# A Statistical Mechanics Description of Environmental Variability in Metabolic Networks

**Jonathan J. Crofts** · **Ernesto Estrada**

**Abstract** Many of the chemical reactions that take place within a living cell are irreversible. Due to evolutionary pressures, the number of allowable reactions within these systems are highly constrained and thus the resulting metabolic networks display considerable asymmetry. In this paper, we explore possible evolutionary factors pertaining to the reduced symmetry observed in these networks, and demonstrate the important role environmental variability plays in shaping their structural organization. Interpreting the returnability index as an equilibrium constant for a reaction network in equilibrium with a hypothetical reference system, enables us to quantify the extent to which a metabolic network is in disequilibrium. Further, by introducing a new directed centrality measure via an extension of the subgraph centrality metric to directed networks, we are able to characterise individual metabolites by their participation within metabolic pathways. To demonstrate these ideas, we study 116 metabolic networks of bacteria. In particular, we find that the equilibrium constant for the metabolic networks decreases significantly in-line with variability in bacterial habitats, supporting the view that environmental variability promotes disequilibrium within these biochemical reaction systems.

**Keywords** Complexity · Metabolic networks · Returnability · Centrality

## 1 Introduction

The set of biochemical metabolic reactions within a living cell form a network of chemical transformations known as a *metabolic network* [3,6]. Metabolic networks

J. J. Crofts

School of Science and Technology, Department of Physics and Mathematics, Nottingham Trent University, Nottingham, NG11 8NS, UK
Tel.: +44(0)115 848 8419
E-mail: jonathan.crofts@ntu.ac.uk

E. Estrada

Department of Mathematics and Statistics, Institute of Complex Systems, University of Strathclyde, Glasgow, G1 1XQ, UK

share many interesting features with other reaction networks. For instance, the products of a metabolic reaction act as reactants for other reactions, thus forming a complex network of metabolic reactions. The *metabolites*, i.e. those chemicals involved in metabolic reactions, are small molecules which are imported/exported and/or synthesised/degraded inside the organism. They can play the role of *substrates* (reactants) or *products*. As with many other chemical reactions, in metabolic ones there exists another important contributor, playing the role of catalyst: the *enzymes*. Last but not least are the *cofactors*, which are small molecules that bind enzymes and can enhance or decrease their catalytic activities. This combination of substrates, products, enzymes and cofactors make metabolic reactions more complex than usual chemical reactions. Representing such a vast complexity of metabolic transformations is a very challenging problem. Typically, a reductionist approach is taken whereby one represents the network of substrates and products linked by their direct transformations into a condensed metabolic network [16]. However, such oversimplification can lead to a number of problems when information concerning specific metabolic pathways or the individual role of metabolites is to be inferred [1]. A global analysis of metabolic networks is less prone to such kinds of errors. For instance, Parter et al. [20] found that certain global characteristics of metabolic networks, e.g. modularity, correlated with the level of variability within an organisms environment for some 117 bacterial species (a result further evidenced in a more comprehensive review of some 325 bacterial species by [15]). In other words, they found that the more variable the environment, the more modular the metabolic network. These studies, however, failed to account for an extremely important characteristic of metabolic and other reaction networks. That is, under determined physiological conditions some reactions are *irreversible*, i.e. they take place in one of only two possible directions. The reversibility of a metabolic reaction is controlled by the thermodynamics, the kinetics and the stoichiometry of the reaction.

Here we analyse metabolic networks for 116 of the bacterial species previously studied by Parter et al. [20]. These species live in a broad range of habitats such as oceans, salt lakes, thermal vents, soil and within hosts. Using a classification of these species according to variability in their environments we investigate the extent to which the reversibility of an organisms metabolic network impacts on its adaptability. That is, we consider the directed network of metabolic reactions in which the directionality of a given reaction is indicated by an arrow from substrate to product. Our method consists of formulating a statistical mechanics approach to derive a hypothetical equilibrium constant between the real metabolic network and one in which all reactions are reversible. The equilibrium constant measures how far from a globally reversible state, i.e one in which all reactions are reversible, a metabolic network is. Calculating the free energy of the system with respect to this hypothetical equilibrium provides a measure of the 'energetic cost' of the evolutionary process giving rise to the corresponding metabolic network. Using this approach we conclude that bacteria living in more variable environments display on average a higher 'global reversibility' (or returnability) in their reactions. In contrast, the metabolic networks of those living in very specialized media display significantly reduced levels of returnability. Maintaining the reversibility of all possible metabolic reactions comes at an extremely high cost. Thus, increased returnability in metabolic networks can

be considered as an extreme evolutionary strategy due to pressures imposed by the potential lack of certain nutrients within highly variable environments. On the other hand, when the media is specialised, a constant stream of nutrients is more readily available, allowing the returnability of the metabolic network to be kept at a minimum thus saving considerable energy during the evolutionary process.

## 2 On Directionality of Metabolic Networks

Directionality or lack of it in metabolic networks indicates whether a metabolic reaction is reversible or not, respectively. Strictly speaking all chemical reactions are reversible and so they should be represented by undirected (or bidirected) links. This means that all reactions have finite standard Gibbs energies and consequently they have finite equilibrium constants. However, it is well-known that some metabolic reactions have very large equilibrium constants, e.g. of the order of 1000 or larger, which means that the consideration of the reverse reaction has no practical consequences for cell physiology. In these cases it is widely accepted to consider such reactions as irreversible. There are other situations in which the reactions are considered irreversible despite the fact that the equilibrium constants are relatively small. This happens when it is known that the concentrations of the reactants is always much greater than those of the products in normal physiological conditions. The reader interested in the details of these considerations can consult the excellent review by Cornish-Bowden and Cárdenas [4].

Using a combination of chemical and physiological information Ma and Zeng [17] have reviewed the reaction direction for all metabolic maps in the KEGG reaction database. They have concluded that some of these reactions have the wrong direction and have proposed to consider 11 different types of metabolic reactions as being irreversible under physiological conditions. These include, for instance, oxygen consumption reactions, most of the production of CO2, NH3, phosphate, many reactions in which ATP is consumed, and several hydrolization reactions, amongst others [17]. This corrected database which contains 1969 irreversible metabolic reactions is widely used in modelling metabolic networks [11].

It is important to remark that even in the case when the aim of the study is to model quantitatively the kinetics of a metabolic reaction the consideration of irreversible reactions plays a fundamental role. Cornish-Bowden and Cárdenas [4,5] have studied the influence of considering some irreversible reactions in modelling the kinetics of metabolic reactions. They have clearly concluded that the important aspect of these models is to allow for information about metabolite concentrations to flow from the end of the model to the beginning. When such loops are allowed it is usually "unnecessary to allow for the small degree of reversibility of reactions with large equilibrium constants in favour of the forward direction". In other words, irreversible reactions can be considered if certain degree of returnability of the flow is allowed. Because returnability of metabolic reactions is the topic of the current work we remark here another conclusion from the same authors about the importance of this concept "If there is a feedback loop that allows communication around an irre-
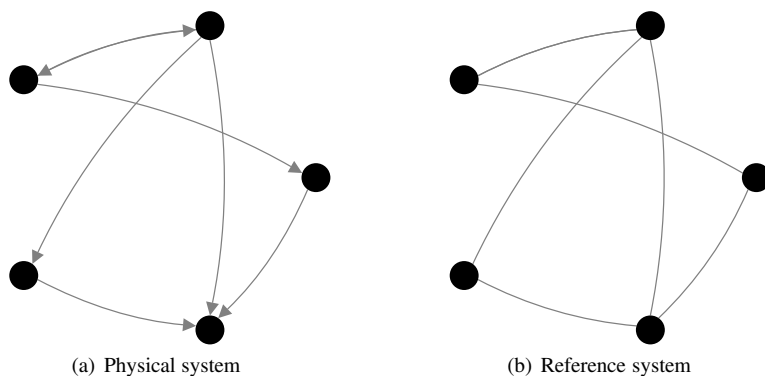
(a) Physical system                                (b) Reference system

**Fig. 1** Reaction network (left) and its reference system (right) consisting of the same set of nodes but with every link being undirected.

versible step then the behaviour is virtually identical whether the reversibility of the nearly irreversible step is allowed for or not".

## 3 An Equilibrium Constant for Complex Networks

We begin by reformulating the concept of network returnability, as introduced by Estrada and Hatano [8], in terms of the equilibrium of a directed reaction graph and its hypothetical reference system. Thus providing a quantitative measure for investigating the extent to which disequilibrium within chemical reaction systems may be considered a product of their natural environments. To start, let us consider that for each reaction network there exists a hypothetical reference system in which all reactions are reversible. In the language of graph theory, the reference system of a set of chemical reactions is nothing other than the underlying, undirected graph for the corresponding reaction graph (see Figure 1). The reference system considers all reactions to be reversible and thus may be considered as an ideal equilibrium. Note, that a network in which all links are reversible is represented as an undirected graph.

In order to determine how far from this ideal equilibrium a real chemical reaction network lies, we consider the difference in free energy between the physical and reference systems. Let $F_p$ and $F_r$ denote the free energies of the reaction and reference networks respectively. Then

$$\Delta F = F_p - F_r = -\beta^{-1} \ln \frac{Z_p}{Z_r}, \tag{1}$$

where $Z_{\{p,r\}}$ is the partition function of the corresponding (physical or reference) system and $\beta = 1/kT$ denotes the inverse temperature[1]. Here temperature may be considered more of a metaphor than a 'real' physical parameter, and quantifies the extent to which the network under consideration is subject to external stresses. For

---

[1]  $k$ is the Boltzmann constant and $T$ the temperature.

instance, it could represent physiological or extreme environmental conditions under which the metabolic network functions.

Note that the system consisting of only reversible reactions, i.e. the reference system, can be considered as being more stable than the physical system for which only a fraction of reactions are reversible. Now, by considering the ratio of the two partition functions in (1), we are able to determine the relative returnability [8] of a reaction network with respect to its hypothetical system

$$K = \frac{Z(D)}{Z(G)} = e^{-\beta \Delta F},$$

where here, we have used the fact that the chemical reaction system can be represented by the digraph $D = (V, E_D)$ and the corresponding reference system by its underlying graph $G = (V, E_G)$.

Recently, Estrada and Hatano [7] showed that the partition function of an undirected graph is given by

$$Z(A) = \text{trace}\left(e^{\beta A}\right), \tag{2}$$

where $A$ is the adjacency matrix of the graph. The quantity in (2) defines the so-called Estrada index of a network [21]. Expanding (2) in a Taylor series gives

$$Z(A) = \text{trace}(I) + \beta \text{trace}(A) + \frac{\beta^2}{2!}\text{trace}(A^2) + \cdots + \frac{\beta^k}{k!}\text{trace}(A^k) + \cdots.$$

Here $I$ denotes the identity matrix. It is well known that the $\text{trace}(A^k)$ counts the number of closed (self-returning) walks of length $k$ within a graph. Thus, $\text{trace}(A^k) > 0$ for any network containing at least one cycle of length $k$. The partition function $Z(G)$ can then be considered as a weighted sum over all closed walks (CWs) within the undirected graph – longer CWs being more heavily penalized by the factor $1/k!$.

Similarly, one may define the partition function of a directed network as

$$Z(D) = \text{trace}(e^{\beta D}) = \text{trace}(I) + \beta \text{trace}(D) + \frac{\beta^2}{2!}\text{trace}(D^2) + \cdots \tag{3}$$
$$+ \frac{\beta^k}{k!}\text{trace}(D^k) + \cdots.$$

Here, $\text{trace}(D^k)$ counts the number of returnable walks of length $k$ starting and ending at the same node. Again, $\text{trace}(D^k) > 0$ if and only if the network has a returnable cycle of length $k$. Now, because $\text{trace}(I)$ is equal to the number of vertices in the graph, and since we are are not interested in the influence network size has on the relative 'returnability' of a reaction network, we simply remove this term from the partition functions. Hence, we obtain the following augmented ratio

$$K_r(\beta) = \frac{Z(D,\beta) - n}{Z(G,\beta) - n}.$$

We call the quantity $K_r(\beta)$ the returnability of a directed network.

The returnability of a directed network is bounded as $0 \leq K_r(\beta) \leq 1$. The lower bound being attained for acyclic digraphs, whilst the upper bound is attained by symmetric digraphs. For the directed cycle on $n$ nodes $C_n$, we have that $K_r \to 0$ as $n \to \infty$ [8]. Note that for $\beta = 0$ the returnability is undefined since $Z(D, \beta) = Z(G, \beta) = 0$. This represents the situation in which the network is at infinite temperature with every node in isolation and thus we may set $K_r = 0$ in this case. Note, that we define the *thermodynamic equilibrium constant* of a network as

$$pK_r = -\log K_r. \tag{4}$$

A large value for this index corresponds to a network that lies far from the ideal equilibrium in which all reactions are reversible, or, in other words, larger values of $pK_r$ are associated with less returnable networks.

For completeness, we note that the differences between the returnability metric as defined here and the network reciprocity [10] are substantial. The latter being the fraction of links in a network that are reciprocal (bidirected). For example, a directed cycle clearly has zero reciprocity yet it is straightforward to realize that the returnability is non-zero in this case. On the other hand, the existence of reciprocal links does not guaranty a high level of returnability within the network. The interested reader is directed to the reference [8] for further details.

We define the returnability of a given chemical (substrate or product) in a reaction graph analogously to that of the global returnability. That is, given a chemical $i$ we define

$$K_r(\beta, i) = \frac{(e^{\beta D})_{ii} - 1}{(e^{\beta G})_{ii} - 1}, \tag{5}$$

to be the returnability of all chemical reactions in which this chemical takes place. In this way, we may consider Equation (5) to be a new, directed centrality measure given by the ratio of *subgraph centralities* [9] for the directed and underlying networks respectively. In the current context, we may interpret this as follows: metabolites that are involved in a larger number of pathways, and particularly shorter, more energy efficient ones, should be considered as being more important to the metabolic process.

## 4 Metabolic Networks of Bacteria

To illustrate our approach, we begin by investigating whether or not a correlation exists between environmental variability and the returnability-equilibrium constant as defined above. To do this, we analyse metabolic networks for some 116 bacterial species, each of which can be categorised according to their natural environment (see Table 1 for details). The organisms abide in a variety of habitats, ranging from highly specialised environments, with little, if any, contact with the outside world – symbiotic bacteria such as *buchnera* for example – to those living in extremely heterogeneous media such as soil, and as such, have evolved under very different selective pressures[2].
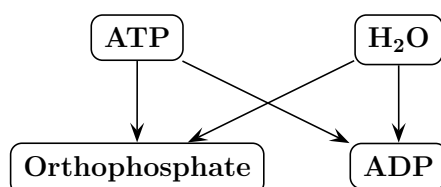
---

[2] The taxonomy used to rank environmental variability is based upon the NCBI classification for bacterial lifestyle (see [20] and references therein for further details).

| Network Statistic | Nodes | | | Edges | | |
|---|---|---|---|---|---|---|
| | min | median | max | min | median | max |
| Obligate (35) | 21 | 89 | 280 | 23 | 102 | 322 |
| Specialised (5) | 206 | 235 | 299 | 222 | 258 | 336 |
| Aquatic (4) | 224 | 264 | 386 | 251 | 290 | 440 |
| Facultative (41) | 40 | 333 | 473 | 42 | 385 | 574 |
| Multiple (28) | 193 | 454 | 360 | 214 | 431 | 553 |
| Terrestrial (3) | 282 | 385 | 421 | 329 | 463 | 498 |
| Total (116) | 21 | 282 | 473 | 23 | 322 | 574 |

**Table 1** Network statistics for the reaction graphs of the bacterial species studied in this work grouped according to the NCBI classification scheme. According to the NCBI, obligate bacteria have the most constant environment, followed by specialised and aquatic bacteria, and then facultative, multiple and terrestrial bacteria in that order (see [20] and references therein for further details).

$$\boxed{\text{ATP}} \quad + \quad \boxed{\text{H}_2\text{O}} \quad \longrightarrow \quad \boxed{\text{Orthophosphate}} \quad + \quad \boxed{\text{ADP}}$$
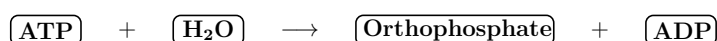
(a) **Hypothetical reaction.**



(b) **Corresponding substrate-product graph.**

**Fig. 2** Illustration of a hypothetical reaction and the corresponding representation as a substrate-product graph.

These metabolic networks are the same as those studied in [20] which were constructed using the KEGG database [13]. Note that, there are a number of ways of representing the set of metabolic reactions of an organism as a simple graph [12], the most popular of which is a *substrate-product graph* whereby each node corresponds to a metabolite, and a connection is made between reacting substances (substrates) and the products of the reaction. One potential caveat of such an approach, is that it can lead to the erroneous identification of metabolic pathways. Consider, for example, the hypothetical reaction given in Figure 2a. From the corresponding graphical representation (Figure 2b) one may identify a path of length 1 leading from ATP to Orthophosphate; however, any chemist would immediately object, noting that such a reaction is not possible as ATP must react with $H_2O$ in order to produce Orphophosphate.

Much of the criticism of network based studies of metabolism to date are the result of such misrepresentations of the underlying biology [2, 18]. We note that a correct treatment that avoids such issues is provided via directed hypergraphs [14,
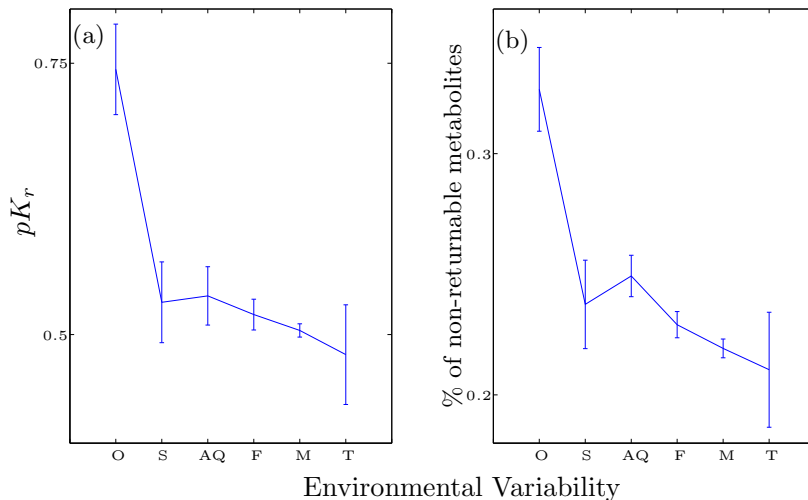
**Fig. 3** Relation between environmental variability and (a) the mean thermodynamic equilibrium constant $\langle pK_r(\beta) \rangle$ ($\beta = 0.25$), (b) the mean number of non-returnable metabolites, i.e. those metabolites with zero local returnability $K_r(0.25, i)$. In both cases, the six bacterial habitats under investigation are ordered along the *x*-axis in accordance with the predicted levels of variability provided by [20]: **O**bligate, **S**pecialised, **AQ**uatic, **F**acultative, **M**ultiple and **T**errestrial. Error bars represent standard errors.

24]. However, since we are not performing a path analysis and in order to allow for ease of comparison with related studies [20, 22, 25], we consider the substrate-product representation of metabolism in all of our experiments. Note also, that we remove so-called currency metabolites, e.g., ATP, NADH and $H_2O$, as they tend not to be involved in higher order functions. Additionally, we simplify our analysis by considering only the giant connected component of each reaction graph.

We calculate the mean thermodynamic index for 116 of the metabolic networks studied by Parter et al. [20]. In all calculations the inverse temperature was fixed to $\beta = 0.25$. Ideally, $\beta$ should be calibrated for each metabolic network according to individual environmental pressures. For example, metabolic networks functioning in extreme conditions should exhibit a value of $\beta \approx 0$ (high temperature) reflecting the increased environmental stresses endured by such systems. However, as the exact conditions for fitting the parameter $\beta$ are unknown, we assume an average value for all environments, selected as that value for which the different bacterial habitats are best discriminated. Figure 3(a) shows a plot of the mean thermodynamic index $\langle pK_r \rangle$ versus environmental variability for the different bacterial species. As can be clearly seen, the highest $\langle pK_r \rangle$ value was obtained for those bacteria in the obligate class, then came the specialised and aquatic bacteria, followed by a steady decrease for facultative, multiple and terrestrial classes. The group differences observed in Figure 3(a) were found to be significant (P-value: $p < 10^{-5}$) according to the Kruskal-Wallis (KW) test.

Additionally, we found that networks from those species residing in more constant environments contained significantly (P-value: $p < 10^{-5}$ using the KW test) more non-returnable nodes, i.e. nodes such that $K_r(\beta, i) = 0$. Figure 3(b) shows the

fraction of non-returnable metabolites versus environmental variability. Note that we can distinguish two subgroups of particular significance within the set of non-returnable metabolites of a reaction graph: (i) the so-called external metabolites, i.e. substrates that are not produced internally by the reaction network; and (ii) those product metabolites that are not fed back into the system. Such metabolites are often referred to respectively as *sources* and *sinks*. Clearly, we would expect to see a reduction both in the number of external metabolites and the number of products not being consumed via feedback mechanisms as environmental variability increases.

We remark that this idea relates to previous studies that have found a correlation between genome length and variations in bacterial lifestyle [20, 23], in that, constant environments that provide a steady supply of external metabolites lead to redundancy amongst certain genes, which, in turn, leads to genome reduction during the evolutionary process. Further evidence supporting this view was recently provided by Zhou and Nakhleh [25], who found that microbes inhabiting varied, heterogeneous environments displayed a larger metabolome than those leading a more specialised lifestyle. For example, Párez-Brocal et al. [19] have shown that the obligate symbiont *Buchnera aphidicola* displays a significantly reduced genome, some 200 kilobases, when compared against previously sequenced strains.

Figure 4 provides an illustration of the above in the case of *Buchnera*. Sink and source nodes are highlighted using red triangles and blue squares respectively (colour on-line); whilst nodes with non-zero local returnability are scaled proportionately. The first interesting observation is the high number of non-returnable nodes, approximately 30% of all metabolites. However, perhaps more interesting is the fact that some 65% of these metabolites are obtained from the environment, via *Buchnera-Aphid* symbiosis in this case, thus highlighting the central role played by the host in maintaining functional integrity within these systems.

Note, that in this work we follow the same environmental scheme as the one proposed by Parter et al. [20]. However, when considering our results for specialized and aquatic bacteria we want to make a call for prudence concerning this particular classification scheme. Whilst it is true that some specialised bacteria live in more restricted environments than aquatic ones, Figure 3 deals with the average thermodynamic equilibrium constant across both specialised and aquatic bacteria. It can clearly be seen that specialized bacteria display a significantly larger variance in their values of $pK_r$ than aquatic ones. The reason for this apparent anomaly can be understood as follows. The specialized bacteria considered in this study are all so-called thermophiles, i.e. *A. aeolicus*, *C. tepidum*, *T. Tengcongensis*, *T. Elongatus* and *T. Maritima*. However, these bacteria vary significantly with respect to several other environmental factors. For instance, *A. aeolicus* requires oxygen to survive whilst *C. tepidum* is an anaerobic phototrophic bacterium. Habitats for this last bacterium are quite variable including anoxic and sulphide-rich waters, mud, sediments, microbial mats, and even microbial consortia. The range of temperatures in which these bacteria live also varies significantly, ranging from an optimum growth temperature of $55\,^{\circ}\mathrm{C}$ for *T. Elongatus* to water temperature of $80\,^{\circ}\mathrm{C}$ for *T. Maritima*. On the other hand, the aquatic bacteria studied here (*C. crescentus*, *Prochlorococcus*, *Synechococcus* and *Synechocystis* sp.) live in globally less variable environments. The main difference being that *C. crescentus* and *Synechocystis* sp. live in fresh water environments while the others live in
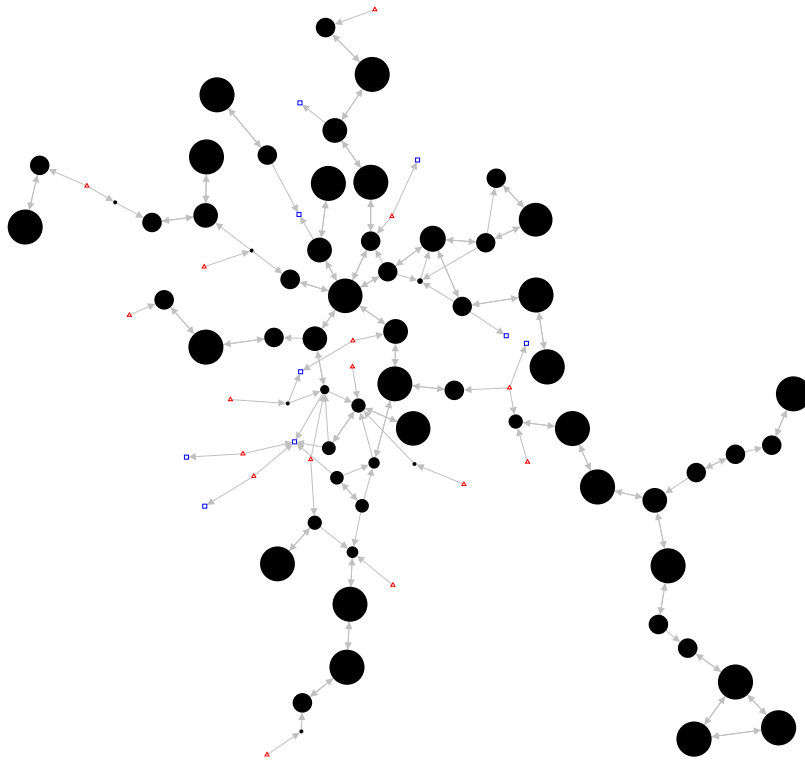
**Fig. 4** Visualisation of the metabolic network for the bacterium *Buchnera aphidicola*. Nodes are proportional to the local returnability metric given in (5) with $\beta = 0.25$. Source and sink nodes have been highlighted using red triangles and blue squares respectively (colour on-line).

marine ones. However, many species of *Synechococcus* have also been observed in freshwaters. In order to obtain a sense of the level of 'specialization' of some aquatic bacteria we mention further the fact that *Synechococcus* are also found in oceanic areas which are nutrient depleted, such as the central gyres. Altogether, these facts support our results that indicate (i) a lack of any significant differences between $pK_r$ values for specialised and aquatic groups; and (ii) a significantly increased variability in the values of $pK_r$ for specialised bacteria.

Next, we analyse the role of individual metabolites by considering the local returnability metric for the 116 metabolic networks studied above. The local returnabilty is given by Equation (5) and represents a new type of directed centrality measure. First, we analyse the extent to which the information contained in this new descriptor is reproduced by other well-known centrality measures. In Figure 5 we plot the Spearman rank correlation coefficients between the local returnability and the betweenness, eigenvector, in- and out-degree centralities. In all cases we have considered only directed versions of these network measures. As can be clearly seen, the correlations are low indicating that the information contained in the local returnability metric is not reproduced by any of the other centrality measures. In particular,
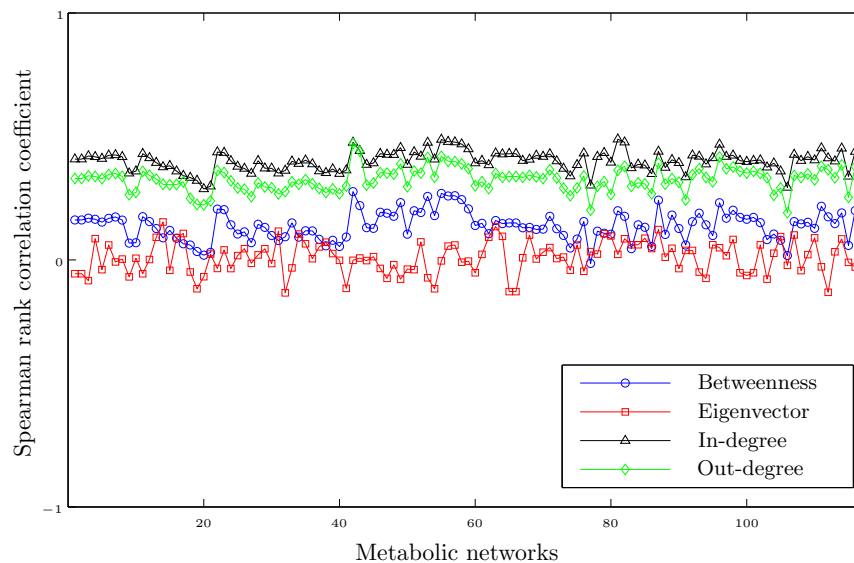
**Fig. 5** Spearman rank correlation between the local returnability metric $K_r(\beta, i)$, with $\beta = 0.25$, and the centrality measures of betweenness, eigenvector, in-degree and out-degree for 116 bacterial metabolic networks.

the eigenvector centrality shows extremely poor correlations, indicating that the local returnability provides a unique spectral measure of centrality.

In Figure 6 we plot the local returnability profiles for illustrative bacterial networks from the six environmental lifestyles. Only metabolites that occur across all six organisms (68 in total) are displayed; in all cases, metabolites are ordered according to the ranking provided by *Buchnera*. As can be seen there are a number of universal characteristics shared by these metabolites. Firstly, a small number of metabolites are found to be highly returnable across all six networks, e.g. L-tryptophan, indole and indoleglycerol phosphate. Furthermore, a significant overlap, approximately 60%, was found between non-returnable metabolites across the six featured networks. In addition to these similarities, a number of substantial differences are readily observed through the plots in Figure 6. Perhaps the most striking of which, is the large number of non-returnable metabolites (24 of a possible 68) displayed by the obligate bacterium *Buchnera*. Overall, we see that despite their differences, dramatic changes in the profiles of these common metabolites are not present. This can be understood due to the existence of a certain core of common, or universal, metabolites that have evolved in a 'standard way' across the different bacteria. Similarity between returnability scores across this common core can be easily calculated using the Pearson correlation coefficient, and we have observed (data not shown) that these correlations are always greater than 0.87, and in some cases as high as 0.98, clearly supporting our hypothesis that variability in returnability across these core metabolites is significantly reduced. Thus, the main differences in local and global variability can be attributed to a smaller set of metabolites that are more or less specific to the different kinds of bacteria (according to their environments). Such variations in the local re-
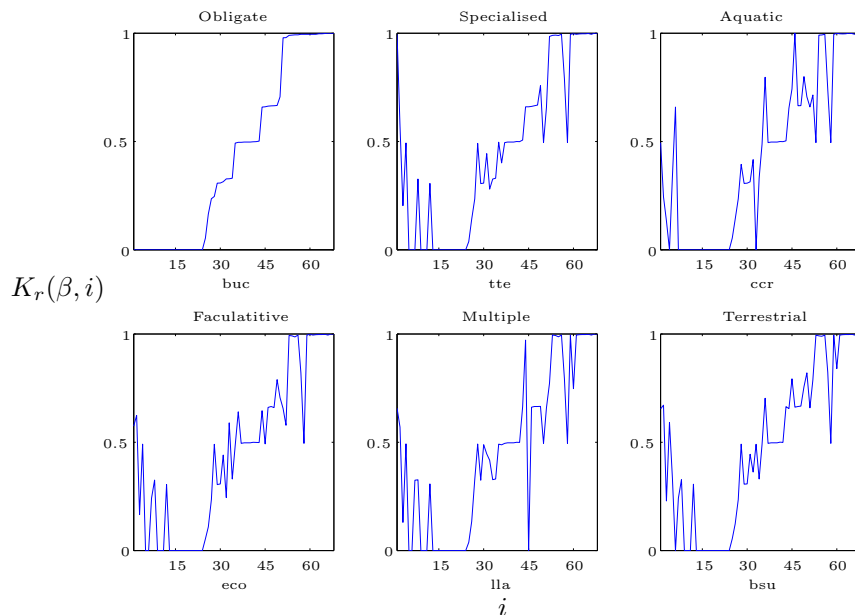
$$K_r(\beta, i)$$



**Fig. 6** Local returnability profile for networks from each of the six bacterial lifestyles. Values of $K_r(0.25, i)$ are given for those metabolites that occur in each of the networks (68 in total): *B. aphidicola* (buc); *T. tengcongensis* (tte); *C. crescentus* (ccr); *E. coli* (eco); *L. lactis* (lla) and *B. subtilis* (bsu).

turnability are well described by the values of $K_r(\beta, i)$ which can be further used as an indicator of adaptability and evolution for specific metabolites.

## 5 Conclusions

In this paper, we considered the difference in free energies, between a chemical reaction network and its 'hypothetical' equilibrium, to be a proxy for the extent to which a metabolic system is in disequilibrium. In particular, we found that on average 'global reversibility' increases significantly in line with environmental variability, supporting the view that organism adaptability leads to increased complexities in the resultant metabolic networks. In addition, we have proposed a new, directed centrality measure for characterising nodes (metabolites) in terms of the returnable pathways they participate. For the 116 metabolic networks studied here, the new measure does not show strong correlations with other directed centrality measures, providing a distinctly different ranking of metabolites. Moreover, we used the aforementioned local returnability metric to analyse the role of individual metabolites for illustrative networks across the six different lifestyles studied. We found that common metabolites across these networks were extremely highly correlated, indicating that differences displayed in local and global returnability may be attributed to a small set of environment specific metabolites.

# References

1. Arita, M.: The metabolic world of eschericjia coli is not small world. Proceedings of the National Academy of Sciences of the United States of America **101**, 1543–1547 (2004)
2. Bourguignon, P.Y., Helden, J., Ouzounis, C., Schchter, V.: Computational analysis of metabolic networks. In: D. Frishman, A. Valencia (eds.) Modern Genome Annotation, pp. 329–351. Springer Vienna (2008)
3. Buchanan, M., Caldarelli, G., De Los Rios, P., Rao, F., Vendruscolo, M.: Networks in Cell Biology. Cambridge University Press (2010)
4. Cornish-Bowden, A., Cárdenas, M.L.: Irreversible reactions in metabolic simulations: how reversible is irreversible? In: J.H.S. Hofmeyr, J.M. Rohwe, J.L. Snoep (eds.) Animating the Cellular Map, pp. 65–71. Stellenbosch University Press, South Africa (2000)
5. Cornish-Bowden, A., Cárdenas, M.L.: Information transfer in metabolic pathways. Effects of irreversible steps in computer models. European Journal of Biochemistry **268**(24), 6616–6624 (2001)
6. Estrada, E.: The Structure of Complex Networks. Oxford University Press (2011)
7. Estrada, E., Hatano, N.: Statistical-mechanical approach to subgraph centrality in complex networks. Chemical Physics Letters **439**(1-3), 247–251 (2007)
8. Estrada, E., Hatano, N.: Returnability in complex directed networks (digraphs). Linear Algebra and Its Applications **430**(8-9), 1886–1896 (2009)
9. Estrada, E., Rodriguez-Veláquez, J.A.: Subgraph centrality in complex networks. Physical Review E **71**, 056,103 (2005)
10. Garlaschelli, D., Loffredo, M.: Patterns of link reciprocity in directed networks. Physical Review Letters **93**(26), 268,701 (2004)
11. Guimerá, R., Medina, M.A., Nunes Amaral, L.A.: Functional cartography of complex metabolic networks. Nature **433**(7028), 895–900 (2005)
12. Holme, P.: Model validation of simple-graph representations of metabolism. Journal of the Royal Society Interface **6**(40), 1027–1034 (2009)
13. Kanehisa, M., Goto, S.: Kegg: Kyoto encyclopedia of genes and genomes. Nucleic acids research **28**(1), 27–30 (2000)
14. Klamt, S., Haus, U., Theis, F.: Hypergraphs and cellular networks. PLoS Computational Biology **5**(5), e1000,385 (2009)
15. Kreimer, A., Borenstein, E., Gophna, U., Ruppin, E.: The evolution of modularity in bacterial metabolic networks. Proceedings of the National Academy of Sciences **105**(19), 6976–6981 (2008)
16. Lacroix, V., Cottret, L., Thebault, P., Sagot, M.F.: An introduction to metabolic networks and their structural analysis. Computational Biology and Bioinformatics, IEEE/ACM Transactions on **5**(4), 594–617 (2008). DOI 10.1109/TCBB.2008.79
17. Ma, H.W., Zeng, A.P.: Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics **19**(2), 270–277 (2003)
18. Montañez, R., Medina, M.A., Solé, R.V., Rodríguez-Caso, C.: When metabolism meets topology: Reconciling metabolite and reaction networks. BioEssays **32**(3), 246–256 (2010)
19. Párez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., Latorre, A.: A small microbial genome: The end of a long symbiotic relationship? Science **314**(5797), 312–313 (2006)
20. Parter, M., Kashtan, N., Alon, U.: Environmental variability and modularity of bacterial metabolic networks. BMC Evolutionary Biology **7**(169) (2007)
21. De la Peña, J., Gutman, I., Rada, J.: Estimating the Estrada index. Linear Algebra and its Applications **427**(1), 70–76 (2007)
22. Takemoto, K., Borjigin, S.: Metabolic network modularity in archaea depends on growth conditions. PloS One **6**(10), e25,874 (2011)
23. Xu, J.: Invited review: Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Molecular Ecology **15**(7), 1713–1731 (2006)
24. Zhou, W., Nakhleh, L.: Properties of metabolic graphs: biological organization or representation artifacts. BMC Bioinformatics **12**(132) (2011)
25. Zhou, W., Nakhleh, L.: Convergent evolution of modularity in metabolic networks through different community structures. BMC Evolutionary Biology **12**(181) (2012)