Size Dependent Growth in Metabolic Networks

Henry Dorrian[1*], Kieran Smallbone[2†] and Jon Borresen[1‡]
[1] *Dept of Computing and Mathematics, Manchester Metropolitan University, Manchester UK M1 5GD*
[2] *School of Mathematics, The University of Manchester, Manchester UK M60 1QD*
(Dated: October 10, 2012)

Accurately determining and classifying the structure of complex networks is the focus of much current research. One class of network of particular interest are metabolic pathways, which have previously been studied from a graph theoretical viewpoint in a number of ways. Metabolic networks describe the chemical reactions within cells and are thus of prime importance from a biological perspective.

Here we analyse metabolic networks from a section of microorganisms, using a range of metrics and attempt to address anomalies between the observed metrics and current descriptions of the graphical structure. We propose that the growth of the network may in some way be regulated by network size and attempt to reproduce networks with similar metrics to the metabolic pathways using a generative approach. We provide some hypotheses as to why biological networks may evolve according to these model criteria.

PACS numbers: 89.75.Hc, 89.75.Fb, 84.35.Sn, 05.65.+b

The graphical structure of metabolic pathways has been extensively studied [1–4] and describing the structure should give insight into functionality [5]. Many of the salient features of such networks have been investigated, particularly any scale free nature [6] although the scale free model is currently the subject of some debate [7, 8]. Of particular interest are the high clustering coefficients observed in metabolic pathways which have previously been explained using concepts of topological hierarchy [9] and modularity [10]. Other concepts describing the structure such as memory [11] and decomposition into functional modules [12] have also been proposed.

A more recent model by Schneider et al. [13] uses network depletion, where a fully connected network is degraded according to the degree of the nodes. This gives rise to the high clustering coefficient and lower degree distribution values found in metabolic pathways, but the model cannot agree with how these biological networks would have evolved (i.e it is not plausible that metabolic networks have evolved from being fully connected and slowly losing connections as, for one thing, not all metabolites can react with one another).

Previous work by Dorogovtsev and Mendes [14] explains the presence of high degree decay rates in evolving networks with accelerating growth rates, where it is shown that for networks where the growth is accelerating and where the decay rate of the degree distribution $\gamma > 2$, the probability distribution for preferential attachment must be non-stationary.

It has also been shown that the probability of a metabolite reacting with k other metabolites decays as $P(k) \sim k^{-2.2}$ [2, 3]. However the metabolic pathways investigated here demonstrate $\gamma < 2.2$ in all cases. This leads to the biologically plausible idea that the growth of such networks may be slowing and that the probability of connection to existing nodes may not be static.

Initially, we conduct a graphical analysis of eight bacterial metabolic pathways, concentrating on their clustering coefficients and mean path lengths. Here microorganisms have been chosen, as the metabolic pathways are observed to be less modular than higher organisms allowing greater illustration of the concept. Further to this, a growth model, whereby the rate of growth decays as a function of network size, is used to demonstrate that size dependent growth may provide a suitable explanation as to many of the structural features of metabolic networks.

I. GRAPHICAL ANALYSIS

Eight microbial metabolic pathways were considered, they were as follows:
Escherichia Coli, Escherichia Coli iAF1260, Escherichia Coli iJR904, Helicobacter Pylori, Methanosarcina Barkeri, Staphylococcus Aureus, Mycobacterium Tuberculosis and Saccharomyces Cerevisiae.

Three formulations of E Coli are chosen to ensure the results are independent of the methodology used to initially determine the network.

The eight metabolic reconstructions are downloaded in SBML format [15] from the BiGG database [16]. The models were imported to Matlab using libSBML [17].

We consider each metabolic pathway as an undirected graph with the adjacency matrix of the graph being the boolean representation of the chemical interactions. For

*H.Dorrian@mmu.ac.uk
†Kieran.Smallbone@Manchester.ac.uk
‡J.Borresen@mmu.ac.uk

each model, nodes i and j are defined as adjacent if metabolite i appears as a reactant and j as a product, or indeed i as a product and j as a reactant in any reaction. Although this is a highly simplified representation this has been shown to be a useful tool in analysing such systems [18].

Note: There are a variety of network constructions available, both including and excluding subcellular compartmentalisation, and with/without considering the role of water and protons. We consider a formulation with subcellular compartmentalisation, as this most accurately represents the structural nature of the biochemical processes within the cells and excluding the presence of water and protons as is customary in such studies as these [1–4]. The results obtained here are applicable to the other available formulations with some small modification to the size dependent decay constant C described in Section I 4.

1. Clustering Coefficients

The clustering coefficient of a network is a measure of transitivity - how the nodes in the network tend to cluster together. We consider a network average global clustering coefficient [19] of the form.

$$< c >= \frac{1}{n}\sum_{i=1}^{n} c_i, \tag{1}$$

where n is the number of nodes in the network,

$$c_i = \frac{2e_i}{k_i(k_i-1)}, \tag{2}$$

where k_i is the degree of node i and e_n is the number of connected pairs between the nodes to which node i is connected.

The observed clustering coefficients $< c >$ for the metabolic data show that the networks are highly clustered and that the clustering coefficient is independent of network size (see figure 1).

2. Mean Minimum Path Lengths

The minimum path length (or geodesic length) is a measure of the smallest number of nodes between any two nodes, and represents the shortest route along the network between them. For our metabolic pathways, the average path length is surprisingly low given both the size of the networks and their average degree. This demonstrates a very strong small world effect [20]. Conversely, the highest of the minimum path lengths are somewhat greater than one would expect, given such a small world effect. For instance, the H. pylori metabolic network with 562 nodes has an average path length of 2.8 and a maximum path length of 8. For a non-directed network of this size, such a high maximum path length suggests structural qualities not in keeping with other small world networks.

3. Scale Free Structure

It has been previously observed that the graphical structures of metabolic pathways have much in common with many other complex networks, particularly with respects to their possible 'scale-free' nature [2, 21].

In a scale-free network, the probability, $P(k)$, of a node in the system having k connections follows a power law distribution of the form $P(k) \sim k^{-\gamma}$ [22]. This is in contrast to the much studied random graphs which follow a Poisson or Binomial distribution [23] but are similar to the social networks Milgram [20] described.

As can be observed (see figure 2), the metabolic networks display some scale-free property, however they cannot be considered truly scale free as there are too few connections with low degree and too many with high degree for an accurate fit of the form $P(k) \sim k^{-\gamma}$ to be valid. Approximations for the value γ (above) for the metabolic pathways via least squares fits, obtain values in all cases of $\gamma < 2$ - although such a logarithmic fit of the data yields high least square errors.

Networks which obey a power law distribution of the form $P(k) \sim k^{-\gamma}$ can be artificially generated. For instance Barabási and Albert [22] describe a network growing via *preferential attachment* of new nodes to existing nodes with a higher degree, the probability that a new node connects to an existing node is calculated using equation 3.

$$P(\textit{New node connects to node } i) = \frac{K_i}{\sum_j (K_j)} \tag{3}$$

where K_i is the degree of node i and the sum is over all pre-existing nodes j.

Artificially generated networks of this type typically have $\gamma \in [2, 3]$ - greater than the observed metabolic data. The clustering coefficients are also considerably lower than those observed for the metabolic networks (typically lower than 0.1 for any randomly generated Barabási Albert model (BA model)) and are observed to decrease with increasing network size.

4. A Size Dependent Generative Model

Although the BA model (equation 3) does not provide good fits to the metabolic networks, a generative model using preferential attachment would seem to have much to offer when considering the growth of metabolic networks. Here promiscuous metabolites within the network

with high degree ie those which are present as reactants in higher numbers of reactions will be those which are more likely to form new connections, whereas co-factors which are more specific in their function will form less connections.

As such, a generative model similar to the BA model appears as a strong candidate for describing some of the features which develop in the growth of metabolic networks. However, as a variety of studies have demonstrated [9, 13, 24], the scale-free model alone is not sufficient to fully describe the networks when a range of graph metrics are considered.

In considering a generative model approach to recreating networks similar to the metabolic networks of microorganisms, the concept of size dependent growth was considered. As such we attempt to introduce some concept of limiting factors on the generative model. Essentially we attempt to model the growth of the network such that, initially, it is very easy for new connections to form and as the network grows the probability of new metabolites attaching is reduced. We have chosen a linear model for simplicity.

Networks are grown according to the following:

$$P(\textit{New node connects to node } i) = \frac{K_i}{\sum_j (K_j)} \times \frac{C}{n} \quad (4)$$

where C is some constant and n is the number of nodes in the existing network. Such a model will produce a globally connected network for $n^2 + 1 \leq C$. For $n^2 + 1 > C$ the preferential attachment model begins, however, the probability of attachment is initially high. As the network grows and $\sum j \sim C$, new nodes attach in a manner identical to the original BA model and for $\sum j > C$ the network ceases to grow any further. (Note, this model assumes that nodes are not self connected and restricts the probability of any new attachment to $P \leq 1$, for $n^2 + 1 \leq C$).

The effect of modifying the BA model thusly has two effects: Firstly the probabilities of attachment are not static and are rescaled as each new node enters the network, secondly the probability that any new node will attach is decreasing as the network grows.

The size to which any network will grow in finite time, is essentially determined by the value C. When $n > C$, the probability of new nodes attaching is very small and these are new nodes are rejected. Simulations were conducted until the network had grown to a specified size, attempting to attach new nodes until this was successful.

Repeated simulations for network growth according to model 4 were performed for varying constant C. It was observed that for $C = N/2$, where N is the size to which the network is grown, the average path length and the clustering coefficients of the generated networks fitted the metabolic pathway data better than any single previous structural description of metabolic networks (see figure 1).

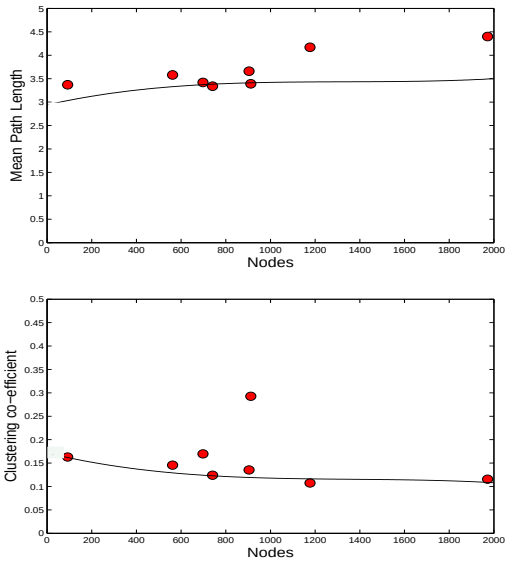

FIG. 1: Mean minimum path lengths (top) and clustering coefficients(bottom) for eight microbial metabolic pathways (red circles) and (black line) average minimum path lengths and clustering coefficients for randomly generated size dependent networks with $C = N/2$ (equation 4).

It should be noted that the networks generated are not truly scale-free, being somewhere a hybrid of both exponential and scale free type distributions which is not unlike the original metabolic networks.

The overall effect of such a size dependent modification to the original BA model is to initially produce a highly connected 'hub' which then grows via preferential attachment giving rise to high clustering coefficients and a very strong small world effect resulting in low average path lengths.

It is straightforward to amend the size dependent model 4 such that no globally connected initial stage occurs by rescaling the size dependent modification in Equation 4 to $C/n+\sqrt{C}$. This produces a network which has a more pronounced scale free structure, while retaining the high clustering coefficients of the original formulation.

5. *Example - S Aureus*

The bacteria Staphylococcus Aureus has a metabolic network of 741 nodes, a clustering coefficient of $< c >= 0.124$, a maximum geodesic length of 9 and a mean geodesic length of 3.34.

The (mean) average values for 100 trials of a network of 741 nodes, generated according to equation 4 with $C = 37.5$ are: clustering coefficient $< c >= 0.125$, maximum path length 7.84 and mean geodesic length 3.53.

The distribution of geodesic lengths over the whole of

the S Aureus network and an example size dependent network are shown in figure 2 in addition to the degree distributions.

Although somewhat circumstantial as evidence, it is often useful to illustrate the similarities and differences between graphical structures using pictorial representations. Such comparison between the original BA model, S. Aureus and the size dependent growth model demonstrate the similarity between the size dependent model and the metabolic pathways (see figure 3). Here the highly connected initial growth stage, although omitting the underlying compartmentalisation of the S. Aureus metabolic network provides a good approximation of the structure. Additionally the presence of some 'dead end' metabolites, which appear as nodes connected to only 1 other node (thus giving rise to higher maximum path lengths than may otherwise be observed) is not modelled by this approach.

II. DISCUSSION

This investigation has demonstrated that the metabolic networks of microorganisms are more accurately modelled with a network growth model in which network size is a modifying factor, a concept which has surprisingly not previously been considered. This model, as well as fitting the graphical measures examined, also demonstrates a one possible mechanism which may be significant in the network evolution. This suggests that when the metabolic networks are growing (or evolving) the size of the network causes it to be increasingly unlikely for a new metabolite to join the network and participate in the reactions. Due to the size dependent growth networks having a densely connected cluster, these networks will have an increased resilience to targeted attacks than that of the BA model, which are known to be devastated by targeted attacks [25]. This suggests that the structure of metabolic pathways gives them a greater resilience to targeted attacks than if they were examples of scale-free networks, modelled by preferential attachment.

The value of the rate at which the probabilities of attachment decay, C, has in our modelling been chosen as a single value to fit all metabolic networks, however we envisage that for specific networks a better fit would be available using particular values.

We have presented a model which is both simple and biologically plausible. Due to the fact it does not require any specific seed network it allows for a generic model which can be used to model various metabolic pathways with only the network size being known. This allows for various graphical measures to be estimated for any given metabolic network, without them needing to be individually analysed.

Clearly the concept of size dependent growth may not be confined to evolving metabolic networks but may be applicable to a variety of networks where growth rates may be affected by limited resources. One example where this may be applicable is that of the London Underground, where an initially a highly connected network has grown and now as the network has become larger and more complex it has become increasingly difficult for new stations and connections to be added to the network.

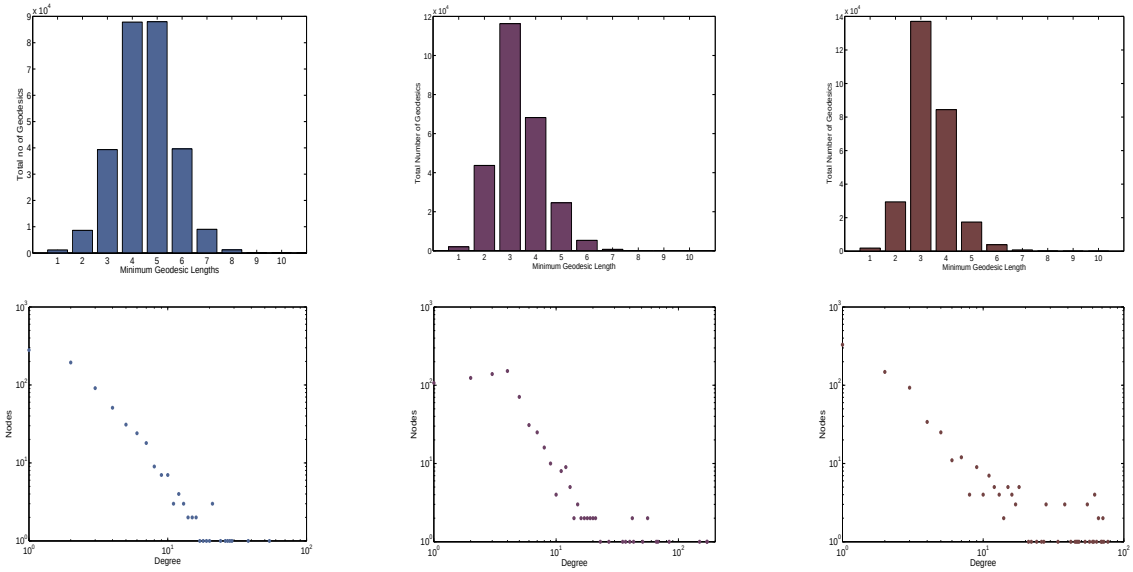

FIG. 2: Minimum geodesic lengths (top) and degree distribution plots (bottom) for S Aureus and example size dependent network of 741 nodes. Note the geodesic distribution for the size dependent network matches closely the metabolic network. The degree distribution for the metabolic network is not entirely scale free due to fewer nodes of degree 1 and 2 than would be expected. The size dependent growth model demonstrates a degree distribution which is in many respects closer that of the metabolic pathways. 3

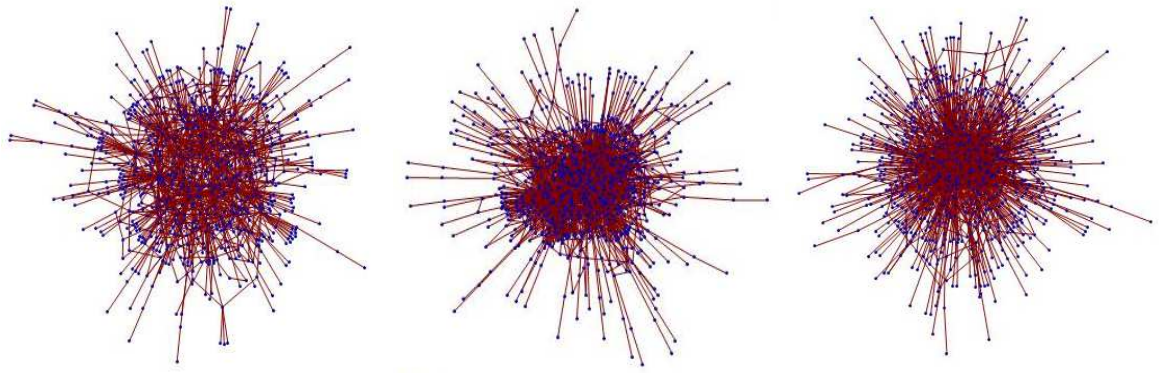

FIG. 3: Graphical representations of artificially generated Barabási Albert Scale-free network (741 nodes), S Aureus isB619 (741 nodes) and artificially generated size dependent network (741 nodes). All figures are generated using the Mathematica spring embedding algorithm.

Acknowledgements

KS is grateful for the financial support of the EU FP7 (KBBE) grant 289434 "BioPreDyn: New Bioinformatics Methods and Tools for Data-Driven Predictive Dynamic Modelling in Biotechnological Applications".

[1] V. Lacroix, L. Cottret, P Thébault, and M-F. Sagot. An introduction to metabolic networks and their structural analysis. *Ieee/Acm Transactions On Computational Biology And Bioinformatics*, 5(4):594–617, 2008.

[2] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[3] A. Wagner and D.A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1478):1803,

2001.

[4] S. Schuster and B.H. Junker. Topological Analysis of Metabolic and Regulatory Networks. *Modeling in Systems Biology*, pages 209–224, 2011.

[5] J. Stelling, S. Klamt, K. Bettenbrock, Schuster S., and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, November 2002.

[6] A.L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1-4):69–77, 2000.

[7] R. Khanin and E. Wit. How scale-free are biological networks. *Journal Of Computational Biology*, 13(3):810–818, 2006.

[8] M. P. H. Stumpf, C. Wiuf, and R. M. May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *PNAS*, 102(12):4221–4224, 2005.

[9] E. Ravasz and A.L. Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.

[10] Kazuhiro Takemoto. Metabolic network modularity arising from simple growth processes. *Phys. Rev. E*, 86:036107, Sep 2012.

[11] Konstantin Klemm and Víctor M. Eguíluz. Highly clustered scale-free networks. *Phys. Rev. E*, 65:036123, Feb 2002.

[12] Hong-wu ma, xue-ming zhao, ying-jin yuan and an-ping zeng. *Bioinformatics*, 20(12):1870–1876, 2004.

[13] C.M. Schneider, L. de Arcangelis, and H.J. Herrmann. Scale free networks by preferential depletion. *Arxiv preprint arXiv:1103.1396*, 2011.

[14] S. N. Dorogovtsev and J. F. F. Mendes. Effect of the accelerating growth of communications networks on their structure. *Phys. Rev. E*, 63:025101, Jan 2001.

[15] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, Cuellar A. A. Cornish-Bowden, A., S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19:524–531, 2003.

[16] J. Schellenberger, J.O. Park, T.M. Conrad, and B.Ø. Palsson. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC bioinformatics*, 11(1):213, 2010.

[17] B. J. Bornstein, S. M. Keating, A. Jouraku, and M. Hucka. LibSBML: an API library for SBML. *Bioinformatics*, 24:880–881, Mar 2008.

[18] S. Schuster, D.A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature biotechnology*, 18(3):326–332, 2000.

[19] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[20] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[21] D.A. Fell and A. Wagner. The small world of metabolism. *Nature Biotechnology*, 18(11):1121–1122, 2000.

[22] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[23] P. Erds and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist*, 38(4):343–347, 1961.

[24] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551, 2002.

[25] M.E.J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.