Complexity and robustness in hypernetwork models of metabolism

Nicole Pearcy, Nadia Chuzhanova, Jonathan J Crofts*

School of Science and Technology, Department of Physics and Mathematics, Nottingham Trent University, Nottingham, NG11 8NS, UK

Abstract

Metabolic reaction data is commonly modelled using a complex network approach, whereby nodes represent the chemical species present within the organism of interest, and connections are formed between those nodes participating in the same chemical reaction. Unfortunately, such an approach provides an inadequate description of the metabolic process in general, as a typical chemical reaction will involve more than two nodes, thus risking over-simplification of the the system of interest in a potentially significant way. In this paper, we employ a complex hypernetwork formalism to investigate the robustness of bacterial metabolic hypernetworks by extending the concept of a percolation process to hypernetworks. Importantly, this provides a novel method for determining the robustness of these systems and thus for quantifying their resilience to random attacks/errors. Moreover, we performed a site percolation analysis on a large cohort of bacterial metabolic networks and found that hypernetworks that evolved in more variable environments displayed increased levels of robustness and topological complexity.

Keywords: Complexity, Hypernetworks, Metabolism, Evolution

1. Introduction

Many biological systems can be described in terms of their interaction patterns (Buchanan, 2010) and thus are naturally modelled using the tools of network science (Newman, 2010; Estrada, 2011). Such models typically

---

*Corresponding author: Tel: +44(0)115 8488419
Email address: jonathan.crofts@ntu.ac.uk (Jonathan J Crofts)

take the form of a simple or directed graph, in which the vertices represent the different components of the system under investigation, and a link is formed between vertex pairs if they interact in some way. However, some biological processes consist of more complex relations involving more than two interacting components, in which case, such an approach is likely to provide an inadequate description of the underlying biology (Klamt et al., 2009; Montañez et al., 2010). Metabolic networks, for example, comprise many different biochemical reactions, each involving multiple substrates and products, and thus capturing precisely the full complexity of these metabolic transformations requires a more general framework.

Complex hypernetworks provide an attractive alternative since they allow for more general interactions consisting of multiple nodes (Johnson, 2013). For metabolic modelling, such a formalism provides a natural setting in which metabolites are modelled as vertices and chemical reactions as hyperedges (Klamt et al., 2009). One caveat of such an approach, however, is the relative paucity of available hypernetwork measures with which to investigate these more complicated objects. Recently, a handful of studies have proposed definitions for the hypernetwork analogue of some of the most popular complex network measures, such as the degree-distribution (Latapy et al., 2008), clustering coefficient (Estrada and Rodriguez-Velazquez, 2006; Zhou and Nakhleh, 2011; Gallagher and Goldberg, 2013) and measures of centrality (Estrada and Rodriguez-Velazquez, 2006; Pearcy et al., 2014). Unfortunately, these measures are often accompanied by increased algorithmic complexities, or are not well-posed, in the sense that a variety of different author-dependent definitions exist. In addition to network measures, other notable attempts to apply complex network reasoning to hypernetworks include extensions of popular network models such as Erdös-Rényi and Barabási-Albert (Guillaume and Latapy, 2004; Wang et al., 2010); the use of random walks to infer information flow and network architecture (Bellaachia and Al-Dhelaan, 2013; Ducournau and Bretto, 2014); and novel community detection algorithms for determining modular hypernetwork structure (Vazquez, 2009; Michoel and Nachtergaele, 2012).

In this paper, we adapt widely studied percolation-based approaches (Bollobás and Riordan, 2006) in order to probe complex metabolic hypernetwork topology, and to quantify the robustness and fragility of these systems. Historically, a number of studies have investigated the resilience of metabolic networks to random mutations (and targeted attacks), typically by measuring the effect on network connectedness of the random, or targeted, removal

of nodes or edges (Wilhelm et al., 2004; Smart et al., 2008; Iyer et al., 2013). The critical point at which the network breaks down into isolated components, and hence can no longer function, is regarded as a proxy measure for network robustness (Callaway et al., 2000; Karrer et al., 2014). Here, we extend the notion of *site percolation* to the case of undirected hypernetworks, and use these techniques to investigate relations between hypernetwork topology and environmental variability for a large cohort of bacterial species. We find that species inhabiting more varied environments are more robust, in the sense that the transition between the non-percolating and percolating regimes occurs faster. Additionally, we find increased complexities in the hypernetwork topology of those organisms inhabiting harsher environments when compared against 'equivalent' random surrogates.

The remainder of the paper is organized as follows. In §2 we begin by describing the metabolic data used in this study, and provide a brief description detailing the construction of the metabolic hypernetworks. We then introduce the required theoretical prerequisites regarding hypernetworks, and provide algorithmic details outlining the extension of the site percolation approach to hypernetworks. In §3 we present and discuss the results of applying the new approach to investigate a cohort of some 115 bacterial species, each of which can be classified according to the variability within their natural habitats, as well as a more detailed analysis on two well-studied model organisms in *E. coli* and *B. aphidicola*. We conclude in §4 by briefly summarising our results and highlighting possible future directions of study.

2. Methods

2.1. Description of the datasets

The metabolic data employed in this study was acquired from the KEGG database on 12th October 2013 (Kaneshisa, 2008). The organism specific reaction lists were derived using the reaction.lst file from the KEGG ftp site, which include full chemical reaction equations, including both stoichiometric coefficients and currency metabolites. Note that we retain these high-degree nodes (e.g. H_2O, ATP, NADH, etc.) in our investigations since their role reflects more closely the biology when considered from a hypernetwork point-of-view. More specifically, KEGG XML files were used to extract the set of reactions for an organism, and then reaction.lst was used to obtain the full chemical equations for these reactions. These reaction lists are described

Table 1: Network statistics for the reaction graphs of the 115 bacterial species studied in this work classified according to environmental variability. According to the NCBI, obligate bacteria have the most constant environment, followed by specialised and aquatic, and then facultative, multiple and terrestrial bacteria in that order. In the first column, numbers in brackets denote the number of networks in each class.

Environment	Nodes			Hyperedges		
	min	median	max	min	median	max
Obligate (34)	224	441	979	143	337	883
Specialised (5)	643	695	743	554	627	651
Aquatic (4)	754	851	1014	645	747	896
Facultative (41)	244	947	1308	153	883	1199
Multiple (28)	631	900	1226	545	821	1143
Terrestrial (3)	890	942	955	832	912	936
Total (115)	224	748	1308	143	695	1199

with reaction IDs, metabolic map IDs and the chemical equation whose compounds are represented as the KEGG compound IDs. The following is an example from Butanoate metabolism:

R00212: 00620: C00024 + C00058 <=> C000010 + C00022

Note that these reaction lists have been derived by curating several chemical pathway maps from the KEGG database. Thus, since a reaction may be present within multiple metabolic maps there exist some reactions that are repeated within the list. Any repeat reactions with the same reaction ID are thus removed from the reaction list. However, due to errors within the KEGG database some of these repeats are not identical. This is due to the fact that chemical equations in different chemical pathways maps are catalysed by the same enzyme and thus have the same reaction ID, yet sometimes the reactions differ. These non-trivial cases were treated by taking the most comprehensive equation.

We constructed metabolic hypernetworks for 115 bacterial species (see Table 1 for an overview of their network properties) each of which can be classified according to the variability in their natural habitat using the NCBI classification for bacterial lifestyle (Entrez-Genome-Project, 2015). The classification includes six classes: Obligate bacteria that are obligately associated with a host, either intracellulary or extracellulary. Specialized bacteria that

live in specialized environments such as marine thermal vents. Aquatic bacteria, that live in fresh or seawater environment, and are not associated with hosts. Facultative bacteria, free living bacteria such as *E. coli* that often associate with a host. Multiple bacteria, that live in multiple different kinds of environments such as bacteria with a wide host range, and Terrestrial Bacteria, that live in the soil.

2.2. Hypernetwork preliminaries

A complex hypernetwork can be described by a pair of objects $H = (V, E)$, where $V = \{v_1, v_2, ..., v_n\}$ is a set of n vertices and $E = \{E_1, ..., E_m\}$ the corresponding edge set. Each hyperedge consists of subsets of V, such that $\bigcup_i E_i = V$ and $E_i \neq \emptyset$. We say that two vertices v_i and v_j are *adjacent* if they are contained within the same hyperedge, *i.e.* $v_i, v_j \in E_k$.

A hypernetwork can be represented by a variety of different matrices (Gallo et al., 1993), the most popular of which is the *incidence matrix*, an $n \times m$ matrix $C(H)$ representing the relationships between the n nodes and m hyperedges. The entries of the matrix $C(H)$ are given by

$$C_{ij} = \left\{ \begin{array}{ll} 1, & \text{if } v_i \in E_j, \\ 0, & \text{otherwise,} \end{array} \right.$$

that is, C_{ij} equals 1 if v_i belongs to the jth hyperedge. Importantly, given the incidence matrix it is straightforward to compute the *adjacency matrix* for a hypernetwork as follows

$$A(H) = C(H)C(H)^T.$$

The ijth entry of the adjacency matrix, $A(H)$, is given by the cardinality of the set of hyperedges containing both nodes i and j.

2.3. Percolation in hypernetworks

In a standard site percolation process network nodes are referred to as *sites* and they can exist in one of two states: *active* or *inactive*. Such a process starts from an initial formation in which all states are inactive, sites are then turned on at random, and networks edges added whenever two adjacent nodes become active; this process is continued until the system achieves full activation. In this way we can observe the formation of so-called percolation clusters which form as the proportion, p, of active network nodes

increases. In practice a *percolation threshold*, $p = p_c$, exists at which point a phase transition occurs and the system goes from a *non-percolating phase*, containing lots of small microscopic clusters, to a *percolating phase* in which a single dominant cluster, comparable to system size n, forms. Importantly, the critical point, $p = p_c$, separating the two different phases can be considered a proxy for network robustness (Callaway et al., 2000; Karrer et al., 2014) – the birth of a single dominant cluster is an indication that the metabolic hypernetwork is capable of performing its intended function.

Note that in the network science literature it is common to consider the reverse process to that described above, that of so-called *inverse percolation* (Newman, 2010; Barabási, 2016), in which one measures the impact on network integrity of the removal of a fraction, f say, of nodes. The point at which the network disintegrates into isolated components, *i.e.* $f = f_c$, is then considered to be a measure of network robustness. Importantly, just as in the case of standard networks, these two processes can be shown to be equivalent, being related by the expression $p_c = 1 - f_c$.

The key difference between our approach and the standard one (described above) lies in the criteria by which network nodes/edges are added. In the standard approach nodes are activated at random and edges placed between activated, adjacent node pairs. In the case of hypernetworks, we impose the more stringent requirement that all nodes within a hyperedge must be activated before any links are added. Or in terms of metabolism, all substrates and products of a reaction must be present before a reaction occurs. Below we provide algorithmic details for site percolation in an undirected hypernetwork; see Figure 1 for a graphical illustration of our approach.

1. Starting from an empty hypernetwork in which all nodes are inactive (i.e. $p = 0$), set S, the relative size of the giant connected component (GCC), equal to 0.
2. Activate a randomly chosen node, i say.
3. Loop through all hyperedges containing node i adding those hyperedges for which all nodes are active.
4. Compute the relative size of the GCC, S.
5. Repeat steps 2 to 4 until all sites are activated (i.e. $p = 1$, $S = 1$).

The above steps constitute a single realisation of our algorithm. To generate a statistically reliable estimate of $S(p)$ we repeat the process a large number, M say, of times.

Note that we use an adaptation of the Newman-Ziff algorithm (Newman and Ziff, 2000) in all the computations we perform which is significantly faster than the usual breadth-first search, and that the GCC can be computed using standard network algorithms applied to the adjacency matrix of the hypernetwork. To determine the percolation threshold we employ the network *susceptibility* function as defined in (Radicchi, 2015), which is given by

$$\chi = \frac{\langle S^2 \rangle - \langle S \rangle^2}{\langle S \rangle}. \tag{1}$$

The peak of the susceptibility function, $\chi_{\max}$, is indicative that the network is undergoing a phase transition, and so importantly the value of p that corresponds to $\chi_{\max}$ provides an estimate of the percolation threshold, p_c. Here, we consider these two values to be equivalent in the sense that we refer to the point at which the susceptibility is maximised as the percolation threshold.

2.4. Null model comparison

We identify organisational effects on metabolic robustness by comparing metabolic hypernetworks to appropriate null models. The null models employed are constructed using a randomisation procedure forwarded by Zhou and Nakhleh (2011) which uses a rewiring algorithm to generate 'equivalent' random hypernetworks. In this algorithm the hyperedges of a metabolic hypernetwork are randomly rewired, whilst preserving both the size of the hypernetwork (i.e. number of nodes and hyperedges) and the hyperedge degree distribution. In all of our experiments random surrogates were constructed by applying 20,000 hyperedge swaps, and network statistics were computed using ensemble averages over some 100 realisations. Note, that the randomisation procedure described above can result in networks that are disconnected, and in this case we restrict to the giant connected component; in practice this can lead to a small discrepancy (at most 2-3% in our experiments) in network size between metabolic hypernetworks and their random surrogates, however, the effects on the results presented here are expected to be minimal.

3. Results

In this section we perform a site percolation analysis in order to investigate the hypernetwork topology of two model organisms: *E. coli* and *B.*

aphidicola. We then extend this analysis to a large cohort of some 115 bacterial species for which detailed knowledge concerning the environments in which they evolved is available.

3.1. Two model organisms

The first experiment considers the effects of a site percolation on two well-studied organisms:

(i) *E. coli*: a facultative bacteria that is free-living within a variety of different hosts, and thus its environment is relatively varied, requiring the system to be highly adaptable. The hypernetwork of *E. coli* consists of 1097 vertices (metabolites) and 1117 hyperedges (reactions).

(ii) *B. aphidicola*: a symbiotic bacteria that associates with one host, and thus lives within a very controlled environment. The hypernetwork of *B. aphidicola* consists of 444 nodes (metabolites) and 332 hyperedges (reactions).

Figure 2 plots both the mean size of the GCC, $\langle S \rangle$, and the susceptibility function, $\chi(p)$, versus the proportion of activated network nodes, p, for the two bacterial species. Perhaps the first point of note is the difference between the percolation thresholds of the two organisms: $p_c = 0.53$ for *E. coli* and $p_c = 0.61$ for *B. aphidicola*. Note that the earlier appearance of the GCC in *E. coli* as opposed to *B. aphidicola*, is suggestive of a more robust network structure in the sense that *E. coli* would seem to be less susceptible to random attacks or errors. This may be considered a consequence of the greater evolutionary pressures/competition associated with the more variable habitat that *E. coli* has evolved in, as compared to *B. aphidicola*.

Another interesting outcome of Figure 2, is that both organisms would appear to be less robust then there random counterparts, with the difference being slightly greater for *B. aphidicola*. Note that whilst this result might sound counter-intuitive, suggesting as it does, that these organisms are less resilient to node failure than 'matching' random graphs, similar results have been reported for standard networks. One possible explanation might be due to the increased clustering observed in metabolic hypernetworks (data not shown), as compared to their random surrogates, which, when coupled with high levels of heterogeneity, has been found to induce a so-called *core-periphery structure* (Holme, 2005). In such a structure, the network is organised into a highly connected core, whilst the remainder of

the nodes form the periphery. A high amount of clustering within the network leads to a large highly entangled core that is very difficult to break down, and therefore decreases the percolation threshold (see, for example, Kiss and Green (2008); Serrano and Boguná (2006); Colomer-de Simón and Boguñá (2014) and references therein). High clustering within the periphery of the network, however, leads to small sparsely interconnected cliques, which are very fragile to random mutations, and thus increases the percolation threshold (Colomer-de Simón and Boguñá, 2014). The balance between these two effects can significantly alter the position of the percolation threshold. Importantly, a number of recent studies have suggested that metabolic networks are organised into a highly modular core-periphery type structure, such that the core module connects the central metabolites and carries out basic metabolic functions, whilst periphery modules perform highly specific functions with minimum interactions with other modules (Zhao et al., 2007; Rosa da Silva et al., 2008).

3.2. Cohort study

Further evidence supporting the idea that network resilience, as measured using percolation thresholds, is correlated with the variability within an organisms environment is provided by Figure 3. Here, we partition the bacterial species into the 6 different environmental classes (see Table 1): obligate, specialised, aquatic, facultative, multiple and terrestrial. We then computed the mean percolation threshold, $\langle p_c \rangle$, for each class and plotted it against increasing environmental variability. As can be readily seen, we find that the average percolation threshold decreases with increased variability, backing up our previous investigations of *E. coli* and *B. aphidicola*. More specifically, we find that the obligate class has a significantly larger percolation threshold than the other five classes, providing further evidence that host-associated bacteria are more vulnerable to random failures. This is perhaps a consequence of the symbiotic lifestyle of the obligate bacteria, where a metabolite-rich environment is provided by the host. Bacteria inhabiting such an environment, are believed to have experienced a genome reduction throughout evolution, such that only essential genes necessary for survival within the host are retained (Moran, 2002; McCutcheon and Moran, 2012). Thus any random error occurring within such a bacteria is likely to be highly detrimental to network functionality. The next interesting observation is that the percolation thresholds for the specialised and aquatic classes are almost identical. This is perhaps not too surprising, however, since these two

9

classes are often considered to be equivalent in terms of their environmental variability (Parter et al., 2007; Crofts and Estrada, 2014; Pearcy et al., 2015). Therefore, the bacteria from these two classes are likely to have a similar tolerance towards random errors, despite these bacteria being exposed to quite different conditions. We then observe a relatively large decrease in the percolation threshold for the facultative and multiple classes, suggesting a higher resilience to random failures. Again, this comes as no surprise, since the bacteria within these two classes live in a variety of different environments, and therefore are required to maintain functionality in conditions where metabolite availability is uncertain. Finally, the smallest percolation threshold is observed for the terrestrial class, as expected, due to the highly heterogeneous conditions that bacteria living in soil are exposed to.

Note that the group differences shown in Figure 3 are significant by the Kruskal-Wallis (KW) test (p-value $< 10^{-10}$). To control for the effect of network size, we also computed the correlation between the percolation threshold and environmental variability conditioned both on the size of the networks as well as the total number of reactions. Importantly, we found that the Spearman's partial correlation between average percolation thresholds and variability remained significant when accounting for both network size ($c = -0.36$; p-value $< 10^{-5}$) and the total number of reactions ($c = -0.32$; p-value $< 10^{-4}$).

4. Summary and outlook

In this work we have adapted widely used percolation techniques (i.e. the process by which nodes are randomly (de)activated within a network) to a hypernetwork formalism, as a method for quantifying metabolic hypernetwork robustness/vulnerability to random failures. The key difference between our approach and standard percolation analyses, lies in the criteria by which network nodes/edges are activated: a hyperedge only becomes activated in the network if all nodes (metabolites) involved in the hyperedge (reaction), are currently active (available). Using this new percolation-based approach, we have provided evidence for the pivotal role that environmental pressures have in shaping the biochemical reaction networks of bacteria. In addition, our investigation suggests that the previously reported core-periphery structure of metabolic networks remains evident when employing the more biologically plausible hypernetwork framework.

Note that our approach can be refined in a number of ways. From a biological point-of-view, for example, it would be interesting to further probe the effect of currency metabolites on our analysis, since recent studies (Zhou and Nakhleh, 2011) employing a hypernetwork formalism have been shown to be sensitive to their inclusion/exclusion. One possible way to isolate this effect, that does not simply discard such high-degree metabolites (which is clearly not a good long-term strategy, particularly given the lack of consensus for what actually constitutes a currency metabolite), is to consider node activation according to some probability distribution, as opposed to activating metabolites uniformly at random as in this study. Such a distribution could be theoretical, based on a structural feature of interest such as hypernetwork degree, or could reflect biological knowledge of metabolite abundance across the different environments. Note that these ideas have important implications for the core periphery structures alluded to in this study, since many chemical reactions consist of a mixture of both high and low degree nodes, and as such, potentially form a bridge between core metabolic processes and smaller subsystems found at the periphery. Further investigation into this, as well as more general structural properties, such as network density and clustering which have been shown to impact percolation thresholds for standard networks in a non-trivial manner (Newman, 2010; Serrano and Boguná, 2006), is an important area of future research. Finally, the extension of the percolation process forwarded here to consider the effects of more general cascading events will provide a better understanding of how hypernetwork topology influences the dynamic capabilities of metabolism.

Acknowledgement

References

Barabási, A., 2016. Network Science. Cambridge University Press.

Bellaachia, A., Al-Dhelaan, M., 2013. Random walks in hypergraph, in: Proc. of the 2013 International Conference on Applied Mathematics and Computational Methods, Venice, Italy. pp. 187–194.

Bollobás, B., Riordan, O., 2006. Percolation. Cambridge University Press.

Buchanan, M., 2010. Networks in Cell Biology. Cambridge University Press.

Callaway, D.S., Newman, M.E.J., Strogatz, S.H., Watts, D.J., 2000. Network robustness and fragility: percolation on random graphs. Physical Review Letters 85, 5468–5471.

Crofts, J.J., Estrada, E., 2014. A statistical mechanics description of environmental variablity in metabolic networks. Journal of Mathematical Chemistry 52, 675–688.

Ducournau, A., Bretto, A., 2014. Random walks in directed hypergraphs and application to semi-supervised image segmentation. Computer Vision and Image Understanding 120, 91–102.

Entrez-Genome-Project, 2015. `http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi`.

Estrada, E., 2011. The Structure of Complex Networks: Theory and Applications. Oxford University Press.

Estrada, E., Rodriguez-Velazquez, J.A., 2006. Subgraph centrality and clustering in complex hyper-networks. Physica A 364, 581–594.

Gallagher, S.R., Goldberg, D.S., 2013. Clustering coefficients in protein interaction hypernetworks, in: Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, New York, NY, USA. pp. 552–560.

Gallo, G., Longo, G., Pallotino, S., 1993. Directed hypergraphs and applications. Discrete Applied Mathematics 42.

Guillaume, J., Latapy, M., 2004. Bipartite structure of all complex networks. Information Processing Letters 90, 215–221.

Holme, P., 2005. Core-periphery organisation of complex networks. Physical Review E 72, 046111.

Iyer, S., Killingback, T., Sundaram, B., Wang, Z., 2013. Attack robustness and centrality of complex networks. PLoS ONE 8, e59613.

Johnson, J.H., 2013. Hypernetworks in the science of complex systems. Imperial College Press (London).

Kaneshisa, M., 2008. The KEGG Database. John Wiley and Sons, Ltd. pp. 91–103.

Karrer, B., Newman, M.E.J., Zdeborová, L., 2014. Percolation on sparse networks. Physical Review Letters 113, 208702.

Kiss, I.Z., Green, D.M., 2008. Comment on "properties of highly clustered networks". Physical Review E 78, 048101.

Klamt, S., Haus, U.U., Theis, F., 2009. Hypergraphs and cellular networks. PLoS Computational Biology 5, e1000385.

Latapy, M., Magnien, C., Vecchio, N.D., 2008. Basic notions for the analysis of large two-mode networks. Social Networks 30, 31–48.

McCutcheon, J.P., Moran, N.A., 2012. When metabolism meets topology: Reconciling metabolite and reaction networks. Nature Reviews Microbiology 10, 13–26.

Michoel, T., Nachtergaele, B., 2012. Alignment and integration of complex networks by hypergraph-based spectral clustering. Physical Review E 86, 056111.

Montañez, R., Medina, M.A., Solé, R.V., Rodrguez-Caso, C., 2010. When metabolism meets topology: Reconciling metabolite and reaction networks. BioEssays 32, 246–256.

Moran, N.A., 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108, 583–586.

Newman, M.E.J., 2010. Networks: An Introduction. Oxford University Press.

Newman, M.E.J., Ziff, R.M., 2000. Efficient Monte Carlo algorithm and high precision results for percolation. Physical Review Letters 85, 4104–4107.

Parter, M., Kashtan, N., Alon, U., 2007. Environmental variability and modularity of bacterial metabolic networks. BMC Evolutionary Biology 7, 169.

Pearcy, N., Chuzhanova, N., Crofts, J.J., 2015. Network motif frequency vectors reveal evolving metabolic network organisation. Molecular Biosystems 11, 77–85.

Pearcy, N., Crofts, J.J., Chuzhanova, N., 2014. Hypergraph models of metabolism. International Journal of Biological, Veterinary, Agricultural and Food Engineering 8, 732–736.

Radicchi, F., 2015. Predicting percolation thresholds in networks. Physical Review E 91, 010801.

Serrano, M.A., Boguná, M., 2006. Clustering in complex networks. ii. percolation properties. Physical Review E 74, 056115.

Rosa da Silva, M., Ma, H., Zeng, A.P., 2008. Centrality, network capacity, and modularity as parameters to analyse the core-periphery structure in metabolic networks. Proceedings of the IEEE 96, 1411–1420.

Colomer-de Simón, P., Boguñá, M., 2014. Double percolation phase transition in clustered complex networks. Physical Review X 4, 041020.

Smart, A.G., Amaral, L.A.N., Ottino, J.M., 2008. Cascading failiure and robustness in metabolic networks. Proceedings of the National Acadamies of Science USA 105, 13223–13228.

Vazquez, A., 2009. Finding hypergraph communities: a bayesian approach and variational solution. Journal of Statistical Mechanics: Theory and Experiment 77, P07006.

Wang, J.W., Ring, L.L., Deng, Q.H., Y., Z.J., 2010. Evolving hypernetwork model. The European Physical Journal B 77, 493–498.

Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. Systems Biology 1, 114–120.

Zhao, J., Ding, G.H., Tao, L., Yu, H., Yu, Z.H., Luo, J.H., Cao, Z.W., Li, Y.X., 2007. Modular co-evolution of metabolic networks. BMC Bioinformatics 8.
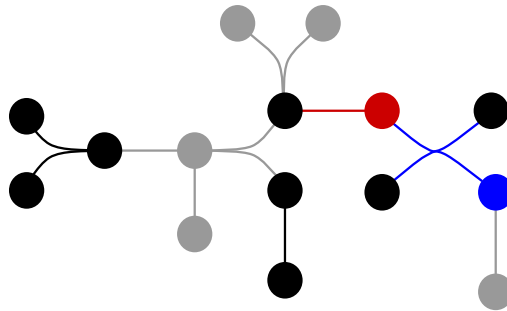
Zhou, W., Nakhleh, L., 2011. Properties of metabolic graphs: biological organization or representation artifacts? BMC Bioinformatics 12.
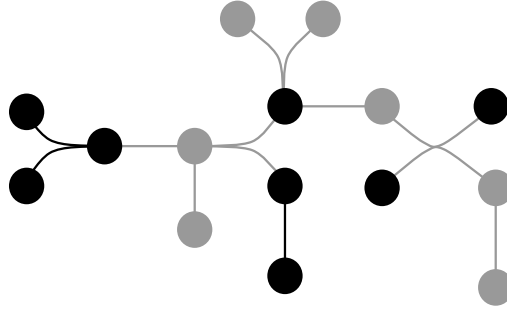
(a)
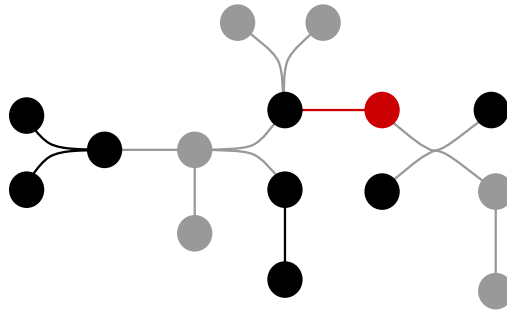
(b)

(c)

Figure 1: Figures (a)-(c) highlight two iterations of the site percolation process for a toy hypernetwork with $n = 15$ and $m = 9$ starting from a configuration with $p = 0.5$.
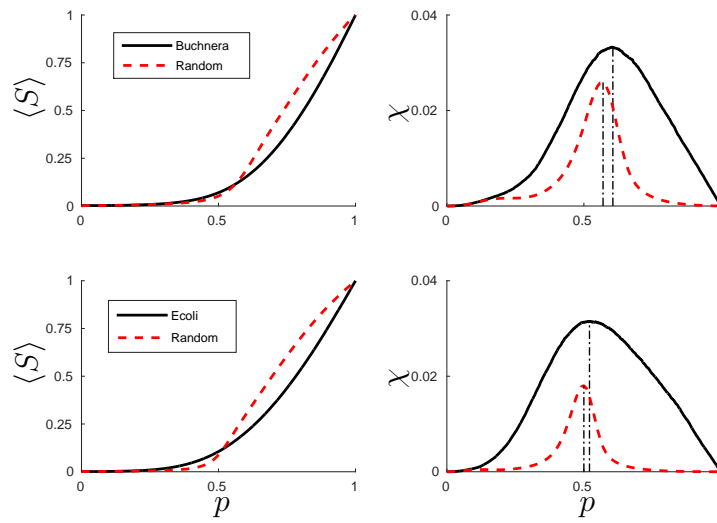
Figure 2: Site percolation for the two bacterial hypernetworks: *B. aphidicola* (top row) and *E. coli* (bottom row). Here, we compare the percolating properties S and χ of the original hypernetworks (black solid lines) against an ensemble of 100 rewired hypernetworks (red dashed lines). Note that the dashed lines indicate percolation thresholds of the corresponding hypernetwork: $p_c = 0.53$, for *E.coli* ($\langle p_c^{\mathrm{rand}} \rangle = 0.5 \pm 0.0026$) and $p_c = 0.61$ for *B. aphidicola* ($\langle p_c^{\mathrm{rand}} \rangle = 0.57 \pm 0.0048$).
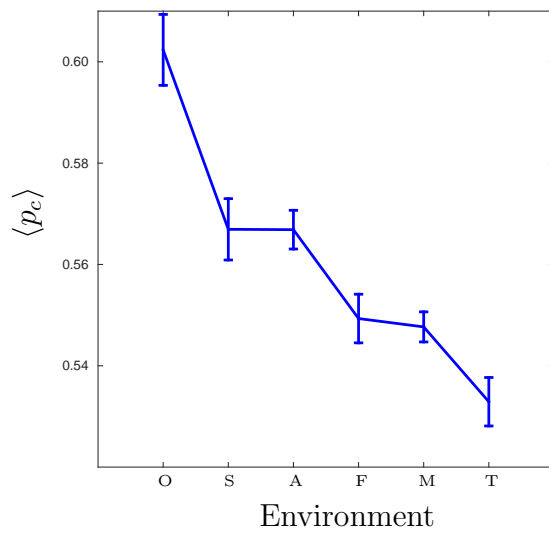
Figure 3: Relationship between the average percolation threshold, $\langle p_c \rangle$, and environmental variability. Note that the six bacterial habitats along the x-axis are in order of environmental variability: **O**bligate, **S**pecialised, **A**quatic, **F**acultative, **M**ultiple and **T**errestrial. The vertical bars represent the standard error of the mean.