Structure and dynamics of human disease-complication network

Xiong-Fei Jiang^{1,2*}, Long Xiong³, Ling Bai², Jie Lin¹, Jing-Feng Zhang¹, Kun Yan¹, Jia-Zhen Zhu¹, Bo Zheng^{3,4,5*}, Jian-Jun Zheng^{1*}

1 Hwa Mei Hospital, University of Chinese Academy of Sciences, China, Ningbo 315010, PRC

2 College of Finance and Information, Ningbo University of Finance and Economics, Ningbo 315175, PRC

3 Department of Physics, Zhejiang University, Hangzhou 310027, PRC

4 School of Physics and Astronomy, Yunnan University, Kunming, 650091, PRC

5 Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, PRC

These authors contributed equally to this work.

* jiangxiongfei@nbufe.edu.cn, zhengbo@zju.edu.cn, zhengjianjun@ucas.ac.cn

Abstract

A complication is an unanticipated disease arisen following, induced by a disease, a treatment or a procedure. We compile the Human Disease-Complication Network from the medical data and investigate the characteristics of the network. It is observed that the modules of the network are dominated by the classes of diseases. The relations between modules are unveiled in detail. Three nontrivial motifs are identified from the network. We further simulate the dynamics of motifs with the Boolean dynamic model. Each motif represents a specific dynamic behavior, which is potentially functional in the disease system, such as generating temporal progressions and governing the responses to fluctuating external stimuli.

Author summary

Advances in molecular biology lead to a new discipline of network medicine, investigating human diseases in a networked structure perspective. Recently, clinical records have been introduced to the research of complex networks of diseases. An important available medical dataset that has been overlooked so far is the complications of diseases, which are vital for human beings. We compile the Human Disease-Complication Network, representing the causality between the upstream diseases and their downstream complications. This work not only helps us to comprehend why certain groups of diseases appear collectively, but also provides a new paradigm to investigate the dynamics of disease progression. For clinical applications, the investigation of complications may yield new approaches to disease prevention, diagnosis and treatment.

Introduction

Advances in molecular biology lead to a new discipline of network medicine, which has multiple potential biological and clinical applications to human disease [1–5]. Much

attention has been paid to molecular networks, including protein interaction networks [6–9], regulatory networks [10], metabolic networks [11–14], RNA networks [15–17], and kinase-substrate networks [18], whose nodes are molecules linked to each other by physical interactions. In parallel, an increasing number of studies focus on phenotypic networks, including co-expression networks [19,20], and genetic networks [21–23].

The highly interconnected property of the interactome suggests that diseases are not independent of one another, whether at the molecular level or phenotypic level. From psychiatric patient records, comorbidities are determined by the co-occurring frequency of disease pairs in patients more often than expected [24]. Khan et. al. investigate the chronic disease progression pattern of Type 2 Diabetes by the comorbidity network drawn from the Australian healthcare records [25]. The phenotypic disease network with comorbidity patterns is extracted from more than 30 million medicare records, and particularly the dynamic progression patterns are investigated [26]. Further, the genetic overlap between diseases is estimated with the comorbidity correlation, which is extracted from the disease history of 1.5 million patients [27]. Similarly, Zhou et. al. observe that the symptom-based similarity between two diseases positively correlates with the number of shared genes from the human disease network [28]. The projection of such network-based dependencies between disease genes and disease phenotypic features has brought the concept of diseasesome [1,29–31].

To go beyond the global features, the motif is utilized to characterize the local properties of those biological networks. The pioneering work by Milo et al. identifies motifs in networks from biochemistry, neurobiology, ecology, and engineering [32]. Then motifs have been observed in many biological networks [33–35], and their dynamics have been investigated further [34,36].

A critical medical dataset that has been overlooked so far is the complications of diseases, which are vital for patients in clinical practice. For example, complications may affect clinical decisions for physicians in certain circumstances [37]. A complication is an unanticipated disease arisen following, induced by a disease, a treatment or a procedure. Diseases form a networked structure with their complications. For example, COVID–19 generates numerous complications, such as venous thromboembolism and acute kidney injury, etc [38]. We compile the Human Disease-Complication Network, representing the causality between the upstream diseases and their downstream complications. We systematically investigate the structure of the network, and pay attention to the disease modules. Taking into account the complexity of network dynamics, we further study the motifs, with which the dynamics can be depicted in the disease system. This work not only helps us understand how different medical subdisciplines organize, but also provides a comprehensive understanding of why certain groups of diseases appear collectively.

Materials and methods

Materials

We collected data from the Clinical Medicine Knowledge Database, including 6715 diseases. The complications of diseases are extracted from the descriptions of diseases in the database. A node denotes a disease, and a directed link from the i -th node to the j -th node is drawn if the i -th disease generates the j -th complication. Then, we compile the directed Human Disease-Complication Network (HDCN), containing 3471 nodes and 9164 directed links. The HDCN is represented by the adjacency matrix $\mathbb{G} = [G_{ij}]$, where $G_{ij} = 1$ if there is a directed link pointing from the i -th node to the

j -th, otherwise $G_{ij} = 0$. The giant connected component includes 3390 nodes and 9116 directed links. The data is available in the supporting information.

Methods

k -core

The k -core analysis is able to uncover a nucleus set of nodes, i.e., a set of nodes with a high degree connected to each other. It has been widely used in networks to identify the kernels more robustly than simply through the ranking of centrality measures [39, 40]. The k -core of a network consists of nodes i with the degree $k_i \geq k$, and the k -core could be extracted by the iterative removal of all nodes i with degrees $k_i < k$ when $k > 0$.

Motif

The motifs are defined as local patterns occurring in the real network significantly more frequently than in randomized networks with the same degree sequence [32, 34]. For any given network, the occurrence number N_g of the g -th connected subsets is related to the network size and the degree distribution. For measuring the statistical significance, the randomized ensemble of networks is generated as a null model [32]. The statistical significance Z score is defined as

$$z_g = \frac{N_g - \langle N_g \rangle}{\sigma}, \quad (1)$$

where N_g is the appearance number of the g -th subset in the real HDCN, $\langle N_g \rangle$ and σ are the average appearance number and standard deviation in the randomized ensemble of networks. The motifs are those subsets with significantly higher frequency in the real network than randomized ones, measured by the Z score. In this paper, we mainly focus on the 3-node and 4-node connected subsets.

Boolean dynamics

The Boolean dynamics is a common model in gene regulatory networks and signal transduction networks [41, 42]. Each node in a Boolean network represents a sub-cellular component such as protein, gene, transcription factor or metabolite. The states of input nodes i are described by a binary value X_i . $X_i = 1$ represents that the component i is active or expressed, $X_i = 0$ means that it is inactive or not expressed. The state of node j at time $t + 1$, $X_j(t + 1)$, is determined by a logic operation together with the current state of its upstream regulators $X_i(t)$. The logic operation is the Boolean update function, denoted by the logic operators or a weighted sum of the inputs to an activation threshold. The output nodal dynamics is described by a differential equation like

$$\frac{dX_j}{dt} = F(X_1, T_{1j}; \dots; X_i, T_{ij}) - \alpha X_j, \quad (2)$$

where $F(X_1, T_{1j}; \dots; X_i, T_{ij})$ is the Boolean update function.

Here, we introduce the Boolean dynamic model to simulate the complication progression in the HDCN. The disease progression in the HDCN is compared to other biological processes, such as gene regulatory networks. A complication is generated by the collective effect of upstream diseases, and it will presumably self-cure after the upstream diseases are cured. It is reasonable to utilize the Boolean dynamic model to investigate the complication dynamics. Therefore, X_i represents the activation of the

i -th upstream disease, T_{ij} is the activation threshold of the i -th disease to the j -th complication, α is the lifetime of the cured disease, and $F(X_1, T_{1j}; \dots; X_i, T_{ij})$ denotes the summarized effect of overall i upstream diseases on the j -th complication.

Results

Qualitatively, the HDCN is formed by very few disconnected components and a large giant connected component (see Fig 1), including approximately 93.64% of all diseases. The majority of the nodes exhibit low clustering coefficients ($0 < c_i \leq 0.1$), representing a sparse local structure, while the minority demonstrate high clustering coefficients ($0.1 \leq c_i < 1.0$), indicating a dense local structure. The disease-complication relations in the HDCN are specific, 16%, 32% and 21% of the diseases generate 1, 2 and 3 complications respectively, while only 31% diseases cause more than 3 complications. Vice versa, 73% of the diseases are only caused by a single disease, 10% are caused by two. We calculate the probability distribution of the degrees, which are distinguished between in-degree, i.e. the number of edges pointed to some vertex, and out-degree, i.e. the number of edges pointing away from it. As shown in Fig 2, the probability distribution of out-degree exhibits an approximative power-law, while the one of in-degree decays in an exponentially. The out-degree distribution of the HDCN indicates that most diseases generate only a few other diseases, whereas a few phenotypes such as Scleroderma ($k^{out} = 23$) and Renal Damage due to Hyperthyroidism ($k^{out} = 19$) generate a large number of distinct diseases. The in-degree distribution decays faster than the out-degree distribution, but still significantly deviates from the Poisson distribution expected for a random graph.

In the HDCN, a disease presents a streaming structure: generating complications, and as a complication caused by other diseases. Therefore, we introduce in-degree k^{in} and out-degree k^{out} to quantitatively categorize the diseases into these three structural levels, i.e., upstream, intermediate and downstream. The diseases with highest k^{out} and k^{in} are listed in Table 1. A node with high k^{out} and low k^{in} is an upstream disease such as Acute Lymphoblastic Leukemia ($k^{out} = 16, k^{in} = 1$), since it could lead to other diseases yet be hardly produced by others. On the contrary, a node with low k^{out} and high k^{in} is a downstream disease such as Pneumonia ($k^{out} = 0, k^{in} = 181$), which almost results from the other diseases, nevertheless could hardly produce others. A node, which could generate other diseases and be produced by others, is an intermediate disease. In this case, the terms of upstream, intermediate and downstream are not absolute, but relative. Although it is less likely to do so, a downstream disease could cause complications. Qualitatively, diseases with the highest k^{in} are usually general diseases, while nodes with the highest k^{out} are frequently associated with specific organs.

To quantify the correlation between the in-degree and out-degree of nodes, we compute the Spearman rank correlation coefficient SC between these two ranking [43]. A negative value of $SC = -0.26$ indicates that the in-degree and out-degree in the same node are significantly asymmetric, i.e., the disease connecting with more upstream diseases usually results in fewer downstream complications, and vice versa.

With the k -core method [39, 40], we identify a small well-connected nucleus, consisting of 98 diseases with the maximal coreness= 7 (See S1 Table). The components of the nucleus mainly belong to three classes of diseases, i.e., Cardiovascular, Hepatology, and General Diseases. It suggests that these diseases play a crucial role in the progression of complications.

Fig 1. Human Disease-Complication Network. Each node corresponds to a distinct disease, colored based on the disease class to which it belongs. Names of the 14 disease classes are indicated on the bottom.

Disease Modules

Although the HDCN layout is generated without any priori knowledge on disease classes, it is naturally and visibly clustered according to major disease classes. In order to quantitatively understand this clustering nature, we identify the community structure of the HDCN. In a complex network, the community structure is the grouping of nodes into clusters with a high density of internal links, while including a relatively low density of links between clusters [44,45].

As shown in Fig 1, the diseases in the HDCN form several functional modules dominated by the organs in which the diseases occur. We identify 14 modules, including Orthopedics, Gastroenterology, Cancer, Neurosurgery, Dermatology, Urology, Cardiovascular, Respiratory, Ophthalmology, Infectious Disease, General Disease, Hepatology, Gynecology-Nephrology-Urology (GNU) and Endocrine-Metabolic-Hematology-Rheumatology (EMHR). Hepatology includes Hepatobiliary, Pancreatic and Splenic Diseases. It should be noted that the Cancer module only includes a small fraction of cancers. The majority of cancers are

Fig 2. Degree distribution. The out-degree and in-degree distributions $P(k)$ of the HDCN. In-degree is the number of edges pointed to some vertex, and out-degree is the number of edges pointing away from it. the probability distribution of out-degree exhibits an approximative power-law, while the one of in-degree decays in an exponentially.

distributed in the related organs, rather than form a single module as in the gene-disease network [30]. This difference is rooted in that the cancers in the former one usually cause complications in the related organs, while the latter one may share the same genes and generate dense connections between each other. Gynecology, Nephrology and Urology are grouped into a single module, since all three belong to the genitourinary system in which the related organs are physiologically close and interlinked by the blood supply and some meatus. In a similar vein, the diseases in Endocrine, Metabolic, Hematology and Rheumatology form a single module, which has a global effect on the whole human disease system.

To quantify the influence of modules in the progression of complications, we define the complicating ability of the p -th module as

$$C_p = \frac{\kappa_p^{out} - \kappa_p^{in}}{\kappa_p^{out} + \kappa_p^{in}}, \quad (3)$$

where κ_p^{out} and κ_p^{in} are the total out-links and in-links of the p -th module connecting to other modules. The high value of C_p suggests that the module can trigger downstream complications with high possibility, while the high absolute value of negative C_p means that the module is prone to be caused by upstream diseases. As shown in Table 2, Orthopedics, Dermatology, and Cancer are the modules with the highest C_p . Therefore, these three are the most influential disease modules, which can generate the downstream diseases with great capability. Meanwhile, General Disease, Cardiovascular and Infectious Disease are the modules with the lowest C_p . It implies that they can be triggered by other diseases easily.

To further unveil the causal relations between the disease modules, we define the density of directed links from the i -th to the j -th module, $\rho_{pq} = \frac{2L_{pq}}{N_p N_q}$, where L_{pq} is the number of directed links from the p -th module to the q -th module, N_p is the number of diseases in the p -th module, and $\rho_{pq} \neq \rho_{qp}$. The average density of links for the entire network is 6.8×10^{-4} , and the one within the modules is 2.0×10^{-2} . We set the average density of links between the modules, $\bar{\rho}_{pq}^{be} = 5.4 \times 10^{-4}$, as the threshold to

Table 1. Diseases with the highest out-degree k^{out} and in-degree k^{in} in the HDCN. RDH: Renal Damage due to Hyperthyroidism, PPNS: Pediatric Primary Nephrotic Syndrome, ALL: Acute Lymphoblastic Leukemia, PNH: Paroxysmal Nocturnal Hemoglobinuria, SRHD: Senile Rheumatic Heart Disease, CACCR: Cardiac Arrest and Cardiopulmonary-Cerebral Resuscitation, CAHD: Coronary Atherosclerotic Heart Disease, SSH: Spontaneous Subarachnoid Hemorrhage, SIE: Subacute Infective Endocarditis, ATH: Alimentary Tract Hemorrhage.

Disease	k^{out}	Disease	k^{in}
Scleroderma	23	Shock	302
RDH	19	Heart Failure	261
Chronic Cor Pulmonale	18	Arrhythmia	211
Hyperthyroidism	18	Pneumonia	181
PPNS	18	Septicemia	179
Viral Hemorrhagic Fever	16	Hypertension	163
Epidemic Hemorrhagic Fever	16	Dehydration	136
ALL	16	Jaundice	132
PNH	15	Ileus	129
SRHD	15	Diarrhea	111
CACCR	14	Myocarditis	95
Acute Infective Endocarditis	13	Diabetes Mellitus	91
Cholelithiasis	13	Ascites	91
Pediatric Acute Pancreatitis	13	Pulmonary Edema	88
Senile Hyperthyroidism	13	Hepatic Cirrhosis	86
Hepatic Cirrhosis	12	ATH	86
CAHD	12	Thrombosis	68
SSH	12	Congestive Heart Failure	65
Mitral Stenosis	12	Pleurisy	58
SIE	12	Myocardial Infarction	58

Table 2. Modules in the HDCN. Here N is the number of diseases in modules, m is the number of links connecting nodes within modules, M^{out} and M^{in} are the total numbers of out-links and in-links connecting nodes in other modules. Hepatology includes Hepatobiliary, Pancreatic and Splenic Diseases. GNU: Gynecology-Nephrology-Urology, EMHR: Endocrine-Metabolic-Hematology-Rheumatology.

	N	m	M^{out}	M^{in}	r
Orthopedics	77	84	29	11	0.45
Gastroenterology	184	279	177	179	-0.01
Cancer	30	37	26	11	0.41
Neurosurgery	322	477	231	203	0.06
Dermatology	140	168	145	55	0.45
Urology	65	72	47	39	0.09
Cardiovascular	413	921	280	438	-0.22
Respiratory	248	432	255	181	0.17
GNU	230	341	166	149	0.05
Hepatology	353	670	359	282	0.12
Ophthalmology	177	267	61	50	0.10
EMHR	306	469	369	260	0.17
Infectious Disease	297	445	254	342	-0.15
General Disease	356	531	274	473	-0.27

determine the backbone of the HDCN. In other words, only the links whose ρ is higher than ρ_{pq}^{be} remain.

Fig 3 outlines the relations and directions between the disease modules. EMHR and Dermatology are the two modules with the strongest influence. For example, EMHR generates almost half of the complication modules, including Cancer, Hepatology, GNU, Cardiovascular, and General Disease. On the other hand, Infectious Disease and General Disease modules are the two most attractive modules. Infectious Disease is possibly caused by Dermatology, Respiratory, Gastroenterology, Hepatology, and Neurosurgery with high possibilities. Ophthalmology and Orthopedics modules are relatively independent in the HDCN. Ophthalmology only causes Neurosurgery, and Orthopedics only leads to General Diseases. It implies that Ophthalmology and Orthopedics have a specific impact on the complication progression. Gastroenterology and Hepatology have the strongest bi-directed causalities in the HDCN since both of them belong to the digest system. Besides, EMHR may lead to Cardiovascular, and Dermatology may result in Infectious Disease with high probabilities, respectively. Notably, Cancer module does not have a very high complication effect on other modules. The ultimate reason is that there are only 32 cancers included in Cancer module (40 diseases in total), and most cancers are classified into the corresponding modules of organs.

Fig 3. Causal connections between disease modules. The node size represents the number of diseases in the module, while the width of the line suggests the density of the connection between two modules.

Motif

To mitigate the complexity of the HDCN, we investigate the local interactions among small connected subsets, i.e., motifs. The statistical significance of the network motifs is evaluated by comparison with randomized networks having the same characteristics as the HDCN. We identify three kinds of motifs, i.e., the feedforward loop, the bi-fan, and the overlapping feedforward loop, as shown in Table 3.

Feedforward loop

The feedforward loop (FFL) is formed by three diseases: U (upstream disease) generates M (mediated disease) and D (downstream disease), while M also causes D , as illustrated in Table 3. For instance, pediatric food allergy (U), allergic purpura (M) and arrhythmia (D) consist of a typical FFL. To deeply understand the medical function of the FFL, we search all compositions of diseases in the motif. The FFL appears a total 1733 times in the HDCN. Thereinto, U , M and D nodes contain 722, 274, 193 unique diseases, respectively. The diseases with top 10 appearing times in the three nodes are displayed in S2 Table. Most of the diseases in U belong to the upstream diseases, the diseases in M are generally the intermediate diseases, while the diseases in D tend to be the downstream diseases.

Table 3. Network motifs detected in the HDCN. For each motif, the numbers of appearances in the real network N_{real} , and in the randomized networks $N_{rand} \pm \sigma$ are listed. The P value of all motifs is $P < 0.01$, as determined by comparison to 1000 randomized networks. The Z score = $(N_{real} - N_{rand}) / \sigma$, as a measure of statistical significance. U : upstream disease, M : mediated disease, D : downstream disease.

		N_{real}	$N_{rand} \pm \sigma$	Z score
FFL		1733	569 ± 39	30
Bi-fan		43340	23136 ± 966	21
OFFL		10976	937 ± 279	36

We calculate the average in-degree $\bar{k}^{in}$ and out-degree $\bar{k}^{out}$ for U , M , D nodes in the FFL, listed in Table 4. It is observed that $\bar{k}_U^{out} > \bar{k}_M^{out} > \bar{k}_D^{out}$, whereas $\bar{k}_U^{in} < \bar{k}_M^{in} < \bar{k}_D^{in}$. These results quantitatively confirm that U , M and D nodes in the FFL tend to be the upstream, intermediate and downstream diseases, respectively.

Bi-fan

The bi-fan is comprised of four nodes, classified into two layers between which there exist overlapping interactions, as demonstrated in Table 3. For instance, group A streptococci (U_1), brucellosis (U_2), pneumonia (D_1) and endocarditis (D_2) consist of a bi-fan. We search all diseases appeared in the bi-fan as well. U_1 , U_2 , D_1 , D_2 nodes encompass 1399, 1401, 317 and 327 unique diseases, respectively. The diseases with top 10 appearing times in the four nodes are listed in S3 Table. Most of the diseases in U_1 and U_2 belong to upstream diseases, while D_1 and D_2 tend to be downstream diseases. Hence, we call U_1 and U_2 as upstream disease layer, while D_1 and D_2 as downstream disease layer. $\bar{k}^{out}$ and $\bar{k}^{in}$ of four nodes in the bi-fan are shown in Table 4.

The average in(out)-degree of U_1 and U_2 are almost the same, while the average in(out)-degrees of D_1 and D_2 are similar, since the structure of the bi-fan is

symmetric. Then $\bar{k}_{U_1}^{out} \approx \bar{k}_{U_2}^{out} > \bar{k}_{D_1}^{out} \approx \bar{k}_{D_2}^{out}$, while $\bar{k}_{U_1}^{in} \approx \bar{k}_{U_2}^{in} < \bar{k}_{D_1}^{in} \approx \bar{k}_{D_2}^{in}$. The results ascertain that U_1 and U_2 nodes tend to be the upstream diseases, meanwhile D_1 and D_2 nodes would be the downstream diseases in the bi-fan.

Overlapping feedforward loop

Further, we identify a novel motif called overlapping feedforward loop (OFFL), which has not been observed in other networks. As shown in Table 3, the OFFL consists of two FFLs, sharing M and D nodes. For instance, viral hemorrhagic fever (U_1), cholelithiasis (U_2), septicemia (M) and jaundice (D) form an OFFL. There are 548, 546, 132 and 97 unique diseases included in U_1 , U_2 , M and D nodes respectively. The diseases with top 10 appearing times in the four nodes are listed in S4 Table . Most of the diseases in U_1 and U_2 nodes can be categorized as upstream, whilst diseases in M as intermediate, and diseases in D as downstream.

Table 4 displays $\bar{k}^{out}$ and $\bar{k}^{in}$ for the four nodes in the OFFL. The average in(out)-degrees of U_1 node are almost the same as those of U_2 node. $U_{1,2}$ nodes have a high $\bar{k}^{out}$ with low $\bar{k}^{in}$, while the D nodes have a high $\bar{k}^{in}$ with low $\bar{k}^{out}$. Besides, $\bar{k}^{in}$ and $\bar{k}^{out}$ of M are intermediate between $U_{1,2}$ and D . It quantitatively supports that U_1 and U_2 in the OFFL tend to be upstream diseases, M are supposed to be intermediate diseases, while D would be downstream diseases. This result consolidates the earlier proposition that OFFL is formed with two FFLs, which share the same intermediate and downstream diseases.

Table 4. Average in-degree $\bar{k}^{in}$ and out-degree $\bar{k}^{out}$ of nodes in the motifs. The values are computed from the weighted average. U : upstream disease, M : mediated disease, D : downstream disease.

FFL	U	M	D	
$\bar{k}^{out}$	8.55	6.59	1.41	
$\bar{k}^{in}$	1.90	52.86	124.64	
Bi-fan	U_1	U_2	D_1	D_2
$\bar{k}^{out}$	9.18	8.89	1.12	0.90
$\bar{k}^{in}$	3.12	1.77	16.94	156.90
OFFL	U_1	U_2	M	D
$\bar{k}^{out}$	7.82	7.58	5.04	0.51
$\bar{k}^{in}$	1.75	1.84	148.25	204.59

Dynamics

The frequency of motifs appeared significantly higher than in a random network, implying that the motifs may represent medical functions. The dynamics of motifs offers an insight to understand the progression of diseases further. Here we introduce the Boolean dynamic model to simulate the dynamics of the motifs in the HDCN, and only consider the effect of several typical mediated pathways on the dynamics in this paper.

Feedforward loop

In the FFL motif, the effect of the upstream disease U is transferred to the downstream disease D via two pathways, a direct one of $U \rightarrow T$ and a mediated one of

$U \rightarrow M \rightarrow D$. U and M may act in an AND-gate or OR-gate manner to control D .

For the AND-gate, the differential equations are written as

$$\frac{dM}{dt} = F(U, T_{um}) - \alpha M, \quad (4)$$

and

$$\frac{dD}{dt} = F(U, T_{ud}) F(M, T_{md}) - \alpha D, \quad (5)$$

where $F(u, T) = (u/T) / (1 + (u/T))$. In the remainder of this paper we will adopt this function for all AND-gate, unless stated otherwise. The dynamic behaviors of the AND-gate are shown in Fig 4.a for different parameters. The upstream disease U generates the intermediate disease M , then U and M jointly stimulate the downstream disease D . When the activation of U is transient, the morbidity of D is low. In other words, the stimulus can not be transduced through the circuit easily. The downstream disease D will be activated only if U is a persistent stimulus. Meanwhile, M will decrease rapidly, once the stimulus of U vanishes. The FFL motif may resist stochastic perturbations in the environment, and prevent the wild disease progression in the disease system. In fact, the FFL is also a common motif detected in transcriptional gene regulation networks and neuronal connectivity networks [34]. Therefore, the AND-gate of the FFL is probably a general mechanism to protect biological functions.

For the OR-gate, Eq 5 is substituted by

$$\frac{dD}{dt} = G(U, T_{ud}, M, T_{md}) + G(M, T_{md}, U, T_{ud}) - \alpha D, \text{ where}$$

$G(u, T_u, v, T_v) = (u/T_u) / (1 + (u/T_u) + (v/T_v))$. If not stated otherwise, the function will be adopted for all OR-gate. The results are exhibited in Fig 4.b. In the OR-gate, U and M can stimulate the downstream disease D independently. The response rate of D in the OR-gate is higher than the one in the AND-gate. In contrast, the relaxation rate of D in the OR-gate is lower than that in the AND-gate. D would be generated even if the stimulus of U is transient, since the causing pathway of $U \rightarrow D$ and $M \rightarrow D$ are independent. Besides, the relaxation time of D in the OR-gate is much longer than that in the AND-gate when the stimulus of U vanishes. The pathway of $M \rightarrow D$ would enhance the morbidity of D , whenever the stimulus of U is persistent or transient. Hence the circuit of OR-gate may aggravate the disease progression.

Fig 4. Dynamic features of the feedforward loop motif. $T_{ud} = 0.5$, $T_{wd} = 0.3$, and $\alpha = 1.0$.

Bi-fan

In the bi-fan motif, upstream diseases $U_{1,2}$ have two affecting pathways, $U_1 \rightarrow D_{1,2}$ and $U_2 \rightarrow D_{1,2}$. Likewise, U_1 and U_2 may act in an AND-gate or OR-gate manner to control $D_{1,2}$.

For the OR-gate, both $D_{1,2}$ could be activated by U_1 and U_2 independently, i.e., $U_1 \rightarrow D_{1,2}$ and $U_2 \rightarrow D_{1,2}$. The differential equations of the bi-fan are written as

$$\frac{dD_1}{dt} = [G(U_1, T_{u_1d_1}, U_2, T_{u_2d_1}) + G(U_2, T_{u_2d_1}, U_1, T_{u_1d_1})] - \alpha D_1, \quad (6)$$

and

$$\frac{dD_2}{dt} = [G(U_1, T_{u_1d_2}, U_2, T_{u_2d_2}) + G(U_2, T_{u_2d_2}, U_1, T_{u_1d_2})] - \alpha D_2, \quad (7)$$

As shown in Fig 5, for the transient signals, if there is only one upstream disease (U_1 or U_2) with a transient signal, the downstream diseases can not be activated. Only when U_1 and U_2 co-occur simultaneously will the morbidity of D_1 and D_2 be high. To some extent, the FFL motif provides a temporal mechanism to prevent the wild disease progression of complications, while the bi-fan motif offers a spatial protection mechanism. Intriguingly, the dynamics of M with the type-1 stimulus are different from those with the type-2 stimulus. The latter one is sensitive to the variation of T_{u_1d} , whereas the former is not. Since U_1 remains persistent in type-2 stimulus, while it becomes transient in type-1 stimulus Hence, $D_{1,2}$ nodes response sensitively to the persistent stimulus.

For the AND-gate, both $D_{1,2}$ should be activated by U_1 and U_2 jointly. Eq 6 and 7 is substituted by $\frac{dD_1}{dt} = F(U_1, T_{u_1d_1}) F(U_2, T_{u_2d_1}) - \alpha D_1$ and $\frac{dD_2}{dt} = F(U_1, T_{u_1d_2}) F(U_2, T_{u_2d_2}) - \alpha D_2$, respectively. The results are shown in Fig 5. It is evident that D_1 and D_2 can be activated only if U_1 and U_2 activated simultaneously. The response rates of $D_{1,2}$ in the OR-gate are higher than the ones in the AND-gate, while the relaxation rates of $D_{1,2}$ in the OR-gate are slower than the ones in the AND-gate.

Overlapping feedforward loop

The OFFL has a more complicated structure than the other two. For simplicity, we set that U_1 and U_2 always act in an OR-gate manner to control M and D . However, M still acts with U_1 and U_2 in an AND-gate or OR-gate manner to control D . We will discuss this AND-gate or OR-gate below.

For a mixed model of AND-gate and OR-gate, the differential equations of the OFFL are written as

$$\frac{dM}{dt} = [G(U_1, T_{u_1m}, U_2, T_{u_2m}) + G(U_2, T_{u_2m}, U_1, T_{u_1m})] - \alpha M, \quad (8)$$

and

$$\begin{aligned} \frac{dD}{dt} = & [G(U_2, T_{u_2d}, U_1, T_{u_1d}) + G(U_1, T_{u_1d}, U_2, T_{u_2d})] F(M, T_{md}) \\ & - \alpha D. \end{aligned} \quad (9)$$

The results are shown in Fig 6. The progression level of D disease in the OFFL is significantly lower than the bi-fan, under the same stimulus circumstance of U_1 and U_2 . It is counter-intuitive to some extent, since there is an extra pathway from M to D in the OFFL. Two pathways activating $D_{1,2}$ in the bi-fan are direct, and therefore U_1 and U_2 can activate $D_{1,2}$ with transient stimulus. By contrast, besides two direct

Fig 5. Dynamic features of the bi-fan motif. $T_{u_1d_2} = 0.7$, $T_{u_2d_1} = 0.3$, $T_{u_2d_2} = 0.6$ and $\alpha = 1.0$. (a) The OR-gate with type-1 stimulus of $U_{1,2}$, (b) The AND-gate with type-1 stimulus of $U_{1,2}$, (c) The OR-gate with type-2 stimulus of $U_{1,2}$, (d) The AND-gate with type-2 stimulus of $U_{1,2}$. Type-1 stimulus: two transient stimuli in U_1 , while one transient and one persistent stimuli in U_2 . Type-2 stimulus: one transient and one persistent stimuli in U_1 , while two transient stimuli in U_2 .

pathways, there is an indirect pathway mediated by M in the OFFL. Hence the contribution of pathways of $U_1 \rightarrow D$ and $U_2 \rightarrow D$ are lowered by the pathway of $M \rightarrow D$. Meanwhile, the pathway $M \rightarrow D$ should be activated by a persistent stimulus of U_1 . Therefore, it is reasonable that the progression level of D decreases instead in the OFFL. The loop structure in the OFFL is vital to mediate the disease progression in the HDCN. Similar to the FFL, the AND-gate in the OFFL may reduce the morbidity of complications.

For the OR-gate, Eq 9 are written as $\frac{dD}{dt} = G(U_2, T_{u_2d}, U_1, T_{u_1d}, M, T_{md}) + G(U_1, T_{u_1d}, U_2, T_{u_2d}, M, T_{md}) + G(M, T_{md}, U_1, T_{u_1d}, U_2, T_{u_2d}) - \alpha D$. where $G(u, T_u, v, T_v, w, T_w) = (u/T) / (1 + (u/T_u) + (v/T_v) + (w/T_w))$. The results are shown in Fig 6. Likewise, D in the OR-gate presents a dramatic response and longer relaxation time than in the AND-gate. Hence, it is validated that the AND-gate could aggravate the disease progression. Similar to the bi-fan motif, D in the OFFL is sensitive to the variation of T_{u_1d} .

Discussion and future work

Complications may affect clinical decisions for physicians under high pressure and great uncertainty. Singh shows that physicians' delivery mode decisions (i.e., to perform a

Fig 6. Dynamic features of the OFFL motif. $T_{u_1d_2} = 0.7$, $T_{u_2m} = 0.3$, $T_{u_2d} = 0.6$, $T_{md} = 0.6$, and $\alpha = 1.0$. (a) The OR-gate with type-1 stimulus of $U_{1,2}$, (b) The AND-gate with type-1 stimulus of $U_{1,2}$, (c) The OR-gate with type-2 stimulus of $U_{1,2}$, (d) The AND-gate with type-2 stimulus of $U_{1,2}$. Type-1 stimulus: two transient stimuli in U_1 , while one transient and one persistent stimuli in U_2 . Type-2 stimulus: one transient and one persistent stimuli in U_1 , while two transient stimuli in U_2 .

vaginal versus a cesarean delivery) are influenced by such a heuristics [37]. If the prior patient had complications in one delivery mode, the physician will be more likely to switch to the other, and likely inappropriate-delivery mode for the subsequent patient, regardless of patient's indicators. More importantly, this strategy presents small but significantly negative effects on patient health outcomes and increases resource use. There is no clinical reason why the delivery decisions for two separate patients should be causally related to each other. In applications, physicians may overcome the shortcoming of biased decision-making and make appropriate clinical decisions under high pressure, if they have a comprehensive and accurate complication map to evaluate the disease progression quantitatively. Therefore, our work provides a potential framework to aid complex decision-making in clinical practice.

Limitations. There are several limits in the current work. First, we assume nodes are homogeneous, i.e., all diseases have the same capability to cause the downstream complications in Boolean dynamics. However, different diseases have various likelihood of generating complications in practice. For example, COVID-19 has a high likelihood of generating acute kidney injury, a medium likelihood of generating acute liver injury, while a low likelihood of generating immune thrombocytopenia [38]. Second, the update rule of Boolean dynamics in this study is synchronous. However, the complication progression is obviously asynchronous in clinical practice. Third, the

dynamical model is only considered in a small scale of motif system.

As part of the future work, we plan to address some of the aforementioned limitations. In particular, we plan to conduct a more comprehensive study on the dynamics of a large network following heterogeneous likelihood of generating complications, with asynchronous update rule. It may initiate a quantitative framework to trace back to the possible original diseases in terms of the current diseases, and predict the progression of complications according to the current diseases. The results from such study will be valuable for diagnosing and developing accurate treatments against complicated diseases, such as COVID-19 and diabetes, which have widespread effects on the whole disease system.

In parallel, many methods have been introduced to the classification of human diseases, such as machine learning [46], integration of phenotypic similarity with genomics [47], pathway-based classification [48] and consensus-based technique [49]. However, contemporary approaches usually do not consider the interactions among diseases [50]. This failure partly comes from the focused nature of medical training, and the reductionist paradigm in modern medicine. To overcome this shortcoming, the network framework is applied to define human disease [29, 51]. In our work, the disease modules clustered in the HDCN may further provide complementary information to classify the human disease more accurately.

Conclusion

In this paper, the HDCN is constructed from the medical data. We investigate the topological characteristics, including the degree distribution, clustering coefficient and k -core of the HDCN. Further, we identify the disease modules which are dominated by the classes of diseases. The relations between modules are unveiled in detail. The 3-node and 4-node motifs are extracted from the HDCN, among which three motifs are significant. We simulate the dynamics of motifs with the Boolean dynamic model. Each motif represents a specific dynamic behavior, which is potentially functional in the disease system, such as generating temporal progressions and governing the responses to fluctuating external stimuli. For clinical applications, the investigation of complications may yield new approaches to disease prevention, diagnosis and treatment.

Acknowledgments

This work was supported in part by Ningbo Municipal Non-Profit Fund for Applied Research under Grant No. 2019F1033, Ningbo Medical and Health Brand Discipline under Grant No. PPXK2018-05, NNSF of China under Grant No. 11505099, 12175193 and 11775186, Zhejiang Provincial Natural Science Foundation of China under Grant No. LY20G010015.

References

1. Kwang-Il Goh and In-Geol Choi. Exploring the human diseasome: the human disease network. *Briefings in Functional Genomics*, 11(6):533–542, 2012.
2. Joseph Loscalzo and Albert-László Barabási. Systems biology and the future of medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3:619–627, 2011.

3. Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: A network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011.
4. Chuang Liu, Yifang Ma, Jing Zhao, Ruth Nussinov, Yi-Cheng Zhang, Feixiong Cheng, and Zi-Ke Zhang. Computational network biology: data, models, and applications. *Physics Reports*, 846:1–66, 2020.
5. Jifan Shi, Kazuyuki Aihara, and Luonan Chen. Dynamics-based data science in biology. *National Science Review*, 8:nwab029, 2021.
6. David E Gordon, Joseph Hiatt, Mehdi Bouhaddou, Veronica V Rezelj, Svenja Ulferts, Hannes Braberg, Alexander S Jureka, Kirsten Obernier, Jeffrey Z Guo, Jyoti Batra, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science*, 370(6521), 2020.
7. Taibo Li, Rasmus Wernersson, Rasmus B Hansen, Heiko Horn, Johnathan Mercer, Greg Slodkowitz, Christopher T Workman, Olga Rigina, Kristoffer Rapacki, Hans H Stærfeldt, et al. A scored human protein–protein interaction network to catalyze genomic interpretation. *Nature methods*, 14(1):61–64, 2017.
8. Jean-François Rual, Kavitha Venkatesan, Tong Hao, Tomoko Hirozane-Kishikawa, Amélie Dricot, Ning Li, Gabriel F Berriz, Francis D Gibbons, Matija Dreze, Nono Ayivi-Guedehoussou, et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062):1173–1178, 2005.
9. Ulrich Stelzl, Uwe Worm, Maciej Lalowski, Christian Haenig, Felix H Brembeck, Heike Goehler, Martin Stroedicke, Martina Zenkner, Anke Schoenherr, Susanne Koeppen, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6):957–968, 2005.
10. Pea Carninci, T Kasukawa, S Katayama, J Gough, MC Frith, Norihiro Maeda, Rieko Oyama, T Ravasi, B Lenhard, C Wells, et al. The transcriptional landscape of the mammalian genome. *Science*, 309(5740):1559–1563, 2005.
11. Christopher E Lawson, Sha Wu, Ananda S Bhattacharjee, Joshua J Hamilton, Katherine D McMahon, Ramesh Goel, and Daniel R Noguera. Metabolic network analysis reveals microbial community interactions in anammox granules. *Nature Communications*, 8(1):1–12, 2017.
12. Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.
13. David A Fell and Andreas Wagner. The small world of metabolism. *Nature Biotechnology*, 18(11):1121–1122, 2000.
14. Natalie C Duarte, Scott A Becker, Neema Jamshidi, Ines Thiele, Monica L Mo, Thuy D Vo, Rohith Srivas, and Bernhard Ø Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782, 2007.
15. Benjamin Kleaveland, Charlie Y Shi, Joanna Stefano, and David P Bartel. A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell*, 174(2):350–362, 2018.

16. Benjamin P Lewis, Christopher B Burge, and David P Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20, 2005.
17. Angela Reynolds, Devin Leake, Queta Boese, Stephen Scaringe, William S Marshall, and Anastasia Khvorova. Rational siRNA design for RNA interference. *Nature Biotechnology*, 22(3):326–330, 2004.
18. Rune Linding, Lars Juhl Jensen, Adrian Pasculescu, Marina Olhovsky, Karen Colwill, Peer Bork, Michael B Yaffe, and Tony Pawson. Networkin: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Research*, 36(suppl.1):D695–D699, 2007.
19. Ben Readhead, Jean-Vianney Haure-Mirande, Cory C Funk, Matthew A Richards, Paul Shannon, Vahram Haroutunian, Mary Sano, Winnie S Liang, Noam D Beckmann, Nathan D Price, et al. Multiscale analysis of independent alzheimers cohorts finds disruption of molecular, genetic, and clinical networks by human herpesvirus. *Neuron*, 99(1):64–82, 2018.
20. Joshua M Stuart, Eran Segal, Daphne Koller, and Stuart K Kim. A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302(5643):249–255, 2003.
21. Rajesh Kumar Singh, Jay P Maurya, Abdul Azeez, Pal Miskolczi, Szymon Tylewicz, Katja Stojković, Nicolas Delhomme, Victor Busov, and Rishikesh P Bhalerao. A genetic network mediating the control of bud break in hybrid aspen. *Nature Communications*, 9(1):1–10, 2018.
22. Charles Boone, Howard Bussey, and Brenda J Andrews. Exploring genetic interactions and networks with yeast. *Nature Reviews Genetics*, 8(6):437–449, 2007.
23. Pedro Beltrao, Gerard Cagney, and Nevan J Krogan. Quantitative genetic interactions reveal biological modularity. *Cell*, 141(5):739–745, 2010.
24. Francisco S Roque, Peter B Jensen, Henriette Schmock, Marlene Dalgaard, Massimo Andreatta, Thomas Hansen, Karen Søby, Søren Bredkjær, Anders Juul, Thomas Werge, et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS Computational Biology*, 7:e1002141.
25. Arif Khan, Shahadat Uddin, and Uma Srinivasan. Comorbidity network for chronic disease: A novel approach to understand type 2 diabetes progression. *International Journal of Medical Informatics*, 115:1–9, 2018.
26. César A Hidalgo, Nicholas Blumm, Albert-László Barabási, and Nicholas A Christakis. A dynamic network approach for the study of human phenotypes. *PLoS Computational Biology*, 5:e1000353.
27. Andrey Rzhetsky, David Wajngurt, Naeun Park, and Tian Zheng. Probing genetic overlap among complex human phenotypes. *Proceedings of the National Academy of Sciences*, 104:11694–11699, 2007.
28. XueZhong Zhou, Jörg Menche, Albert-László Barabási, and Amitabh Sharma. Human symptoms-disease network. *Nature Communications*, 5:1–10, 2014.

29. Arda Halu, Manlio De Domenico, Alex Arenas, and Amitabh Sharma. The multiplex network of human diseases. *NPJ Systems Biology and Applications*, 5(1):1–12, 2019.
30. Kwangil Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences*, 104:8685–8690, 2007.
31. Md Shahriare Satu, Md Imran Khan, Md Rezanur Rahman, Koushik Chandra Howlader, Shatabdi Roy, Shuvo Saha Roy, Julian MW Quinn, and Mohammad Ali Moni. Diseasesome and comorbidities complexities of SARS-COV-2 infection with common malignant diseases. *Briefings in Bioinformatics*, 22(2):1415–1429, 2021.
32. Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298:824–827, 2002.
33. Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8:450–461, 2007.
34. Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, 31:64–68, 2002.
35. Janis Klaise and Samuel Johnson. The origin of motif families in food webs. *Scientific Reports*, 7:1–11, 2017.
36. Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100:11980–11985, 2003.
37. Manasvini Singh. Heuristics in the delivery room. *Science*, 374(6565):324–329, 2021.
38. BMJ Best Practice. Complications of coronavirus disease 2019. <https://bestpractice.bmj.com/topics/en-us/3000168/complications>, 2021.
39. Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.
40. Xiongfei Jiang, Long Xiong, Ling Bai, Na Zhao, Jiu Zhang, Ke Xia, Kai Deng, and Bo Zheng. Quantifying the social structure of elites in ancient china. *Physica A: Statistical Mechanics and its Applications*, 573:125976, 2021.
41. Harley H McAdams and Adam Arkin. Simulation of prokaryotic genetic circuits. *Annual Review of Biophysics and Biomolecular Structure*, 27:199–224, 1998.
42. Naama Barkai and Stan Leibler. Robustness in simple biochemical networks. *Nature*, 387:913–917, 1997.
43. Charles Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 100:441–471, 1987.
44. X.F. Jiang, T.T. Chen, and B. Zheng. Structure of local interactions in complex financial dynamics. *Sci. Rep.*, 4:5321, 2014.

45. Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.
46. Sadaf Malik, Nadia Kanwal, Mamoon Naveed Asghar, Mohammad Ali A Sadiq, Irfan Karamat, and Martin Fleury. Data driven approach for eye disease classification with machine learning. *Applied Sciences*, 9(14):2789, 2019.
47. Mikhail G Dozmorov. Disease classification: from phenotypic similarity to integrative genomics and beyond. *Briefings in Bioinformatics*, 20(5):1769–1780, 2019.
48. Ivan Y Iourov, Svetlana G Vorsanova, and Yuri B Yurov. Pathway-based classification of genetic diseases. *Molecular Cytogenetics*, 12(1):1–5, 2019.
49. Clara DM van Karnebeek, Daniël Beumer, Colleen Pawliuk, Helly Goetz, Sara Mostafavi, Gail Andrews, Rose Steele, and Harold Siden. A novel classification system for research reporting in rare and progressive genetic conditions. *Developmental Medicine and Child Neurology*, 61(10):1208–1213, 2019.
50. Frank Emmert-Streib, Shailesh Tripathi, Ricardo de Matos Simoes, Ahmed F Hawwa, and Matthias Dehmer. The human disease network: Opportunities for classification, diagnosis, and prediction of disorders and disease genes. *Systems Biomedicine*, 1(1):20–28, 2013.
51. Joseph Loscalzo, Isaac Kohane, and Albert-László Barabási. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Molecular Systems Biology*, 3(1):124, 2007.

Supporting information

S1 Table. Components of the k -core.

S2 Table. Diseases with the most occurring times of each node in the FFL motif.

S3 Table. Diseases with the most occurring times of each node in the bi-fan motif.

S4 Table. Diseases with the most occurring times of each node in the OFFL motif.

Data availability. The data used for this article are available in the supporting information. The source of the directed network is shown in DATA.xlsx, in which the names of nodes are chinese.