys a power law similar to the diagnosis frequency distribution of hospital data. Why do these two distributions resemble each other? The answer to this question may be that medical texts continuously reflect new knowledge in clinical practice. Therefore, a disease that occurs frequently in clinical practice is likely to be described many times in texts. Similarly, a disease that is mentioned frequently in texts is easy for a doctor to recall in clinical practice. The doctor obtains medical knowledge from texts or medical articles; therefore, the knowledge in the doctor's mind would have a network structure similar to that of the texts. This mutual influence of medical knowledge and clinical practice may determine the similar structures and distributions. That is, the similar structures may emerge from this mutual influence.

It used to be thought that diagnosis frequency is an objective index existing in the world; however, given the similarity between the diagnosis frequency of a hospital and the diagnosis degree of the MKN, diagnosis frequency could be influenced by medical knowledge. Medical knowledge influences clinical practice, and this practice influences the frequency of diagnosis. Because doctors diagnose through clinical practice, clinical practice is thought to be the observation of disease. Therefore, the fact that clinical practice influences diagnosis frequency implies that diagnosis frequency is not an objective index but, to a certain degree, a subjective index, the value of which varies somewhat with the observation of disease.

7. Future research directions

The world of diseases is remarkably complicated: diverse elements such as diagnoses, symptoms, and results of examinations are mutually connected, and these connections construct the integrated space (i.e., the MKN). Clarification of the structure of the MKN is an important problem in actual clinical practice. One attempt to clarify the connections among the elements is the classification of disease based on criteria like the ICD [2]. However, complex relations among much of the information make it difficult to clarify these connections by simple classification. Therefore, the mathematical analysis of the networks proposed in this study is needed to understand this structure.

In this study, we evaluate a certain MKN constructed from a well-known textbook of internal medicine. There are many other textbooks of medicine, and these textbooks change with time. This means that a number of different MKNs can exist. To clarify the common and different properties among many MKNs is very important for understanding the change and evolution of medical service. Moreover, by comparing many different MKNs, we may reveal the relation between the MKN and medical results.

Because mathematical analyses of MKNs have just begun, it is important to try new analysis techniques for networks. In particular, details of hierarchical and community structures [56, 57] may be useful for clinical practice. Detecting communities in the MKN will make it possible to create a new method for classifying diseases. Almost all methods of classifying diseases use some predefined criteria or multivariate statistics such as cluster analysis. However, from mathematical studies based on network analysis, a new classification method may emerge that reflects the complicated relationships among diverse clinical elements. Our preliminary study shows that the classification by detecting communities in a network from hospital database is different from that by cluster analysis.

Other than the analysis of the MKN, the network analysis of data in a hospital information system would be useful for hospital administration and management. A hospital information system involves data such as treatments, examinations, and nursing care for patients. By using the network

analysis of these data, we may classify patients according to clinical features.

Finally, we propose a new field in mathematical medicine[5] that can be called "medical complex systemology". We consider that medical service is a complex system composed of multiple interconnected elements such as patients, medical staff including doctors, medical knowledge, and the results of clinical practice (medical results). It is plausible that some features may emerge from the interactions among these elements. One example is the power law observed in the diagnosis frequency distribution. Dealing with medical service first requires expressing the relationships among these elements mathematically. In this study, we mathematically express medical knowledge by constructing networks (i.e., the MKN). Although the significance of medical knowledge in medical service is nothing special, we could not previously analyze the influence of medical knowledge on medical service. However, network analysis of the MKN brings forth the possibility to evaluate its influence on medical service.

We can also express medical practice (e.g., by the DDN). Medical results, such as diagnoses and outcomes from therapy, were previously expressed by piecemeal data. However, combining these data by a network analysis makes it possible to deal with them as one mathematical entity. As a result, we can analyze complex relations among diagnoses and therapies and their results, which is impossible when examining them individually. Moreover, we showed that the DDN and MKN share a common structure. The common structure between the medical results and the MKN is thought to be one that is emergent in the complex system. From this consideration, network analysis of the MKN and DDN, or "medical complex systemology", is a promising tool for evaluating medical service.

Previously, we could not but understand medical service empirically. However, cooperation between mathematics and medicine may make it possible to comprehend and demonstrate intrinsic functions and structures in medical service. Moreover, we can construct an actual and reasonable

[5] Mathematical medicine is a new area of study that utilizes mathematical methods to understand medicine.

medical system and manage it appropriately on the basis of this comprehension. This mathematical approach and its methods are worthy of a new mathematical area. The Mathematics Subject Classification (MSC) formulated by the American Mathematical Society[6] includes the category of game theory, economics, and social and behavioral sciences (code 91). We hope that the new division of mathematical medical service[7], including the subdivision of "medical complex systemology", will be added to this category and that many mathematical professionals in this area will work in medical faculties.

For developing this new area, the following problems should be investigated:

1) Clarifying the common and different properties among multiple MKNs.

2) Clarifying how MKNs change.

3) Clarifying the relationships between MKN and medical results.

4) Clarifying the history of MKN change.

5) Revealing the evolution of medical service.

6) Describing the overall picture of medical service.

By doing so, we may begin to answer the questions, "What is medical service?" and "What is medicine?".

[6] http://www.ams.org/home/page

[7] We also propose "mathematical medical service" as a new area of study that utilizes mathematical methods to understand medical service.

Acknowledgements

This work is based on an article [1] completed in cooperation with Mr. Hiroaki Iwanaga and Dr. Takashi Tahara, to both of whom I would like to express my sincere appreciation. I would also like to express my special thanks to Professors Ichiro Tsuda, Michiko Yuri, Kenji Matsumoto, and Takao Namiki of Hokkaido University for many helpful suggestions during the course of this work.

Original Article

The original article [1] on which this thesis is based is Y. Tachimori, H. Iwanaga, T. Tahara, The networks from medical knowledge and clinical practice have small-world, scale-free, and hierarchical features, *Physica* **A 392** (2013) 6084-6089.

References

[1] Y. Tachimori, H. Iwanaga, T. Tahara, The networks from medical knowledge and clinical practice have small-world, scale-free, and hierarchical features, Physica A 392 (2013) 6084-6089.

[2] WHO, ICD-10: The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines, World Health Organization, 1992.

[3] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440-442.

[4] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286(1999) 509-512.

[5] A.-L. Barabási, E. Ravasz, T. Vicsek, Deterministic scale-free networks, Physica A 299 (2001) 559–564.

[6] S. N. Dorogovtsev, A. V. Goltsev, J. F. Mendes, Pseudofractal scale-free web, Phys. Rev. E 65 (2002) 066122.

[7] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks, Nature 407 (2000) 651-654.

[8] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (2002) 1551-1555.

[9] E. Ravasz, A.-L. Barabási, Hierarchical organization in complex networks, Phys. Rev. E 67 (2003) 026112.

[10] R. F. Cancho, R. V. Solé, The small-world of human language, Proc. R. Soc. Lond. B 268 (2001) 2261-2266.

[11] R. V. Solé, B. C. Murtra, S. Valverde, L. Steels, Language networks: their structure, function and evolution, Complexity 15 (2010) 20-26.

[12] R. Solé, Syntax for free?, Nature 434 (2005) 289.

[13] A. E. Motter, A. P. S. de Moura, Y. -C. Lai, P. Dasgupta, Topology of the conceptual network of language, Phys. Rev. E 65 (2002) 065102(R).

[14] P. Erdös, A. Rényi, On random graphs, Publ. Math. 6 (1959), 290-297.

[15] P. Erdös, A. Rényi, On The Evolution of Random Graphs, Publ. Math. Inst. Hunga. Acad. Sci. 5 (1960), 17–61.

[16] E. N. Gilbert, Random Graphs, The Annals of Mathematical Statistics 30(4) (1959), 1141—1144.

[17] N. Konnno, Y. Ide, Introduction to Complex Graphs and Networks (Japanese), Kodansha Scientific (2008).

[18] A. Barrat, M. Barthélemy, A. Vespignani, Dynamical Processes on Complex Networks, Cambridge University Press, Cambridge, 2008.

[19] M. E. J. Newman, Networks An Introduction, Oxford University Press, Oxford, New York, 2010.

[20] B. Bollobás, Random Graphs (Second Editon), Cambridge University Press (2001).

[21] V. Batagelj, A. Mrvar, Pajek – Program for Large Network Analysis.
http://pajek.imfm.si/doku.php

[22] M. Markošová, Network model of human language, Physica A 387 (2008) 661-666.

[23] R. Albert, H. Jeoung, A. L. Barabási, Diameter of the World Wide Web, Nature 401 (1999) 130-131.

[24] B. Bollobás, The Diameter of Random Graph, Trans. Amer. Math. Soc. 267 (1981b), 41-52.

[25] V. Klee, D. Larman, Diameters of Random Graph, Can. J. Math. 33(3) (1981), 618-640.

[26] B. Bollobás, O. Riordan, The diameter of a scale-free random graph, Combinatorica, 24 (2004), 5-34.

[27] F. Chung, L. Lu, The Diameter of Random Sparse Graphs, Advances in Applied Mathematics 26 (2001), 257–279.

[28] M. E. J. Newman, Random graphs as models of networks, arXiv:cond-mat/0202208

[cond-mat.stat-mech].

[29] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, Y. Åberg, The web of human sexual contacts, Nature 411 (2001) 907-908.

[30] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46 (2005) 323-351.

[31] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (2009) 661-703.

[32] A.-L. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks, Physica A 272 (1999) 173-187.

[33] L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, Classes of small- world networks, PNAS 97 (2000) 11149-11152.

[34] S. Mossa, M. Barthélémy, H. E. Stanley, L. A. N. Amaral, Truncation of power law behavior in "scale-free" network models due to information filtering, Phys. Rev. Lett. 88 (2002) 138701.

[35] J. Lee, N. K. Park, Scale-free Networks in the Presence of Constraints: An Empirical Investigation of the Airline Route Network, Seoul Journal of Business 13 (2007), Number 1 (June 2007) 3-19.

[36] R. Cohen, S. Havlin, Scale-Free Networks are Ultrasmall, Phys. Rev. Lett. 90 (2003), 058701.

[37] C. Song, S. Havlin, H. A. Makse, Self-similarity of complex networks, Nature, 433, (2005), 392-395.

[38] G.K. Zipf, Human Behavior and the Principle of Least Effort, Cambridge Mass, Addison-Wiesley, 1949.

[39] K. Okuyama, M. Takayasu, H. Takayasu, Zipf's law in income distribution of companies, Pyhsica A 269(1999), 125-131.

[40] R.N. Mantegna, S.V. Buldyrev, A.L. Goldberger, Linguistic features of noncoding DNA sequences, Phys. Rev. Lett. 73 (1994) 3169-3172.

[41] J.S. Nicolis, I. Tsuda, On the Parallel between Zipf's Law and 1/f Processes in Chaotic Systems Possessing Coexisting Attractors, Prog. Theor. Phys. 82(1989) 254-274.

[42] B.B. Mandelbrot, The stable Paretian income distribution when the apparent exponent is near zero, Int. Econ. Rev. 4(1963) 111-115.

[43] A. Sato, H. Takayasu, Dynamic numerical models of stock market price: from microscopic determinism to macroscopic randomness, Physica A 250 (1998) 231-252.

[44] Y. Tachimori, T. Tahara, Clinical Diagnoses Following Zipf's Law, Fractals 10 (2002) 341-351.

[45] W. Fink, V. Lipatov, M. Konitzer, Diagnoses by general practitioners: Accuracy and reliability, International Journal of Forecasting 25 (2009) 784-793.

[46] M. Kouchi, Y. Tachimori, M. Hoshi, Difference in distribution of duration of hospitalization between general hospitals, long-term care facilities and sanatorium type medical care facilities (Japanese), JCMI 25 (2005) 245-248.

[47] O. Moriyama, H. Itoh, S. Matsushita, M. Matsushita, Long-Tailed Duration Distributions for Disability in Aged People. JPSJ 70(10) 2003: 2409-2412

[48] L.A. Lipsitz and A.L. Goldberger, Loss of 'Complexity' and Aging Potential Applications of Fractals and Chaos Theory to Senescence, JAMA 267(1992) 1806-1808.

[49] P. Bak, C. Tang, K. Wisenfeld, Self-organized criticality, Phys. Rev. A. 38 (1988) 364-374.

[50] R. F. Cancho, R. V. Solé, Least effort and the origins of scaling in human language, PNAS 100 (2003) 788-791.

[51] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, Nature 406 (2000) 378-382.

[52] S. N. Dorogovtsev, J. F. F. Mendes, Evolution of reference networks with aging, Phys. Rev. E 62 (2000) 1842.

[53] T. Maschberger, P. Kroupa, Estimators for the exponent and upper limit, and goodness-of-fit

tests for (truncated) power-law distributions, Mon. Not. R. Astron. Soc. 395 (2009) 931-942.

[54] R. F. Cancho, R. V. Solé, R. Köhler, Patterns in syntactic dependency networks, Phys. Rev. E 69 (2004) 051915.

[55] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, PNAS 99 (2002) 7821-7826.

[56] M. E. J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (2004) 066133.

Appendix 1 Three types of plot

A. Rank-size plot

When data are arranged in order of size, a plot of sizes of quantities vs. their rank order is called a "rank-size" or "Zipf's" plot. We can find many examples, such as rank orders of word frequencies, diagnosis frequencies, city sizes, and degree sizes of a network. Because "size" means "frequency" in word frequency and diagnosis frequency distributions, this plot is also called a "rank-frequency" plot.

In a power-law distribution, the size S_n of the quantity of the rank n is expressed by

$$S_n = \frac{A}{n^{\zeta}}, \qquad (A1-1)$$

where A is a constant. In this equation ζ is referred to as the Zipf exponent.

B. Size-frequency plot

This is a plot of frequencies vs. their size. That is, this plot is essentially the same as a probability density function $p(x)$. In a word frequency distribution, the number of words that have the same incidence in a text is plotted against the incidence. It is the same in a diagnosis frequency distribution.

In a power-law distribution, the frequency $f(x)$ of the quantity x is expressed by

$$f(x) = \frac{B}{x^{\alpha}} \quad (x > x_{min}), \qquad (A1-2)$$

where B is a constant.

C. CCD and complementary cumulative frequency plot

The CCD $P(x)$ is defined as

$$P(x) = \Pr(X \geq x), \qquad (A1-3)$$

where $x > x_{min}$.

However, we frequently plot a histogram of data instead of the CCD. Because this histogram is essentially the same as the CCD, we plot the histogram instead of the CCD without notification.

In a power-law distribution, the CCD, $P(x)$, of the quantity x is expressed by

$$P(x) = \Pr(X \geq x) = \int_x^{\infty} p(t)dt = \frac{D}{x^{\beta}}, \qquad (A1-4)$$

where D is a constant and obviously $\beta = \alpha - 1$.

The power exponent α in eq. (A1-2) is also expressed by the Zipf exponent ζ as follows:

$$\alpha = \frac{\zeta + 1}{\zeta} \ . \qquad (A1-5)$$

This equation is derived in the following way.

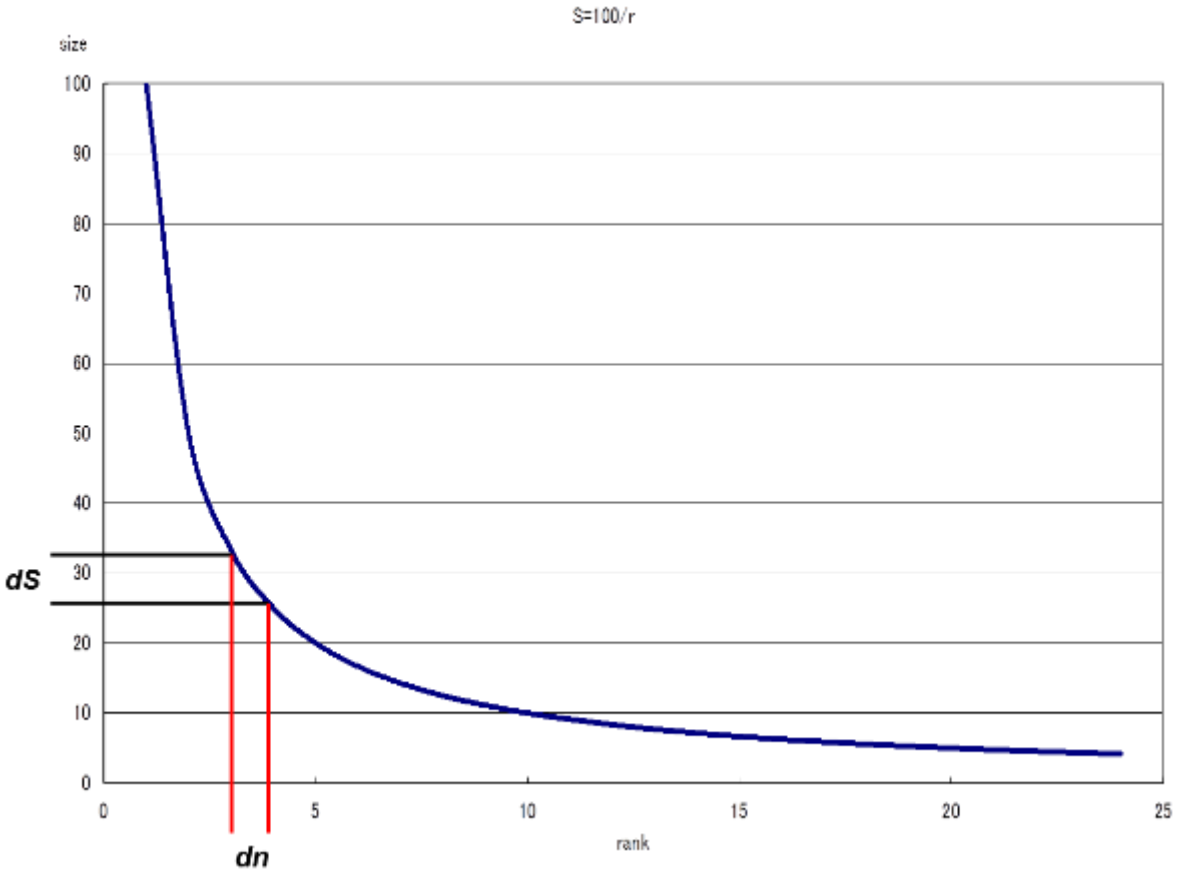


Figure A1-1. Rank-size plot (Zipf plot).

Figure *A*1-1 shows a rank-size plot of a power-law distribution. Let us suppose that F_n is a continuous function of *n*. Then the size *S* of the rank *n* element is expressed as

$$S = \frac{A}{n^{\zeta}} = s(n)\,. \qquad (A1-6)$$

Let the number of elements that exist between *S* and *S*+*dS* be |*dn*|, then

$$|dn| = |s^{-1}(S + dS) - s^{-1}(S)|\,. \qquad (A1-7)$$

Differentiating eq. (*A*1-5) with respect to *n* yields

$$dS = -\zeta \frac{A}{n^{\zeta+1}} dn\,. \qquad (A1-8)$$

That is,

$$|dn| = \frac{n^{\zeta+1}}{\zeta A} |dS|\,. \qquad (A1-9)$$

From eq. (*A*1-5), we obtain

$$n = \left(\frac{A}{S}\right)^{\frac{1}{\zeta}}. \qquad (A1-10)$$

Substituting eq. (*A*1-9) into eq. (A1-8) gives

$$|dn| = \frac{A^{\frac{1}{\zeta}}}{\zeta}|dS|\frac{1}{S^{\frac{\zeta+1}{\zeta}}} = \frac{B}{S^{\frac{\zeta+1}{\zeta}}}\ , \qquad (A1-11)$$

where $B = \frac{A^{\frac{1}{\zeta}}}{\zeta}|dS|$ is a constant. In the size-frequency plot, |*dn*| denotes the frequency *f* of the size *S*.

Then

$$f = |dn| = \frac{B}{S^{\frac{\zeta+1}{\zeta}}}\ . \qquad (A1-12)$$

Comparing eq. (*A*1-12) to eq. (*A*1-2), we obtain

$$\alpha = \frac{\zeta+1}{\zeta}\ .$$

Appendix 2　Estimating the power exponent by regression analysis

There are three types of plots that can express a set of data: rank-size plot, size-frequency plot, and complementary cumulative frequency plot (or CCD). If the data follow a power law, on these three types of plots with logarithmic scales, the data appear along a straight line, and we can estimate the power exponent α by regression analysis (Appendix 1). Although all three plots are expected to yield the same value of α, in practical use the CCD leads to the most accurate value.

Figure *A*2-1 shows the distribution of the degree of diagnosis (see Fig. 5-3a) on the three types of plots. Table *A*2-1 presents the values of the slopes and α derived from regression analysis. These results show that size-frequency plots have large variation, and α derived from this plot is estimated to be less than from the other plots. In fact, Newman [30] showed, by numerical analysis, that the estimate from the size-frequency plot is less accurate than from the other plots.

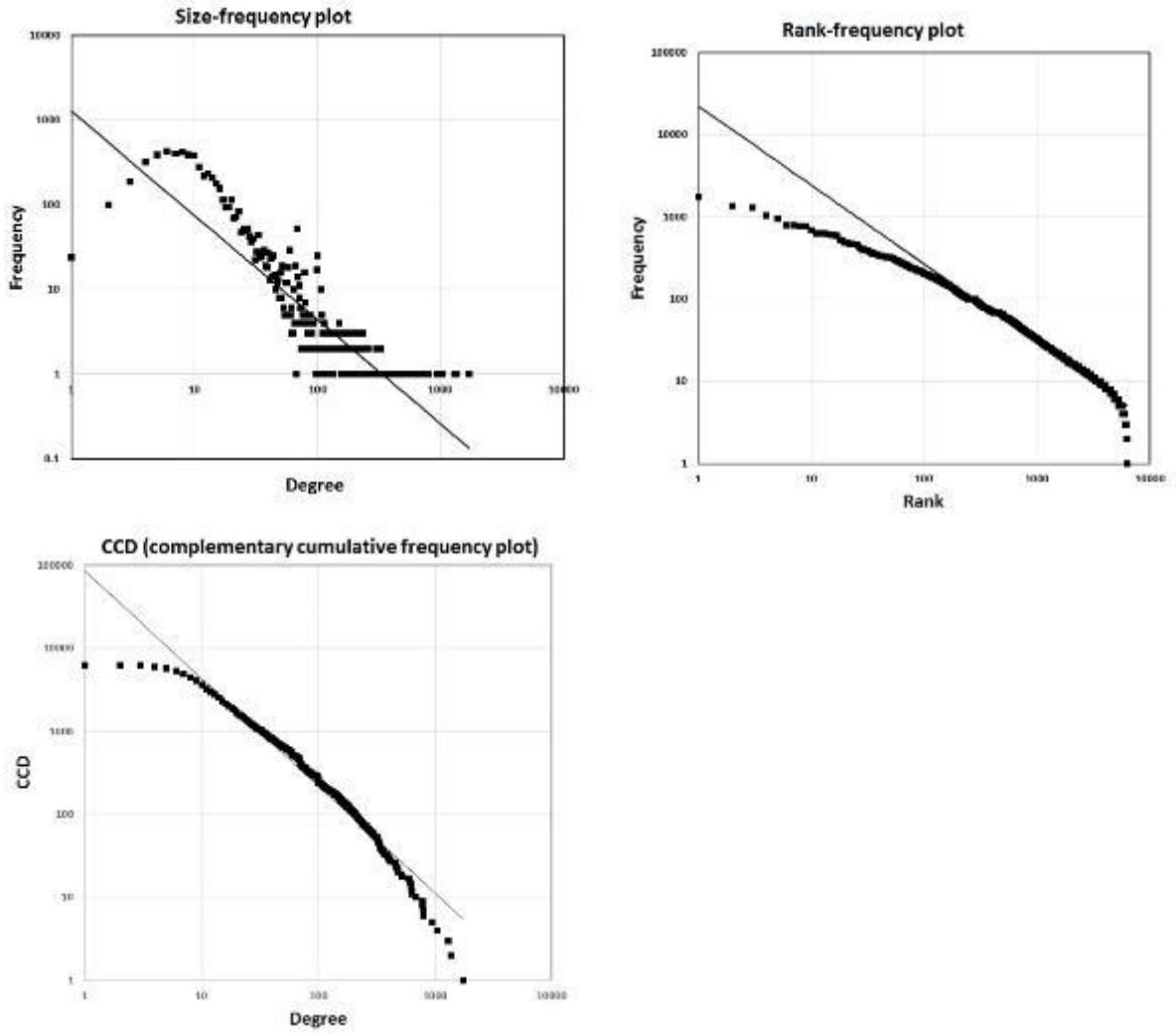


Figure A2-1. Three types of plots of the degree distribution.

Table A2-1. Results of regression analyses for three types of plots.

	Slope	R^2	Power exponent α derived from slope
Rank-size plot	0.96±0.003	0.94	2.001
Size-frequency plot	1.23±0.04	0.76	1.23
CCD	1.29±0.018	0.95	2.29
Maximum likelihood	–	–	2.10

R^2 is the coefficient of determination in linear regression analysis used to derive the slopes. The slopes are derived from linear regression analysis, and the power exponents are derived from the slopes (Appendix 1).

Appendix 3 Maximum likelihood estimate of the power exponent [30, 31]

Let $p(x)$ be the power-law distribution

$$p(x) = \frac{C}{x^{\alpha}} \qquad (x > x_{min}), \tag{A3-1}$$

where C is a constant. The constant C is given by

$$1 = \int_{x_{min}}^{\infty} \frac{C}{x^{\alpha}} dx = \frac{C}{1-\alpha}\left[\frac{1}{x^{\alpha-1}}\right]_{x_{min}}^{\infty} . \tag{A3-2}$$

If $\alpha > 1$, then

$$C = (\alpha - 1)x_{min}^{\alpha-1} . \tag{A3-3}$$

Therefore

$$p(x) = \frac{\alpha - 1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha} . \tag{A3-4}$$

Given a set of n values $x_i \geq x_{min}$ that are supposed to follow a power-law distribution with a power exponent α, the likelihood of the data set is defined as follows:

$$P(x|\alpha) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}}\left(\frac{x_i}{x_{min}}\right)^{-\alpha} . \tag{A3-5}$$

The value of α that maximizes this function is the maximum likelihood estimate of α. Because this value is equal to that which maximizes the logarithm of $P(x|\alpha)$, we can calculate the most likely value of α by maximizing the log likelihood. Setting $\partial \ln P(x|\alpha)/\partial\alpha = 0$, we obtain the maximum likelihood estimate

$$\hat{\alpha} = 1 + n\left[\sum_{i=1}^{n} \ln\frac{x_i}{x_{min}}\right]^{-1} . \tag{A3-6}$$

The standard error of $\hat{\alpha}$ is estimated as follows [31]:

$$\sigma = \frac{\hat{\alpha} - 1}{\sqrt{n}} + O(1/n) . \tag{A3-7}$$