Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

Article

Optimal Microbiome Networks: Macroecology and Criticality

Jie Li [1,2‡], **Matteo Convertino** [1,2‡*]

[1] Nexus Group, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, JP
[2] GI-CORE Global Station for Big Data and Cybersecurity, Hokkaido University, Sapporo, JP
* Correspondence: matteo@ist.hokudai.ac.jp; Tel.: +1-781-645-6070
‡ These authors contributed equally to this work.

Abstract: The human microbiome is an extremely complex ecosystem considering the amount of bacterial species, their interactions, and its variability over time. Here we untangle the complexity of the human microbiome for the Irritable Bowel Syndrome (IBS) that is the most prevalent functional gastrointestinal disorder in human populations.
Based on a novel information theoretic network inference model we detect species interaction networks that are functionally and structurally different for healthy and unhealthy individuals. Healthy networks are characterized by a neutral symmetrical pattern of species interactions and scale-free topology versus random unhealthy networks. We detect an inverse scaling relationship between species total outgoing information flow, meaningful of node interactivity, and relative species abundance (RSA). The top ten interacting species are also the least relatively abundant for the healthy microbiome and the most detrimental. These findings support the idea about the diminishing role of network hubs and hubs should be defined considering the total outgoing information flow rather than the node degree. Macroecologically, the healthy microbiome is characterized by the highest total species diversity growth rate, the lowest species turnover, and the smallest variability of RSA for all species. This result challenges current views that posit a universal association between healthy states and the highest absolute species diversity in ecosystems. Additionally, we show how the transitory microbiome is unstable and microbiome criticality is not at the phase transition between healthy and unhealthy states. We stress out the importance of considering interacting pairs versus single node dynamics when characterizing the microbiome and of ranking these pairs in terms of their dynamics. Interactions (i.e. species collective behavior) shape transition from healthy to unhealthy states. The macroecological characterization of the microbiome is useful for diagnostic purposes and disease etiognosis, while species-specific analyses can detect species that are more beneficial leading to personalized design of pre- and pro-biotic treatments and microbiome engineering.

Keywords: microbiome, complex networks, species diversity, criticality, RSA, information flow, transitions

1. Introduction

1.1. Microbiome Dynamics and Health

Microbial ecology has become an important topic for health sciences and other sciences such as biology, ecology, forensic and agriculture. Recent work has shown how each person maintains a fairly unique microbial fingerprint, and that microbial dysbioses are often associated with shifts in health-status. These shifts are typically associated with the gut that is the most diverse part of

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

the human body considering the bacteria holobiont [20,82]. We recognize that our microbiota are highly dynamic, and that these dynamics are linked to environmental and individual states [82]. The field is still in its infancy and it is not yet settled upon whether gut microbial community structure varies continuously or if it jumps between "discrete" community states, and whether these states are shared across individuals. In particular, some researchers suggest that gut communities can be binned into discrete enterotypes [3], while others argue that gut communities vary along multidimensional continua without any universality [45]. If the ultimate goal of microbiome research is to improve human health by engineering the ecology of the gut, and other applications are also of interest, we must first understand how and why our microbiota varies in time, whether these dynamics are consistent across humans, and whether we can define stable or healthy dynamics. This line of research is primarily missing how microbial diversity is organized considering all its facets and how this diversity changes when species interaction networks change. For instance the same level of diversity can be achieved via different network topologies that may lead to different health states [12].

1.2. Microbiome Diversity and Functional Network Organization

In order to determine the network organization of the microbiome and associate that to healthy or unhealthy states we consider the Irritable Bowel Syndrome (IBS) as the template syndrome to characterize microbiome dynamics [60,74]. IBS shows common symptoms of cramping, abdominal pain and diarrhea related to altered gut flora. Previous research has found that the microbiome in people with IBS differs from that in healthy people [60]; however, nobody demonstrated how the microbiome network is different for these healthy and unhealthy groups (i.e. "states" generally speaking when not focused on a particular subpopulation) and how the transition from one to another occurs. By exploring this topic we propose novel network inferential models for gathering microbiome networks from species big-data; these models are based on the principle of maximum entropy that tries to gather the most informative set of variables about stable state patterns [31,52,59,73]. An example can be about sets of species for predicting a diverse set of species networks. Big-data is not necessarily related to the size of the data used but also to the number of calculations required to infer the underlying network structure. These computations increase exponentially with the number of species/nodes n considered beyond the geometrical criteria, where the number of connections is $n(n-1)$, because of the undirected topology of the network. A directed topology is for instance found when species interaction networks are non-symmetrical which means that the direct influence of two species does not have the same magnitude for different directions of interactions [48]. A variety of different models have been proposed to infer network structures from small and large datasets. For biological systems in particular, the inference of causal interactions among systems' components is a daunting task because not all interactions are known, or the "true" magnitude of interactions, considering the data used to assess these interactions. For instance microbiome networks are in principle different if the used input data are species occurrence, relative species abundance (RSA), geographic range or other features. Also for this motivation, we employ assumption free inference models that consider the whole probability distribution of species dynamics and that are validated considering their ability to predict population patterns over time. We extract optimal microbiome networks as optimal information networks (OINs) [73] for healthy, transitory and unhealthy groups to investigate general patterns and drivers underlying microbiome stability and the interactions among different species in terms of network topology, magnitude and preferential direction. Additionally, we characterize macroecological functions α, β and γ-diversity which describe the temporal organization of microbiome biodiversity considering point time intertemporal and total diversity. We show how these functions are related to microbiome network features and different topologies emerge for different diversity/health states. The linkage between microbiome networks and macroecology is unique and offers additional insights into the ecology and the evolution of the microbiome with

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

relevance to ecosystem health.

1.3. Microbiome Inference, Neutrality and Criticality

Speculations about the underlying processes of ecosystems' organization have been moved in the past considering diversity patterns and models able to predict these patterns such as neutral models [4,15,61,91] and other models [75]. Neutral models posit that biological diversity is driven solely by ecological drift without a strong interference of environmental biases that lead to preferential dynamics ("niche") for some species versus others. From neutral to niche states a critical transition is typically observed where species network organization exhibits scale-free behavior [17,28,29,41,46,58]. This scale-free behavior was though to occur only at the critical transition point but recent evidence also shows that criticality [9] does exist also for stable states where system's component organization is optimal due to optimal information sharing among components and the environment [37,61]. This result was already found for geophysical networks and coupled ecological networks [7,16] for instance where energy dissipation tends to a global minimum.

Some indications that microorganism cooccurrence patterns are shaped by species interactions that are altered from niche to neutral is available (e.g. [26,41]). This has also conceptual and numerical confirmation when thinking and simulating species that are just responding to local resources and species that are somehow "equal" and responding to fundamental speciation-dispersal processes. The former are interacting more randomly with limited dispersal ranges while the latter are interacting with much larger dispersal ranges. Without introducing any model (but with the knowledge of the underlying potential processes) these changes in network topologies have been observed for large scale ecosystems [19].

However, these models of microbiome characterization are typically driven by some assumptions about the species interaction network which may lead to erroneous conclusions about the fitted patterns: in other words, predictability (under some assumptions) of biological patterns does not imply causality considering the hypothesized and implemented processes [76]. Model of microbiome network inference exist (see e.g. [5] and [75]) but they simply infer species co-occurrence network without assessing the magnitude and directionality of potential species interdependence. A different approach is achieved by pattern-oriented models charactering systems' dynamics [25,33,36] such as the one here proposed, that do not assume any mechanism a priori and consider the whole information content in data (via probability distributions and their relevance to predict patterns via entropic functions [59]) to claim underlying processes. In this sense we move our discussion of the problem of understanding microbiome dynamics toward which information is critical, and how that model criticality [59,62] is associated to biological criticality [9] also considering the neutrality of diversity dynamics. Therefore, rather than trying to untangle biological complexity via fitting some biologically inspired models we use all data available to check their information content to define all possible microbiome states and associated diversity patterns. In this information theoretic framework, in particular we show how criticality coincides with neutrality and optimal microbial network organization that lead to healthy states. We also show how criticality corresponds to a scale-free functional networks relating RSA interdependencies even when the functional co-occurrence network of species is not scale-free.

As a caveat, it should be noted that neutral patterns does not necessarily imply neutral processes [14] despite many papers try to define one from the other [41,47,49,51,53]. Furthermore neutral models can predict non-neutral processes (care must be made when considering predictability vs causality) and neutrality might not be present at all scales of biological organization [51]. The focus here is on microbiome pattern detection and predictability which we believe to be extremely important and the starting point for a top-down investigation of the underlying processes. Different patterns are evident for different health states when RSA interdependence networks are considered, and these networks seem to shape microbiome diversity.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

2. Material and Methods

2.1. Microbiome Data

We considered microbiome data originally published by Durbán et al. [24] and later on used by Martí et al. [60] for which species data of six individuals was available over time (30 days). Fine scale species Operational Taxonomic Unit (OTU) RSA data were derived by published 16S rRNA and shotgun metagenomic sequencing (SMS) data pertaining to the gut microbiotas. In Durbán et al. [24], species-level phylotypes were defined at 97% of sequence identity, that is the lowest taxonomic rank that is used to identify differences in biological states of interest (e.g. healthy and unhealthy). Two individuals suffered from IBS. Two others were healthy and the other two individuals, one was treated with antibiotics and the other was at the verge to be unhealthy. The latter individuals are considered to be part of a "transitory" group between healthy and unhealthy. In Durbán et al. [24] the healthy subjects were considered as those individuals who did not suffer from lab-confirmed IBS, while the patients who had this disease were taken as perturbations. In the dataset [24] the healthy period is from time points before the IBS triggering event altering the microbiome.

More specifically, the datasets are composed by 2 healthy individuals (A and B individuals in the original datasets [24,60]), 2 transitory individuals (C and C1), and two patients (P1 and P2). The length of data for these individuals is for A (RSAs for 30 day), B (15 days), C (15 days), C1 (9 days), P1 (9 days), and P2 (14 days).

2.2. Time Series Reconstruction

The raw data available present the challenge of individuals sampled for different time lengths. Computationally, in order to have datasets with the same length we used the concept of Least Common Multiple (LCM) [21]. LCM extends time series at their maximum by preserving their probability distribution functions (pdfs); in our case the pdfs are associated to each RSA. LCM guarantees to have the largest dataset representative of the stochastic dynamics analyzed. We calculated LCM considering the number of data for each health group. This implies to have more data for smaller groups and to maintain the data length for the data-rich group. LCM between individuals A and B is 30; yet, the length of the abundance time series for A is unchanged while B becomes 30 (B is repeated twice). This is done by copying the data in B till the 30th day. LCM for C and C1 is 45; yet, both C and C1 time series were extend to 45. LCM for P1 and P2 is 126; yet, both time series have been expanded to the 126th day. These examples show that data rich sample are more kept as they are while data poor samples are extended. In order to create pdfs of RSA representative of each group we considered the average values of RSA for common species. If for individuals belonging to the same group different species were found, the pdf of RSA was just based on the time series as they were. This choice is dictated by the desire to emphasize common dynamics for each group when possible.

2.3. Probabilistic Characterization of the Microbiome

We characterize probabilistically the distribution of microbiome macroecological and species interaction network variables (generally indicated as Y as for a generic random variable) considering the following general exceedance probability distribution function (see Convertino et al. [18]):

$$P(Y \geq y) \sim \begin{cases} e^{-\lambda_1 y} & \text{for } y < Y^* \\ y^{-\epsilon+1} f\left(\frac{y}{m}\right) e^{-\lambda_2 y} & \text{for } y \geq Y^* \end{cases}, \quad (1)$$

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

where Y^* is the truncation point ("hard truncation") for which the transition in the regime of the probability distribution is observed from exponential to power-law. We refer to "hard truncation" when the pdf clearly exhibits two regimes (for $y < Y^*$ and $y > Y^*$) in which two diverse pdf can be identified. λ factors are scale factors for the exponential distribution, either above or below the lower/upper cutoff defining the scale-free regime with power-law distribution. m is the upper cutoff after which exponential finite size effects occur. We introduce the function $f(y/m)$ to give more generality to the cutoff (or homogeneity) function [18]. $y^{-\epsilon+1}$ is the scaling function where ϵ is the scaling exponent of the power-law distribution. Note that the probability distribution function $p(y)\ y^{-\epsilon}$ scales with ϵ only. ϵ dictates how the mean and the variance behave. For $\epsilon = 2$ the pdf is the Zipf's law that is found for many socio-ecological systems [18,40].

2.4. Network Inference and Dynamical Species Characterization

2.4.1. Information Balance and Exchange

To infer species interaction networks based on microbial RSA data we ground our approach on the model developed in Servadio and Convertino [73] as well as on previous computational efforts [54,84]. We consider the microbiome as a dynamic network of species interactions where the total free energy changes over time. Codes are available at the GitHub account https://github.com/HokudaiNexusLab/Microbiome. The pdf of each RSA for each group is derived by putting together the RSA time series for all individuals; in this network the RSA is treated as a random variable meaningful of the group and each individual is offering one realization of the same random variable. The RSA matrix was created with compositions in mind and therefore the sum of each sample was constrained [80]. Considering information entropy as the energy's counterpart, the total network entropy can be written as:

$$H(N) \approx \sum_i H(x_i) + \sum_i \sum_{j \neq i} TE_i(x_i, x_j) + \sigma(N) \tag{2}$$

where x_i denote the $i - s$ variables that contribute to the total information of the network N. In our case x is the RSA of species. In this equation, $H(x_i)$ denotes Shannon entropy, and $TE(x_i, x_j)$ denotes Transfer Entropy from the first variable to the second variable [11,54,69,73,88]; in our case both variables are the RSA of two different species. Eq. 2 represents a fundamental principle of information balance independently of the chosen entropy [34] and forms the general basis of sensitivity analyses. Eq. 2 states that the total network entropy can be decomposed into the entropy of each individual node plus the entropy of interactions. The sum of absolute TEs is a proxy of the Mutual Information (MI) of a variable, thus it considers the whole set of variable interdependencies; in Eq. 2 we consider the sign of TE because H(N) should consider the typology of interactions with their sign. $\sigma(N)$ is a noise term that captures the unexplained variability of N related to variables not considered and other discretization factors. Eq. 2 can also be extended in space if spatially explicit calculations are needed as in Servadio and Convertino [73].

The computation of TE is based on the distributions of the two variables of interest (i.e., RSA) conditioned on their histories. Comparing the conditional probability of the variable on its own history with the conditional probability of the variable on both its own history and the history of a predictor variable provides asymmetry in determining predictive abilities of one variable onto another. Thus a directed network can be inferred. Directed TE of two time series variables, denoted as X_i and X_j, was calculated as

$$TE_{X_i \to X_j} = \sum p(X_{j,t}, X_{j,\tau}, X_{i,\tau}) \cdot \log \left(\frac{p(X_{j,t} | X_{j,\tau}, X_{i,\tau})}{p(X_{j,t} | X_{j,\tau})} \right) \tag{3}$$

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

where $X_{i,\tau}$ and $X_{j,\tau}$ denote the respective histories of X_i and X_j at time t as well as considering all past values for the period $t - \tau$. Here we consider the same memory lag for X_i and X_j but in principle historical dependencies can be different when considering other variables and the variable itself. In our microbiome study X_i and X_j are RSA of species i and j.

The definition of TE can assume that the processes analyzed obeys a Markov model, that is suitable for memoryless stochastic process. This implies that future states depend only on the current state and not on events that occurred before it. Thus, in a Markov process it is assumed that $\tau = 1$. Most of the time, this is true, especially for rapidly varying processes (such as for microbial RSA); however, this constraint can be relaxed by choosing temporal lags that are small enough to focus on short-term interdependencies which are not related to long dependencies in the underlying processes. In our case study RSA values of two randomly selected species did not correlate with RSA values for $\tau = 1$; thus, memory processes are relevant and as in Villaverde et al. [84] we selected the τ that maximizes the interdependency between two species assessed by the functional distance (see Eq. 12).

2.4.2. Maximum Entropy Networks

Among all values of TE the question remains on which value is the most informative about the potential causal relationship between two variable. As in Servadio and Convertino [73] we proposed to select TEs that lead to the maximum entropy for the inferred network. This corresponds to maximize the Fisher information matrix [10]. MaxEnt [31] favors probability distribution functions with maximum entropy as the most general distributions that fit the observed data [35]. This theory can be applied to a functional network where edge weights are based on TE. The network with the greatest total entropy can be similarly favored as the most general network structure that fits the observed data. The method considers all possible pairs of variables in both directions for predicting a pattern of interest. The edges that comprise the network with the greatest total TE are then included. Selecting the edges that contribute to the greatest amounts of TE, according to the MaxEnt theory, produces the network that most accurately describes "causal" patterns among the included variables.

A utility function is needed in order to establish the function where MaxEnt is applied. The utility function can be thought as a systemic (network) value function $\sum_{i,j} f_{i,j}(X)\, w_{i,j}$ (potentially multiplied by weight factors $w_{i,j}$) where value functions $f_{i,j}$ are TEs among RSAs. These TEs as in Eq. 3 assess the potential causal interaction between species pairs. Thus, the utility function is the total network entropy $H(N)$ (Eq. 2) that needs to be optimized in order to define necessary and sufficient TEs with the maximum entropy. The optimization can be subjected to feasibility constraints, for instance related to the ability to control certain species or data limitations. In the context of the present goal of creating a microbiome network indicator, the value functions $f_{i,j}$ are defined as:

$$f_{i,j}(X) = \begin{cases} TE_{X_i \to X_j}, & \text{for } \{X_i, X_j\} \in E_{MENet} \\ 0, & \text{for } \{X_i, X_j\} \notin E_{MENet} \end{cases}, \tag{4}$$

where $\{X_i, X_j\}$ represents the directed edge connecting X_i to X_j, and MENet (Maximum Entropy Network) represents the set of directed edges in the network with the maximum total network entropy $H(N)$. The selection of edges to be included in the network is determined by finding the network with the greatest total entropy as in Eq. 2. In the present study, the utility function is defined as the total TE of the network (plus Shannon entropies of each RSA but those turn out to be second or third order factors that can be neglected), and it was maximized by selection of the $f_{i,j}$ functions. To the best of our knowledge, this is one of the the first times that TE was framed in a decision analytical model via a network threshold entropy criteria that defines MENets.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

2.4.3. Optimal Information Networks

To reduce redundancy in creating a MENet, variables that are strongly predicted by other variables (hypothetically establishing a strong causality if prediction accuracy of one decreases quickly when removing the other [76]) can be excluded. This can be done by evaluating the weighted in-degree and out-degree of each node in the network (i.e. TE). Nodes with a greater weighted out-degree than in-degree can be included in the Optimal Information Network (OIN) that one among many MENets with the same average total entropy. These nodes are strongly predicting the variability of other nodes, thus the overall network dynamics. OIN is then the necessary and sufficient MENet for predicting microbiome function. Here, we refer to microbiome function as the information network related to the interdependence between RSA measured by TE; this function is not the "true" biological function but it is likely related to the variability in mutual abundance found in ecological systems [57,70]. This entropy reduction (that does not affect much the total entropy and remove the indirect interactions as in Lizier [55]) can be achieved by introducing functions $g(X_i)$, defined as follows

$$g(X_i) = \begin{cases} 1, & \text{for } \sum_j f_{i,j}(X) > \sum_j f_{j,i}(X) \\ 0, & \text{for } \sum_j f_{i,j}(X) \leq \sum_j f_{j,i}(X) \end{cases}, \tag{5}$$

where $\sum_j f_{i,j}(X) = OTE$ and $\sum_j f_{j,i}(X) = ITE$ where OTE and ITE are the total outgoing and incoming TE for a node. Thus, variable inclusion depends on the comparison of the TE projected by the variable X_i onto the other variables and the TE projected by the other variables onto X_i.

The defined function g was then used to create the total network entropy that can be used to carefully describe the network dynamics:

$$H(N \equiv OIN) = \sum_i H(x_i) \cdot g(x_{i,t}) + \sum_i \sum_{j \neq i} TE_i(x_i, x_j) \cdot g(x_{i,t}) + \sigma(Y) \tag{6}$$

which represents the sum of all necessary variables that were included by the structure of MENet in a multi-criteria value function, and the sufficient variable after the redundancy exclusion to form OIN. In this way, the OIN inference was based on information theoretic and topological criteria to screen (i) the necessary information to maximize network entropy H(MENet) (i.e. information content), and (ii) the smallest non-redundant information to sufficiently predict total network function (of maximum entropy H(OIN)). This OIN is the network with the highest accuracy in predicting macroecological patterns of diversity over time that are dependent on fluctuating RSA.

2.4.4. Assessment of Species Importance & Collectivity

Subsequently the inference of OIN, it is possible to quantify the importance of different species considering their variability in isolation and in cooperation with other species for the dynamics of the microbiome. Species first order importance and interaction for reproducing the network dynamics are then calculated considering new indices based on nodal information transfer rather than on Mutual Information Indices (MII) as in Lüdtke et al. [56]. σ_i is describing species interaction and is calculated as the ratio between the total Outgoing Transfer Entropy (OTE) though as information flow ($OTE(j) = \sum_i TE_{j \to i}$) and the total network entropy, while μ_i is describing the species importance as the ratio between the nodal Entropy thought as information content (using Shannon entropy) and the total network entropy. These Transfer Entropy Indices (TEI) are useful when no systemic variable is needed (see Servadio and Convertino [73]), and analytically the are formulated as:

$$TEI = \begin{cases} \sigma_i = \frac{OTE(j) = \sum_i TE_{j \to i}}{H(OIN)} \\ \mu_i = \frac{H(x_i) \cdot g(x_{i,t})}{H(OIN)} \end{cases}, \tag{7}$$

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

When considering a systemic indicator (see e.g. Servadio and Convertino [73]) MII are better suited to identify variable importance because no directional influence is needed. MII use the mutual information (MI) normalized by the entropy of the output variable considering one independent variable or pairs of variables for predicting a dependent variable Y that is in this case undefined. These MII indices are $s_i = \frac{MI(X_i;Y)}{H(Y)}$ and $s_{ij} = \frac{MI(X_i;X_j|Y)}{H(Y)}$, where X_i is any variable (e.g., RSA) and Y is the predicted variable built using the same process of constructing OINs but selecting variable features rather than keeping entropy of species as independent variables. The use of the transfer entropy can give further information about the directionality of the causality (in a predictive sense of the model), and the time-lag of the causality.

2.5. Macroecological Indicators

To characterize the microbiome as an ecosystem we introduce macroecological indicators that aim to describe ecosystems' collective dynamics of diversity locally, within communities or time points, and globally. In this paper we use such macroecological indicators that are time dependent (because space information is not provided and hardly inferable) and of order zero [43]. For a set of unique distinct species $\mathbf{S} = \{S_1, S_2, ..., S_n\}$ whose RSA $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ changes over time, we define the local species diversity, or α-diversity as:

$$\alpha(t) = \sum_{k=1,t}^{n} p_k(t)^0 \tag{8}$$

where $p_k(t)$ is the probability to find one species at time t. Thus, α is the sum of diverse species at any given time during the observation period (30 days). Considering this definition of α is easily noticeable that the sum of the entropy of all RSA $H_\alpha = \sum_k H(x_k) = -\sum_k p_k(t) \log p_k(t)$ is proportional to the Shannon index that is the local species diversity of order one [43].

Leaving aside the controversy about the definition of interspecies diversity over time, i.e. species turnover, we define β-diversity as the complementary variable of species similarity (here introduced via the Jaccard Similarity Index (JSI) as in Convertino et al. [16] and Convertino [15]):

$$\beta(t) = 1 - JSI(t) = 1 - \frac{S_{t,t+1}}{S_t + S_{t+1} - S_{t,t+1}} \tag{9}$$

where $S_{t,t+1} = \sum_{k=1,t}^{n} (p_k(t)^0 + p_k(t+1)^0)/2$ is the number of species present at both time steps if $p_k(t)^0$ and $p_k(t+1)^0$ are $\neq 0$, otherwise $S_{t,t+1} = 1$. $S_t = \sum_{k=1,t}^{n} p_k(t)^0 = \alpha(t)$ is the number of species present at time t (or $t+1$) (Eq. 8).

Note that this definition of β is proportional to the "true" β that is classically defined as the number of diverse species between two samples (either over space or time). β-diversity can also be defined as a second order index where the entropy related to β is $H_\beta = H_\gamma - H_\alpha$ [43] where $H_\gamma = H(N)$ is the total network entropy (Eq. 2). Considering the variation of diversity over time β-diversity is proportional to the complementary of the mutual information $1 - MI_{X_i,X_j} = 1 - \sum p(X_j, X_i) \cdot \log_2 \left(\frac{p(X_j,X_i)}{p(X_j)\, p(X_i)} \right)$. Yet, $1 - \beta(t)$ is proportional to the sum of the TEs.

The total diversity γ is defined as:

$$\gamma(t) = \sum_{k=1,t=t^*}^{S,T} p_k(t)^0 \tag{10}$$

is an event when a species is introduced in the microbiome; this species can be already present or can be a new distinct species that is established over the total number of speciation events M.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

M is the sum of all species at any given time independently of their diversity calculated from time $t = 1$ to the final time of observation T; equivalently, M is the number of events when new or existing species are introduced. A speciation event is an event when a species is introduced in the microbiome; this species can be already present or a new distinct species. The concept of speciation event is introduced because that determines the number of total species introductions independently of the true temporal dimension. Thus, the speciation event focuses on the dynamics of the process independently of time because it counts events. The total number of speciation events, can be related to the number of unique species S (i.e. all distinct species occurred in the time period) as follows.

$$M = \sum_{k=1}^{S} m_i x_i^0 \tag{11}$$

where S is the number of unique species across the whole observation period, x_i is the RSA of the counted species, and m_i is the number of times that species occurs. Considering the validity of the information balance equation (Eq. 2) that leads to the diversity balance equation $H_\gamma = H_\alpha + H_\beta$, the total diversity can also be calculated as $\gamma = \alpha \cdot \beta$ [43].

2.6. Functional and Structural Network Metrics

The organizational topology of the microbiome is characterized via structural and functional complex network metrics. Functional metrics are based on information theoretic quantities that quantify the interactions among species while structural metrics are based on the geometry of the network and can be derived from the former ones.

The functional distance between species is defined as:

$$d_f(X_i, X_j) = \min_\tau \; e^{-MI(X_i(t\pm\tau), X_j(t))} \tag{12}$$

where the value is considering the minimum for all possible time delays τ. X_i and X_j are the RSA of species i and j and MI is the mutual information evaluated for different values of the temporal scale of species dependency τ. The τ that minimizes the distance d_f is chosen for capturing the maximum interdependence MI_{max}. Such distance as in Villaverde et al. [84] quantifies the magnitude of interactions between species: the higher MI the shorter the distance that signifies high levels of interaction without specifying the directionality. Thus, because of the inability of assessing the direction of causality, MI is a metric useful for identifying the most interacting pairs of the microbiome rather than individual species.

The calculation of the structural distance is based on the functional distance and the concept of the shortest path. The structural distance is then defined as the minimum number of steps from one node (species) to another independently of the magnitude of these steps (e.g. in terms of TE). Thus, analytically the structural distance is defined as:

$$d(X_i, X_j) = \text{argmin} \left[\sum_{i,j} d_f(X_i, X_j)^0 \right] \quad \text{if } A_{ij} = 1 \tag{13}$$

where $A_{ij} = TE_{ij}^0$ is the adjacency matrix that can be formulated in terms of TE. The rationale for considering the shortest paths is related to the exponentially large ensemble of distances as a function of the number of nodes and the fact that biological systems always optimize information transmission [6]; yet, Pareto shortest paths are always chosen [72].

In terms of connectivity the functional degree is defined for the directed network as the sum of the weighted in- and out-degree (i.e. TE) elevated to a power exponent equal to zero. Then, analytically the functional degree is:

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

$$k_f = k_{in} + k_{out} = \sum_{i,j} \left[f_{i,j}(X)^0 + f_{j,i}(X)^0 \right] \tag{14}$$

where $\sum f_{i,j}(X) = TE_{ij}$ is the transfer entropy as defined in Eq. 3.

The structural degree is defined by thinking the network as an undirected network, thus

$$k = \sum_i a_{i,j} \tag{15}$$

where $a_{i,j} = 1 = TE^0_{i,j}$ if i and j are connected. Classically, the structural degree considers the number of connections independently of the bidirectional pathways implied by TE. Thus, functional degree is always greater or equal to structural degree.

3. Results

The simplest analysis of the microbiome starts by looking the temporal trajectories of RSA. By a simple cursory analysis it is evident that the average RSA of the healthy microbiome is lower than the average RSA of the unhealthy microbiome independently of the species; however, the maximum RSA is higher for the healthy microbiome and that species is one of the the most beneficial for health. A recent data set with absolute abundances suggests that healthy gut microbiota have higher total abundances than diseased ones [83] but no studies exist about the universality of this abundance-health relationship. By looking into species diversity (Figure 1 A) it is observed that the average number of species at any time point (α) is lower for the healthy microbiome than the unhealthy one. This may seem in contrast with previous findings that report higher diversity for healthy microbiome or in general for healthy ecosystems [46,64,89]. A controversy on the subject is already found in literature [64] and just maximizing total diversity, without considering how that diversity grows and is organized, is not intuitively a necessary and sufficient ingredient to achieve a stable healthy state [42] More importantly, the RSA-rank pattern (Fig. 1B) shows only one dynamical regime, corresponding to the common Zipf-Mandelbrot model for RSA [63], for the healthy microbiome (exponential like) vs. two regimes for the transitory and the unhealthy microbiomes. Figure 1C shows that the decay in richness over RSA is higher for the unhealthy microbiome; this result underlines the fact that higher diversity does not imply stability because of the suboptimal, yet unsustainable distribution of species in the unhealthy microbiome. Stability is related to network topology [20] that is also affecting diversity [39,85] and the systemic fluctuations of the microbiome as shown by the Taylor's law [60] that highlights how variance in RSA abundance changes with the mean. "Optimal" is in this case referring to the healthy state as a reference state because it has the smallest fluctuations for the highest achievable total diversity growth rate γ' (this is the Pareto solution) and the associated network topology is more resilient to random node removal (Fig. S3). The Pareto solution has the largest diversity growth rate and is not by chance accompanied by a Pareto-like species interaction network where interactions are inferred by TE (Fig. 5 middle plot). Figures 1 B and C show the RSA-rank plot and the Preston's plot [38] of species diversity dependent on RSA. The RSA-rank shows two dynamical regimes for the unhealthy and transitory groups: a result that likely confirms the bimodality in local species richness α. By plotting the Preston's plot in log-log a scaling relationship is found showing a faster decay in species richness for the unhealthy group.

Considering the RSA of species in time, from the most to the least relatively abundant, a transition in the pdf of RSA is observed from a pseudo-normal distribution (corresponding to a homogenous spatial distribution) to a Dirac-like distribution (corresponding to a singular point distribution) considering the maximum and minimum RSA. Considering the RSA of all species together (Figure S2) the transition is less dramatic; from an exponential to a log-normal like distribution. Intermediate RSA species, independently of species belonging to the healthy, unhealthy or transitory group, show a scale-free like distribution underlying the fact that these species are fundamentally important in

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

the function of the complex microbiome as highlighted in Lahti et al. [46]. Rare species seem also to display a truncated scale-free behavior (limited by their maximum RSA as finite size factor rather than limited by spatial biological constraints), that also underline their importance for the microbiome organization. These pdfs are a signature of species interaction networks for different RSA groups: pseudo-random, scale-free, and small-world topology for the highest, intermediate and lowest RSA class, respectively. Further results will discuss the connection between RSA and species information flow.

The inferred microbial networks corresponding to the three microbiome groups are shown in Figure 2. Maximum entropy networks are evidencing the different topology in microbiome organization for healthy, unhealthy and transitory group. In the structure of these networks, the size of each node is proportional to the Shannon entropy of the species and the color is proportional to the structural degree. in Fig. S3 we show the networks whose nodal color is proportional to the total outgoing TE (OTE) that is likely more representative of node activity in a collective network sense. The higher the value of structural degree (or OTE in Fig. S3), the warmer the color. The width of each edge is proportional to the TE between pairs and the direction is the corresponding to the directional influence. The MaxEnt networks are OINs; yet, the networks for which the total network entropy is maximized (MENets) and where redundant nodes are removed (see section 2.4.3). The transition in network topology, from random to small-world (tending toward a scale-free network) for the unhealthy and healthy groups, is manifested also by the shift in total entropy pattern (left plot in Fig. 2). The latter is asymmetrical and symmetrical for the random/unhealthy and SW/healthy microbiome. This type of network transitions have been observed for large ecosystem such as in Winemiller [87]. The network entropy plots show that network entropy over information flow is roughly symmetrical for healthy individuals, expressing that the interconnectedness in healthy communities is more dynamically balanced. Figure S3 shows the microbiome networks for high value of the threshold (TE_{ij}) that is establishing the information exchange between species above which links become relevant. Considering the total network entropy and its decomposition, it is observed that the most important nodes in terms of OTE (Eq. 6, and Fig. S6), that is the information flow necessary to predict all other nodes' dynamics, are the dominant factors in making up the total network information (Fig. S5). In other words, the entropy of each single node in isolation $H(x_i)$ is a second or third order factor in determining the total network entropy. Figure S7 shows that most of species interactions (TEs) are positive for the unhealthy microbiome underlying the evidence that mutualistic positive feedback leads to instability; therefore, higher diversity α does not guarantee stability if interactions are predominantly in one direction. The healthy microbiome instead has balanced positive and negative interactions.

Figure 3 shows macroecological indicators of diversity of the microbiome for healthy, unhealthy and transitory individuals. We show that species diversity α, and total species diversity γ are the highest in the unhealthy group (for which average RSA is also the highest) but species similarity 1-β and the the diversity growth rate α' over time are the highest for the healthy group. This is a critical result that shapes microbiome organization around healthy or dysbiotic state. The highest fluctuations in RSA and macroecological indicators (in particular α and γ) are observed for the transitory and unhealthy groups. These results underline the potential conclusion that too high levels of diversity are possibly unsustainable, yet leading to unhealthy unstable states related to the abnormally excessive multiplication of species in the guy ecosystem. It is interesting to note the behavior of the pdf of α that informs about the potential states of the microbiome in each group. The pdf is platykurtic multimodal for the unhealthy microbiome which suggests the presence of multiple unstable states, and it is leptokurtic monomodal for the healthy microbiome which implies one stable state. The transitory microbiome shows an almost symmetrical pdf underlying the fact it exists in between the healthy and unhealthy microbiome. These results underlines the resilience of the microbiome as a whole dictated by the ability to change as a function of external stressors as well as the higher stability of the optimal

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

healthy state. However, the latter seems easy to perturb consider the lower entropy (and probability) to be defined in one state.

Species collective interaction and singular importance is shown in Figure 4 by plotting the information theoretic global sensitivity indices σ_i and μ_i (see Methods). The top 10 interacting species are also the least relatively abundant for the healthy microbiome and the most detrimental; however these species are controlled by other species and the microbiome is organized into a healthy state. The definition of detrimental and beneficial is based on previously published papers. We incorporated this classification in Table S1. For instance, *Lactobacillacea e.* and *AcidobacteriaGp18* are beneficial, while *Neisseriaceae* and *Campylobacter aceae* are detrimental. Of course this is just a rough categorical classification because as we emphasize in this work, for a bacteria being detrimental or not is a function of relative abundance and network topology rather than just being present or not in the microbiome or other independent properties without considering the bacteria collectivity. Microbiome functional network topology defines how all bacteria behave synergistically and that synergy brings to an healthy or unhealthy state. This result sheds some light into a vision where a diminishing role of network hubs (considering total information flow) is reported as found by other studies [67]. The least relatively abundant species for the unhealthy microbiome are the most interactive and the least detrimental. On the contrary, the most relatively abundant species (Fig. S4) for the unhealthy microbiome are the least interactive and the most detrimental. These analyses considering the activity of species show the importance of weak ties (interactions) for the healthy and unhealthy groups. This is in accordance to general dynamical principles such as the Granovetter principle about the strength of weak ties for the systemic dynamics of a system [30]. For the healthy microbiome, the highest RSA species interact the least and these species are the most beneficial. These species-specific analyses, when verified, are useful for detecting species that are more beneficial or detrimental and this knowledge can lead to design probiotic treatment, microbiome transplants [27], and large scale ecosystem microbiome controls [78] for instance.

Figure S7 shows that from the top to the least 10 TE species there is a shift in the pdf of RSA from a bimodal to a monomodal distribution for the healthy microbiome. For the transitory and unhealthy microbiome instead there is a shift from a leptokurtic (Dirac-like) to a platykurtic pdf (uniform-like). The top 10 TE species are the most detrimental bacteria ("antibiotic") but their RSA is small for the healthy microbiome; this means that these bacteria are controlled (in terms of RSA variability) by all other beneficial bacteria. The top 10 TE species are mostly characterized by positive interactions (positive TEs) while the least ten 10 TE species are characterized by negative feedbacks. For characterizing species collectivity or single species dynamics, as well as for predictability, OTE is however better suited than TE that is a link measure. The pdfs of OTE in Figure S6 show more clearly the changes in species dynamics for each health state and overall species activity manifested by the magnitude of OTE. The top 10 OTE species are always characterized by positive feedbacks vs. the least 10 OTE species with negative feedbacks (top and bottom plots of Fig. S6).

The non-linear duality between microbiome structure and function is shown in Figure 5 where structure is considered via the network degree (Fig. S8 and S9) and function is about the nodal information flow OTE. The epdfs show how microbiome function is much more suited to show functional network topology versus microbiome structure. Function is a much more important property than structure which is just based on geometrical analysis of cooccurrence species network (e.g. as in Baldassano and Bassett [5]). This scale-free function may be related to the scale-free behavior of the intermediate RSA species as shown in Figure S2. As shown by Figure 2 visually, the healthy microbiome function is tending toward a scale-free topological organization. Statistics of the functional scale-free network based on TE are in Figure 5. This mild scale-free organization (see e.g. [22] where the authors highlight the difficulty in defining radically the classification for these networks) does not correspond to a scale-free distribution of α-diversity (Fig. 5 bottom plot) that instead is exponential. Additionally some functional network features beyond the inferred RSA-based interdependence (TE and OTE) show a bimodal or Poisson distribution (Fig. S8) characterizing more

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

small world networks rather than scale-free ones. However, we point out how these features are more structural than functional (see Eq. 12 and 14) since they are characterizing species interactions directly. The non-linearity between structure, function and microbiome service (i.e. diversity in this paper) is highlighted when plotting α dependent on functional network degree and distance (Fig. S9). This "non-pure" scale-free organization may confer the ability of the microbiome to adapt to different externally-driven changes and to adapt vs. a more stable scale-free topology. Overall, we suggest to focus on TE as the best indicator of microbiome function, vs. any other indicator, since that is related to species interdependence. As highlighted in recent studies (see Rivett and Bell [70]) abundance determines the functional role of bacterial phylotypes in complex communities; rare and common bacteria are implicated in fundamentally different types of ecosystem functioning [70] Such knowledge could be used, for example, to understand how bacteria modulate biogeochemical cycles, and to engineer bacterial communities to optimize desirable functional processes. Microbiome service is here identified by any microbiome diversity indicator in analogy to how services are also expressed for large scale ecosystems. Certainly it is true that α, β and γ-diversity cannot be "equated" to a large scale ecosystem services (i.e. the benefits that people derive from nature and natural capital), but any diversity measure is a valuable indicator of biological function at any scale of biological organization (see for instance Isbell et al. [39] and Mori et al. [65]) much more than structural indicators as shown in this paper. Therefore, there is a desired ecosystem service-function nexus that is desirable and related to healthy states (which is the benefit individuals get to have the "right" organization of diversity macroecological indicators). α increases for high values of functional connections (Eq. 14) but does not have a clear trend when considering the functional distance (Eq. 12); however, $\alpha(d_f)$ is lower for the unhealthy than the healthy microbiome for the same range of functional distances which highlight the more random distribution of diversity in any dysbiotic state. We observe 72, 378, and 9647 unique values of functional distance for the healthy, transitory and unhealthy group. These values were normalized and the distribution of α over the normalized distance shows a random arrangement for the unhealthy group with respect to the healthy one.

The most interesting results we find is when we combine microbiome service and function indicators, for instance considering total macroecological diversity γ and OTE. Figure 6 shows the relationship between γ and the temporal sampling scale (i.e. the number of speciation events) in analogy to the species-area relationship widely used in macroecology [38]. The plot shows a scaling relationship for two orders of magnitude whose exponent is higher for the healthy than unhealthy group underlying the optimal growth of diversity for the healthy microbiome. Considering this optimal diversity growth relationship it is meaningful how the transitory microbiome has the largest value of γ' leading to a change in diversity from the healthy species "poor" to the unhealthy species "rich" microbiome. Perhaps healthy individuals have larger gradients of speciation events and higher growth rate for γ-diversity because they produce more species (diverse of not) to guarantee necessary/basic biological function and functions related to extreme fluctuations. Not all species need to be present all the time and that is likely the motivation for which the average γ' is lower than unhealthy people as well the average γ is lower for healthy ones. γ' seems to reflect the general dynamical systems' pattern indicated by the Heap's law [79] that regulates the rate of diversity produced by a system. This is associated to the Taylor's law regulating mean and fluctuations and Zipf's law (in our case of RSA which influence macroecological indicators). These results are in synchrony with the power-law decay of species similarity $1-\beta$ over time (Fig 6 bottom left plot). When considering OTE of species as a function of their RSA we find a surprising scaling law for four orders of magnitude; this law with an average exponent close to $1/4$ (very common in biology, for instance the mass-specific Kleiber's law [23]) implies a decay in species interaction for highly relatively abundant species. When comparing γ over OTE a non-linear growth is detected where a common increase in total diversity occurs until a critical species interaction value above which γ slows down or remain stationary as for the healthy group.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

4. Discussion

We employ an information theoretic model for the inference of microbial species interaction networks based on RSA interdependence. The model is used to infer microbial networks associated to different health states and is suitable for predicting selected biodiversity patterns characterizing the space-time organization of bacteria α, β, and γ-diversity. Thus, the primary purpose of the model is not to infer causal (or "true") species-species interactions among bacteria. The computational inference of "real" interactions is always very hard — provided that there is a complete knowledge of the reality on which results can be validated — and any inferred interaction is always dependent on the analytics and data used. For instance, RSA profile may not necessarily contain the information about all species-species interactions aimed to be assessed but still the question remains about what is truly an interactions since any physical or functional interaction may not necessarily reflect any change in RSA, or other biomarker. Additionally, any change in RSA or other biomarkers may be related to other external factors. What is certainly true though, is that if the inference model detects strikingly different patterns for different population groups, then likely those patterns tell something meaningful about different dynamics [25,33]. In this perspective the entropy-based model is focused on the predictability of patterns vs. causal investigation of mechanisms. The proposed model can be applied to both abundance and RSA, or other biomarkers, without any special modification. Theoretically, the pdf of abundance and relative abundance is the same leaving aside numerical artifacts; independently of this, RSA seems better suited for these ecological analysis because it informs about changes of species abundance with respect to the whole community. Abundance and/or RSA seems also the most likely to detect species functional roles and interactions as highlighted by recent studies [57,70]. Constructing a network for each health group is the purpose of studies like ours that try to identify common group dynamics in populations independently of individual variability (see e.g. Bashan et al. [8]). The identified network topologies have a correspondence with the dynamics of RSA, that is a critical dynamics for the scale-free information network associated to the healthy state, and exponential dynamics for the random network associated to the unhealthy state. The total network entropy is the lowest for the healthy microbiome for any threshold of the information flow TE (Fig. 2). This implies higher free energy available to the healthy microbiome and lower information needed to function where information entropy in the physical space can be though as the average interspecies communication/interdependence. The lower entropy in species collective interaction has certainly implications for data collection, potentially implying less data needed for characterizing healthy microbiomes. This is because one single globally stable state is identified for the healthy microbiome (in the entropy pattern of Fig. 2) vs multiple stable states for the unhealthy microbiome (one globally and locally for high and low value of entropy, respectively). The existence of multiple dysbiotic states seems to confirm the previously observed "Anna Karenina effect" [90] where "all healthy microbiome look alike, instead each unhealthy microbiome is diverse in its own way?". More theoretically speaking, the highest entropy is a sign of criticality that is the state toward which any ecosystem tend to [37] where there is a balance of self-organization and environmental influence [25].

The inferred patterns in this paper are representative of confirmed health states where individuals are confirmed representative samples (see Durbán et al. [24] that published the original dataset) for IBS and non-IBS people as confirmed by Martí et al. [60]. Patterns and methods are proposed to highlight what is relevant to look at when describing state transitions and characterization of health states. The number of individuals sampled in a population matters as a function of the pattern changes to expect or found. Reliability is not only dependent on the sample size but also on the consistency and differences within and among samples. In this particular study we find striking differences between potential health states and many times concordant with the reported literature. Further research is required to test the biological universality [8] or local specificity of these patters across a much large population sample. Analyses have been made considering varying data lengths for individuals and this does not change any pattern considered significantly. This means that the dynamics represented in the time series is well contained at least in the smallest data available. The smallest reliable sample is for ten

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

data points that seems in this case the minimum data length to have in order to have representative probability distributions.

Considering the issue of compositionality [13,80,86] the theory suggests that a small numbers of species should increase compositional effects. In our case the number of species is 47 at minimum and that should limit the effect of compositionality. Microbiome sequence data sets are typically high dimensional, with the number of species much greater than the number of samples. The consideration of pdfs limits the issue of compositionality, as well as the focus on group vs. individual statistics limits the issue of data sparsity (considering both rare species and the length of time series). Of course, this macroecological purview does not imply any strict causality inference but rather aims to set up the basis for the predictability of microbiome group features. Despite sophisticated approaches to statistical transformation, the analysis of compositional data may remain a partially intractable problem because RSA is the information that is available. Given these findings, promising work has been done on addressing compositional data as a significant challenge to co-occurrence network inference, but the problem is still not solved [80]. However, TE seems not affected by compositional data precisely because it uses pdfs in network inference and the pdf of RSA and raw abundance is the same.

The entropy/free energy patterns (or "entropy-flow patterns") in Fig. 2 do not show any strong scale invariance as for instance in Servadio and Convertino [73], likely because no pure scale-free networks are observed. In this paper we focus on the total entropy as a utility function versus the value function defined in Servadio and Convertino [73] (based on a systemic indicator) where raw values of the network variables were considered rather than TEs among them. The focus on network variable interdependence (that is between species in this context) rather than nodal values (i.e. RSA for the microbiome) leads to a higher variability in network entropy patterns. However, we believe that the focus should be on network function in order to better characterize networks; this is substantiated by the higher importance of species interactions (OTE) versus species independent dynamics (represented by nodal entropy) as shown in Fig. S5. This figure shows that OTE makes up almost all Network Entropy (H_N) (see Eq. 6) so Nodal Entropy has little importance. Entropy-flow pattern are then useful for detecting scale-invariance in the functional topology of the network. Additionally the entropy-flow patterns can reveal healthy vs. unhealthy states by considering the symmetry of the entropy distribution; if symmetrical positive and negative species interactions (TEs) sum up to zero leading to a healthy neutral state. The asymmetry of unhealthy microbiome can certainly relate to non-neutral states created by strong stressors as highlighted theoretically in Borile et al. [10] and may not allow host individuals to keep the microbiome "on a leash" [68]. The broken symmetry can be indeed manifesting an unhealthy state. The neutral state also coincides with the critical state because of the tendency of the network toward a scale-free organization manifested by the epdf of OTE (Fig. 5), higher functional distances and smaller functional degrees (Fig. S8).

In order to assess the robustness of microbiome networks we consider the network topology for high thresholds values of the interspecies TE. In other words, we consider as meaningful TEs, only those above a certain threshold. According to the 80-20 Pareto principle (that states that 20% of subcomponents make up at least 80% if a system's dynamics [66]) we consider only the 20% of the highest TEs for the inferred networks. These Pareto high threshold networks show that the healthy group maintains the topology while changing TE; this is because healthy networks are more scale-free than unhealthy one (see Fig. 5 middle plot for the epdf of OTE), yet scale-invariance is preserved when changing the threshold defining the scale at which the network is observed. This scale analysis is equivalent to make experiments when random nodes are removed simulating a random attack on networks [1]; thus, we can also claim the higher resilience of the healthy network for the microbiome. This result is however expected considering the known optimality of scale-free networks [6]. The scale-free configuration enhances stability as confirmed by the calculation of the dominant eigenvalue

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

for both the adjacency and TE matrix; the dominant eigenvalue is the smallest for the healthy group which is a signature of stability of the network [20].

In our microbiome data we consider the complementary of β-diversity over time via the Jaccard Similarity Index (JSI) and we show that JSI is higher for the healthy than the unhealthy microbiome over time. This means that the local species richness, α, tends to be more equal to previous values over time; yet, this underlines the stability of α (species organization) in the healthy state. For the unhealthy microbiome the similarity over time is lower (i.e. higher species turnover, or higher β-diversity) such as for the corals in Zaneveld et al. [89] that are evaluated over time as a function of external stressors. In other types of ecosystems, e.g. in coral ecosystems under stress it was found that the true β-diversity increases over time Zaneveld et al. [89]. In macroecology, leaving aside the debates about the many definition of species turnover, in an entropic context β-diversity is the ratio between regional (γ) and local species diversity (α) [43]. This definition is in line with the general information balance equation (Eq. 2) and the more specific diversity balance equation $H_\gamma = H_\alpha + H_\beta$ as in Jost [43]. An increase in β is typically associated with a decrease in α as much as we observe for the healthy microbiome, and this is also associated to fluctuations of α that are smaller than for the unhealthy microbiome. The "proportional species turnover" (i.e. where $\beta_p = 1 - \alpha/\gamma$) that quantifies what proportion of species diversity is not contained in an average representative sample, is also higher. In ecology these quantities are typically evaluated over space and in healthy conditions 1-β has a relatively fast decay but never goes to zero; this means that heterogeneity exists but even communities far apart have species in common. Considering space in unhealthy conditions, typically the "true" β-diversity is smaller than in healthy conditions because much more homogeneity is achieved.

The higher variability of β-diversity in healthy individuals highlight the "Anna Karenina phenomenon" for human microbiomes. The principles underlying the phenomenon states that dysbiotic individuals vary more in microbial community composition than healthy individuals paralleling Leo Tolstoy's dictum that all happy families look alike (each unhappy family is unhappy in its own way). The stability-unimodal pattern of diversity is concordant with current theories looking into β-diversity vs. solely α-diversity for the stability of ecosystems [64]. Convertino et al. [19] had previously found that ecosystem hotspots are those that maximize the Value of Information (of biodiversity) which coincides with those that minimize β-diversity variability over time. The multiplicity of "unhappy/unhealty" states is reflected by the network topology that is random for the unhealthy group, which allows many more potential unhealthy microbiome combinations. We argue the point of previous studies, that Anna Karenina effects are a common and important response of animal microbiomes to stressors that reduce the ability of the host or its microbiome to regulate community composition.

Similarly to other ecosystems we show that scale-invariance (that is occurring for the healthy microbiome) does not arise from an underlying criticality (where fluctuations becomes bigger and bigger causing the system to tip abruptly) nor self-organization at the edge of a phase transition. Instead, it emerges from the fact that perturbations to the system exhibit a neutral drift (also relate to small extrinsic environmental changes) with respect to the endogenous spontaneous dynamics. This *neutral* dynamics, similar to the one in genetics and ecology, shows fluctuations of all sizes simultaneously that likely determine power-law distributed species diversity (as well as power-law information exchange among species). The tipping point that is observed, e.g. between healthy and unhealthy microbiome, is a second-order critical transition where exogenous fluctuations are too large to be assimilated by the system and the microbiome tips from healthy to unhealthy. This transition is evident in the shape of the pdf of microbiome function and diversity but not in the shape of microbiome structure (unless a rescaling in size, for instance for the microbial network degree (see Fig. S5 top plot)).

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

The introduction of new pathogens driven by the environment can lead to the alteration of the whole ecosystem microbiome [60]. In our case study, despite the non explicit consideration of the disturbance agent, we see a transition in IBS individuals from healthy to unhealthy states. This disturbance agent has however been considered in Durbán et al. [24] and Martí et al. [60] that worked on the original dataset. The microbiome is like the gut of any ecosystem: no other species at all scales of biological organization can survive optimally if the microbiome is altered. The microbiome is the linkage between the fundamental genetic organization of life and the stochastic environmental dynamics; in the context of a person's growth it is possible to refer to those two processes as nature and nurture. The proposed information theoretic global sensitivity and uncertainty analyses (Figure 4) allows one to map the dynamics of species considering their interactions and absolute influence, and to see how these quantities vary considering their intrinsic biological variability and environmentally driven variability. One must keep in mind that these interactions are based on mutual RSA interdependence assessed by TE so that might not represent the whole "true" interactions among species; however, recent evidence points to this conclusion [57,70]. In the healthy state many more species are influencing the collective dynamics with a more organized distribution of interactions ("hierarchical" organization), while for the transitory and unhealthy state all species are somehow behaving equally and likely driven by external environmental stimuli ("random" organization). This organization is also reflected by network properties (Fig. S8 and S9) that can be altered for the same set of species/diversity. Previous papers found that cooperation promotes ecosystem biodiversity, that in turn increases its stability without any fine tuning of the species interaction strengths or of the self-interactions (i.e. neutrality) [77,81]. Even small values of TEs manifesting mutualistic interactions among species can stabilize the dynamics, yet stability increases with the ecosystem complexity where the latter is related to the scale-free like organization of bacteria. It is interesting to note that this scale-free cooperation of species leads to Taylor's laws [44,58] between mean and variance of RSA where Taylor's exponent is different for healthy and unhealthy groups [60]. Yet, this reemphasizes the connection between time dynamics, network organization, and ecological patterns of diversity and RSA [29,32,77]. In particular it has been shown that higher-order interactions (e.g. captured by σ_i in our model) have a stabilizing role [32]. These higher-order interactions are all those beyond the simple pairwise interactions whose sum indeed cannot explain the whole composition of ecosystems [50]. We show that these higher-order interactions cannot be prevalent because some species must have an independent dynamics (captured by μ_i) otherwise instability and tendency toward disorganized unhealthy state is very likely (Fig. 4).

Universality in human microbiota dynamics [8], whether present, can be ideally manipulated in a similar or even identical fashion in multiple individuals. Following the discovery of universality and the demonstration of beneficiary effects of specific interventions, microbiome engineering efforts can be applied to large number of people. In this way, microbiome engineering will be highly cost-effective as a public-health based approach. This in sharp contrast to the excessive cost of "precision-medicine" that tries to correct individual microbiome dynamics by thinking as it is a purely individual-based function.

Current frontier topics are also related to the understanding of how the microbiome and functional brain networks "communicate" [2]. The hypothalamic-pituitary-adrenal axis (HPA axis) is a primary mechanism by which the brain can communicate with the gut to help control digestion through the action of hormones [2]. It seems that the nervous system, through its ability to affect gut transit time and mucus secretion, can help dictate which microbes inhabit the gut; this in turns affect emotional response and long term well being beyond short-term health.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

5. Conclusions

An information theoretic model for the inference of microbiome networks and its biodiversity organization over time is proposed. The model consists in the assessment of transfer entropy based species interactions after entropy reduction calculations that remove the spurious direct interactions related to indirect interactions between species as in Lizier [55]. Maximum entropy networks are then extracted considering the highest information content without model overfit; overfitting is avoided by removing the redundant variables for the simplest MENet, that is an Optimal Information Network. The macroecological validation of the model is performed considering the ability to simultaneously predict the pdf of α-diversity, γ-diversity growth, species similarity $(1-\beta)$ decay, and the RSA-rank profile. This validation allows to predict other biodiversity patterns such as the Preston's plot of average species richness dependent on species RSA. Considering the application of the model to healthy and IBS symptomatic individuals the following points are worth mentioning, without possibly lack of generality.

- Directed species interdependencies and phase transitions of the microbiome over time are detected. The healthy microbiome is characterized by balanced positive and negative species interactions vs. the unhealthy microbiome where most species interactions are positive. The balanced interactions are evidenced by the symmetrical pattern of the network total entropy as a function of the pairwise information flow (TE) vs. the positively biased asymmetrical pattern of the dysbiotic microbiome. The healthy symmetrical network entropy pattern underlines the neutral "sum to zero" dynamics of species interactions (based on RSA); the same neutrality is found for biodiversity of large scale ecosystems at stationarity that are driven predominantly by intrinsic ecological stochasticity (ecological drift). On the contrary, unhealthy microbiome entropic patterns are affected by environmental disturbances; the positive bias in information flow (that may related to infections and antibiotics as shown in the original data [24]) causes an overgrowth in RSA of many opportunistic species as well as the generation of new detrimental species. The categorization of beneficial and detrimental species is based on published literature; however, we emphasize how important is to consider bacteria collectivity topology vs. individual bacteria behavior when defining health and disease;
- The healthy state is characterized by the highest total species diversity growth rate γ' (leaving aside the transitory microbiome) and the lowest loss of species similarity over time, i.e. species turnover $((1-\beta)')$. A relationship similar to the species-area relationship for large scale ecosystems [38] is found between γ-diversity and the number of species generations with an exponent equal to 0.20 on average. The fact that the healthy microbiome has the lowest average total diversity (γ) is in contrast to what is observed in large scale ecosystems at stationarity where the highest total diversity correspond to the stable and supposedly healthy state [85]. However, we speculate that an optimal diversity growth is oriented toward maximizing growth rate rather than total diversity (as according to many Pareto portfolio theories). The latter can lead to over-redundancy of microbial interactions and instability as observed for the dysbiotic microbiome. Hence, we tend to challenge the diversity-health-stability hypothesis when for diversity the total systemic diversity γ is solely considered;
- We observe a phase transition of the second order from the healthy to the unhealthy state and vice versa. The transition from healthy to unhealthy is characterized by typical signs of transitions observed in many complex systems [71], i.e. an increase and a decrease in mean and variance of species diversity while approaching the transition ("critical slowing down"). In the unhealthy state the variance of α is higher than in the healthy state and concentrated around two values which underline the likely chaotic-like dynamics of the microbiome. In terms of microbiome functional network topology a transition between the scale-free to the random network topology is observed. The critical state, defined by a scale-free like organization of microbial species interactions, coincides with the neutral state (i.e. for the symmetrical network entropy pattern) emphasizing how criticality does not necessarily occur at critical phase transitions, particularly

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

for second order transitions as in this case. Rather, criticality can coincides with neutrality in open energy dissipative systems as observed in other complex systems [61]. Neutrality implies higher topological complexity but higher dynamical stability considering the scale-free and small-world functional and structural organization of the microbial network.

- A probabilistic linkage is found between microbiome function and services, defined as species interactions and macroecological indicators, respectively. We do not find any correspondence between microbiome structure and function, which emphasizes the non-linearity between the two and the importance to assess function rather than structure in biological networks. We propose the total Outgoing Transfer Entropy (OTE) as the measure to identify the most influential nodes; these nodes are able to predict the behavior of all other connected nodes, yet of the whole microbiome. OTE is largely determining the total entropy of the network compared to the sum of nodal entropies whose contribution is negligible. This emphasizes even more the role of collective behavior vs. individual nodes considered in isolation. The highest OTE nodes have the lowest RSA, and these are the most beneficial and the most detrimental bacteria for the dysbiotic and healthy microbiome. A scaling law is found between OTE and RSA with an exponent close to $1/4$ that is similar to the mass-specific Kleiber's law [23] where the species specific metabolic rate is the OTE and the mass is the RSA. A power-law distribution for the microbiome function (i.e. the sum of nodal OTE) is found for the healthy state (with an exponent ~ 2 that implies finite mean but infinite variance) despite no information (or resolution) invariance is detected in the network entropy pattern (see Servadio and Convertino [73]). The lack of scale invariance in the entropy/free-energy phase space may imply the metastability of the microbiome which can confer its resilience in terms of ability to move quickly from one state to another.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

Glossary

IBS: Irritable Bowel Syndrome
RSA: Relative Species Abundance
α: local time point species diversity
β: intertemporal species diversity
γ: total species diversity
JSI: Jaccard Similarity Index
TE: Transfer Entropy
TEI: Transfer Entropy Indices
OTE: Outgoing Transfer Entropy
MaxEnt: Maximum Entropy
MENets: Maximum Entropy Networks
OIN: Optimal Information Network

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

Author Contributions: Conceptualization, M.C.; Methodology, M.C.; Software, J.L.; Validation, M.C. and J.L.; Formal Analysis, J.L.; Investigation, M.C. and J.L.; Resources, M.C.; Data Curation, J.L.; Writing and Original Draft Preparation, M.C. and J.L.; Writing, Review and Editing, M.C. and J.L.; Visualization, M.C. and J.L.; Supervision, M.C.; Project Administration, M.C.; Funding Acquisition, M.C.; Codes are available at the GitHub account https://github.com/HokudaiNexusLab/Microbiome

Funding: M.C. and J.L. gratefully acknowledge the funding provided by the GI-CORE Global Station for Big-Data and Cybersecurity, as well as funding from the Graduate School of Information Science and Technology at Hokkaido University, Sapporo, JP. M.C. also acknowledges the NIH funded Big Data to Knowledge (BD2K) 2017 Innovation Lab "Quantitative Approaches to Biomedical Data Science Challenges in our Understanding of the Microbiome" managed by the BD2K Training Coordinating Center (TCC).

Acknowledgments: M.C. and J.L. gratefully acknowledge Jose Manuel Marti and Andres Moya at the Institute for Integrative Systems Biology (I2SysBio), Valencia, Spain for sharing the polished version of data and providing some insights.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378, 2000.
2. Andrew P. Allen, Timothy G. Dinan, Gerard Clarke, and John F. Cryan. A psychology of the human brain?gut?microbiome axis. *Social and Personality Psychology Compass*, 11(4):e12309, 2017. doi: 10.1111/spc3.12309. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/spc3.12309.
3. M et al. Arumugam. Enterotypes of the human gut microbiome. *Nature*, 473:174–180, 2011. doi: 10.1038/nature09944.
4. Sandro Azaele, Samir Suweis, Jacopo Grilli, Igor Volkov, Jayanth R. Banavar, and Amos Maritan. Statistical mechanics of ecological systems: Neutral theory and beyond. *Rev. Mod. Phys.*, 88:035003, 2016. doi: 10.1103/RevModPhys.88.035003. URL https://link.aps.org/doi/10.1103/RevModPhys.88.035003.
5. Steven N Baldassano and Danielle S Bassett. Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Scientific reports*, 6:26087, 2016.
6. Jayanth R Banavar, Amos Maritan, and Andrea Rinaldo. Size and form in efficient transportation networks. *Nature*, 399(6732):130, 1999.
7. Jayanth R. Banavar, Francesca Colaiori, Alessandro Flammini, Amos Maritan, and Andrea Rinaldo. Scaling, optimality, and landscape evolution. *Journal of Statistical Physics*, 104:1–48, 2001. ISSN 0022-4715. doi: 10.1023/A:1010397325029. URL http://dx.doi.org/10.1023/A%3A1010397325029.
8. Amir Bashan, Travis E Gibson, Jonathan Friedman, Vincent J Carey, Scott T Weiss, Elizabeth L Hohmann, and Yang-Yu Liu. Universality of human microbial dynamics. *Nature*, 534(7606):259, 2016.
9. Franziska Bauchinger. Self-organized criticality in the gut microbiome. *Interdisziplinares Masterstudium Environmental Sciences*, 2015. URL http://othes.univie.ac.at/40027/.
10. C. Borile, M. A. Muñoz, S. Azaele, Jayanth R. Banavar, and A. Maritan. Spontaneously broken neutral symmetry in an ecological system. *Phys. Rev. Lett.*, 109:038102, Jul 2012. doi: 10.1103/PhysRevLett.109.038102. URL https://link.aps.org/doi/10.1103/PhysRevLett.109.038102.
11. Terry Bossomaier, Lionel Barnett, Michael Harré, and Joseph T Lizier. An introduction to transfer entropy. *Cham, Germany: Springer International Publishing. Crossref*, 2016.
12. Robert Caesar, Valentina Tremaroli, Petia Kovatcheva-Datchary, Patrice D. Cani, and Fredrik Backhed. Crosstalk between gut microbiota and dietary lipids aggravates wat inflammation through tlr signaling. *Cell Metabolism*, 22(4):658 – 668, 2015. ISSN 1550-4131. doi: https://doi.org/10.1016/j.cmet.2015.07.026.
13. Anne Chao, Robin L Chazdon, Robert K Colwell, and Tsung-Jen Shen. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology letters*, 8(2):148–159, 2005.
14. Ryan A. Chisholm and Stephen W. Pacala. Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proceedings of the National Academy of Sciences*, 107(36):15821–15825, 2010. ISSN 0027-8424. doi: 10.1073/pnas.1009387107. URL https://www.pnas.org/content/107/36/15821.

15. M. Convertino. Neutral metacommunity clustering and sar: River basin vs. 2-d landscape biodiversity patterns. *Ecological Modelling*, 222(11):1863 – 1879, 2011. ISSN 0304-3800. doi: http://dx.doi.org/10.1016/j.ecolmodel.2011.03.015.
16. Matteo Convertino, Rachata Muneepeerakul, Sandro Azaele, Enrico Bertuzzo, Andrea Rinaldo, and Ignacio Rodriguez-Iturbe. On neutral metacommunity patterns of river basins at different scales of aggregation. *Water Resources Research*, 45(8), 2009. ISSN 1944-7973. doi: 10.1029/2009WR007799.
17. Matteo Convertino, Adam Bockelie, Gregory A. Kiker, Rafael Muñoz-Carpena, and Igor Linkov. Shorebird patches as fingerprints of fractal coastline fluctuations due to climate change. *Ecological Processes*, 1(1), 2012. ISSN 2192-1709. doi: 10.1186/2192-1709-1-9. URL https://doi.org/10.1186/2192-1709-1-9.
18. Matteo Convertino, Filippo Simini, Filippo Catani, Igor Linkov, and Gregory A Kiker. Power-law of aggregate-size spectra in natural systems. *ICST Transactions on Complex Systems*, 2013.
19. Matteo Convertino, Rafael Muñoz-Carpena, Gregory A. Kiker, and Stephen G. Perz. Design of optimal ecosystem monitoring networks: hotspot detection and biodiversity patterns. *Stochastic Environmental Research and Risk Assessment*, 29(4):1085–1101, 2015. doi: 10.1007/s00477-014-0999-8.
20. Katharine Z. Coyte, Jonas Schluter, and Kevin R. Foster. The ecology of the microbiome: Networks, competition, and stability. *Science*, 350(6261):663–666, 2015. ISSN 0036-8075. doi: 10.1126/science.aad2602. URL http://science.sciencemag.org/content/350/6261/663.
21. Richard Crandall and Carl B Pomerance. *Prime numbers: a computational perspective*, volume 182. Springer Science & Business Media, 2006.
22. Gábor Csányi and Balázs Szendrői. Fractal–small-world dichotomy in real-world networks. *Physical Review E*, 70(1):016122, 2004.
23. John P. DeLong, Jordan G. Okie, Melanie E. Moses, Richard M. Sibly, and James H. Brown. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences*, 107(29):12941–12945, 2010. ISSN 0027-8424. doi: 10.1073/pnas.1007783107. URL http://www.pnas.org/content/107/29/12941.
24. Ana Durbán, Juan J Abellán, Nuria Jiménez-Hernández, Alejandro Artacho, Vicente Garrigues, Vicente Ortiz, Julio Ponce, Amparo Latorre, and Andrés Moya. Instability of the faecal microbiota in diarrhoea-predominant irritable bowel syndrome. *FEMS microbiology ecology*, 86(3):581–589, 2013.
25. Karoline Faust, Franziska Bauchinger, Béatrice Laroche, Sophie De Buyl, Leo Lahti, Alex D Washburne, Didier Gonze, and Stefanie Widder. Signatures of ecological processes in microbial community time series. *Microbiome*, 6(1):120, 2018.
26. Charles K. Fisher and Pankaj Mehta. The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences*, 111(36):13111–13116, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1405637111. URL https://www.pnas.org/content/111/36/13111.
27. Beatriz García-Jiménez, Tomás de la Rosa, and Mark D Wilkinson. Mdpbiome: microbiome engineering through prescriptive perturbations. *Bioinformatics*, 34(17):i838–i847, 2018.
28. Christopher L. Gentile and Tiffany L. Weir. The gut microbiota at the intersection of diet and human health. *Science*, 362(6416):776–780, 2018. ISSN 0036-8075. doi: 10.1126/science.aau5812. URL http://science.sciencemag.org/content/362/6416/776.
29. Didier Gonze, Katharine Z Coyte, Leo Lahti, and Karoline Faust. Microbial communities as dynamical systems. *Current Opinion in Microbiology*, 44:41 – 49, 2018. ISSN 1369-5274. doi: https://doi.org/10.1016/j.mib.2018.07.004.
30. Mark Granovetter. The strength of weak ties: A network theory revisited. 1983.
31. L. Gresele and M. Marsili. On maximum entropy and inference. *Entropy*, 19(12), 2017. URL http://www.mdpi.com/1099-4300/19/12/642.
32. J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548:210–213, 2017. doi: 10.1038/nature23273.
33. Volker Grimm, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M. Mooij, Steven F. Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L. DeAngelis. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310(5750):987–991, 2005. doi: 10.1126/science.1116681. URL http://science.sciencemag.org/content/310/5750/987.
34. R. Hanel and S. Thurner. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *EPL (Europhysics Letters)*, 93(2):20006, 2011.

35. Rudolf Hanel, Stefan Thurner, and Murray Gell-Mann. How multiplicity determines entropy and the derivation of the maximum entropy principle for complex systems. *Proceedings of the National Academy of Sciences*, 111(19):6905–6910, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1406071111.
36. Alan Hastings, Karen C. Abbott, Kim Cuddington, Tessa Francis, Gabriel Gellner, Ying-Cheng Lai, Andrew Morozov, Sergei Petrovskii, Katherine Scranton, and Mary Lou Zeeman. Transient phenomena in ecology. *Science*, 361(6406), 2018. ISSN 0036-8075. doi: 10.1126/science.aat6412. URL http://science.sciencemag.org/content/361/6406/eaat6412.
37. Jorge Hidalgo, Jacopo Grilli, Samir Suweis, Miguel A. Muñoz, Jayanth R. Banavar, and Amos Maritan. Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1319166111. URL https://www.pnas.org/content/111/28/10095.
38. S.P. Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2001.
39. Forest Isbell, Vincent Calcagno, Andy Hector, John Connolly, W Stanley Harpole, Peter B Reich, Michael Scherer-Lorenzen, Bernhard Schmid, David Tilman, Jasper Van Ruijven, et al. High plant diversity is needed to maintain ecosystem services. *Nature*, 477(7363):199, 2011.
40. Charlotte James, Sandro Azaele, Amos Maritan, and Filippo Simini. Zipf's and taylor's laws. *Phys. Rev. E*, 98: 032408, 2018. doi: 10.1103/PhysRevE.98.032408.
41. Patricio Jeraldo, Maksim Sipos, Nicholas Chia, Jennifer M Brulc, A Singh Dhillon, Michael E Konkel, Charles L Larson, Karen E Nelson, Ani Qu, Lawrence B Schook, et al. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proceedings of the National Academy of Sciences*, 109(25):9692–9698, 2012.
42. Katerina V.-A. Johnson and Philip W. J. Burnet. Microbiome: Should we diversify from diversity? *Gut Microbes*, 7(6):455–458, 2016. doi: 10.1080/19490976.2016.1241933. URL https://doi.org/10.1080/19490976.2016.1241933.
43. Lou Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006. doi: 10.1111/j.2006.0030-1299.14714.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2006.0030-1299.14714.x.
44. A. M. Kilpatrick and A. R. Ives. Species interactions can explain Taylor's power law for ecological time series. *Nature*, 422:65–68, 2003. doi: 10.1038/nature01471.
45. Dan Knights, Tonya L. Ward, Christopher E. McKinlay, Hannah Miller, Antonio Gonzalez, Daniel McDonald, and Rob Knight. Rethinking "enterotypes". *Cell Host & Microbe*, 16(4):433 – 437, 2014. ISSN 1931-3128. doi: https://doi.org/10.1016/j.chom.2014.09.013. URL http://www.sciencedirect.com/science/article/pii/S1931312814003461.
46. L. Lahti, J. Salojärvi, A. Salonen, M. Scheffer, and W. M. de Vos. Tipping elements in the human intestinal ecosystem. *Nature Communications*, 5:4344, 2014. doi: 10.1038/ncomms5344.
47. Guillaume Latombe, Cang Hui, and Melodie A McGeoch. Beyond the continuum: a multi-dimensional phase space for neutral–niche community assembly. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821): 20152417, 2015.
48. Mehdi Layeghifard, David M. Hwang, and David S. Guttman. Disentangling interactions in the microbiome: A network perspective. *Trends in Microbiology*, 25(3):217 – 228, 2017. ISSN 0966-842X. doi: https://doi.org/10.1016/j.tim.2016.11.008. URL http://www.sciencedirect.com/science/article/pii/S0966842X16301858.
49. Mathew A Leibold, Mark C Urban, Luc De Meester, Christopher A Klausmeier, and Joost Vanoverbeke. Regional neutrality evolves through local adaptive niche evolution. *Proceedings of the National Academy of Sciences*, page 201808615, 2019.
50. JM Levine, J Bascompte, PB Adler, and S Allesina. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, 546:56–64, 2017. doi: 10.1038/nature22898.
51. Roie Levy and Elhanan Borenstein. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, 110(31):12804–12809, 2013.
52. Timothy R Lezon, Jayanth R Banavar, Marek Cieplak, Amos Maritan, and Nina V Fedoroff. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proceedings of the National Academy of Sciences*, 103(50):19033–19038, 2006.
53. Lianwei Li and Zhanshan Sam Ma. Testing the neutral theory of biodiversity with human microbiome datasets. *Scientific reports*, 6:31448, 2016.

54. Joseph T. Lizier. Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11, 2014. ISSN 2296-9144. doi: 10.3389/frobt.2014.00011. URL https://www.frontiersin.org/article/10.3389/frobt.2014.00011.
55. Joseph T Lizier. Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11, 2014.
56. Niklas Lüdtke, Stefano Panzeri, Martin Brown, David S Broomhead, Joshua Knowles, Marcelo A Montemurro, and Douglas B Kell. Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. *Journal of The Royal Society Interface*, 5(19):223–235, 2008. doi: 10.1098/rsif.2007.1079.
57. L Ma and OX Cordero. Solving the structure-function puzzle. *Nature microbiology*, 3(7):750–751, 2018.
58. Zhanshan (Sam) Ma. Power law analysis of the human microbiome. *Molecular Ecology*, 24(21):5428–5445, 2015. doi: 10.1111/mec.13394. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13394.
59. Matteo Marsili, Iacopo Mastromatteo, and Yasser Roudi. On sampling and modeling complex systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(09), 2013. URL http://stacks.iop.org/1742-5468/2013/i=09/a=P09003.
60. Jose Manuel Martí, Daniel Martínez-Martínez, Teresa Rubio, César Gracia, Manuel Peña, Amparo Latorre, Andrés Moya, and Carlos P. Garay. Health and disease imprinted in the time variability of the human microbiome. *mSystems*, 2(2), 2017. doi: 10.1128/mSystems.00144-16. URL https://msystems.asm.org/content/2/2/e00144-16.
61. Matteo Martinello, Jorge Hidalgo, Amos Maritan, Serena di Santo, Dietmar Plenz, and Miguel A. Muñoz. Neutral theory and scale-free neural dynamics. *Phys. Rev. X*, 7:041071, 2017. doi: 10.1103/PhysRevX.7.041071. URL https://link.aps.org/doi/10.1103/PhysRevX.7.041071.
62. Iacopo Mastromatteo and Matteo Marsili. On the criticality of inferred models. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(10):P10012, 2011. URL http://stacks.iop.org/1742-5468/2011/i=10/a=P10012.
63. Thomas J Matthews and Robert J Whittaker. On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*, 52(2):443–454, 2015.
64. C. Mellin, C. J. A. Bradshaw, D. A. Fordham, and M. J. Caley. Strong but opposing beta-diversity–stability relationships in coral reef fish communities. *Proceedings of the Royal Society of London B: Biological Sciences*, 281 (1777), 2014. doi: 10.1098/rspb.2013.1993.
65. Akira S Mori, Forest Isbell, and Rupert Seidl. β-diversity, community assembly, and ecosystem functioning. *Trends in ecology & evolution*, 2018.
66. Vilfredo Pareto et al. Manual of political economy. 1971.
67. Rick Quax, Andrea Apolloni, and Peter MA Sloot. The diminishing role of hubs in dynamical processes on complex networks. *Journal of The Royal Society Interface*, 10(88):20130568, 2013.
68. Kevin R. Foster, Jonas Schluter, Katharine Z. Coyte, and Seth Rakoff-Nahoum. The evolution of the host microbiome as an ecosystem on a leash. *Nature*, 548:43–51, 08 2017. doi: 10.1038/nature23292.
69. Fatimah Abdul Razak and Henrik Jeldtoft Jensen. Quantifying "causality" in complex systems: Understanding Transfer Entropy. *PLoS ONE*, 2014. doi: https://doi.org/10.1371/journal.pone.0099462.
70. Damian W Rivett and Thomas Bell. Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature microbiology*, 3(7):767, 2018.
71. Marten Scheffer, Stephen R Carpenter, Timothy M Lenton, Jordi Bascompte, William Brock, Vasilis Dakos, Johan Van de Koppel, Ingrid A Van de Leemput, Simon A Levin, Egbert H Van Nes, et al. Anticipating critical transitions. *Science*, 338(6105):344–348, 2012.
72. Luis F. Seoane and Ricard Sole. Phase transitions in pareto optimal complex networks. *Phys. Rev. E*, 92:032807, 2015. doi: 10.1103/PhysRevE.92.032807. URL https://link.aps.org/doi/10.1103/PhysRevE.92.032807.
73. Joseph L. Servadio and Matteo Convertino. Optimal information networks: Application for data-driven integrated health in populations. *Science Advances*, 4(2), 2018. doi: 10.1126/sciadv.1701088.
74. Stanislav Sitkin, Timur Vakhitov, and Elena Demyanova. Microbiome, gut dysbiosis and inflammatory bowel disease: That moment when the function is more important than taxonomy. *Almanac of Clinical Medicine*, 46: 396–425, 2018. doi: 10.18786/2072-0505-2018-46-5-396-425.
75. Richard R Stein, Vanni Bucci, Nora C Toussaint, Charlie G Buffie, Gunnar Rätsch, Eric G Pamer, Chris Sander, and Joao B Xavier. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS computational biology*, 9(12):e1003388, 2013.

76. George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012. ISSN 0036-8075. doi: 10.1126/science.1227079.
77. S. Suweis, J. Grilli, J. R. Banavar, S. Allesina, and A. Maritan. Effect of localization on the stability of mutualistic ecological networks. *Nature Communications*, 6:10179, 2015. doi: 10.1038/ncomms10179.
78. Hirokazu Toju, Kabir G Peay, Masato Yamamichi, Kazuhiko Narisawa, Kei Hiruma, Ken Naito, Shinji Fukuda, Masayuki Ushio, Shinji Nakaoka, Yusuke Onoda, et al. Core microbiomes for sustainable agroecosystems. *Nat. Plants*, 4:247–257, 2018.
79. Francesca Tria, Vittorio Loreto, and Vito Servedio. Zipf's, Heaps' and Taylor's laws are determined by the expansion into the adjacent possible. *Entropy*, 20(10):752, 2018.
80. Matthew CB Tsilimigras and Anthony A Fodor. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of epidemiology*, 26(5):330–335, 2016.
81. Chengyi Tu, Samir Suweis, Jacopo Grillib, Marco Formentin, and Amos Maritan. Reconciling cooperation, biodiversity and stability in complex ecological communities. *arXiv e-prints*, 2018.
82. Maarten van de Guchte, Hervé M. Blottière, and Joël Doré. Humans as holobionts: implications for prevention and therapy. *Microbiome*, 6(1):81, 2018. ISSN 2049-2618. doi: 10.1186/s40168-018-0466-8.
83. Doris Vandeputte, Gunter Kathagen, Kevin D?hoe, Sara Vieira-Silva, Mireia Valles-Colomer, João Sabino, Jun Wang, Raul Y Tito, Lindsey De Commer, Youssef Darzi, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, 551(7681):507, 2017.
84. Alejandro F. Villaverde, John Ross, Federico Morán, and Julio R. Banga. Mider: Network inference with mutual information distance and entropy reduction. *PLOS ONE*, 9(5):1–15, 2014. doi: 10.1371/journal.pone.0096732.
85. Shaopeng Wang and Michel Loreau. Biodiversity and ecosystem stability across scales in metacommunities. *Ecology letters*, 19(5):510–518, 2016.
86. Sophie Weiss, Will Van Treuren, Catherine Lozupone, Karoline Faust, Jonathan Friedman, Ye Deng, Li Charlie Xia, Zhenjiang Zech Xu, Luke Ursell, Eric J Alm, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME journal*, 10(7):1669, 2016.
87. Kirk O Winemiller. Spatial and temporal variation in tropical fish trophic networks. *Ecological monographs*, 60 (3):331–367, 1990.
88. Patricia Wollstadt, Ulrich Meyer, and Michael Wibral. A graph algorithmic approach to separate direct from indirect neural interactions. *PloS one*, 10(10):e0140530, 2015.
89. J. R. Zaneveld, D. E. Burkepile, A. A. Shantz, C. E. Pritchard, R. McMinds, J. P. Payet, R. Welsh, A. M. S. Correa, N. P. Lemoine, S. Rosales, C. Fuchs, J. A. Maynard, and R. V. Thurber. Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales. *Nature Communications*, 7: 11833, 2016. doi: 10.1038/ncomms11833.
90. Jesse R Zaneveld, Ryan McMinds, and Rebecca Vega Thurber. Stress and stability: applying the anna karenina principle to animal microbiomes. *Nature microbiology*, 2(9):17121, 2017.
91. Tommaso Zillio, Jayanth Banavar, Jessica Green, John Harte, and Amos Maritan. Incipient criticality in ecological communities. *Proceedings of the National Academy of Science*, 105(48), 2008.

Peer-reviewed version available at Entropy 2019; doi:10.3390/e21050506

Figure Captions

Figure 1. RSA trajectories, RSA-rank, and Relative Species Abundance. Blue, green and red curves refer to the healthy, transitory and unhealthy microbiome. The healthy microbiome shows smaller fluctuations in species diversity α vs. RSA and one regime when considering the RSA-rank profile. An inverse scaling law is detected for the RSA.

Figure 2. Network entropy patterns and inferred Maximum Entropy Networks. Network entropy dependent on the pairwise information flow (TE) (left patterns) and extracted Maximum Entropy Networks on the right. The size of each node is proportional to the Shannon Entropy of the species; the color of the node is proportional to the structural degree (in Fig. S3 the color of each node is proportional to the sum of total outgoing TEs of each node (OTE); the higher OTE, the warmer the color); the distance is proportional to $exp(-MI(X,Y))$ where $MI(X,Y)$ is the mutual information between species RSA x and y; the width of each edge is proportional to the pairwise Transfer Entropy; and the direction is related to $TE(i->j)$; the direction of this edge is from i to j.

Figure 3. Macroecological indicators of microbiome networks and probabilistic characterization. Average α, species similarity $1-\beta$, and total diversity γ are plotted as a function of time. Their probability distribution is shown on the right.

Figure 4. Importance and interaction of microbial species, and top 10 most active species species. σ is describing species interaction and is calculated as the ratio between the total outgoing Information Flow (OTE) ($OTE(j) = \sum_i TE_{j \to i}$) and the Total Network Entropy, while μ is describing the species importance as the ratio between the Nodal Entropy (Shannon Entropy) and the Total Network Entropy. The continuous line in each σ-μ plot shows the critical edge that describes a state between regularity and chaos. On the right the top 10 most active species in terms of OTE (and least relatively abundant) are ranked. These species are the most detrimental for the healthy group and the most beneficial for the unhealthy one.

Figure 5. Exceedance probability distribution of microbiome structure, function, and service. Network degree, total outgoing transfer entropy (OTE) of each node, and α-diversity over time characterize the structure, function and service of the microbiome network.

Figure 6. Macroecological scaling patterns and predicted species interactions. Left plots are about the scaling of total γ-diversity and species similarity $1-\beta$ dependent on the number of speciation events (that is the number of new and existing species introduced until the time considered); speciation time is a proxy of the sampling area over time. Right plots are about the scaling of OTE vs. RSA and γ-diversity vs. OTE.

Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

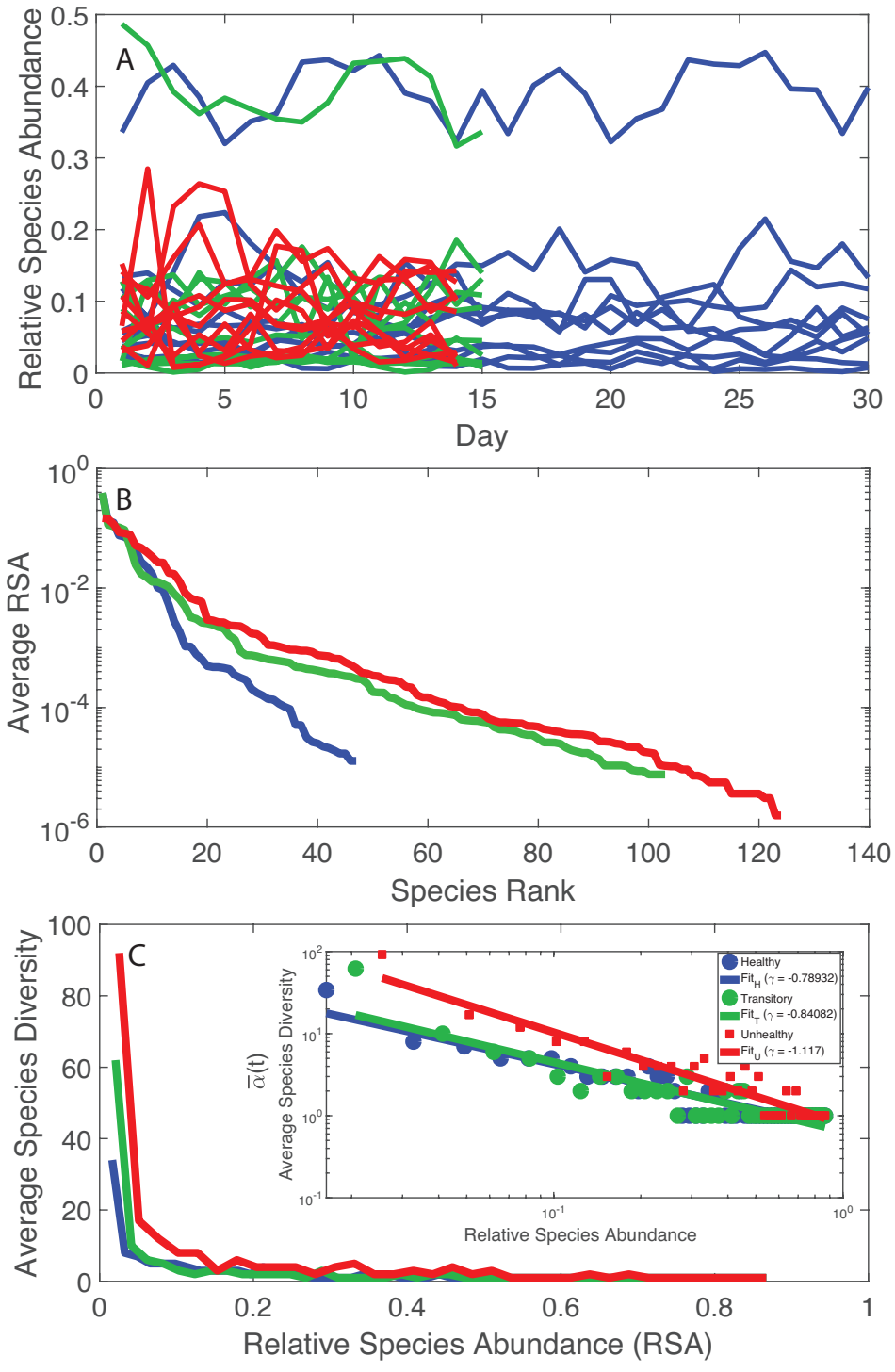

Figure 1

Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

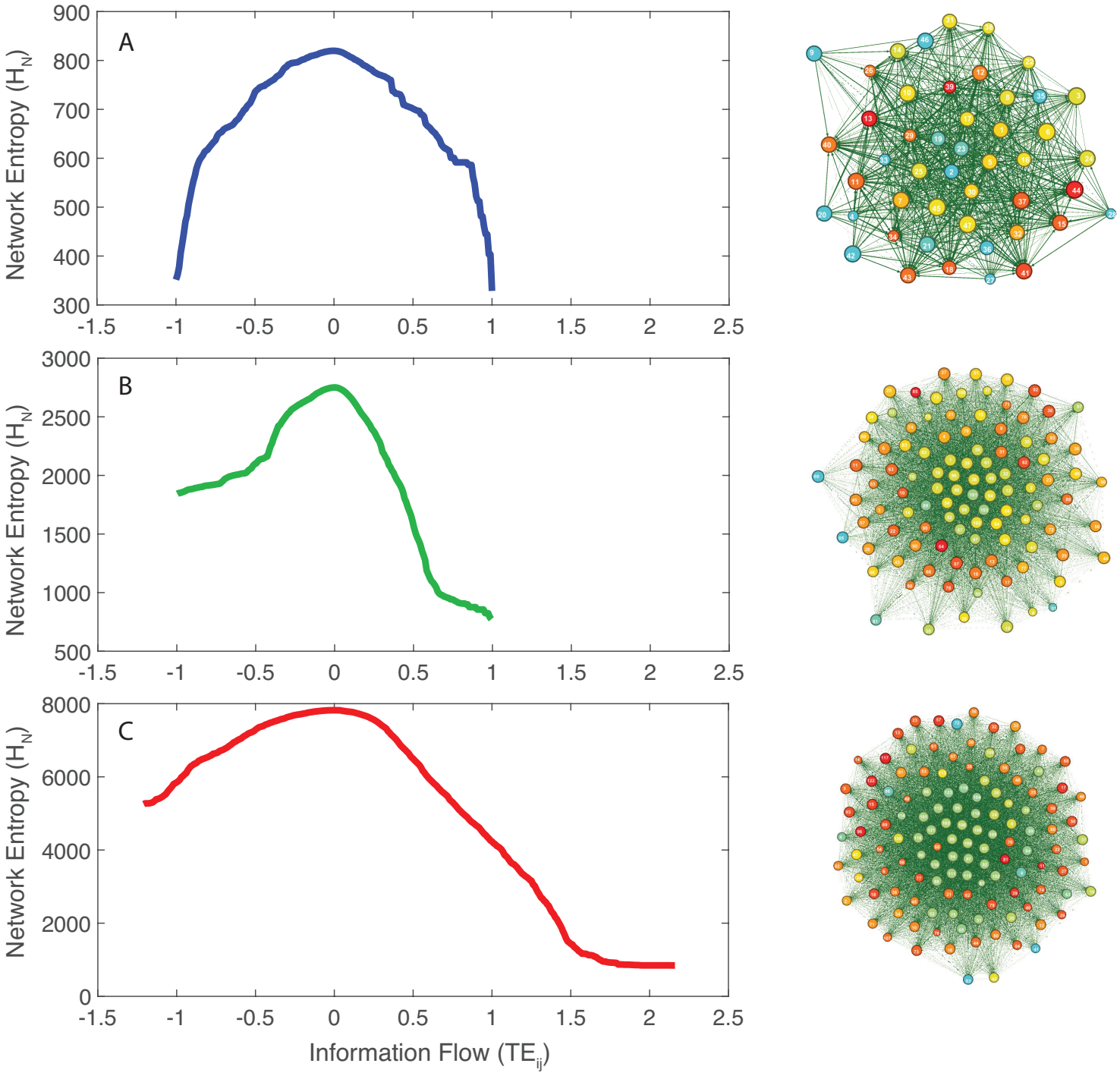

Figure 2

Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

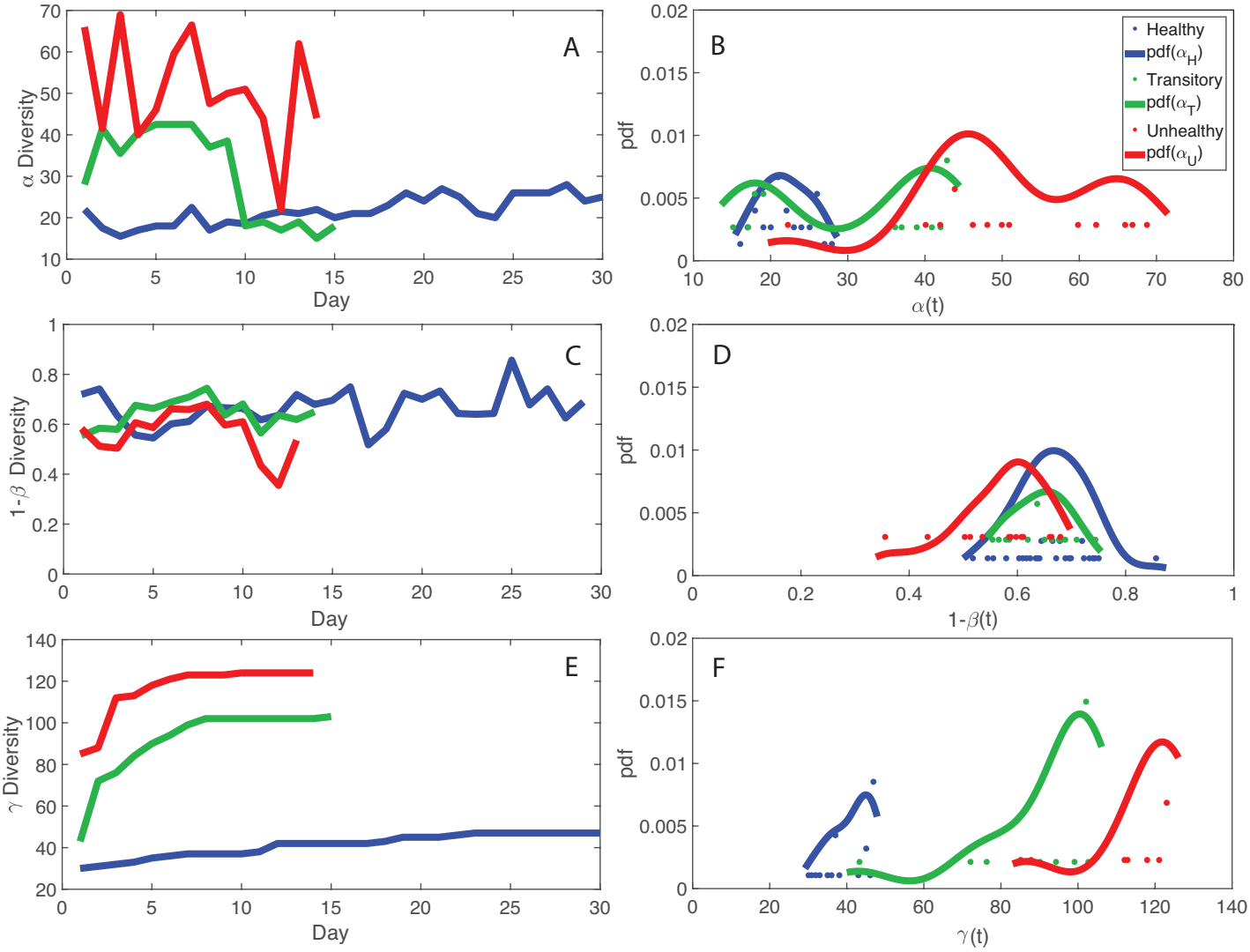

Figure 3

Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

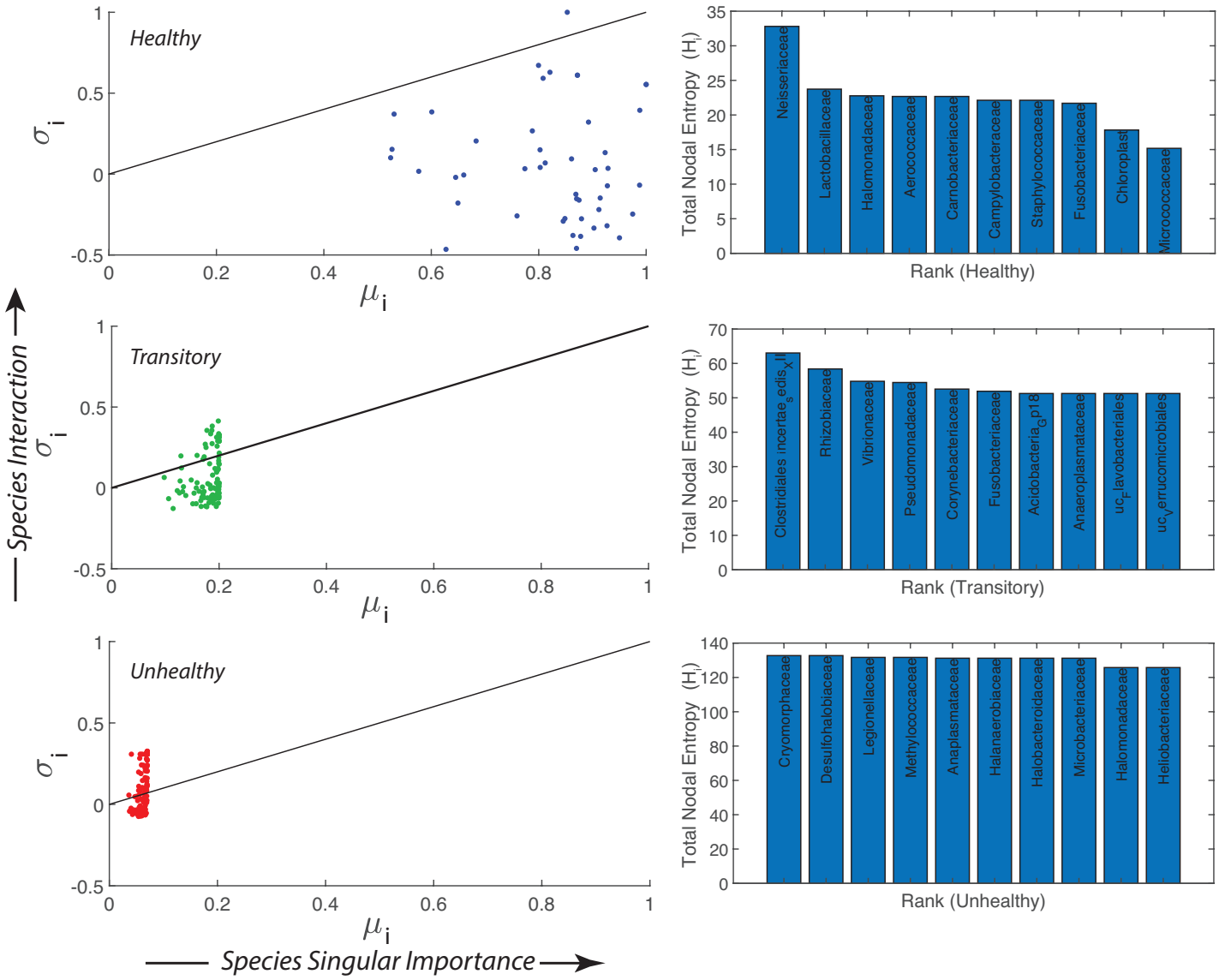

Figure 4

Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

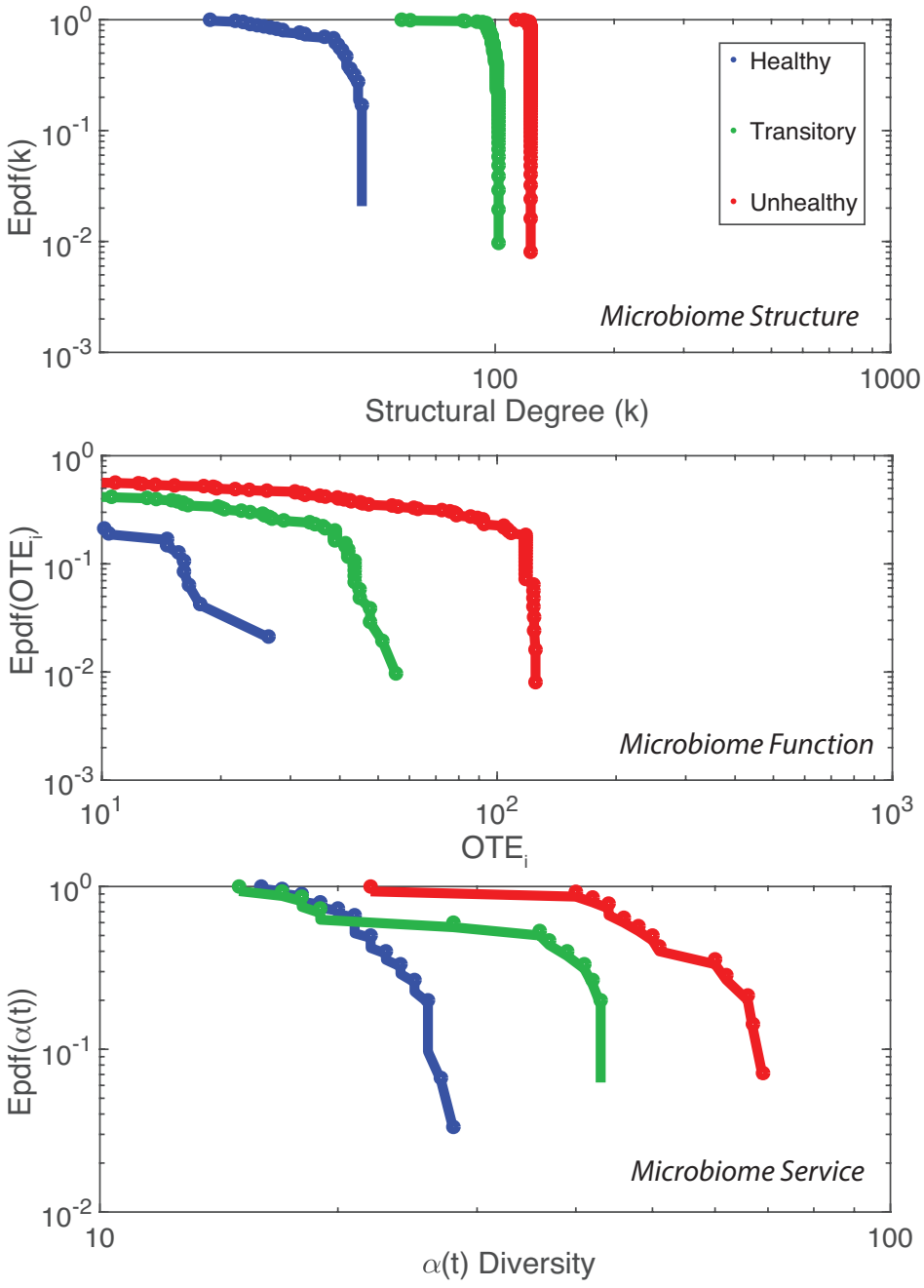

Figure 5

Peer-reviewed version available at *Entropy* **2019**; doi:10.3390/e21050506

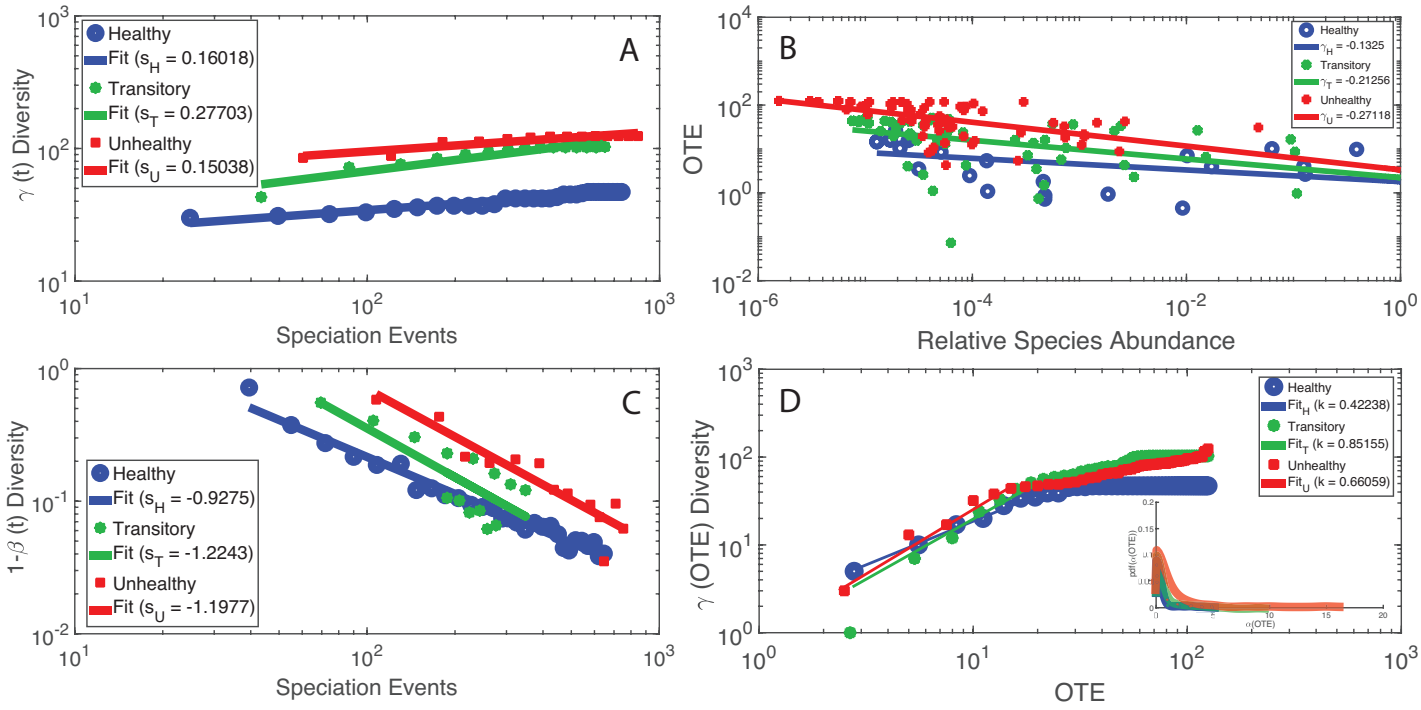

Figure 6