Optimal Microbiome Networks: Macroecological Characterization and Criticality

J Li[a]**, M Convertino**[a,b*]

[a] Nexus Group, Laboratory of Information Communication Networks, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, JP

[b] GI-CORE Global Station for Big Data and Cybersecurity, Hokkaido University, Sapporo, JP

December 14, 2018

Corresponding author: * M. Convertino, 9 Chome, Kita 14, Nishi 9, Kita-ku, room 11-11, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan, 060-0814, phone: +81 011-706-6491, email: matteo@ist.hokudai.ac.jp

Keywords: microbiome, complex networks, species diversity, criticality, transitions

Abstract

The human microbiome is extremely complex considering the amount of species, their interactions, and its variability over time as a function of environmental drivers. Here we untangle the complexity of the human microbiome for the Irritable Bowel Syndrome (IBS) that is the most prevalent functional gastrointestinal disorder linked to many causes. Based on a novel information theoretic network inference model (that considers conditional entropy reduction till the maximum entropy is not reduced further) 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 interaction and small-world features for functional node degree and distance versus random unhealthy networks. We detect an inverse scaling relationship between species total outgoing information flow ("active flow") and abundance. The top 10 interacting species are also the least abundant for the healthy microbiome and the most detrimental; however these species are controlled by other species (via negative feedbacks) and the microbiome is self-organized into a healthy state. On the contrary, the most abundant species for the unhealthy microbiome are the least interactive and the most detrimental. These findings support the idea about a diminishing role of network hubs and hubs should be defined considering total outgoing information flow. The healthy microbiome is characterized by high diversity growth rate, small species similarity decay over time (i.e. low species turnover), and small variability in the abundance of all species. This result challenges current views that posit an association between health states and the highest diversity in ecosystems rather than the highest biodiversity growth as in this study. In a network perspective the healthy microbiome is configured as a small-world network with a tendency toward a critical scale-free network while the unhealthy one is organized as a random network with many more interacting species. We show how the transitory microbiome at the edge of the healthy and unhealthy ones is unstable and criticality of the healthy microbiome is not at the phase transition (or second order) between order and chaos but in a meta-stable state (on the contrary of other critical systems where energy and entropy grow in the same direction and criticality is at the transition). We stress out the importance of considering interacting pairs versus single node dynamics when characterizing the microbiome nexus 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 to humans leading to personalized design of pre- and pro-biotic treatments and engineered microbiome transplants as two examples.

2

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 the human body considering the bacteria holobiont (Coyte et al., 2015; van de Guchte et al., 2018). We recognize that our microbiota are highly dynamic, and that these dynamics are linked to environmental and individual states (van de Guchte et al., 2018). The field is still in its infancy and it 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 (Arumugam, 2011), while others argue that gut communities vary along multidimensional continua without any universality (Knights et al., 2014). 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 (Caesar et al., 2015).

1.2   Microbiome Diversity and Funtional 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 (Martí et al., 2017; Sitkin et al., 2018). 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 (Martí et al., 2017); however, nobody demonstrated how the microbiome network is different for these healthy and unhealthy groups (i.e. "states" generally speaking

85 when not focused on a particular subpopulation) and how the transition from one to another

86 occurs. By exploring this topic we propose novel network inferential models for gathering

87 microbiome networks from species big-data; these models are based on the principle of max-

88 imum entropy that tries to gather the most informative set of variables about stable state

89 patterns (Marsili et al., 2013; Servadio and Convertino, 2018). An example can be about sets

90 of species for predicting a diverse set of species networks. Big-data is not necessarily related

91 to the size of the data used but also to the number of calculations required to infer the un-

92 derlying network structure. These computations increase exponentially with the number of

93 species/nodes n considered beyond the geometrical criteria, where the number of connections

94 is $n(n-1)/2$, because of the directed topology of the network. A directed topology is for

95 instance found when species interaction networks are non-symmetrical which means that the

96 direct influence of two species does not have the same magnitude for different directions of

97 interactions (Layeghifard et al., 2017). A variety of different models have been proposed to

98 infer network structures from small and large datasets. For biological systems in particular,

99 the inference of causal interactions among systems' components is a daunting task because

100 not all interactions are known or the "true" magnitude of interactions considering the data

101 used to assess these interactions. For instance microbiome networks are in principle different

102 if the used input data are species occurrence, abundance, geographic range or other features.

103 Also for this motivation, we employ assumption free inference models that consider the whole

104 probability distribution of species dynamics and that are validated considering their ability to

105 predict population patterns over time. We extract optimal microbiome networks as optimal

106 information networks (OINs) (Servadio and Convertino, 2018) for healthy, transitory and

107 unhealthy groups to investigate general patterns and drivers underlying microbiome stability

108 and the interactions among different species in terms of network topology, magnitude and

109 preferential direction. Additionally, we characterize macroecological functions α, β and γ

110 diversity which describe the temporal organization of microbiome biodiversity considering

111 point time, intertemporal and total diversity. We show how these functions are related to

112 microbiome network features and different topologies emerge for different diversity/health

113 states. The linkage between microbiome networks and macroecology is unique and offers

114 additional insights into the ecology and the evolution of the microbiome with relevance to

115 ecosystem health.

116

1.3   Microbiome 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 (Zillio et al., 2008; Convertino, 2011; Azaele et al., 2016; Martinello et al., 2017). 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 (Convertino et al., 2012; Lahti et al., 2014; Ma, 2015; Gentile and Weir, 2018; Gonze et al., 2018). This scale-free behavior was though to occur only at the critical transition point but recent evidence also shows that criticality does exist also for stable states where system's component organization is optimal due to optimal information sharing among components and the environment (Hidalgo et al., 2014; Martinello et al., 2017). This result was already found for geophysical networks and coupled ecological networks (Banavar et al., 2001; Convertino et al., 2009) for instance where energy dissipation tends to a minimum. However, these models are typically driven by some assumptions 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 (Sugihara et al., 2012). A different approach is achieved by pattern-oriented models (Grimm et al., 2005) 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 (Marsili et al., 2013)) 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 (Mastromatteo and Marsili, 2011; Marsili et al., 2013) is associated to biological criticality 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.

2   Material and Methods

2.1   Microbiome Data

We considered microbiome data from Martí et al. (2017) for which species data for six individuals was available over time (30 days). Species Operational Taxonomic Unit (OTU) abundance data were derived by published 16S rRNA and shotgun metagenomic sequencing (SMS) data pertaining to the gut microbiotas; these data are from Martí et al. (2017). 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.

2.2   Probabilistic Characterization of Microbiome Variables

We characterize probabilistically the distribution of macroecological and microbiome network variables (generally indicated as Y) considering the following general exceedance probability distribution function:

$$P(Y \geq y) \sim \begin{cases} \mathrm{e}^{-\lambda_1 \, y} & \text{for } y < Y^* \\ y^{-\epsilon+1} \, f\left(\frac{y}{m}\right) \, \mathrm{e}^{-\lambda_2 \, y} & \text{for } y \geq Y^* \end{cases} , \qquad (2.1)$$

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. $y^{-\epsilon+1}$ 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 (James et al., 2018).

2.3 Network Inference and Dynamical Species Characterization

2.3.1 Information Balance and Exchange

To infer species interaction networks based on microbial abundance data we base our approach on the model developed in Servadio and Convertino (2018) based on previous efforts (Lizier, 2014; Villaverde et al., 2014). We consider the microbiome as a dynamic network of species interactions where the total free energy changes over time. Considering information entropy as the energy's counterpart, the total network entropy can be written as:

$$H(N) \approx \sum_i H(x_i) + \sum_i \sum_{j \neq i} TE_i(x_i, x_j) + \sigma(N) \tag{2.2}$$

where x_i denote the $i-s$ variables that contribute to the total information of the network N. In our case x is the abundance 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 (Razak and Jeldtoft Jensen, 2014; Lizier, 2014; Servadio and Convertino, 2018); in our case both variables are the abundance of two different species. Eq. 2.2 represents a fundamental principles of information balance independently of the chosen entropy (Hanel and Thurner, 2011) and forms the general basis of sensitivity analyses. Eq. 2.2 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. 2.2 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. 2.2 can also be extended in space if spatially explicit calculations are needed as in Servadio and Convertino (2018).

The computation of TE is based on the distributions of the two variables of interest (species abundance) 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

$$TE_{X_i \to X_j} = \sum p(X_{j,t}, X_{j,\tau}, X_{i,\tau}) \cdot \log \left(\frac{p(X_{j,t}|X_{j,\tau}, X_{i,\tau})}{p(X_{j,t}|X_{j,\tau})} \right) \tag{2.3}$$

7

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 species abundances 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 abundance); 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 abundance values of two randomly selected species did not correlate with abundance values for $\tau = 1$; thus, memory processes are relevant and as in Villaverde et al. (2014) we selected the τ that maximizes the interdependency between two species assessed by the functional distance (see Eq. 2.12).

2.3.2 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 (2018) we proposed to select TEs that lead to the maximum entropy for the inferred network. This corresponds to maximize the Fisher information matrix (Borile et al., 2012). MaxEnt favors probability distribution functions with maximum entropy as the most general distributions that fit the observed data (Hanel et al., 2014). 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 species abundances. These TEs as in Eq. 2.3 assess the potential causal interaction between species pairs. Thus, the utility function is the total network entropy $H(N)$ (Eq. 2.2) 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:

$$
f_{i,j}(X) = \begin{cases} TE_{X_i \rightarrow X_j}, & \text{for } \{X_i, X_j\} \in E_{MENet} \\ 0, & \text{for } \{X_i, X_j\} \notin E_{MENet} \end{cases}, \tag{2.4}
$$

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. 2.2. In the present study, the utility function is defined as the total TE of the network (plus Shannon entropies of each species abundance 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.

2.3.3 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 (Sugihara et al., 2012)) 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. OIN is then the necessary and sufficient MENet for predicting microbiome function. These nodes are strongly predicting the variability of other nodes, thus the overall network dynamics. This entropy reduction (that does not affect much the total entropy) can be achieved by introducing functions $g(X_i)$, defined as follows

$$g(X_i) = \begin{cases} 1, & \text{for } \sum_j f_{i,j}(X) > \sum_j f_{j,i}(X) \\ 0, & \text{for } \sum_j f_{i,j}(X) \leq \sum_j f_{j,i}(X) \end{cases}, \qquad (2.5)$$

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:

$$H(N \equiv OIN) = \sum_i H(x_i) \cdot g(x_{i,t}) + \sum_i \sum_{j \neq i} TE_i(x_i, x_j) \cdot g(x_{i,t}) + \sigma(Y) \qquad (2.6)$$

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 nonredundant 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 species abundance.

2.3.4   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 mcirobiome. 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. (2008). 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

284 and analytically the are formulated as:

$$TEI = \begin{cases} \sigma_i = \frac{OTE(j) = \sum_i TE_{j \to i}}{H(OIN)} \\ \mu_i = \frac{H(x_i) \cdot g(x_{i,t})}{H(OIN)} \end{cases}, \tag{2.7}$$

285 When considering a systemic indicator (see e.g. Servadio and Convertino (2018)) MII are

286 better suited to identify variable importance because no directional influence is needed. MII

287 use the mutual information normalized by the entropy of the output variable considering one

288 independent variable or pairs of variables for predicting a dependent variable Y that is in

289 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

290 is any variable (e.g. species abundance) and Y is the predicted variable built using the same

291 process of constructing OINs but selecting variable features rather than keeping entropy of

292 species as independent variables. The use of the transfer entropy can give further information

293 about the directionality of the causality (in a predictive sense of the model), and the time-lag

294 of the causality.

295

2.4   Macroecological Indicators

297 To characterize the microbiome we introduce macroecological indicators that aim to describe

298 ecosystems' collective variability of diversity locally, within communities or time points, and

299 globally. In this paper we use such macroecological indicators that are time dependent

300 (because space is non relevant nor provided) and of order zero (Jost, 2006). For a set of

301 unique species $\mathbf{S} = \{S_1, S_2, ..., S_n\}$ whose abundance $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ changes over

302 time, we define the local species diversity, or α diversity as:

$$\alpha(t) = \sum_{k=1, t}^{n} p_k(t)^0 \tag{2.8}$$

303 where $p_k(t)$ is the probability to find one species at time t. Thus, α is the sum of di-

304 verse species at any given time during the observation period (30 days). Considering this

305 definition of α is easily noticeable that the sum of the entropy of all species abundance

306 $H_\alpha = \sum_k H(x_k) = -\sum_k p_k(t) \log p_k(t)$ is proportional to the Shannon index that is the

307 local species diversity of order one (Jost, 2006).

308

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)):

$$\beta(t) = 1 - JSI(t) = 1 - \frac{S_{t,t+1}}{S_t + S_{t+1} - S_{t,t+1}} \tag{2.9}$$

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. 2.8).

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 $H_\beta = H_\gamma - H_\alpha$ (Jost, 2006) where $H_\gamma = H(N)$ is the total network entropy (Eq. 2.2). 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.

322

The total diversity γ is defined as:

$$\gamma(t) = \sum_{k=1,t=t^*}^{S,T} p_k(t)^0 \tag{2.10}$$

that is established over the total number of speciation events M (that is the sum of all species at any given time independently of their diversity) or equivalently from time $t = 1$ to the final time of observation T. Considering the validity of the information balance equation (Eq. 2.2) that leads to the diversity balance $H_\gamma = H_\alpha + H_\beta$, the total diversity can also be calculated as $\gamma = \alpha \cdot \beta$ (Jost, 2006).

The total number of speciation events, that is the number of events where new or existing species can be introduced, can be related to the number of unique species as follows.

$$M = \sum_{k=1}^{S} m_i x_i^0 \tag{2.11}$$

where S is the number of unique species across the whole observation period, x_i is the abun-

12

332 dance of the counted species, and m_i is the number of times that species occurs.

333

2.5 Functional and Structural Network Metrics

335 The organizational topology of the microbiome is characterized via structural and functional

336 complex network metrics. Functional metrics are based on information theoretic quanti-

337 ties that quantify the interactions among species while structural metrics are based on the

338 geometry of the network and derived from the former ones.

339 The functional distance between species is defined as:

$$d_f(X_i, X_j) = \min_\tau \; e^{-MI(X_i(t \pm \tau), X_j(t))} \tag{2.12}$$

340 where the value is considering the minimum for all possible time delays τ. X_i and X_j are

341 the abundance of species i and j and MI is the mutual information evaluated for different

342 values of the temporal scale of species dependency τ. The τ that minimizes the distance d_f is

343 chosen for capturing the maximum interdependence MI_{max}. Such distance as in Villaverde

344 et al. (2014) quantifies the magnitude of interactions between species: the higher MI the

345 shorter the distance that signifies high levels on interaction without specifying the direction-

346 ality. Thus, because of the inability of assessing the direction of causality MI is a metric

347 useful for identifying the most interacting pairs of the microbiome rather than individual

348 species.

349 The calculation of the structural distance is based on the functional distance and the

350 concept of the shortest path. The structural distance is then defined as the minimum number

351 of steps from one node (species) to another independently of the magnitude of these steps

352 (e.g. in terms of TE). Thus, analytically the structural distance is defined as:

$$d(X_i, X_j) = \text{argmin} \left[\sum_{i,j} d_f(X_i, X_j)^0 \right] \quad \text{if } A_{ij} = 1 \tag{2.13}$$

353 where $A_{ij} = TE_{ij}^0$ is the adjacency matrix that can be formulated in terms of TE. The

354 rationale for considering the shortest paths is related to the exponentially large ensemble of

355 distances as a function of the number of nodes and the fact that biological systems always

356 optimize information transmission; yet, Pareto shortest paths are always chosen (Seoane and

13

Sole, 2015).

In terms of connectivity the functional degree is defined for the directed network as 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:

$$k_f = k_{in} + k_{out} = \sum_{i,j} \left[f_{i,j}(X)^0 + f_{j,i}(X)^0 \right] \tag{2.14}$$

where $\sum f_{i,j}(X) = TE_{ij}$ is the transfer entropy as defined in Eq. 2.3.

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

$$k = \sum_i a_{i,j} \tag{2.15}$$

where $a_{i,j} = 1 = TE_{i,j}^0$ 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

By a simple cursory analysis it is evident that the average abundance of the healthy microbiome is lower than the average abundance of the unhealthy microbiome independently of the species; however, the maximum abundance is higher for the healthy microbiome and that species is one of the the most beneficial for health. By looking into species diversity (Figure S1) 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 (Mellin et al., 2014; Lahti et al., 2014; Zaneveld et al., 2016). A controversy on the subject is already found in literature (Mellin et al., 2014) 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 (Johnson and Burnet, 2016) More importantly, the rank-abundance pattern (Fig. 1B) shows only one dynamical regime for the healthy microbiome (exponential like) vs. two regimes for the transitory and

14

the unhealthy microbiomes. Figure 1C shows that the decay in richness over abundance 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. Figure 1 B and C shows the rank-abundance plot and the Preston's plot (Hubbell, 2001) of species diversity dependent on abundance. The rank-abundance shows two dynamical regimes for the unhealthy and transitory groups: a result that likely confirm 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 abundance of species in time, from the most abundant to the least abundant, a transition in the pdf of abundance 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 abundance. Considering the abundance of all species together (Figure S2) the transition is less dramatic; from an exponential to a log-normal like distribution. Intermediate abundance 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 the function of the complex microbiome as highlighted in Lahti et al. (2014). Rare species seem also to display a truncated scale-free behavior (limited by their maximum abundance 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 abundance groups: pseudo-random, scale-free, and small-world topology for the highest, intermediate and lowest abundance class, respectively. Further results will discuss the connection between abundance 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), 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 2.3.3). 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. 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. 2.6, 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 abundance 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 abundance 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 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 healthy state. The latter

however seems easy to perturb consider the lower entropy (and probability) to be defined in one state.

Species collective interaction and singular importance is show 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 abundant for the healthy microbiome and the most detrimental; however these species are controlled by other species and the microbiome is organizied into a healthy 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. The least abundant species for the unhealthy microbiome are the most interactive and the least detrimental. On the contrary, the most abundant species (Fig. S4) for the unhealthy microbiome are the least interactive and the most detrimental. For the healthy microbiome, the most abundant species interact the least and these species are the most beneficial. These species-specific analyses are useful for detecting species that are more beneficial or detrimental and this knowledge can lead to design probiotic treatment or microbiome transplants for instance. Figure S7 shows that from the top to the least 10 TE species there is a shift in the pdf of abundance 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 dangerous bacteria ("antibiotic") but their abundance is small for the healthy microbiome; this means that these bacteria are controlled by all other good bacteria.

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 network topology versus microbiome structure. As show by Figure 2 the healthy microbiome function is a tending toward a scale free organization. This mild scale free organization however does not correspond to a scale free distribution of α diversity (Fig. 5 bottom plot) that instead is exponential. The non-linearity between structure, function and microbiome service (diversity) is highlighted when plotting α dependent on functional network degree and distance. α increases for high value of functional connections (Eq. 2.14) but does not have a clear trend when considering the functional distance (Eq. 2.12); 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 (Hubbell, 2001). 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. 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 abundance 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 (DeLong et al., 2010)) implies a decay in species interaction for highly 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.

4 Discussion

We employ a information theoretic model for the inference of the microbial species interaction network. The model is used to infer a microbial network suitable for predicting selected biodiversity patterns characterizing space-time organization of α, β, and γ diversity. Thus, the purpose of the model is not to infer causal (or "true") species-species interactions among bacteria. The "exact" computational inference of these interactions is always very hard — provided that there is a complete knowledge of the reality on which results can be validated — and dependent on the analytics and data used. For instance, abundance profile may not necessarily contain the information about all species-species interactions aimed to be assessed. In this perspective the entropy-based model is focused on the predictability of patterns vs. causal investigation of mechanisms. 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

18

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. The lower entropy in species collective interaction has certainly implications for data collection, potentially implying less data needed for characterizing healthy microbiomes. More theoretically speaking, the highest entropy is a sign of criticality that is the state toward which any ecosystem tend to (Hidalgo et al., 2014) where there is a balance of self-organization and environmental influence.

We do not find any strong scale invariance as for instance in Servadio and Convertino (2018), 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 (2018) 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. abundance 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 versus species independent dynamics as shown in Fig. S5. 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. (2012) and may not allow host individuals to keep the microbiome "on a leash" (R. Foster et al., 2017). The broken symmetry can be indeed manifest an unhealthy state. The neutral state also coincides with the critical state because of the tending scale-free organization of the network manifested by the epdf of OTE (Fig. 5), higher functional distances and smaller functional degrees (Fig. S8). High threshold show that healthy group maintain topology while changing TE; this is because healthy networks are more scale free than unhealthy one. This configuration enhances stability as confirmed by the calculation of the dominant eigenvalue for both the adjacency and TE matrix; the dominant eigenvalue is the smallest for the healthy group that is a signature of stability of the network.

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

19

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. (2016) 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. (2016). 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) (Jost, 2006). This definition is in line with the general information balance equation (Eq. 2.2) and the more specific diversity balance equation $H_\gamma = H_\alpha + H_\beta$ as in Jost (2006). 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 variation 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 (Mellin et al., 2014). Convertino et al. (2015) had previously found that ecosystem hotspots are those that maximize the Value of (biodiversity) Information 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 transitions 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)).

The introduction of new pathogens driven by the environment can lead to the alteration of the whole ecosystem microbiome. 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. 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 external variability. 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) (Suweis et al., 2015; Tu et al., 2018). Even small value 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 in-

teresting to note that this scale-free cooperation of species leads to Taylor's laws (Kilpatrick and Ives, 2003; Ma, 2015) between mean and variance of abundance where Taylor's exponent is different for healthy and unhealthy groups (Martí et al., 2017). Yet, this reemphasizes the connection between time dynamics, network organization, and ecological patterns of diversity and abundance (Suweis et al., 2015; Grilli et al., 2017; Gonze et al., 2018). In particular it has been shown that higher-order interactions (e.g. captured by σ_i in our model) have a stabilizing role (Grilli et al., 2017). These higher-order interactions are all those beyond the simple pairwise interactions whose sum indeed cannot explain the whole composition of ecosystems (Levine et al., 2017). We show that these higher-order interactions cannot be prevalent because some species must have an independent dynamics (capture by μ_i) otherwise instability and tendency toward disorganized unhealthy state is very likely (Fig. 4).

Universality in human microbiota dynamics 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" (Allen et al., 2017). 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 (Allen et al., 2017). 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.

5 Conclusions

A 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. Maximum entropy networks are then extracted considering the highest information without model overfit. The macroecologi-

cal validation of the model is performed considering the ability to simultaneously predict the pdf of α-diversity, γ-diversity growth over time, species similarity $(1 - \beta)$ decay, and the rank-abundance profile. This validation allows to predict other biodiversity patterns such as the Preston's plot of average species richness dependent on species abundance. Considering the application of the model to healthy and Irritable Bowel Syndrome symptomatic individuals the following points are worth mentioning.

- 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 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; 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 (e.g. related to infections and antibiotics) causes an overgrowth in abundance of many opportunistic species as well as the generation of new detrimental species;

- 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 is found between γ diversity and the number of species generations (over time) with an exponent equal to 0.20 on average. The fact that the healthy microbiome has the lowest average diversity 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. 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) that can lead to over-redundancy and instability as observed for the dysbiotic microbiome. Hence, we tend to challenge the diversity-health-stability hypothesis if for diversity is considered the total systemic diversity γ;

- 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, i.e. an increase and a

23

decrease in mean and variance of species diversity while approaching the transition. In the unhealthy state the variance of α is higher than in the healthy state and concentrated around two values which underline the chaotic-like dynamics of the microbiome. In terms of functional network topology a transition between the SW-like to the random network topology is observed for the microbiome network. 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 for second order transitions as in this case); rather criticality can coincides with neutrality in open energy dissipative systems. Neutrality implies higher topological complexity but higher dynamical stability considering the small-world 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 need 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. OTE is largely determining the total entropy of the network compared to the sum of nodal entropies whose contribution is negligible. The highest OTE nodes have the lowest abundance, and are the most beneficial and the most detrimental bacteria for the dysbiotic and healthy microbiome. A scaling law is found between OTE and abundance with an exponent close to $1/4$ that is similar to the mass-specific Kleiber's law where the species specific metabolic rate is the OTU and the mass is the abundance. 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.

24

Acknowledgments

Author Contribution

J.L. made all calculations and contributed to the writing of the paper. M.C. idealized the study and wrote the manuscript as well as supervised J.L. in making all calculations. Codes are available at the GitHub account https://github.com/HokudaiNexusLab/Microbiome

References

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.

M et al. Arumugam. Enterotypes of the human gut microbiome. Nature, 473:174–180, 2011. doi: 10.1038/nature09944.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

S.P. Hubbell. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, 2001.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Chengyi Tu, Samir Suweis, Jacopo Grillib, Marco Formentin, and Amos Maritan. Reconciling cooperation, biodiversity and stability in complex ecological communities. arXiv e-prints, 2018.

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.

Alejandro F. Villaverde, John Ross, Federico Morn, 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.

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.

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.

Figure Captions

Graphical Abstract.

Figure 1. Abundance trajectories, rank-abundance, 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 abundance and one regime when considering the rank-abundance 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 $exp(-I(X, Y))$ where $I(X, Y)$ is the mutual information between species abundance 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. put $\langle\alpha\rangle$ in the first top left plot

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) ($OTE(j) = \sum_i TE_{j\rightarrow i}$) 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 abundant) are ranked. These species are the most detrimental for the healthy group and the most beneficial for the unhealthy one.

921   **Figure 5. Exceedance probability distribution of microbiome structure, func-**

922   **tion, and service.** Network degree, total outgoing transfer entropy (OTE) of each node,

923   and α diversity over time characterize the structure, function and service of the microbiome

924   network.

925   **Figure 6. Macroecological scaling patterns and predicted species interactions.**

926   Left plots are about the scaling of total γ diversity and species similarity $1 - \beta$ dependent on

927   the number of speciation events (that is the number of new and existing species introduced

928   until the time considered); speciation time is a proxy of the sampling area over time. Right

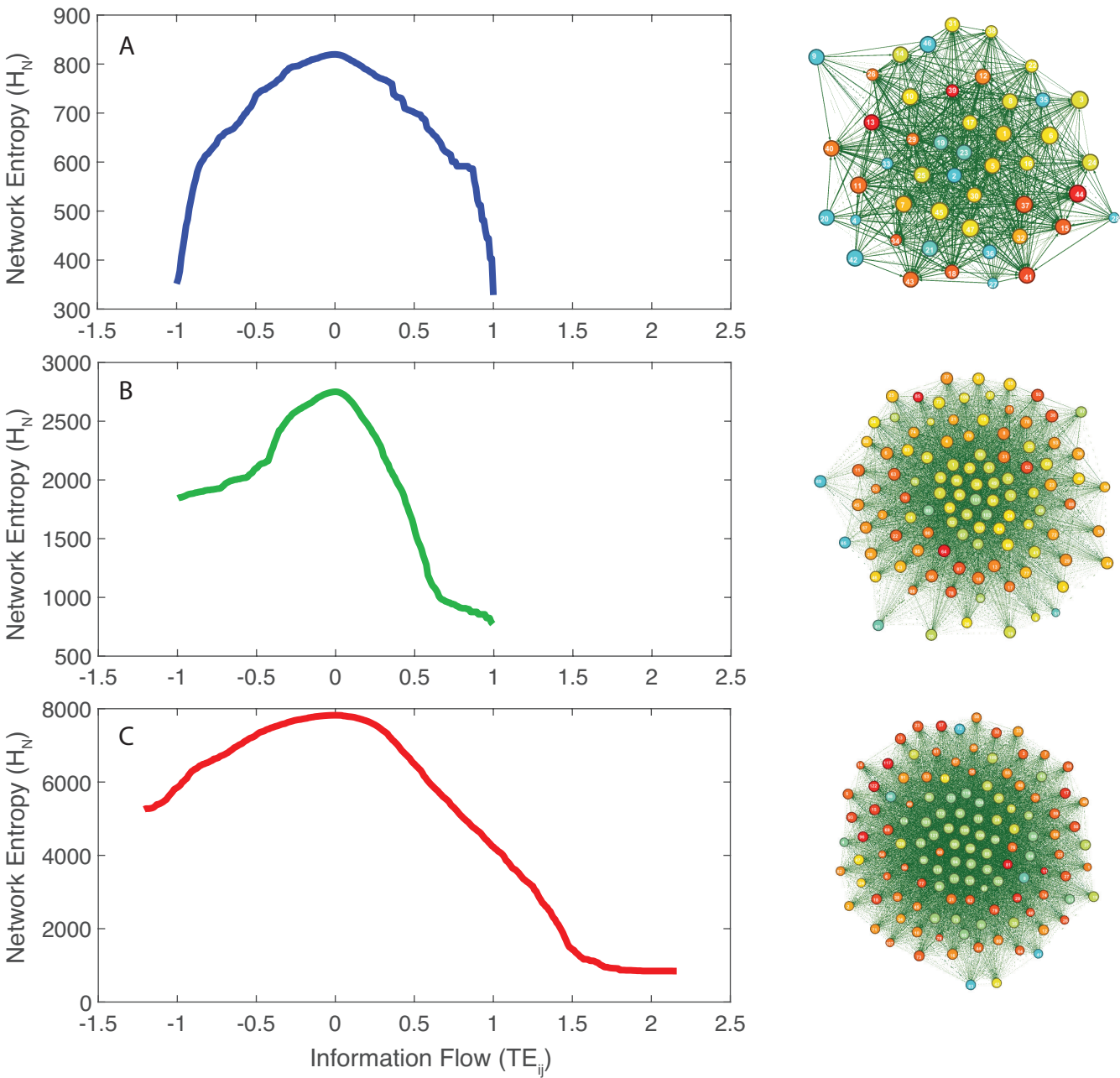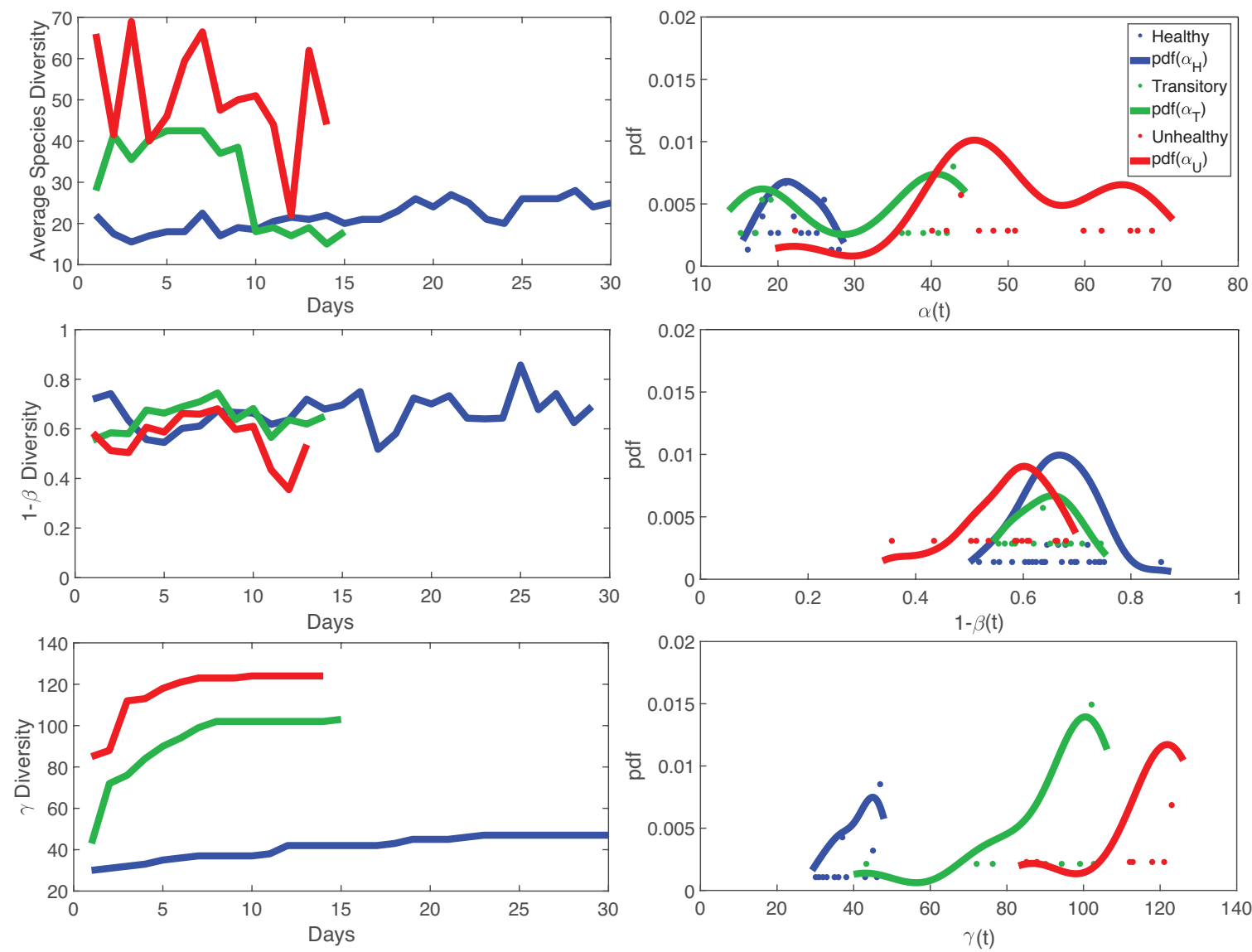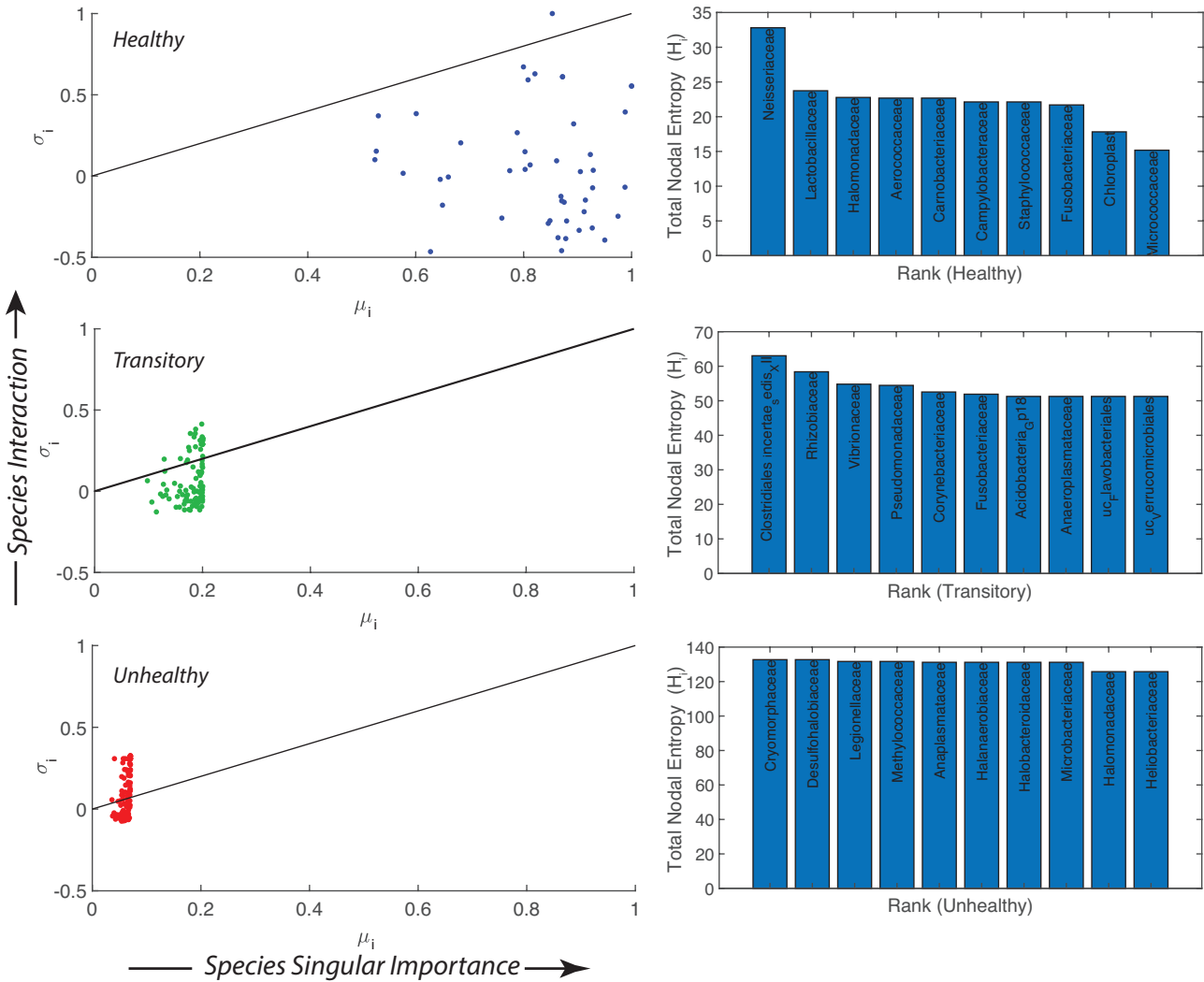929   plots are about the scaling of OTE vs. abundance and γ diversity vs. OTE.

930
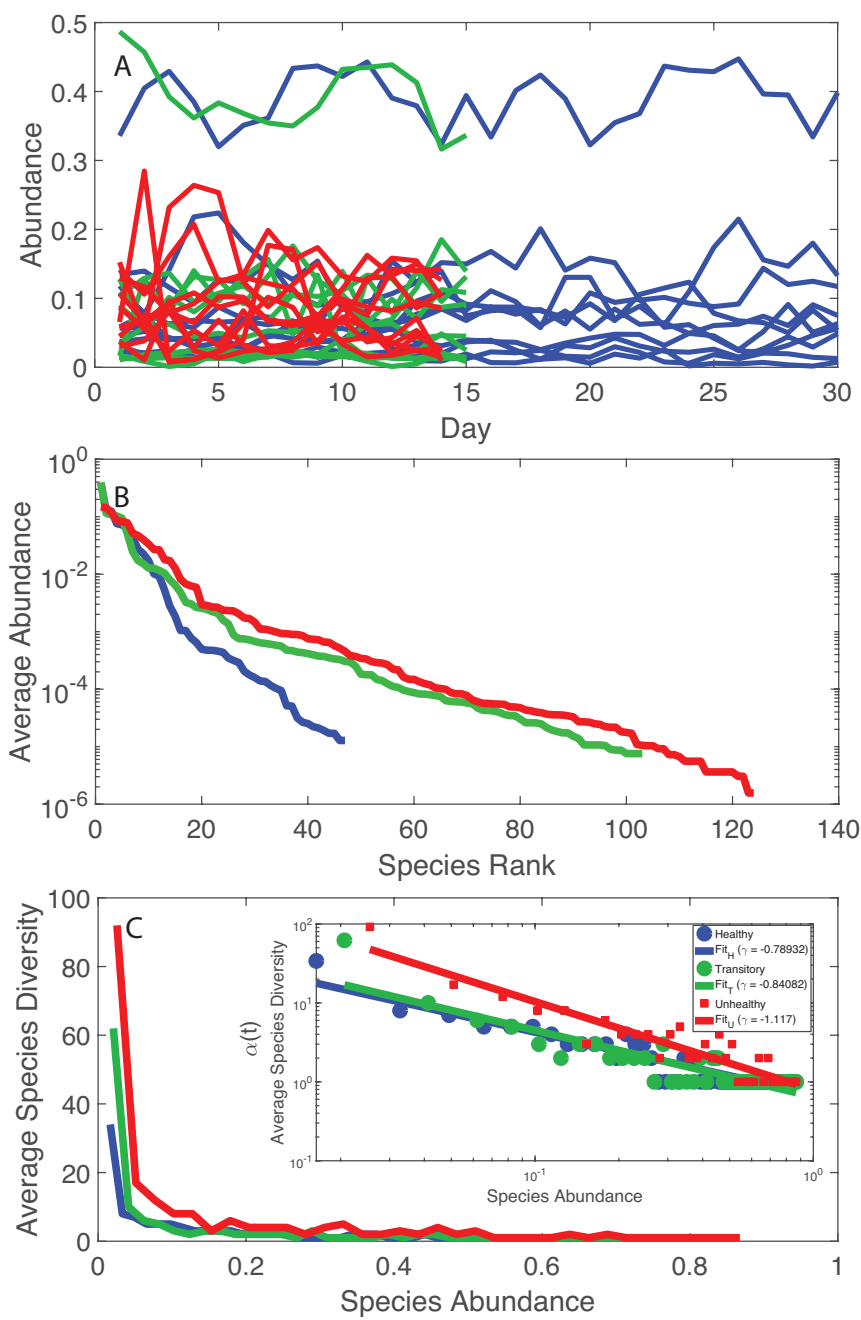
Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6: