

RESEARCH PAPER

 OPEN ACCESS 

## Modeling spatial interaction networks of the gut microbiota

Xiaocang Cao<sup>a\*</sup>, Ang Dong<sup>b\*</sup>, Guangbo Kang<sup>c</sup>, Xiaoli Wang<sup>a</sup>, Liyun Duan<sup>a</sup>, Huixing Hou<sup>a</sup>, Tianming Zhao<sup>a</sup>, Shuang Wu<sup>b</sup>, Xinjuan Liu<sup>d</sup>, He Huang<sup>c</sup>, and Rongling Wu<sup>e</sup>

<sup>a</sup>Department of Gastroenterology and Hepatology, Tianjin Medical University General Hospital, Tianjin Medical University, Tianjin, China; <sup>b</sup>Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China; <sup>c</sup>School of Chemical Engineering and Technology, Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China; <sup>d</sup>Department of Gastroenterology, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China; <sup>e</sup>Center for Statistical Genetics, Departments of Public Health Sciences and Statistics, The Pennsylvania State University, Hershey, PA, USA

### ABSTRACT

How the gut microbiota is organized across space is postulated to influence microbial succession and its mutualistic relationships with the host. The lack of dynamic or perturbed abundance data poses considerable challenges for characterizing the spatial pattern of microbial interactions. We integrate allometric scaling theory, evolutionary game theory, and prey-predator theory into a unified framework under which quasi-dynamic microbial networks can be inferred from static abundance data. We illustrate that such networks can capture the full properties of microbial interactions, including causality, the sign of the causality, strength, and feedback loop, and are dynamically adaptive along spatial gradients, and context-specific, characterizing variability between individuals and within the same individual across time and space. We design and conduct a gut microbiota study to validate the model, characterizing key spatial determinants of the microbial differences between ulcerative colitis and healthy controls. Our model provides a sophisticated means of unraveling a complete atlas of how microbial interactions vary across space and quantifying causal relationships between such spatial variability and change in health state.

### ARTICLE HISTORY

Received 3 March 2022  
Revised 3 June 2022  
Accepted 21 July 2022

### KEYWORDS






Gut microbiota; spatial biology; evolutionary game theory; microbial interaction network; quasi-dynamic ordinary differential equation

## Introduction


Microorganisms residing in the human gut have direct and indirect impacts on health and disease.<sup>1</sup> Disruptions to the community of gut microbes can contribute to the risk and severity of a number of medical conditions, such as obesity, cancer, and autism among others.<sup>2–5</sup> While most studies focus on the determination of microbe-health relationships using homogenized samples, such as stools,<sup>6–10</sup> there is growing appreciation for how the influence of microbes on host physiology is driven by the spatial heterogeneity of microbiota along the length of the digestive tract.<sup>11–18</sup>

Microbiota data collected at a series of gut locations are the key to chart the organizational and functional atlas of the microbial community across

space,<sup>18</sup> but this will critically rely on how these spatial data are modeled and analyzed. Trillions of microbes populate a narrow gut, forming the most complex and densest ecosystem on Earth. It is unlikely that microbes exert their influences independently of each other on host traits, rather they do so interactively through complex but well-orchestrated networks.<sup>19–23</sup> More importantly, such interactions may occur among microorganisms not only at the same taxon level but also across taxonomic groups, given the co-occurrence of phylogenetically diverse microbes. As such, reconstructing large-scale microbial interaction networks across spatial gradients is a crucial choice to interrogate the elaborate crosstalk among gut microbes and between microbes and their hosts.

**CONTACT** Rongling Wu  [rwu@phs.psu.edu](mailto:rwu@phs.psu.edu)  Center for Statistical Genetics, Departments of Public Health Sciences and Statistics, the Pennsylvania State University, Hershey, PA 17033, USA; Xiaocang Cao  [caoxc@tmu.edu.cn](mailto:caoxc@tmu.edu.cn)  Department of Gastroenterology and Hepatology Tianjin Medical University General Hospital, Tianjin Medical University General Hospital, Tianjin Hebei China; Xinjuan Liu  [liuxinjuan@mail.ccmu.edu.cn](mailto:liuxinjuan@mail.ccmu.edu.cn)  Department of Gastroenterology Beijing Chaoyang Hospital, Capital Medical University, Beijing Hebei China; He Huang  [huang@tju.edu.cn](mailto:huang@tju.edu.cn)  School of Chemical Engineering and Technology Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin Hebei China

\*These authors contributed to this work equally

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19490976.2022.2106103>

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

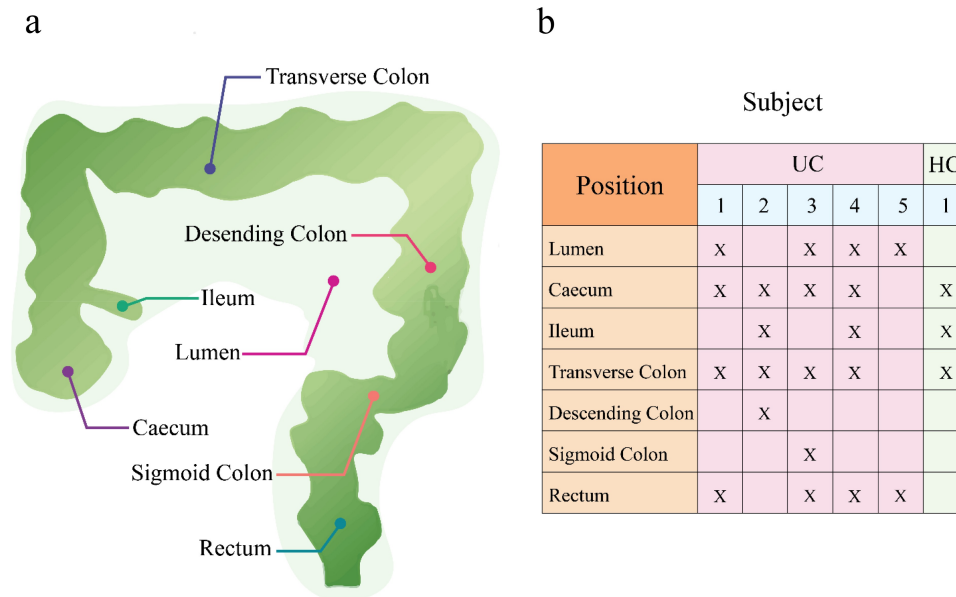
Here, we implement a computational model for inferring space-specific microbial interaction networks. Existing approaches reconstruct an overall microbial network from a large number of samples,<sup>24</sup> failing to characterize sample-specific topological differences. Correlation-based approaches can estimate the strength of microbe-microbe interaction, but cannot infer the direction of the interaction.<sup>25</sup> Bayesian networks can identify causality, but lack a capacity to find feedback cycles.<sup>26</sup> High-density longitudinal data are powerful for reconstructing informative networks filled with bidirectional, signed, and weighted interactions, but such data are hardly available for the gut microbiota,<sup>27</sup> especially at multiple locations of the gut. Our model that integrates elements of different disciplines can overcome all the above limitations to reconstruct informative, dynamic, omnidirectional, and personalized networks (idopNetwork) from static abundance data.<sup>28–30</sup> The idopNetwork can reveal the position-dependent change of microbial networks and infer its spatially causal relationships with health state. As a proof of concept, we validate the usefulness of idopNetwork by designing and conducting a spatial mapping study of the gut microbiota. We identify

key biogeographically-varying microbial interactions that distinguish diseased guts from healthy counterparts.

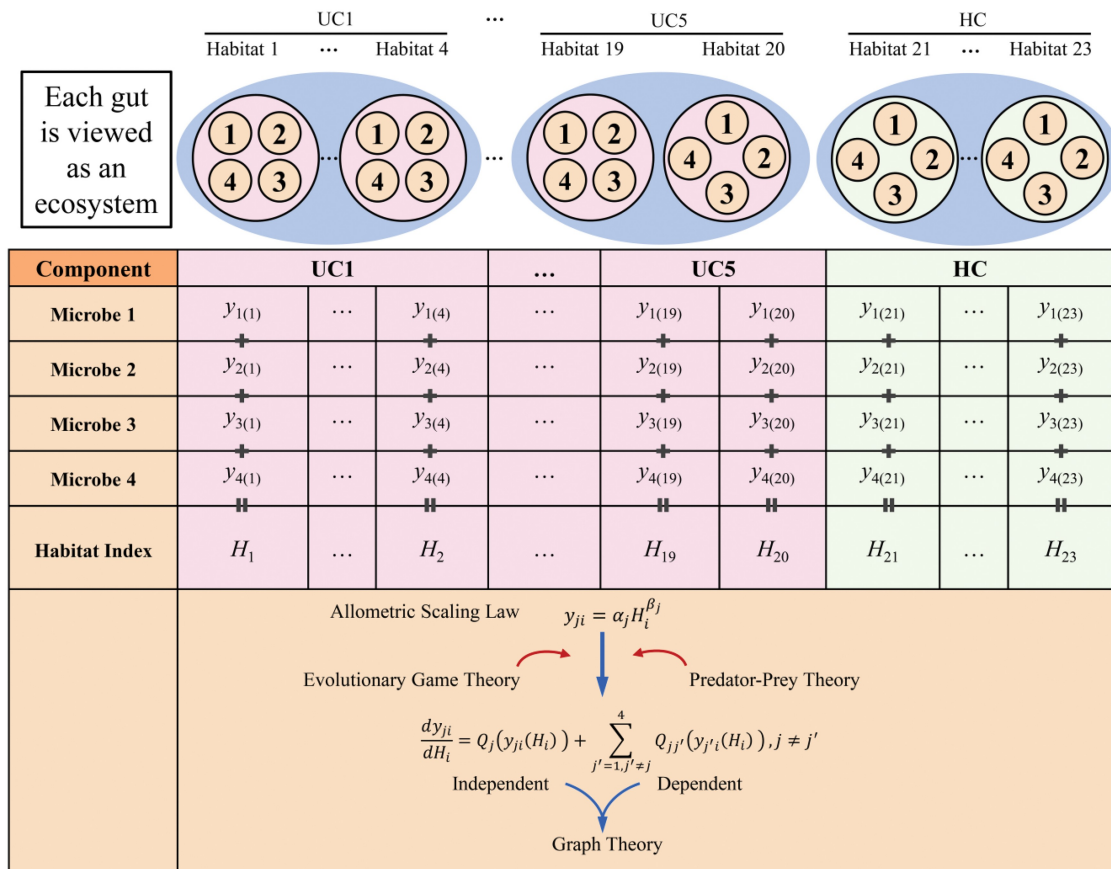
## Results

### *Inferring ecologically functional microbial networks across gut space*

We sample seven spatially different positions along and inside the guts from five patients infected with ulcerative colitis (UC) and one healthy control (HC) (Figure 1) and measure microbial abundance levels at different taxa from a total of 23 positions, i.e., habitats. The diagram in Figure 2 illustrates the workflow of how we reconstruct microbial networks from this spatial data. We identify the 16 most abundant phyla and the remaining phyla (attributed to the other) and plot the abundance level (i.e., niche index) of each phylum against habitat index (HI), defined as the summed abundance level of all microbes at a gut position. We find that niche index-HI change can be well fit by the power equation described in equation (1) (Figure 3). A majority of phyla increase their abundance with habitat index in an exponential manner



**Figure 1.** Sampling strategy of studying the spatial distribution of the gut microbiota. (a) Six sampled positions include rectum, sigmoid colon, descending colon, transverse colon, ileum, and cecum along the gut and lumen inside the gut. (b) Sample distribution of spatial microbiota from six guts, five infected by ulcerative colitis (UC) and one being a healthy control (HC), totaling 23 independent samples.

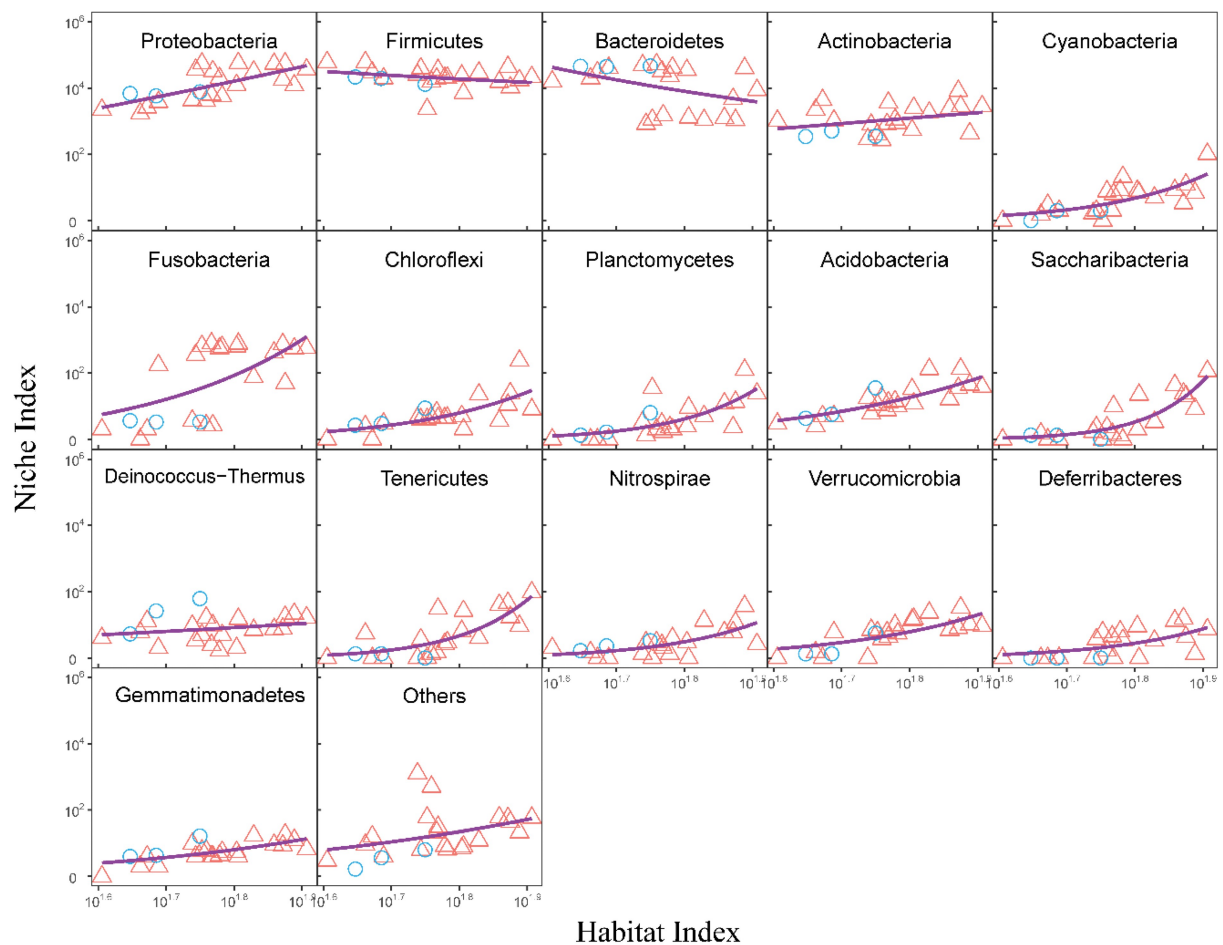


**Figure 2.** Workflow of idopNetwork reconstruction from spatial mapping data of the gut microbiota as collected from the study of Figure 1. Each gut can be viewed as an ecosystem in which different positions represent natural habitats of microbes. As a toy example, we assume four microbes residing at each gut position, whose abundance is monitored to form a data structure. By taking the sum of abundance of all microbes at each position, we calculate habitat index for this position. The power equation is used to model the allometric relationship between the abundance of individual microbes and habitat index, which establishes a foundation for converting static data into its quasi-dynamic representation crucial for the integration of evolutionary game theory and predator-prey theory into graph theory.

( $R^2 = 0.80\text{--}0.90$ ), although the slope of increase varies from phylum to phyla. Only two rich phyla, Firmicutes and Bacteroidetes, follow a reverse pattern of change, i.e., decreasing exponentially with habitat index. We plot the niche index of each species against habitat index and find that their relationship can be similarly fitted by the power equation, but with the slope depending on species (Fig. S1). Taken together, the power equation can not only explain the part-whole relationship of individual microbes at various taxonomic levels with the total microbes, but also provides a means of modeling microbial diversity and interactions across spaces.

In the Materials and methods section, we describe the derivation of the idopNetwork model by building quasi-dynamic ordinary differential equations (qdODE) based on niche index-HI

relationship (Figure 2). The qdODE decompose the observed abundance level of a microbe into its independent component due to its own capacity and dependent component resulting from the influence of other microbes, as shown by equations (2) and (7). We use the independent and dependent abundance components estimated by an optimization technique to reconstruct spatially-varying idopNetworks at the phylum level for UC and HC groups (Figure 4). The topological organization of microbial networks displays both similarities and differences among gut positions. For example, *Deferribacteres* acts as a leader (i.e., those with more and stronger outgoing links than other microbes) that exerts influences on *Cyanobacteria*, *Fusobacteria*, *Chloroflexi*, *Planctomycetes*, and *Acidobacteria* in all networks, regardless of their gut positions and gut health states. There is also



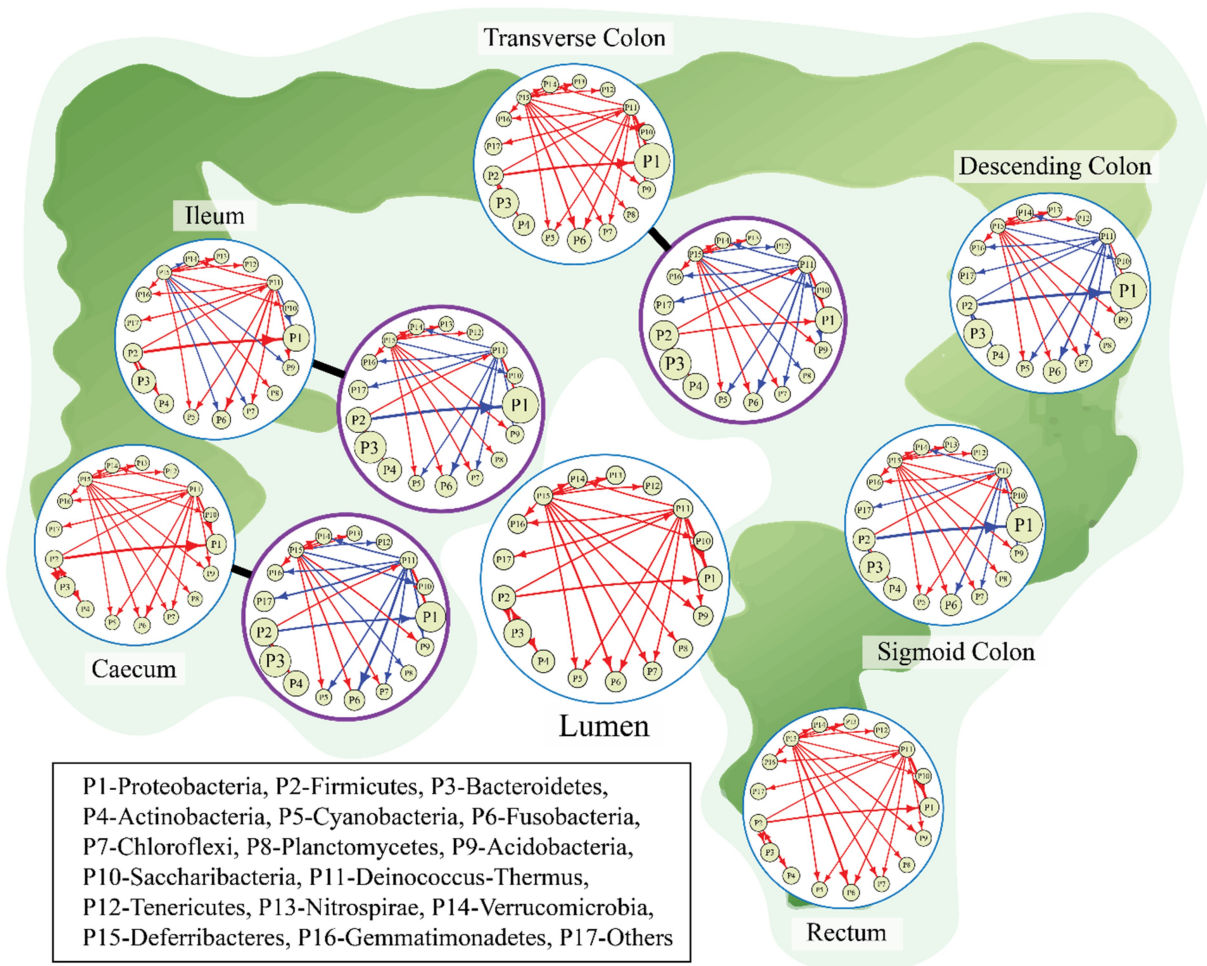
**Figure 3.** The fit of the power equation to the relationship between the abundance of individual phyla (niche index) and the total abundance of all phyla (habitat index) across UC (red triangles) and HC samples (blue circle).

gut position-dependent variability not only expressed in total microbial abundance and microbial composition (Figure 3), but also in the topological structure and organization of microbe-microbe interactions (Figure 4). Position-dependent networks of the UC group can be clearly classified into two types; one involving lumen, rectum, transverse colon, ileum and cecum networks and the second including sigmoid colon and descending colon networks. In the second type, Firmicutes and *Deinococcus-Thermus* bring strong amensalism to bear on Proteobacteria and Fusobacteria, respectively, whereas in the first type, the former is commensalistic toward the latter. Detailed position-dependent differences can be observed in network organization from the same type. In the first type, the strength of this commensalism decreases from the cecum and ileum networks to the transverse colon network to the rectum and lumen networks. *Deferribacteres* is

commensalistic for Fusobacteria in the lumen, rectum, transverse colon, and cecum networks, but this relationship changes as amensalism in the ileum network. In the second type, compared to the sigmoid colon network, the descending colon network is encoded by amensalism at a higher frequency.

Only three positions, transverse colon, ileum, and cecum, were sampled for the HC group, and these positions display distinct network topologies (Figure 4). Each of these three HC networks is tremendously different from the UC network at the same position. In general, the HC networks are well mixed by commensalism and amensalism, whereas the UC networks tend to be dominated by commensalism, suggesting that a healthy network can better balance different types of microbes than a diseased network. At the cecum and ileum positions, Firmicutes establishes a commensalistic relationship toward Proteobacteria in the UC





**Figure 4.** Spatial microbial networks at the phylum level along and inside the gut. Arrow lines with warm and cold colors represent commensalism and amensalism, respectively, with the thickness of lines proportional to the strength of microbe-microbe interactions.

networks, but an amensalistic relationship between these two phyla is detected in the HC networks. In the transverse colon network, although commensalism occurs from Firmicutes to Proteobacteria in both UC and HC groups, the strength of this interaction is larger in the former than in the latter. Also, *Deinococcus-Thermus* is commensalistic toward *Fusobacteria* in the UC networks, but this relationship is amensalistic in the HC networks. Taken together, the strength, pattern and architecture of microbial interactions vary spatially across the biogeographic positions of the gut, with the degree of variation depending on healthy state. Network analysis on the three commonly measured positions suggests that healthy networks display a greater position-dependent variability than diseased

networks. A series of spatially reconstructed networks can more precisely characterize key microbial interactions that interrogate why and how a healthy state becomes unhealthy.

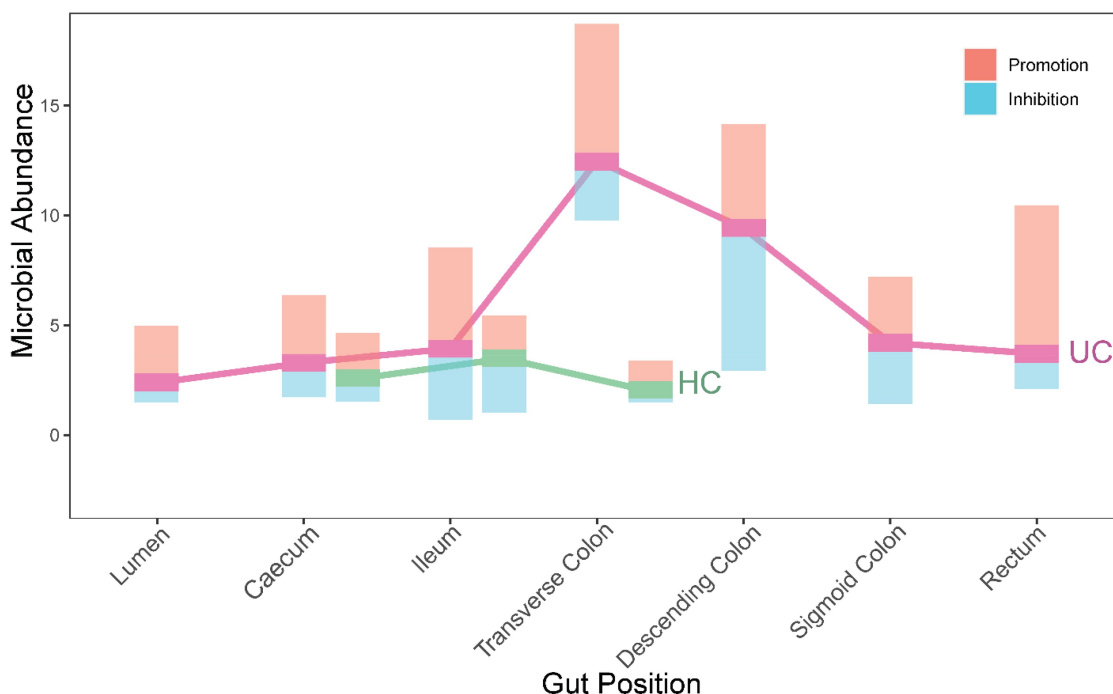
Our qdODE model decomposes the observed abundance level of a microbe into its independent component (expressed when this microbe is socially in isolation) and dependent component (due to the accumulated (positive and negative) influence of other microbes on this microbe) at each gut position. This decomposition is illustrated in Fig. S2, where the independent and dependent abundance of each phylum are characterized across samples. Figure 5 reveals the position-dependent total amounts of independent abundance, positive dependent (promotion) abundance and negative

dependent (inhibition) abundance over all microbes at the phylum level. For UC guts, lumen, caecum, ileum, sigmoid colon, and rectum has a similar amount of independent abundance, which is much lower than that at transverse colon and descending colon. Among three commonly measured positions, caecum and ileum are similar in independent abundance between UC and HC guts, whereas the independent abundance of transverse colon is strikingly larger in the UC than HC guts. This suggests that the intrinsic capacity of microbes to express themselves in transverse colon is a determinant of the health state of a gut. The amounts of positive dependent abundance (via promotion) and negative dependent abundance (via inhibition) substantially vary among gut positions. Some positions, such as ileum, transverse colon, and rectum, have a large amount of positive dependent abundance, whereas some positions, like descending colon, are dominated by negative dependent abundance (Figure 5). At transverse colon, both positive and negative dependent abundance are much richer for UC guts than HC guts, suggesting that the strength and pattern of microbial interaction at transverse colon are a determinant of gut health state. Taken together, when healthy guts get infected by

ulcerative colitis, the capacity of microbes to be independently expressed increases at certain positions, i.e., transverse colon and descending colon. Also, the total strength of microbial interaction and the relative strength of cooperation and competition are associated with the shift of health state from healthy to diseased and vice versa.

### Reconstructing multilayer and multiplex microbial networks

Microbial networks may occur and function at different taxonomic levels, with larger sizes at lower than higher levels. Modularity theory suggests that a robust and stable large-scale network would be divided into distinct network communities within which entities are more functionally correlated with each other than with those from other modules.<sup>31,31</sup> Thus, when we attempt to reconstruct a microbial network from a large number of microbes, we can break it down into sparsely interconnected network communities. Wu and Jiang<sup>30</sup> proposed a bottom-up approach for identifying network communities by classifying all entities into different modules according to their similarity of sample-dependent variability. As shown by Figure 2, the abundance of



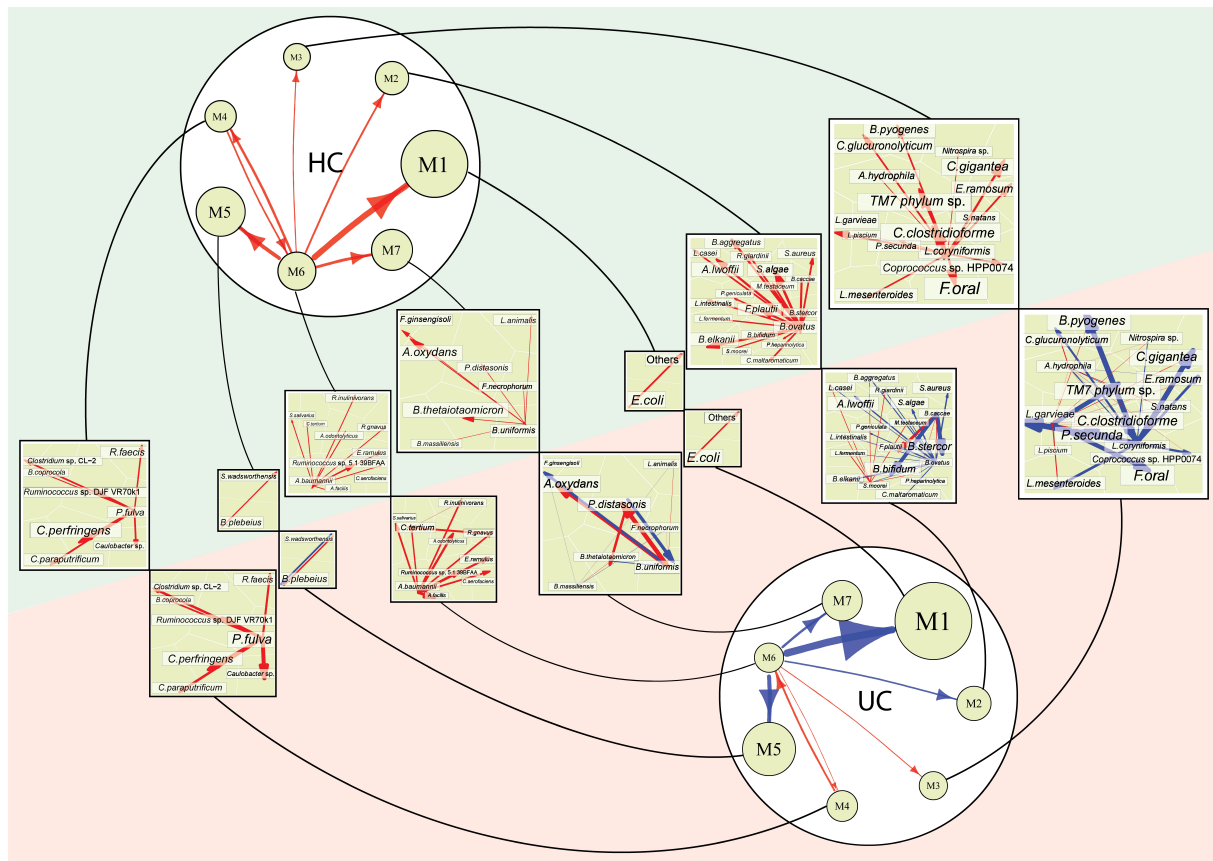
**Figure 5.** Microbial abundance of all phyla at different gut positions, decomposed into the independent component (purple bar for UC, green bar for HC), positive dependent component (due to promotion, warm bar), and negative dependent component (due to inhibition, cold bar).

a microbe changes with habitat index following the power law. We implement the power equation into Kim et al.<sup>31</sup> functional clustering to detect different microbial modules (see the Method).

The mean values of abundance for all microbes within modules are used to reconstruct the coarse-grained module-module idopNetwork or the top-layer microbial network. Microbes within each module form multiple fine-grained microbial networks or bottom-layer microbial networks. The dimension reduction of data by clustering increases statistical and computational efficiency in inferring coarse- and fine-grained networks. However, if the number of microbes within a module is still too large to be used for network reconstruction, we can implement functional clustering to classify this module into distinct submodules and reconstruct cross-submodule networks. If needed, a submodule can be further classified into different sub-

submodules and this process repeats until a trackable number of microbes is obtained. In the end, we reconstruct multilayer microbial networks at different levels of classification. If microbes at a higher classification level interact with those at a lower level, multiplex microbial networks can be reconstructed.

In our spatial mapping study, we identify 65 species, which are classified into seven distinct modules M1-M7 by functional clustering (Fig. S3). Each module contains a different number of species, forming a network community. We reconstruct a seven-node coarse-grained microbial network from modules and seven fine-grained microbial networks from species of the same modules, forming a 2-layer microbial network between and within modules (Figure 6). Such a 2-layer network is inferred for UC and HC groups, respectively. The first-layer network is structurally similar but organizationally different between



**Figure 6.** Two-layer microbial networks at the species level for HC and UC groups. At the top level of the network is the 7-node inter-module (coarse-grained) network and at the bottom level are seven interspecific (fine-grained) subnetworks, visualized by Voronoi treemaps. Arrow lines with warm and cold colors represent commensalism and amensalism, respectively, with the thickness of lines proportional to the strength of microbe-microbe interactions.

the two groups. Module M6, as a leader, exerts a strong commensalism to modules M1 and M5 in HC guts, but this relationship becomes amensalistic in UC guts.

M1 and M5 are highly abundant, but each is only composed of two species. In M1, *Escherichia coli* and those unidentified species establish a commensalistic relationship in both HC and UC groups, but in M5 *Bacteroides plebeius* prompts *Sutterella wadsw* in the HC group, but the latter turns to inhibit the former in the UC group (Figure 6). Considerable differences in both structure and organization are observed for the second-layer networks of the other five modules. Compared to the UC group, the second-layer networks of M2, M3, and M7 contains more amensalistic links than those in the HC group, further suggesting that microbial competition may be a cause of ulcerative colitis. For example, in the second-layer network of M2, *Bacteroides ovatus* is a positive strong leader that promotes many other species in HC guts, but its leadership becomes negative at a reduced strength in UC guts. Also, *Bacteroides caccae* is activated in UC guts, exerting strong commensalism for *Flavonifractor plautii* and strong amensalism for *Bifidobacterium bifidum* and *Bacteroides sterco*.

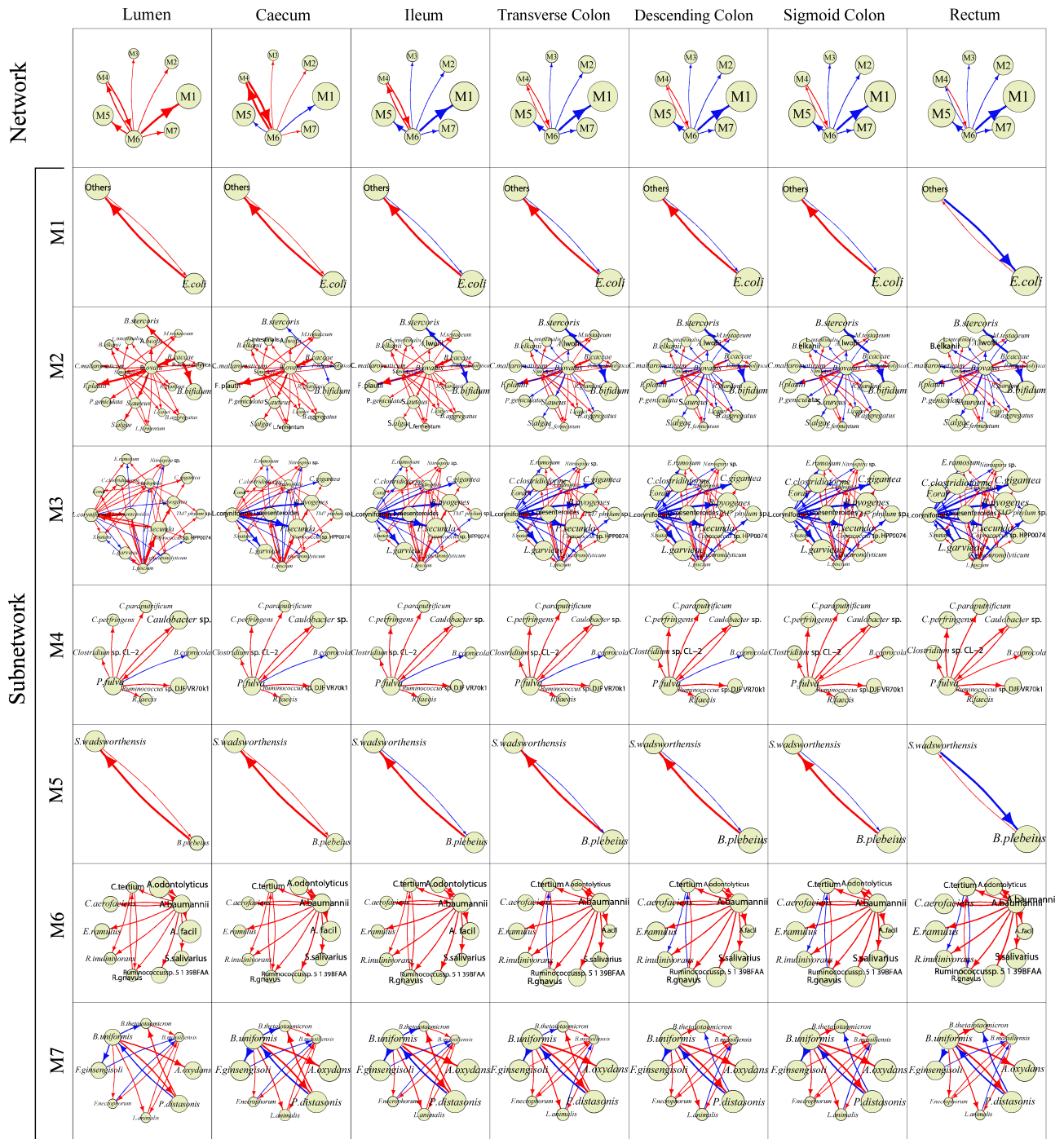
More remarkably, the second-layer network of M3 produces structural and organizational changes from HC to UC groups (Figure 6). *Lactobacillus coryniformis* is a dominant leader that positively impact many other species in the HC group, but it turns to exert strong negative impacts on the same species in the UC group. *Lactococcus piscium* that coexists peacefully with its partners in HC guts becomes aggressive toward *Firmicutes oral* and *phylum sp. Oral* in UC guts. M4 and M6 are filled of commensalism in both HC and UC networks, but the strength of commensalism notably differs between two types of networks. Stronger commensalism exerted by particular species is observed in UC than HC groups. Taken together, multilayer microbial interaction networks provide a detailed roadmap of how each microbial species interacts with every other species through cooperation or competition to determine the change of guts from healthy to unhealthy states.

### Spatial tracing of multilayer networks

We reconstruct multilayer networks at each position from which we can trace how coarse- and fine-grained networks vary along biogeographical gradients (Figures 7, 8). For the UC group, coarse-grained networks are structurally very similar over seven sampled positions, except for the link from modules M6 to M1, which is commensalistic at lumen, cecum, and ileum but amensalistic at transverse colon, descending colon, sigmoid colon, and rectum (Figure 7). Fine-grained networks also vary among gut positions, but with the degree of this difference depending on modules. Network communities for M4 and M6 exhibit a slight increase of commensalism strength from lumen to rectum, but those for M1, M2, M3, M5, and M7 differ dramatically among positions. In those network communities, some species-species links change from commensalism to amensalism. For example, in M5, *Bacteroides plebeius* and *Sutterella wadsworthesis* are mutualistic to each other at lumen, cecum, and ileum, but become increasingly parasitic from transverse colon to descending colon to sigmoid colon to rectum. Similarly, in M3, *Lactobacillus coryniformis* is commensalistic to *Lactococcus garvieae* at lumen and cecum, but this relationship becomes amensalistic with an increasing strength from ileum to transverse colon to descending colon to sigmoid colon to rectum.

The three gut positions sampled for the HC group also exhibit differences in both coarse- and fine-grained networks, although the extent of such differences is module-dependent (Figure 8). In the coarse-grained network, leader M6's commensalism toward M1 and M5 at cecum and ileum becomes amensalism at transverse colon, but the internal workings within M6, M1, and M5 little vary among three positions. Pronounced differences in the sign and strength of certain interspecific interaction are observed in fine-grained networks of M2, M3, and M7. For example, in M2, *Bacteroides caccae* exerts commensalism at cecum and ileum but amensalism at transverse colon for *Bacteroides stercoris* and *Bifidobacterium bifidum*. In M7, *Bacteroides uniformis* is commensalistic toward *Flavisolibacter ginsengisoli* at cecum, but this commensalism changes as amensalism at ileum and transverse colon.



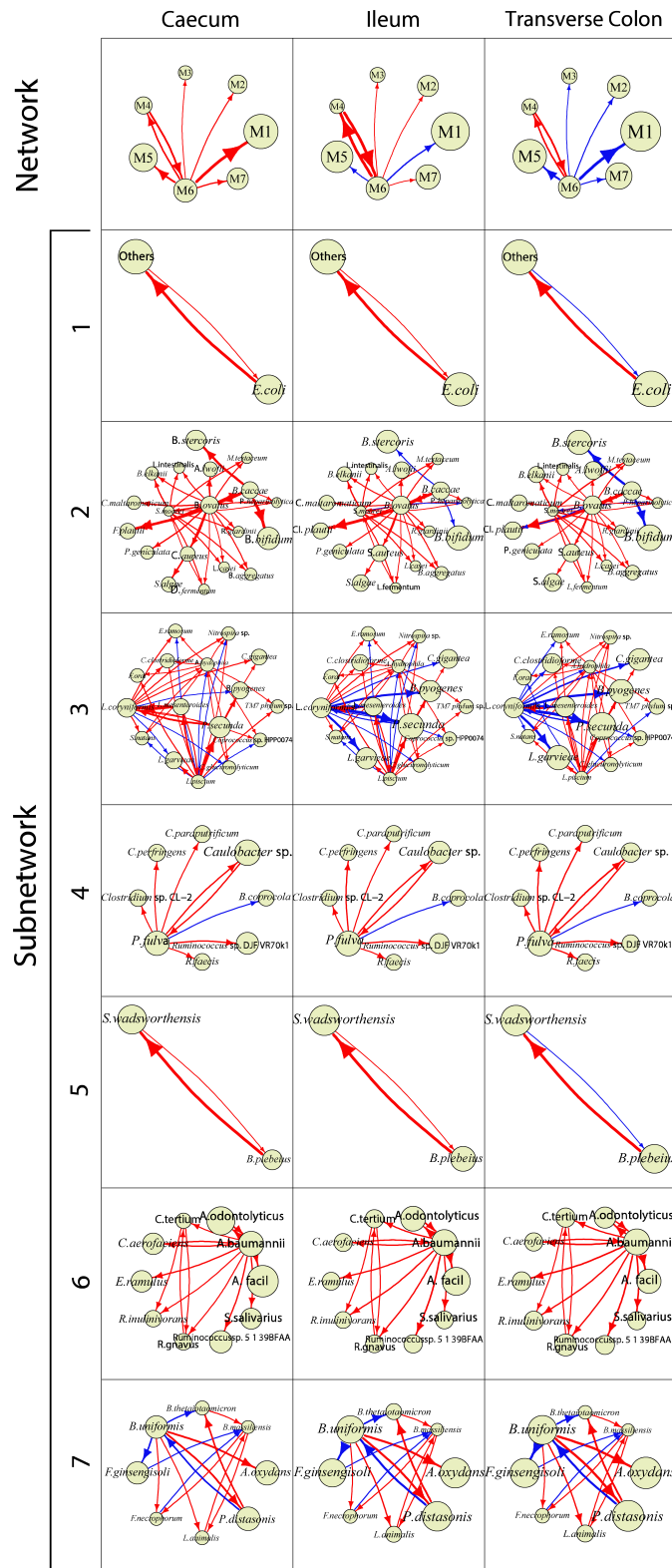


**Figure 7.** Two-layer microbial networks at the species level over different gut positions for the UC group. For each position, at the top level of the network is the 7-node inter-module (coarse-grained) network and at the bottom level are seven interspecific (fine-grained) subnetworks, visualized by Voronoi treemaps. Arrow lines with warm and cold colors represent commensalism and amensalism, respectively, with the thickness of lines proportional to the strength of microbe-microbe interactions.

The two-layer networks provide a detailed atlas of interspecific interactions that determine differences between UC and HC groups (Figure 7 vs. Figure 8). Although most pairwise interactions are consistent in sign between the two groups, there are

a few certain interactions that change their sign from HC to UC guts. In general, UC networks contain more amensalistic interactions than the HC networks, suggesting that microbes tend to turn to be competitive when the guts are infected





**Figure 8.** Two-layer microbial networks at the species level over three different gut positions for the HC group. For each position, at the top level of the network is the 7-node inter-module (coarse-grained) network and at the bottom level are seven interspecific (fine-grained) subnetworks, visualized by Voronio treemaps. Arrow lines with warm and cold colors represent commensalism and amensalism, respectively, with the thickness of lines proportional to the strength of microbe-microbe interactions.

by ulcerative colitis. For example, in M7, *Bacteroides uniformis* at cecum is commensalistic toward *Flavisolibacter ginsengisoli* in the HC group, but becomes amensalistic toward the same species in the UC group. In addition, for some pairs of species, the strength of their interactions may change from HC to UC at the same position. All suggest that certain interspecific interactions may be used as a biomarker to assess health risk.

## Discussion

The network model described in this article adds a previously neglected dimension to fully capture the topology of the vast complexity of the gut microbiota. This additional dimension is the spatial distribution of microbes among physically, chemically and immunologically distinct niches in the gut.<sup>12,32</sup> The biogeography of bacteria in the gut is regulated by nutrient selection and immune activation and, meanwhile, it determines the microbial compositions and heterogeneity along the longitudinal axis of the intestines. However, a precise characterization of microbial interaction networks along gut positions is challenged by unavailability of high-density dynamic or perturbed abundance data collected at each position. In this article, we present a spatial model to disentangle this challenge by extracting dynamic snapshots from static data.

The capacity of the new model to dynamize networks in space with no need of dynamic data results from the seamless integration of multiple disciplines through the implementation of high-dimensional statistical theory. From a dynamic perspective, evolutionary game theory explains the overall abundance of a microbe to be due to its independent component (reflecting how much a microbe can grow in isolation due to its intrinsic capacity) and dependent component (describing the amount of microbial growth due to influence by other co-existing members). The dynamic formulation of evolutionary game theory using static data is achieved by converting them into their quasi-dynamic representation via allometric scaling theory. By introducing predator-prey theory, we construct a system of qdODE to quantify the independent and dependent components for each microbe.<sup>28,29,30</sup> The estimated independent and dependent components by an optimization

technique are coded into graphs, producing one of the most advanced networks – idopNetwork. As compared to the most commonly used Dynamic Bayesian Networks (DBNs) which can identify the causality of interactions from evenly-spaced dynamic data, idopNetwork can extract and capture full properties of microbial interaction networks (including bidirectionality, sign, weight, and feedback loop) from any data domain with any measurement schedule. Through extensive computer simulation studies, idopNetworks were found to perform much better than DBNs in terms of the power and false positive rate of interaction detection.<sup>32</sup> In particular, idopNetworks provide a visualized platform to trace how microbial interaction architecture changes from one location to the next and how these changes impact the reciprocal shift between healthy state to diseased state.

As a proof of concept, we apply idopNetwork to reveal how microbial network structure and organization vary along biogeographical gradients of the gut using the gut microbiota data from a spatial mapping study including ulcerative colitis-infected patients and healthy controls. idopNetwork characterize a dramatic change in network organization (the sign and strength of interactions) from one position to next along and inside the gut (Figures 3, 5, 7, 8). Such an organizational change is much more pronounced in guts inflected by ulcerative colitis than healthy guts, suggesting that the position-dependent change of interaction strength and sign may be a driver of gut shift from a healthy state to an unhealthy state.

An increasing number of studies have showed a great potential to control many diseases by changing the ecological equilibrium of microbes in the gut,<sup>33,34,35</sup> but existing approaches can only treat the whole gut as a mixture, without taking into account position-dependent variability in the gut microbiota. The idopNetwork model can discern spatial changes of microbial interactions and identify the precise gut positions at which microbial functions are most tightly associated with the disease. Spatial data analysis shows that difference between ulcerative colitis guts and healthy guts lies in the shift of ecological interactions between specific microbes from commensalism-dominated

cooperation to amensalism-dominant competition at certain gut positions. Phylum Planctomycetes exerts strong commensalism for phylum Verrucomicrobia at transverse colon and cecum in healthy guts, but this commensalism becomes strong amensalism when the guts are inflected by ulcerative colitis. In clinical practice, by shifting Planctomycetes-Verrucomicrobia amensalism back to Planctomycetes-Verrucomicrobia commensalism at these gut positions via medical interventions, ulcerative colitis may be controlled and eliminated.

The idopNetworks can be reconstructed at any phylogenetic level, including a web of crosstalk among bacteria, fungi, and viruses and networks at any taxonomic level for the same type of microbes. While microbial networks at a higher level of taxon, such as phylum, can generalize a general rule of thumb behind microbial community assembly, those at a lower level, such as species, strain, or even genotype, can precisely characterize the detailed roadmap of how individual microbes interact with each other spatially to determine host health. However, a low taxonomic level may include an enormous number of operational taxonomic units (OTU), which makes it computationally infeasible to reconstruct a large-scale network. To circumvent this issue, we implement power equation-functional clustering to classify all OTUs into distinct size-reduced modules according to developmental modularity theory. In network science, this procedure is a bottom-up approach that breaks down any large-scale network into distinct network communities or subnetworks.<sup>30</sup> We use this approach to classify 65 species into seven modules and reconstruct a two-layer network involving all species, at a higher layer of which is a 7-node module-module network and a lower layer of which are seven species-species networks within modules. Two-layer networks display gut position-dependent variability, healthy state-dependent variability, and position-state interaction variability in topological organization. The shift of commensalism to amensalism between *Bacteroides uniformis* and *Flavisolibacter ginsengisoli* at cecum (within the subnetworks of module M7) could be a driver of healthy guts that turn to ulcerative colitis-infected (Figure 7 vs. Figure 8). Yet, at ileum and transverse colon, these two

species consistently maintain an amensalistic relationship, regardless of health state. Thus, *Bacteroides uniformis*-*Flavisolibacter ginsengisoli* interaction at these two positions cannot be used to distinguish between healthy guts and ulcerative colitis-infected guts.

Although idopNetworks have many advantages, the above results from our spatial mapping study of the gut microbiota should be interpreted with caution. First, the sample size (23 positions) used may be insufficient to make a rigorous inference. Computer simulation shows that at least 20 samples are required to obtain reasonable power for interaction detection, but increasing sample sizes, such as 50 to 100, are necessary if the measurements of microbial abundance contain a certain level of noises.<sup>28,30</sup> Second, as a proof of concept, our data only contain one healthy control. At least three biological replicates for the control are required to reasonably infer about microbial differences between diseased and health groups. Third, the impact of the gut microbiota on host health is determined by genetic, demographic, and environmental factors.<sup>36</sup> The utility of idopNetwork can be best justified only after it is implemented into a genome-wide association studies aimed at characterizing the combined effects of various genetic and environmental factors on the microbiota-health interplay. Despite these limitations, the application of idopNetwork in this study provides a starting point for deciphering how ecological interactions transit along biogeographical locations to shape the resilience, stability and diversity of microbial ecosystems and human health from a large-scale data domain. Furthermore, idopNetwork can be brought into play to unleash a broader microbial community assembly including oral cavity, esophagus, and vaginal microhabitats.

## Conclusions

The spatial variation of microbes determines how the gut microbiota impacts human health. This impact can be better revealed by inferring spatially dynamic interaction networks. In this article, we develop a model for tracing network dynamics along space with no need of dynamic data. Statistical analysis of microbial data according to this model includes four steps: (i) fitting of the power equation to the

abundance of individual microbes over habitat index, (ii) LASSO-based variable selection implemented to choose a small set of the most significant microbes that are linked with a focal microbe, (iii) network reconstruction based on qdODE, and (iv) functional clustering of all microbes into distinct modules based on their pattern of microbial abundance. The last step allows large-scale multilayer sparse microbial interaction networks across gut space to be reconstructed. Our model can discern interaction changes of any microbes from healthy to unhealthy guts to gain new insight into microbial impact on human health and disease.

## Materials and methods

### Allometric scaling law

Consider a spatial mapping study of the gut microbiota involving multiple subjects. For each subject, the abundance of microbes at different taxa is monitored at seven distinct positions, i.e., cecum, ileum, transverse colon, descending colon, sigmoid colon, and rectum along the gut, and lumen inside the gut (Figure 1). If a gut is viewed as an ecosystem, different spatial positions inside and along the gut form multiple ecological habitats. Let  $y_{jkv}$  denote the abundance of microbe  $j$  at position  $k$  from gut  $v$ . Note that a microbe considered here may represent an OTU at any given level, such as individual, strain, species, genus, family, order, class, or phylum. For habitat  $k$  from gut  $v$ , the total amount of abundance of all colonizing microbes (say  $m$ ), i.e.,  $H_{kv} = \sum y_{jkv}$ , reflects its overall capacity to carry and feed the microbes with essential resources for their survival and propagation, which is defined as the habitat index. From an ecological perspective, the habitat index is determined by a mixture of factors including host genes, diet types and life style,<sup>36,37</sup> a concept similar to the environmental index coined to describe the overall quality of site in terms of the accumulative growth of all plants.<sup>38,39</sup> How much a given microbe within a habitat is expressed is the confounding consequence of its intrinsic capacity and its interactions with all possible biotic and abiotic factors. We thus define the abundance of a microbe in a habitat as the niche index.<sup>37</sup>

A habitat forms a microbial community within the gut, with habitat index ( $H_{kv}$ ) describing the whole community behavior and niche indices ( $y_{jkv}$ ) describing components of the community. From a physical perspective, habitat index and niche index establish a part-whole relationship, which can be described by the power function according to allometric scaling theory.<sup>40</sup> Let  $n$  denote the total number of habitats from all subjects (guts). Thus, for a given habitat  $i$  ( $i = 1, \dots, n$ ), this part-whole relationship can be mathematically expressed as

$$y_{ji} = \alpha_j H_i^{\beta_j} \quad (1)$$

where  $\beta_j$  is the scaling exponent and  $\alpha_j$  is an intercept constant of microbe  $j$ , which together determine the scaling shape of individual microbes with habitat index. Previous studies have shown that this power equation fits microbial abundance at any taxonomic levels across habitats.<sup>30</sup>

### Quasi-dynamic evolutionary game theory

Game theory states that a rational player strives to maximize its payoff by choosing an optimal strategy in response to the strategies of other players until the Nash equilibrium is reached.<sup>41,42</sup> As can be seen, such a strategy choice is not arbitrary, but rather includes a rational judgment based on a player's accrued knowledge of the environment affected by other players. However, it seems unreasonable to assume the rationality of microbes in making their decision. We introduce evolutionary game theory, the combined theory of game theory and evolutionary biology,<sup>43</sup> which uses the concept of an evolutionarily stable strategy (ESS) to refine the Nash equilibrium. In an evolving population, any strategy used by a player to maximize its payoff would be constrained by strategies of other players that also strive to maximize their own payoffs and, ultimately, this process through natural selection would optimize the structure and organization of the population, making it reach the maxima of its overall payoff.<sup>44</sup> The dynamic modeling of evolutionary game theory does not specify any individual ESS, but it bears all of the ESS that change in the population,<sup>45,46</sup> in which case the rationality assumption is relaxed.

Although there is no time dimension in our microbiota mapping study, allometric scaling law, described by equation (1), provides a bridge to dynamically model evolutionary game theory by deriving a system of generalized nonlinear predator-prey qdODE with the time derivative replaced by the habitat index derivative.<sup>28,29,30</sup> Each qdODE specifies the overall abundance of a microbe that is determined by its own “strategy” and the “strategies” of its interacting counterparts. The qdODE that quantify the nLV system are expressed as

$$\begin{aligned} \frac{dy_{ji}}{dH_i} &= Q_j(y_{ji}(H_i); \Theta_j) + \sum_{j'=1, j' \neq j}^m Q_{jj'}(y_{ji}(H_i); \Theta_{jj'}), i \\ &= 1, \dots, n; j = 1, \dots, m \end{aligned} \quad (2)$$

where the abundance of any microbe  $j$  expressed at habitat  $i$  is decomposed into the *independent* component ( $Q_j(\cdot)$ ) and *dependent* components ( $Q_{jj'}(\cdot)$ ). The independent component of microbe  $j$  is determined by its intrinsic capacity, which can be fully expressed when it is in isolation, whereas the dependent component of microbe  $j$  is the aggregated effect of influences on it by all other microbe  $j'$  ( $j' = 1, \dots, j-1, j+1, \dots, m$ ). Functions,  $Q_j(\cdot)$  and  $Q_{jj'}(\cdot)$ , may not have explicit forms, but they can be smoothed by a nonparametric approach, such as B-spline or Legendre orthogonal polynomials (LOP),<sup>47,48</sup> where  $\Theta_j$  and  $\Theta_{jj'}$  are the unknown parameters that specify the nonparametric curves. We code the estimates of  $Q_j(\cdot)$  and  $Q_{jj'}(\cdot)$  as nodes and edges, respectively, into quantitative networks as a tool to characterize the structure and organization of microbial community assembly along the gut.

### Sparse microbial networks

In practice, the number of microbes for network reconstruction may be very large, thus if the abundance of each microbe involves the effects of all other microbes, ODEs (2) will quickly become intractable. In biology, it is unlikely that each microbe performs an interaction with every other microbe in the community because a fully connected network is not helpful for the organism to maintain its robustness and stability in response to random perturbations.<sup>49,50,51</sup> Therefore, even if there are a number of microbes within a habitat,

interaction networks they constitute are likely to be sparse. Sparsely connected networks are consistent with Dunbar’s laws. In modeling social networks of non-human primates, Dunbar<sup>52</sup>, found that there is a limit to the number of relationships within a network an individual can stably maintain because of a limited size of its neocortex. Dunbar’s finding on the limit of social relationships, now known as Dunbar’s law, has been considered as a general argument for network reconstruction in a wide range of physical and life sciences.<sup>53</sup>

To choose a subset of the most significant microbes that interact with a given microbe, we formulate a multiple regression model that regresses the abundance of the microbe (response) on the abundance of all other microbes (predictors) across habitats. A regularization-based variable selection approach, such as LASSO<sup>54</sup>, and its variants,<sup>55,56,57</sup> is implemented to shrink the dimension of links possibly owned by a microbe.

**Regression model.** We consider each gut position from each individual (healthy or unhealthy) as a sample. A sample represents a microbial community assembly that contain  $m$  microbes. Let  $\mathbf{y}_j = (y_j(W_1), \dots, y_j(W_n))$  denote a vector of observed abundance values for microbe  $j$  ( $j = 1, \dots, m$ ) in  $n$  samples. Based on the structure of qdODE in equation (2), this microbe’s habitat index-varying abundance level can be described by a multiple regression model, expressed as

$$\begin{aligned} y_j(W_i) &= G_j(y_j(W_i) : \Theta_j) \\ &+ \sum_{j'=1, j' \neq j}^m G_{jj'}(y_{ji}(W_i) : \Theta_{jj'}) + e_j(W_i) \end{aligned} \quad (3)$$

$$= a_j(W_i) + X_j^T \mathbf{b}_j(W_i) + e_j(W_i). \quad (4)$$

In equation (3), and  $G_{jj'}(\cdot)$  are the habitat index-varying independent and dependent abundance of microbe  $j$ , whose derivatives are  $Q_j(\cdot)$  and  $Q_{jj'}(\cdot)$  of equation (1), respectively, and  $e_j(W_i)$  in equation (4) is the residual error of microbe  $j$  at sample  $i$ , obeying a multivariate normal distribution with mean vector  $\mathbf{0}$  and sample-dependent covariance matrix for microbe  $j$ . We assume that the residual errors of microbial abundance are independent among samples so that  $\sum_j$  is structured as  $\sum_j \sigma_j^2 \mathbf{I}_n$  where  $\sigma_j^2$  is



the residual variance of microbe  $j$  at the same sample and  $\mathbf{I}_n$  is the identity matrix. In equation (4), we have  $a_j(W_i) = G_j()$  and  $X_j^T \mathbf{b}_j(W_i) = \sum_{j'=1, j' \neq j}^m G_{jj'}()$ ,

where  $X_j^T$  is the vector containing  $m - 1$  ones and  $\mathbf{b}_j(W_i) = (b_{j1}(W_i), \dots, b_{jm}(W_i))$  is a vector of the dependent value of microbe  $j$  determined by all microbes, except for  $j$ .

**Group LASSO.** LASSO is particularly powerful for the penalty regression analysis of a response on an extremely large number of predictors across a much smaller size of samples. Considering a focal microbe  $j$  as a response, we use nonparametric independent parameters  $\mathbf{a}_j$  to fit its  $G_j()$ . As predictors,  $(m - 1)$  microbes contribute to microbe  $j$ 's dependent component through unknown nonparametric dependent parameters  $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{j1}, \dots, \boldsymbol{\beta}_{j(j-1)}, \boldsymbol{\beta}_{i(j+1)}, \dots, \boldsymbol{\beta}_{jm})$ . Thus, we have  $m - 1$  groups of dependent parameters that reflects the influence of other microbes on the focal microbe. We implemented group LASSO<sup>55</sup> to select those nonzero groups. The group LASSO estimators of dependent parameters, denoted as  $= (\boldsymbol{\beta}_{j1}, \dots, \boldsymbol{\beta}_{j(j-1)}, \boldsymbol{\beta}_{j(j+1)}, \dots, \boldsymbol{\beta}_{jm})$ , where  $d_j$  is the number of the most significant microbes that interact with microbe  $j$ . It can be obtained by minimizing the following penalized weighted least-square criterion,

$$L_1(\boldsymbol{\beta}, \lambda_j) = \left( \mathbf{y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j \right)^T \mathbf{Z}_j \left( \mathbf{y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j \right) + \lambda_{1j} \sum_{j'=1, j' \neq j}^m \|\boldsymbol{\beta}_{jj'}\|_2, \quad (5)$$

where  $\mathbf{y}_j = (y_j(W_1), \dots, y_j(W_n))$ ,  $\boldsymbol{\mu}_j = (\mu_j(W_1), \dots, \mu_j(W_n))$ , and  $\mathbf{b}_j = (\mathbf{b}_j(W_1), \dots, \mathbf{b}_j(W_n))$  determined by  $\boldsymbol{\beta}_j$  through a nonparametric link;  $\lambda_{1i}$  is a penalty parameter determined by BIC or extended BIC; and  $\mathbf{Z}_j = \text{diag}\{z_j(W_1), \dots, z_j(W_n)\}$  where  $z_j(W_i)$  is a prescribed nonnegative weight function on  $[W_1, W_n]$  with boundary conditions  $z_j(W_1) = z_j(W_n) = 0$ . This weight function is used to speed up the rate of convergence.

**Adaptive group LASSO:** In group LASSO, penalty parameters of each group are treated equally, without considering the relative importance of different groups. It has been recognized from traditional linear regression analysis that

the over-penalization of parameters for some predictors may reduce the efficiency of parameter estimation and the continuity of variable selection.<sup>55</sup> To overcome this limit of group LASSO, Wang and Leng<sup>56</sup> integrated it with adaptive LASSO to create adaptive group LASSO. This integrative approach selects significant groups by weighted penalty parameters. Weight  $w_{jj'}$  is obtained as  $\beta_{jj'}^{-1}$  if  $\beta_{jj'} > 0$  and  $\infty$ , otherwise. The adaptive group LASSO estimators of dependent parameters are obtained by minimizing the penalized weighted least-square criterion as follow:

$$L_2(\boldsymbol{\beta}, \lambda_j) = \left( \mathbf{Y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j \right)^T \mathbf{Z}_j \left( \mathbf{Y}_j - \mathbf{a}_j - X_j^T \mathbf{b}_j \right) + \lambda_{2j} \sum_{j'=1, j' \neq j}^m w_{jj'} \|\boldsymbol{\beta}_{jj'}\|_2, \quad (6)$$

where  $\lambda_{2j}$  is a penalty parameter determined by BIC or extended BIC.

After the most significant links (say  $d_j \ll m$ ) for each microbe  $j$  are detected, we substitute them into nLV qdODE in equation (2) to formulate a sparse representation of the full model. The sparse qdODE are written as

$$\frac{dy_{ji}}{dH_i} = Q_j(y_{ji}(H_i); \Theta_j) + \sum_{j'=1, j' \neq j}^{d_j} Q_{jj'}(y_{j'i}(H_i); \Theta_{jj'}), \quad i = 1, \dots, n; j = 1, \dots, m \quad (7)$$

where the notation of the independent and dependent component terms is the same as described in equation (2). By solving equation (7) in which the number of incoming links for a microbe  $j$  is changed from  $m$  to  $d_j$  through variable selection, we can reconstruct an  $m$ -node sparse microbial network.

### Reconstructing microbial networks via maximum likelihood

A number of approaches, including non-linear least-squares and maximum likelihood, can be implemented to solve the shrunk qdODE from

equation (2). Since we argue that microbial community assembly tends to reach its maximum overall payoff guided by evolutionary game theory, a maximum likelihood approach that is founded on the most probable existence of all microbes is chosen for our qdODE solving. Let  $\phi = (\boldsymbol{\mu}; \boldsymbol{\Sigma})$  denote the parameters that explain the regression model. The likelihood function of  $\phi$  given the abundance data is written as

$$\mathcal{L}(\boldsymbol{\mu}; \boldsymbol{\Sigma}) = f(y_1, \dots, y_m | \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m; \boldsymbol{\Sigma}) \quad (8)$$

where  $f(\cdot)$  is the  $n$ -dimensional  $m$ -variate normal distribution for  $m$  microbes across  $n$  samples with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . Specifically, we model the mean vector by equation (7) subject to variable selection, i.e.,

$$\begin{aligned} \boldsymbol{\mu} &= (\boldsymbol{\mu}_1; \dots; \boldsymbol{\mu}_m) = (\mu_1(W_1), \dots, \mu_1(W_n); \dots; \\ &\quad \mu_m(W_1), \dots, \mu_m(W_n)) \\ &= \left( G_1(y_1(W_n) : \Theta_1) + \sum_{j'=2}^{d_1} G_{1j'}(y_{j'}(W_n) : \Theta_{1j'}), \dots, \right. \\ &\quad G_m(y_m(W_1) : \Theta_m) + \sum_{j'=1}^{d_m} G_{mj'}(y_{j'}(W_1) : \Theta_{mj'}), \dots, \\ &\quad \left. G_m(y_m(W_n) : \Theta_m) + \sum_{j'=1, j'}^{d_m} G_{mj'}(y_{j'}(W_n) : \Theta_{mj'}) \right) \end{aligned} \quad (9)$$

Now, we implement a power equation-based LOP nonparametric approach to fit  $G_j(\cdot)$  by parameters  $\Theta_j$  and fit  $G_{jj'}(\cdot)$  by parameters  $\Theta_{jj'}$ . We use a statistically robust approach for modeling the covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sum_1 & \cdots & \sum_{1m} \\ \vdots & \ddots & \vdots \\ \sum_{m1} & \cdots & \sum_m \end{pmatrix} \quad (10)$$

where  $\sum_j$  is the sample-dependent covariance matrix of microbe  $j$ , and  $\sum_{j_1 j_2}$  is the sample-dependent covariance matrix between microbes  $j_1$  and  $j_2$ .  $\sum_{j_1 j_2}$  is structured as  $\sum_{j_1 j_2} \sigma_{j_1 j_2} \mathbf{I}_n$ , where  $\sigma_{j_1 j_2}$  is the residual covariance of microbes  $j_1$  and  $j_2$  at the same sample. If there is a mix of static and temporal data involved, we may implement the first-order autoregressive (AR(1)) model to model autocorrelative structure of the covariance.

Under mean-covariance structure modeling by equations (9) and (10), model parameters  $\phi = (\boldsymbol{\mu}; \boldsymbol{\Sigma})$  become model parameters  $\phi' = [\Theta_j, \Theta_{jj'} (j = 1, \dots, m, j' = 1, \dots, j-1, j+1, \dots, d_j); \sigma_j^2, \sigma_{j_1 j_2} (j_1 \neq j_2 = 1, \dots, m)]$ , whose optimal solution can be obtained, by maximizing the likelihood (8), as

$$\hat{\phi}' \in \left\{ \arg \max_{\phi' \in \Phi'} \mathcal{L}(\phi') \right\} \quad (11)$$

Intuitively, this maximization implies an optimal topological structure and organization by which microbes interact with each other to maximize the overall abundance of microbial community assemble as a whole. The convex optimization formulation under equation (11) ensures the stability and sparsity of the network reconstructed from qdODE of equation (7). Since no constraints are given on the number of outgoing links, the resulting network can be high-dimensional and of a large-scale size.

### Topological dissection of microbial networks

Inferred microbial networks via maximizing the likelihood function (8) meet three essential properties of networks, i.e., causality (derived from directed qdODE), sparsity (due to variable selection), and stability (assured under optimization). Apart from the methodological advantages, these networks have many biological merits. First, they are informative because they can capture the full properties of microbial interactions, including *bidirectionality*, *sign*, *weight*, and *feedback loop*.<sup>28</sup> Second, they are dynamically visualizable by reconstructing a series of quasi-dynamic networks along HI gradient from the instantaneous estimates of independent and dependent components for each sample.

Third, the networks reconstructed by qdODE can omnidirectionally capture ecological interactions that occur among microbes. It can cover all possible types of microbial interactions, including *mutualism* (two microbes promote each other by producing factors that are beneficial for both interacting parties), *antagonism* (two microbes inhibit each other), *commensalism* (one microbe promotes its partner whereas the latter does not affect the former),

*amensalism* (one microbe inhibits the other and the other is neutral), and *parasitism* (one microbe inhibits the other but the latter promotes the former). The opposite to parasitism is *altruism* (one microbe promotes the other but the latter inhibits the former).<sup>25,49,52</sup> A microbe may actively manipulate other microbes (by promoting or inhibiting the latter) and, meanwhile, it may be passively manipulated by other microbes. In an idopNetwork, one can identify the numbers of such active links and passive links for each microbe. If a microbe has more active links than passive links, it is regarded as a leader microbe. If a microbe's active links are more than the average of all microbes (i.e., connectivity), then this microbe is a mighty hub or keystone microbe that is believed to play a pivotal role in maintaining microbial communities. If a microbe has less links, including active and passive, than the average, it is a solitary microbe. The ecological interpretation of these strategies will stimulate researchers to explore the mass, energetic, or signal basis of microbial interactions.<sup>53</sup>

Fourth, taking the means of the estimates of independent and dependent components for a subject from qdODE of equation (7), we can identify a network specifically for this subject. Such personalized microbial networks may provide unique information for designing personalized medicine based on the gut microbiota. Taken together, the qdODE-based networks are informative, dynamic, omnidirectional, and personalized, thus call ideopNetworks.<sup>28</sup>

### Inferring context-specific microbial networks

In general, gut-microbial studies are designed in a case-control cohort fashion, which allows microbial interactions to be compared between a diseased group and healthy group. This can be achieved by reconstructing context-specific networks. Consider a panel of subjects who are classified into  $C$  groups or contexts (such as different races, sexes, use vs. no use of a drug, etc.). Here, we formulate a likelihood function which is the same as equation (8), but the mean vector is now modeled as

$$\begin{aligned} \boldsymbol{\mu} &= (\boldsymbol{\mu}_1; \dots; \boldsymbol{\mu}_m) = (\mu_1(W_1), \dots, \mu_1(W_n); \dots; \mu_m(W_1), \dots, \mu_m(W_n)) \\ &= \left( \sum_{c=1}^C \left[ G_1(y_1(W_n) : \Theta_1^c) + \sum_{j=2}^{d_1} G_{1j}(y_j(W_n) : \Theta_{1j}^c) \right], \dots \right. \\ &\quad \left. , \sum_{c=1}^C \left[ G_m(y_m(W_1) : \Theta_m^c) + \sum_{j=1}^{d_m} G_{mj}(y_j(W_1) : \Theta_{mj}^c) \right], \dots, \right. \\ &\quad \left. \sum_{c=1}^C \left[ G_m(y_m(W_n) : \Theta_m^c) + \sum_{j=1}^{d_m} G_{mj}(y_j(W_n) : \Theta_{mj}^c) \right] \right) \quad (12) \end{aligned}$$

and the covariance matrix modeled similarly as above.

By plugging in the MLEs of mean vectors (9) and (12) into likelihood (8), we obtain the likelihood values  $L_1$  (assuming that there is no difference among  $C$  contexts) and  $L_2$  (assuming that there are differences among  $C$  contexts), respectively. We further estimate the log-likelihood ratio (LR),

$$LR = -2\log(L_0/L_1) \quad (13)$$

as a statistic used to test if  $n$  samples should be sorted into  $C$  contexts. By reshuffling  $n$  samples randomly into  $C$  groups, we calculate the LR value. If this permutation procedure is repeated 1,000 times, we obtain the 95th percentile from 1,000 LR values and use it as a critical threshold.

### Spatial mapping design

#### Patient recruitment and ethics

Six volunteers were recruited for the study at the Department of Gastroenterology and Hepatology of Tianjin Medical University General Hospital. Disease severity in five patients infected with UC was assessed by the modified Mayo endoscopic score. One patient with colonic polyp was marked as a non-UC control (HC). Ethics approval was received from the Tianjin Medical University General Hospital Clinical Research Ethics Committee. All patients signed the informed consent form prior to their operation, and they received polyethylene glycol-based bowel preparation for colonoscopy. Demographic data and clinical characteristics of the UC and non-UC patients are shown in Table 1.

#### Sample collection and brush sampling

Intestinal sampling involves the lumen and six positions long the gut, ileum, cecum, transverse colon, descending colon, sigmoid colon, and rectum (Table 1). The specimens were sampled from three

**Table 1.** Disease status and sampling information of UC and non-UC patients.

Patient No.	Sex	Age	UC Stage	Sampling Site	Sampling Approach
1	F	47	Moderate	Cecum Rectum Transverse colon Lumen	BF,DSB,PSB BF,DSB,PSB BF,DSB,PSB
2	M	32	Severe	Cecum Ileum Transverse colon Descending colon	BF,DSB,PSB BF,DSB,PSB BF,DSB,PSB BF,DSB,PSB
3	F	30	Moderate	Rectum Cecum Transverse colon Sigmoid colon Lumen	DSB,PSB DSB,PSB DSB,PSB BF,DSB,PSB
4	F	60	Severe	Cecum Rectum Ileum Transverse colon Lumen	BF BF,DSB,PSB BF,DSB,PSB BF,DSB,PSB
5	M	62	Moderate	Rectum Lumen	DSB,PSB
6	F	47	Non-UC	Cecum Ileum Transverse colon	BF,DSB,PSB BF,DSB,PSB BF,DSB,PSB

BF: biopsy forceps; DSB: disposable specimen brush; PSB: protected specimen brush

DSB used clinically has no plugging structure and is easily disturbed by body fluids or tissues. To address the defects of the existing technology, we developed a PSB as a contrast, which set up a closed plugging device at the end of the DSB seeker using polyethylene glycol. The plugging portion can be tightly integrated with the inner wall of the sheath. The brush was enclosed in the sheath, introduced through the colonoscope channel, and exposed to the external environment when the brush was sent out to the front of the colonoscope. Specific sampling parts were rubbed and rotated around bristles to obtain superficial microbes. After sampling, the brush was retracted into the sheath, then slowly extracted from the colonoscope channel. Samples were immediately placed on liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until DNA extraction.

to four positions in each patient. A total of 23 specimens were collected from four UC patients and one HC.

#### DNA extraction and amplicon sequencing

The QIAamp DNA Mini Kit (Qiagen, Inc, Hilden, Germany) was used to extract DNA from biopsies and brush specimens, according to the manufacturer's instructions. Primers 515 F and 806 R, each with a barcode and adaptor sequences for high-throughput sequencing, were utilized to amplify the bacterial 16S rRNA gene V4 region; 16S rRNA sequencing was then performed on the Illumina HiSeq2500 platform by Novogene (Beijing, China).

#### Bioinformatics analysis

After removing the barcodes and primers, we merged the reads of each sample using FLASH (version 1.2.7) software to obtain the raw reads.<sup>58</sup> According to the raw read quality control process of Quantitative Insights Into Microbial Ecology (QIIME, version 1.7.0),<sup>59</sup> merged raw reads were qualified to generate

clean reads: raw reads were cut off at the first base of three or more continuous low-quality bases (the default quality threshold was 19); reads with continuous high-quality bases less than 75% were further filtered out. Chimeras were identified and removed through the UCHIME Algorithm against the Gold database to generate effective reads.<sup>60</sup> Uparse was used to cluster all effective reads into OTUs with 97% identity,<sup>61</sup> the highest frequency sequence in each OTU was selected as the representative sequence. The taxonomies were assigned using the Ribosomal Database Project (RDP) classifier<sup>62</sup> against the Greengenes database (the confidence threshold was 80%), then taxonomic information was obtained and the community composition of each sample was summarized at each level.<sup>63</sup> The sequences of all samples were normalized based on the sample with the lowest number of sequences (47,986 sequences). The absolute abundance levels of microbes at different taxonomic levels were inferred by taking the product of relative abundance (assessed by 16S rRNA gene amplicon sequencing) and bacterial

load at each position (measured by broad-range 16S rRNA gene qPCR).<sup>64</sup>

## Acknowledgments

We specially thank our colleague Jing Feng from the Department of Respiratory and Critical Care Medicine for his important advice. We appreciate Jun He from Jiangsu Vedkang Medical Science & Technology Co., Ltd for technology support. We are grateful to our colleagues Li Liang, Lin Fang, Fengying Tian, Lu Xiao, Ying Li, Yan Wang, Ting Li, Xuele Lu, Qi Zhang, Xinyao Zhao (Department of Gastroenterology and Hepatology Endoscopy Center) and Wenjing Song (Department of Pathology), Xin Zhao (Department of Radiology) for their professional help. We sincerely thank the patients and their family for their support and dedication.

## Data and code availability statement

The original nucleotide sequences of this study were deposited to the NCBI Sequence Read Archive under accession number SRP128619. The computer code used in this study can be downloaded at <https://github.com/CCBBeijing/GutMicrobe> or requested from the corresponding author.

## Disclosure statement

There is no potential conflict of interest by the authors.

## Funding

This work was supported by grants from National Key Research and Development Project (No. 2019YFA0905600), Science and Technology Program of Tianjin, China (No. 19YFSLQY00110), and the National Natural Science Foundation of China (82070559).

## Author contributions

X.C. initiated the project and supervised sample and data collection. A.D. and S. W. performed statistical modeling and wrote computer code. G.K., H.H., X.W., L.D., H.H., T. Z. and X.L. collected samples and data. R.W. conceived of the study and wrote the manuscript, with inputs from X.C., H.H., and X.L.

## References

1. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol.* 2021;19(1):55–71. doi:10.1038/s41579-020-0433-9.
2. Li Y, Luo ZY, Hu YY, Bi YW, Yang JM, Zou WJ, YL Song, S Li, T Shen, SJ Li, et al. The gut microbiota regulates autism-like behavior by mediating vitamin B<sub>6</sub> homeostasis in EphB6-deficient mice. *Microbiome.* 2020;8:120. doi:10.1186/s40168-020-00884-z.
3. Muscogiuri G, Cantone E, Cassarano S, Tuccinardi D, Barrea L, Savastano S, Colao, A. Gut microbiota: a new path to treat obesity. *Intl J Obes Supp.* 2019;9:10–19. doi:10.1038/s41367-019-0011-7.
4. Qin N, Yang FL, Li A, Prifti E, Chen YF, Shao L, Guo J, Emmanuelle C, Yao J, LJ Wu, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature.* 2014;513:59–64. doi:10.1038/nature13568.
5. Zitvogel L, Galluzzi L, Viaud S, Vétizou M, Daillère R, Merad M, Kroemer, G. Cancer and the gut microbiota: an unexpected link. *Sci Transl Med.* 2015;7:271. doi:10.1126/scitranslmed.3010473.
6. Christian LM, Galley JD, Hade EM, Schoppe-Sullivan S, Dush KC, Bailey MT. Gut microbiome composition is associated with temperament during early childhood. *Brain Behav Immun.* 2015;45:118–127. doi:10.1016/j.bbi.2014.10.018.
7. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell.* 2012;148:1258–1270. doi:10.1016/j.cell.2012.01.035.
8. Deines P, Bosch TC. Transitioning from microbiome composition to microbial community interactions: the potential of the eukaryote *Hydra* as an experimental model. *Front Microbiol.* 2016;7:1610. doi:10.3389/fmicb.2016.01610.
9. Johnson KV. Gut microbiome composition and diversity are related to human personality traits. *Human Microbiome J.* 2020;15:100069. doi:10.1016/j.humic.2019.100069.
10. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, Gibbons SM, Magis AT. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun.* 2020;11(1):5206. doi:10.1038/s41467-020-18871-1.
11. Donaldson GP, Chou W-C, Manson AL, Rogov P, Abeel T, Bochicchio J, Ciulla D, Melnikov A, Ernst PB, Chu H, et al. Spatially distinct physiology of *Bacteroides fragilis* within the proximal colon of gnotobiotic mice. *Nat Microbiol.* 2020;5(5):746–756. doi:10.1038/s41564-020-0683-3.
12. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol.* 2016;14(1):20–32. doi:10.1038/nrmicro3552.
13. Engevik M, Versalovic J. Taking a closer look at the biogeography of the human gastrointestinal microbiome. *Gastroenterology.* 2019;157:927–929.
14. Mark Welch JL, Hasegawa Y, McNulty NP, Gordon JL, Borisy GG. Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. *Proc Natl Acad Sci USA.* 2017;114:9105–9114. doi:10.1073/pnas.1711596114.



15. Riva A, Kuzyk O, Forsberg E, Siuzdak G, Pfann C, Herbold C, Daims H, Loy A, Warth B, Berry D. A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome. *Nat Commun.* 2019;10:4366. doi:10.1038/s41467-019-12413-0.
16. Shi H, Shi Q, Grodner B, Lenz JS, Zipfel WR, Brito IL, Vlaminc ID. Highly multiplexed spatial mapping of microbial communities. *Nature.* 2020;588:676–681. doi:10.1038/s41586-020-2983-4.
17. Stearns J, Lynch M, Senadheera D, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Hagelsieb GM, Neufeld JD. Bacterial biogeography of the human digestive tract. *Sci Rep.* 2011;1:170. doi:10.1038/srep00170.
18. Tropini C, Earle KA, Huang KC, Sonnenburg JL. The Gut Microbiome: connecting Spatial Organization to Function. *Cell Host Microbe.* 2017;21:433–442. doi:10.1016/j.chom.2017.03.010.
19. Chen L, Collig V, Jaeger M, Munckhof ICL, Vila AV, Kurilshikov A, Gacesa R, Sinha T, Oosting M, Joosten LAB, et al. Gut microbial co-abundance networks show specificity in inflammatory bowel disease and obesity. *Nat Commun.* 2020;11:4018. doi:10.1038/s41467-020-17840-y.
20. Hall CV, Lord A, Betzel R, Zakrzewski M, Simms LA, Zalesky A, Smith GR, Cocchi L. Coexistence of network architectures supporting the human gut microbiome. *iScience.* 2019;22:380–391. doi:10.1016/j.isci.2019.11.032.
21. Jiang LB, Liu XJ, He XQ, Jin Y, Cao YG, Zhan X, Griffin GH, Gragnoli C, Wu RL. A behavioral model for mapping the genetic architecture of gut-microbiota networks. *Gut Microbes.* 2020;1–15. doi:10.1080/19490976.2020.1820847.
22. Layeghifard M, Li H, Wang PW, Donaldson SL, Coburn B, Clark ST, Caballero JD, Zhang Y, Tullis DE, Yau YCW, et al. Microbiome networks and change-point analysis reveal key community changes associated with cystic fibrosis pulmonary exacerbations. *npj Biofilms Microbiomes.* 2019;5:4. doi:10.1038/s41522-018-0077-y.
23. Riera JL, Baldo L. microbial co-occurrence networks of gut microbiota reveal community conservation and diet-associated shifts in cichlid fishes. *Anim Microbiome.* 2020;2:36. doi:10.1186/s42523-020-00054-4.
24. Dohlman AB, Shen X. Mapping the microbial interactome: statistical and experimental approaches for microbiome network inference. *Exp Biol Med.* 2019;244:445–458. doi:10.1177/1535370219836771.
25. Coenen AR, Weitz JS. Limitations of correlation-based inference in complex virus-microbe communities. *mSystems.* 2018;3:e00084–18. doi:10.1128/mSystems.00084-18.
26. Szal M, Mathee K, Ruiz-Perez D, Cickovski T, Narasimhan G. Inferring directional relationships in microbial communities using signed Bayesian networks. *BMC Genomics.* 2020;21:663. doi:10.1186/s12864-020-07065-0.
27. Xiao Y, Angulo MT, Friedman J, Waldor MK, Weiss ST, Liu YY. Mapping the ecological networks of microbial communities. *Nat Commun.* 2017;8:2042. doi:10.1038/s41467-017-02090-2.
28. Chen CX, Jiang LB, Fu GF, Wang M, Wang YQ, Shen BY, Liu ZQ, Wang ZH, Hou W, Berceci SA, et al. An omnidirectional visualization model of personalized gene regulatory networks. *npj. Syst Biol Appl.* 2019;5:38.
29. Griffin CH, Jiang LB, Wu RL. Analysis of quasi-dynamic ordinary differential equations and the quasi-dynamic replicator. *Physica A: Stat Mech Its Appl.* 2020;555:124422.
30. Wu RL, Jiang LB. Recovering dynamic networks in big static datasets. *Phys Rep.* 2021;912:1–57. doi:10.1016/j.physrep.2021.01.003.
31. Kim BR, Zhang L, Berg A, Fan, J, and Wu, RL. A computational approach to the functional clustering of periodic gene-expression profiles. *Genetics.* 2008;180 2 :821–834. doi:10.1534/genetics.108.093690.
32. Nadell CD, Drescher K, Foster KR. Spatial structure, cooperation and competition in biofilms. *Nat Rev Microbiol.* 2016;14:589–600. doi:10.1038/nrmicro.2016.84.
33. Durack J, Lynch SV. The gut microbiome: relationships with disease and opportunities for therapy. *J Exp Med.* 2019;216:20–40. doi:10.1084/jem.20180448.
34. Ford SA, King KC. Harnessing the power of defensive microbes: evolutionary implications in nature and disease control. *PLoS Pathog.* 2016;12:e1005465. doi:10.1371/journal.ppat.1005465.
35. Piewngam P, Zheng Y, Nguyen TH, Dickey SW, Joo HS, Villaruz AE, Glose KA, Fisher EL, Hunt RL, Li B, et al. Pathogen elimination by probiotic *Bacillus* via signaling interference. *Nature.* 2018;562:532–537. doi:10.1038/s41586-018-0616-y.
36. Scepanovic P, Hodel F, Mondot S, Partula V, Byrd A, Hammer C, Alanio C, Bergstedt J, Patin E, Touvier M, et al. A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome.* 2019;7:130. doi:10.1186/s40168-019-0747-x.
37. Pereira FC, Berry D. Microbial nutrient niches in the gut. *Environ Microbiol.* 2017;19:1366–1378. doi:10.1111/1462-2920.13659.
38. Finlay KW, Wilkinson GN. The analysis of adaptation in a plant breeding program. *Aust J Agr Res.* 1963;14:742–754. doi:10.1071/AR9630742.
39. Lobell DB, Roberts MJ, Schlenker W, Braun N, Little BB, Rejesus RM, Hammer GL. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest Science. 2014;344:516–519. doi:10.1126/science.1251423.
40. Shingleton A. Allometry: the study of biological scaling. *Nat Ed Knowl.* 2000;3:2.
41. Nash JF. Equilibrium points in n-person games. *Proc Natl Acad Sci.* 1950;36:48–49. <https://doi.org/10.1073/pnas.36.1.48>

42. Von Neumann J, Morgenstern S. Theory of games and economic behavior. Princeton: Princeton University Press; 1944.
43. Smith JM, Price GR. Logic of animal conflict. *Nature*. 1973;246:15–18. doi:10.1038/246015a0.
44. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. *Nature*. 2017;548:43–51.
45. Hofbauer J, WH S. Stable games and their dynamics. *J Econ Theor*. 2009;144:93–1665. doi:10.1016/j.jet.2009.01.007.
46. Liao D, Tlsty TD. Evolutionary game theory for physical and biological scientists. II. Population dynamics equations can be associated with interpretations. *Interface Focus*. 2014;4:20140038. doi:10.1098/rsfs.2014.0038.
47. Das K, Li J, Wang Z, Tong CF, Fu GF, Li Y, Xu M, Ahn K, Mauger D, Li RZ, et al. A dynamic model for genome-wide association studies. *Hum Genet*. 2011;129:629–639. doi:10.1007/s00439-011-0960-6.
48. Jiang LB, Liu JY, Zhu XL, Ye MX, Sun LD, Lacaze X, Wu RL. 2HiGWAS: a unifying high-dimensional platform to infer the global genetic architecture of trait development. *Brief Bioinform*. 2015;16:905–911. doi:10.1093/bib/bbv002.
49. May RM. Will a large complex system be stable? *Nature*. 1972;238:413–414. doi:10.1038/238413a0.
50. Allesina S, Tang S. Stability criteria for complex ecosystems. *Nature*. 2012;483:205–208. doi:10.1038/nature10832.
51. Busiello DM, Suweis S, Hidalgo J, Maritan A. Explorability and the origin of network sparsity in living systems. *Sci Rep*. 2017;7:12323. doi:10.1038/s41598-017-12521-1.
52. Dunbar RIM. Neocortex size as a constraint on group size in primates. *J Hum Evol*. 1992;22:469–493. doi:10.1016/0047-2484(92)90081-J.
53. Harre M, Prokopenko M. The Social Brain: scale-invariant layering of Erdos-Renyi networks in small-scale human societies. *J Roy Soc Interf*. 2016;13:1–6. doi:10.1098/rsif.2016.0044.
54. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc*. 1996;58:267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
55. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J Roy Stat Soc B*. 2006;68:49–67. doi:10.1111/j.1467-9868.2005.00532.x.
56. Wang H, Leng C. A note on the adaptive group Lasso. *Comput Stat Data Anal*. 2008;52:5277–5286. doi:10.1016/j.csda.2008.05.006.
57. Zou H. The adaptive Lasso and its oracle properties. *J Am Stat Assoc*. 2006;101:1418–1429. doi:10.1198/016214506000000735.
58. Magoc T and Salzberg S L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963. 10.1093/bioinformatics/btr507
59. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–336. doi:10.1038/nmeth.f.303.
60. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27:2194–2200. doi:10.1093/bioinformatics/btr381.
61. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996–998. doi:10.1038/nmeth.2604.
62. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73:5261–5267. doi:10.1128/AEM.00062-07.
63. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72:5069–5072. doi:10.1128/AEM.03006-05.
64. Tettamanti Boshier FA, Srinivasan S, Lopez A, Hoffman NG, Proll S, Fredricks DN, and Schiffer JT. Complementing 16S rRNA gene amplicon sequencing with total bacterial load to infer absolute species concentrations in the vaginal microbiome. *mSystems*. 2020;5:e00777–19. doi:10.1128/mSystems.00777-19.