# Supplementary Material for: Network motif frequency vectors reveal evolving metabolic network organisation

## Contents

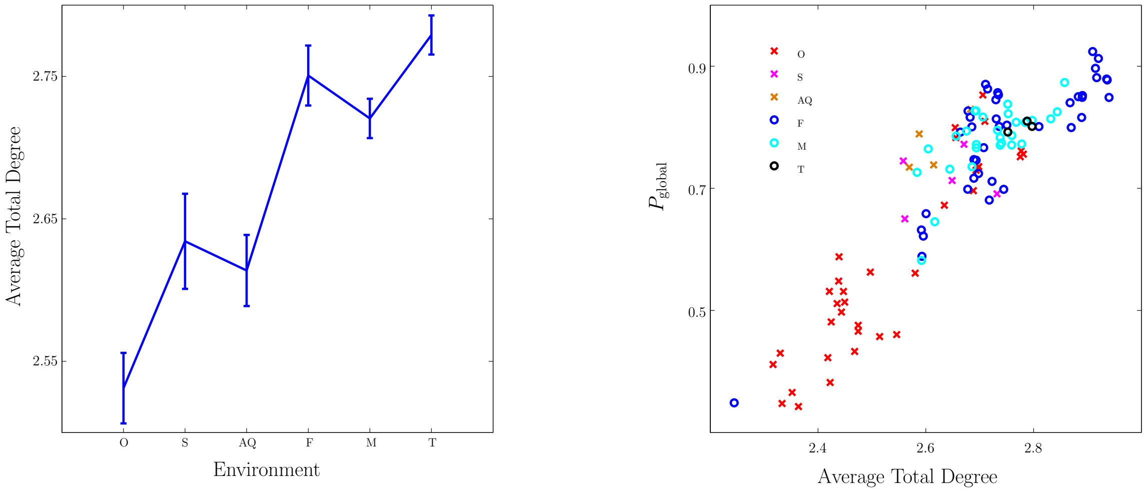
## 1 Relation between environmental conditions and basic structural properties of the metabolic networks

### 1.1 Average total degree

The total degree of a node is the total number of arcs (directed edges) connected to that node, where the arc can be pointing in either direction (i.e. we sum the in-coming and out-going arcs to and from that node). The average total degree $d_j^{\text{tot}}$ for the $j$th network is then calculated by taking the mean over all nodes in the network.

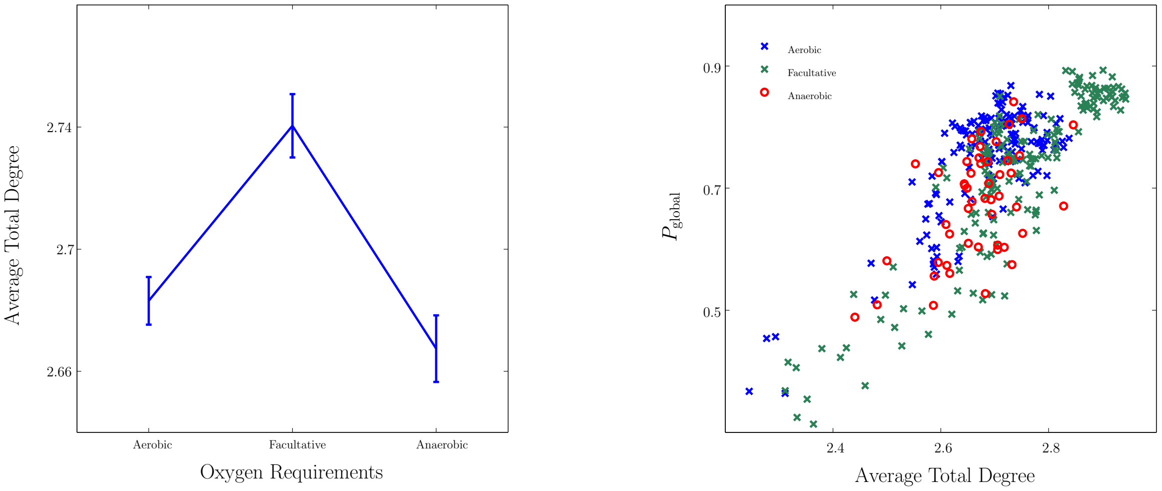
#### Habitat variability

Figure S1 (a) shows that the trend for the average total degree is to increase with variability in the environment, although not monotonically. In Figure S1 (b) we plot the average total degree against our global significance score $P_{\text{global}}$ and find a significant correlation ($r = 0.8983$, $p < 10^{-5}$).



(a) (b)

Figure S1: (a) Relationship between environmental variability and the average total degree for the 6 environmental classes. Note that here we plot the mean value over each environmental class: **O**bligate, **S**pecialised, **AQ**uatic, **F**acultative, **M**ultiple and **T**errestrial. (b) The average total degree plotted against the global significance score $P_{\mathrm{global}}$ for the 115 bacterial networks.



(a) (b)

Figure S2: (a) Relationship between growth conditions, specifically oxygen requirements, and the average total degree. Note that here we plot the mean value over each class and error bars represent standard errors. (b) The average total degree plotted against the global significance score $P_{\mathrm{global}}$ for the 383 bacterial networks.

### Oxygen requirements

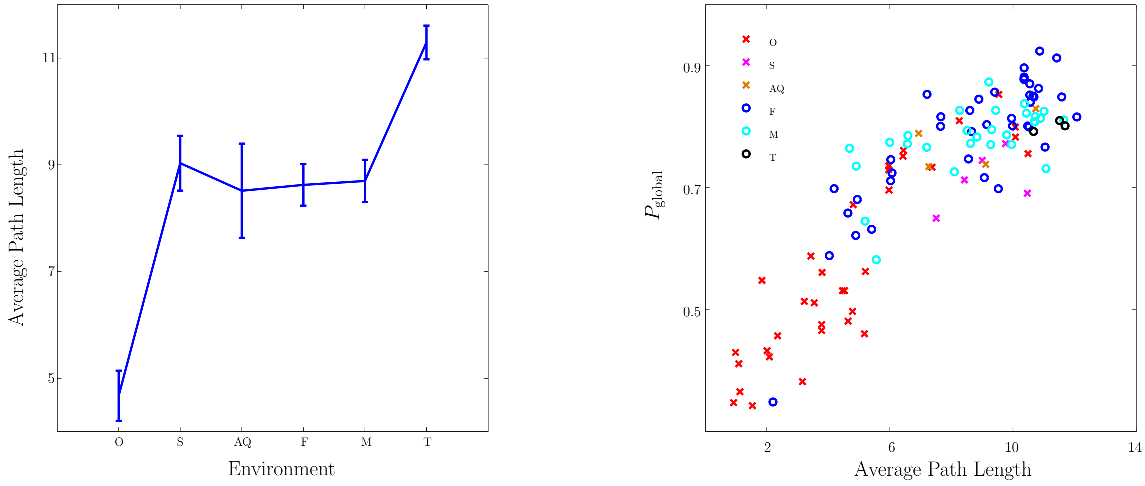
Figure S2 (a) shows that the average total degree does not appear to follow any particular trend regarding oxygen requirements, despite being correlated to the global significance score $P_{\text{global}}$ ($r = 0.7932$, $p < 10^{-5}$) (Figure S2 (b)).

## 1.2 Average path length

The average path length $L$ is defined as the average shortest path length between all pairs of nodes in the network. $L$ is measure of how efficient information can be transported across the network.

### Habitat variability

Figure S3 (a) shows that the average path length is the smallest for the networks within the obligate class (i.e. the most specialised) and the largest for the terrestrial class (i.e. the most varied). The specialised, aquatic, facultative and multiple class however, all have very similar values. In Figure S3 (b) we have plotted the average path length against the global significance score $P_{\text{global}}$ and find a significant correlation ($r = 0.8723$, $p < 10^{-5}$).



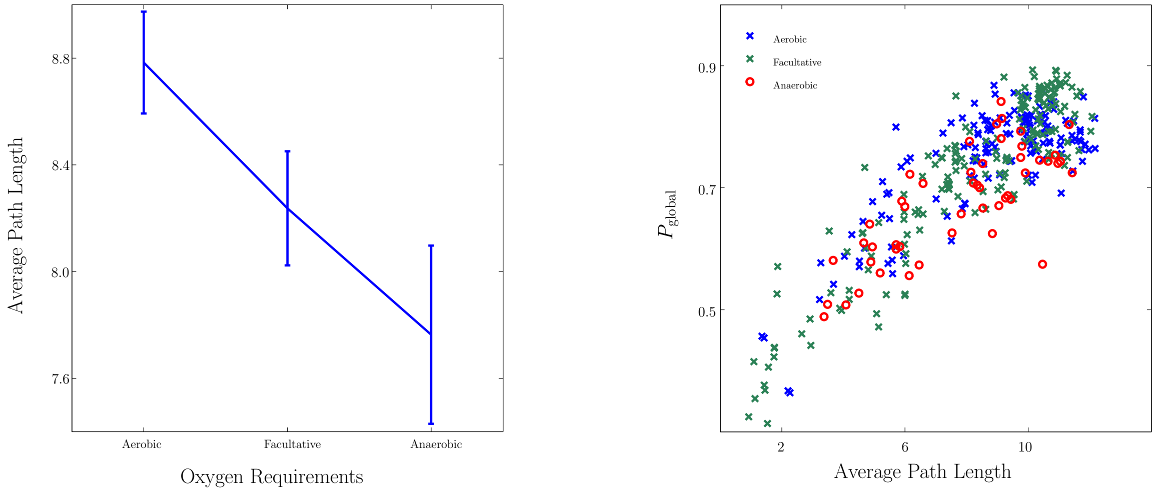
(a)

(b)

Figure S3: (a) Relationship between environmental variability and the average path length for the 6 environmental classes. Note that here we plot the mean value over each environmental class: **O**bligate, **S**pecialised, **AQ**uatic, **F**acultative, **M**ultiple and **T**errestrial. (b) The average path length plotted against the global significance score $P_{\text{global}}$ for the 115 bacterial networks.

### Oxygen requirements

Figure S4 (a) shows that the mean average path length is larger for the metabolic networks that evolved in the presence of oxygen, that is, the aerobic and facultative class. In Figure S4 (b) we have plotted the average path length against the global significance score $P_{\text{global}}$ for the 383 metabolic networks and find a significant correlation ($r = 0.8617$, $p < 10^{-5}$).



(a)

(b)

Figure S4: (a) Relationship between growth conditions, specifically oxygen requirements, and the average path length. Note that here we plot the mean value over each class and error bars represent standard errors. (b) The average path length plotted against the global significance score $P_{\text{global}}$ for the 383 bacterial networks.

## 1.3 Clustering coefficient

The clustering coefficient measures the extent to which the nodes in the network tend to cluster. More formally, the clustering coefficient is defined as the fraction of a nodes neighbours that are also neighbours of each other. We compute the average clustering coefficient by taking the mean value over all nodes within the network.

### Habitat variability

Figure S5 (a) shows that the clustering coefficient does not follow any particular trend as regards environmental habitat. Figure S5 (b) shows that the average clustering coefficient is only weakly correlated with the global significance score $P_{\text{global}}$ ($r = 0.2531$, $p < 0.01$).

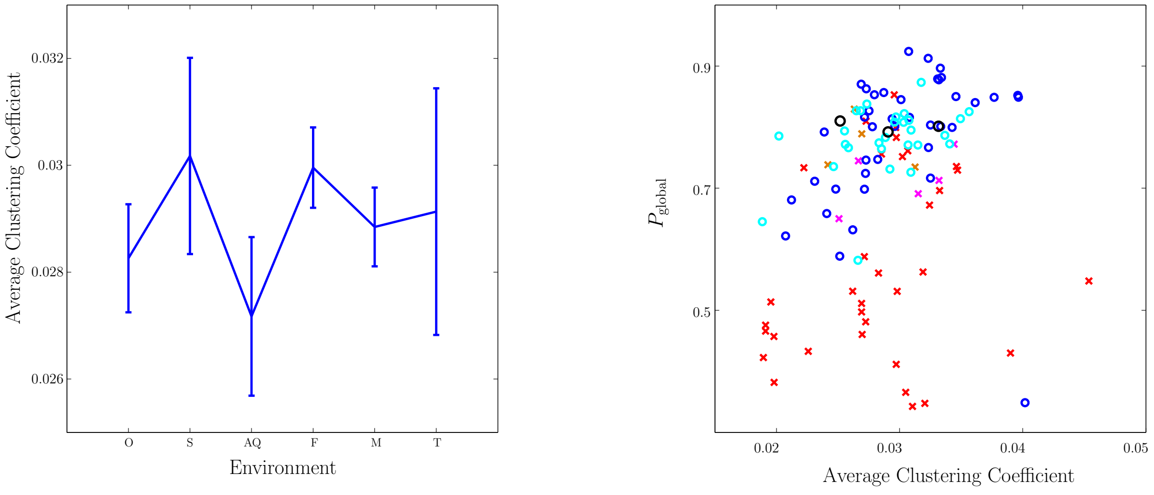
### Oxygen requirements

Figure S6 (a) shows a relationship similar to the average total degree (Figure S2), that is, the facultative class has a significantly larger amount of clustering present than the aerobic and anaerobic classes. Figure S6 (b) shows that the average clustering coefficient and the global significance score $P_{\text{global}}$ are not correlated ($r = 0.0737$, $p = 0.1489$).

See [2] for more details concerning the network measures above.

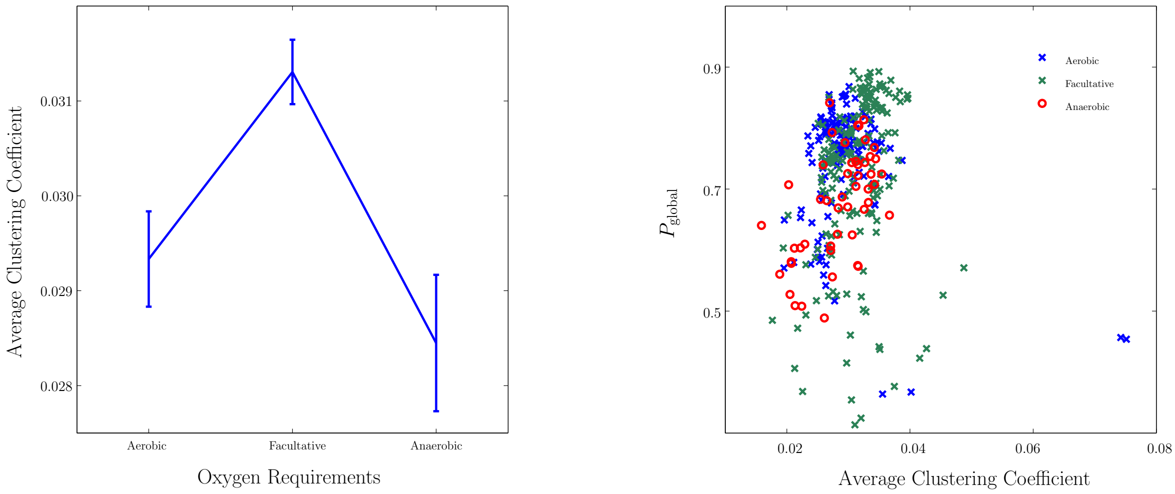
## 1.4 Spearman's partial correlation between $P$ and environment conditioned on basic network measures

Spearman's partial correlation between $X$ and $Y$ conditioned on $Z$ allows one to compute the correlation between $X$ and $Y$, discounting the correlations between $X$ and $Z$ and between $Y$



(a) (b)

Figure S5: (a) Relationship between environmental variability and the average clustering coefficient for the 6 environmental classes. Note that here we plot the mean value over each environmental class: **O**bligate, **S**pecialised, **AQ**uatic, **F**acultative, **M**ultiple and **T**errestrial. (b) The average clustering coefficient plotted against the global significance score $P_{\text{global}}$ for the 115 bacterial networks.



(a) (b)

Figure S6: (a) Relationship between growth conditions, specifically oxygen requirements, and the average clustering coefficient. Note that here we plot the mean value over each class and error bars represent standard errors. (b) The average clustering coefficient plotted against the global significance score $P_{\text{global}}$ for the 383 bacterial networks.

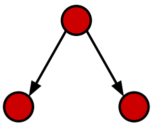
and $Z$ [1]. We computed the correlation between the global motif significance score and both variability and oxygen requirements conditioned on the simpler network metrics considered in the previous section (degree, path-length and clustering) and found that our results remained significant ($c = 1$, $p < 10^{-3}$ in all instances). Note that we use Spearman's correlation since the data consists of a mixture of both ordinal and continuous values; correlations were computed using the `partialcorr` function which is available in the MATLAB Statistics Toolbox.

# 2 Motif dictionary

The motif dictionary provides a graphical description of the 13 3-node subgraphs and the 199 4-node subgraphs used in this study. Here, the top label corresponds to a motif's identification number when using the *mfinder software* (http://www.weizmann.ac.il/mcb/UriAlon/groupNetworkMotifSW.html), whereas the bottom labels correspond to the motif numbering used in this work.

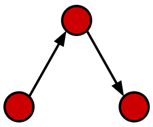
## 2.1 3-node subsgraphs

id6



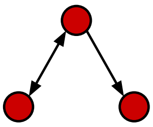
motif 1

id12



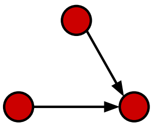
motif 2

id14



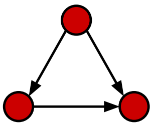
motif 3

id36



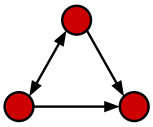
motif 4

id38



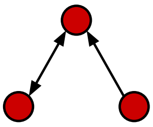
motif 5

id46



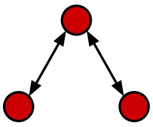
motif 6

id74



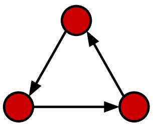
motif 7

id78



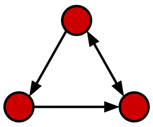
motif 8

id98



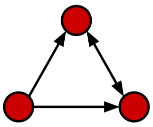
motif 9

id102



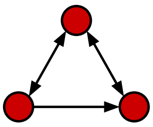
motif 10

id108



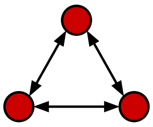
motif 11

id110



motif 12

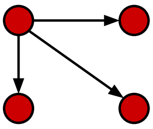
id238



motif 13

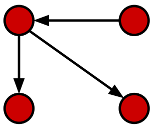
## 2.2 4-node subgraphs

id14



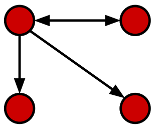
motif 14

id28



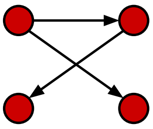
motif 15

id30



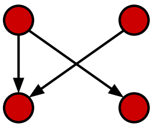
motif 16

id74



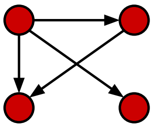
motif 17

id76



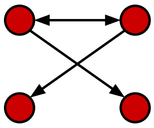
motif 18

id78



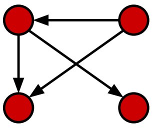
motif 19

id90



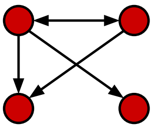
motif 20

id92



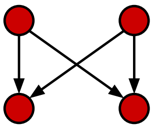
motif 21

id94



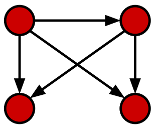
motif 22

id204



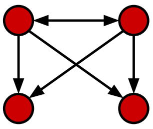
motif 23

id206



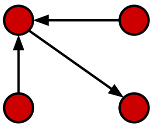
motif 24

id222



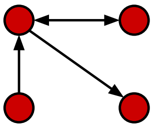
motif 25

id280



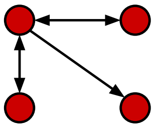
motif 26

id282



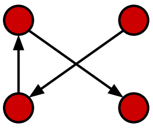
motif 27

id286



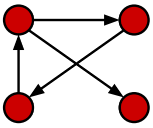
motif 28

id328



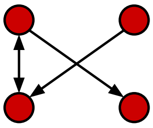
motif 29

id330



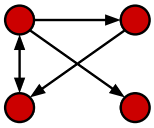
motif 30

id332



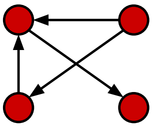
motif 31

id334



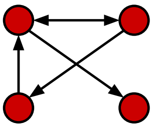
motif 32

id344



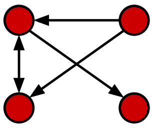
motif 33

id346



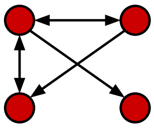
motif 34

id348



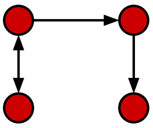
motif 35

id350



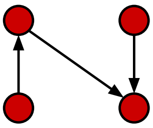
motif 36

id390



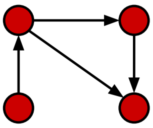
motif 37

id392



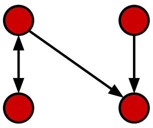
motif 38

id394



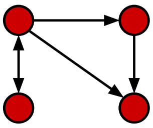
motif 39

id396



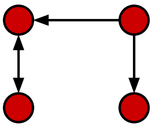
motif 40

id398



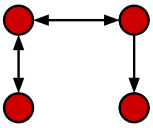
motif 41

id404



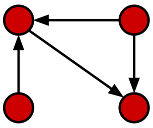
motif 42

id406



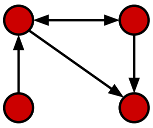
motif 43

id408



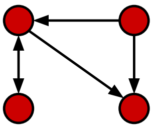
motif 44

id410



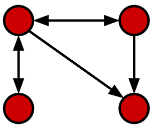
motif 45

id412



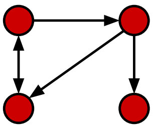
motif 46

id414



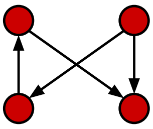
motif 47

id454



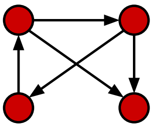
motif 48

id456



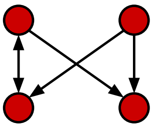
motif 49

id458



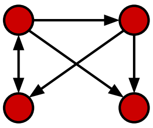
motif 50

id460



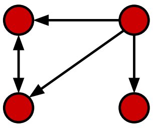
motif 51

id462



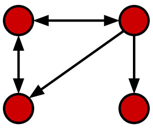
motif 52

id468



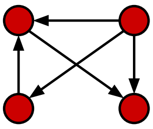
motif 53

id470



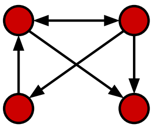
motif 54

id472



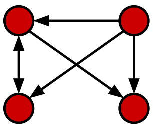
motif 55

id474



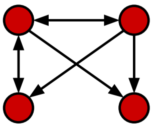
motif 56

id476



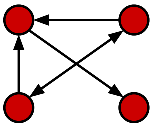
motif 57

id478



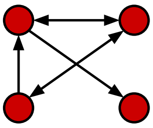
motif 58

id856



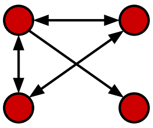
motif 59

id858



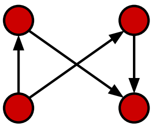
motif 60

id862



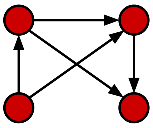
motif 61

id904



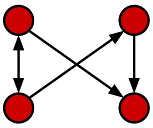
motif 62

id906



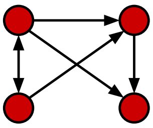
motif 63

id908



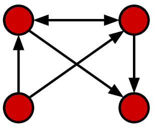
motif 64

id910



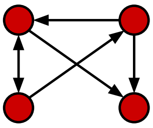
motif 65

id922



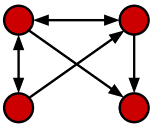
motif 66

id924



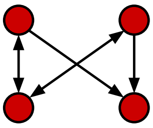
motif 67

id926



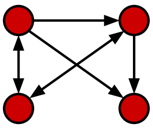
motif 68

id972



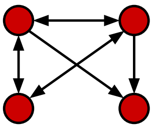
motif 69

id974



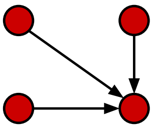
motif 70

id990



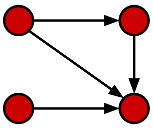
motif 71

id2184



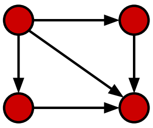
motif 72

id2186



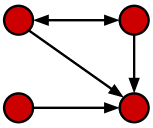
motif 73

id2190



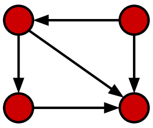
motif 74

id2202



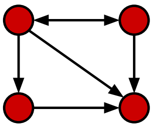
motif 75

id2204



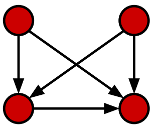
motif 76

id2206



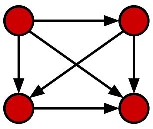
motif 77

id2252



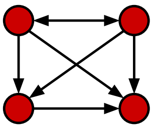
motif 78

id2254



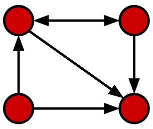
motif 79

id2270



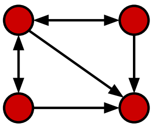
motif 80

id2458



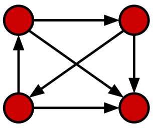
motif 81

id2426



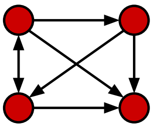
motif 82

id2506



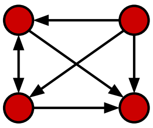
motif 83

id2510



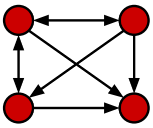
motif 84

id2524



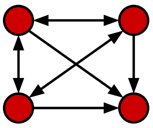
motif 85

id2562



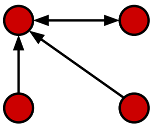
motif 86

id3038



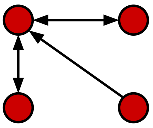
motif 87

id4370



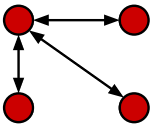
motif 88

id4374



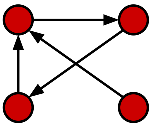
motif 89

id4382



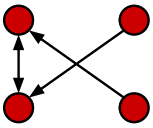
motif 90

id4418



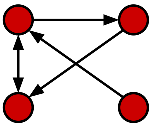
motif 91

id4420



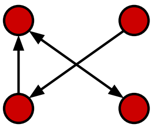
motif 92

id4422



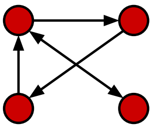
motif 93

id4424



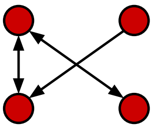
motif 94

id4426



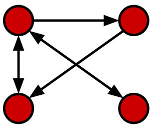
motif 95

id4428



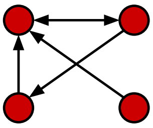
motif 96

id4430



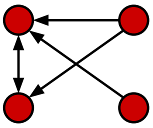
motif 97

id4434



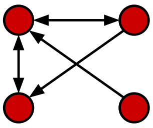
motif 98

id4436



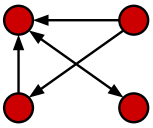
motif 99

id4438



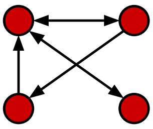
motif 100

id4440



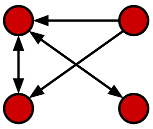
motif 101

id4442



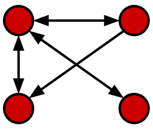
motif 102

id4444



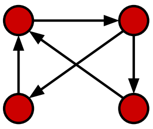
motif 103

id4446



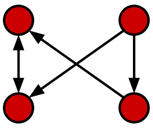
motif 104

id4546



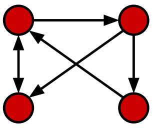
motif 105

id4548



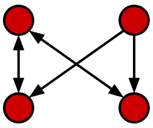
motif 106

id4550



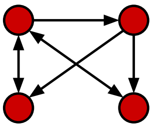
motif 107

id4556



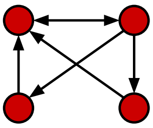
motif 108

id4558



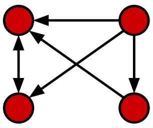
motif 109

id4562



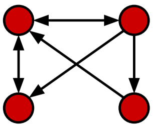
motif 110

id4564



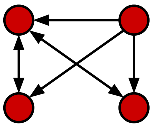
motif 111

id4566



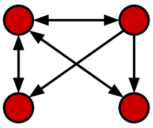
motif 112

id4572



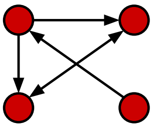
motif 113

id4574



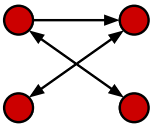
motif 114

id4678



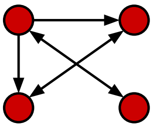
motif 115

id4682



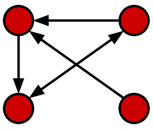
motif 116

id4686



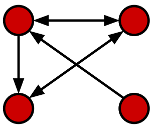
motif 117

id4692



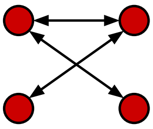
motif 118

id4694



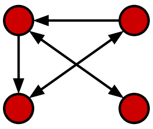
motif 119

id4698



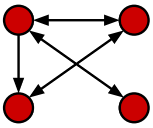
motif 120

id4700



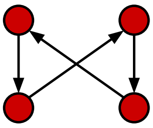
motif 121

id4702



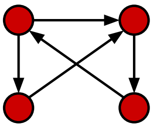
motif 122

id4740



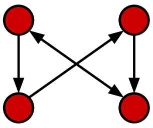
motif 123

id4742



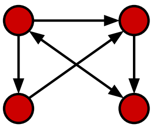
motif 124

id4748



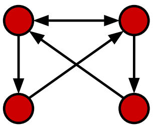
motif 125

id4750



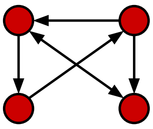
motif 126

id4758



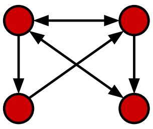
motif 127

id4764



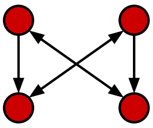
motif 128

id4766



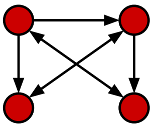
motif 129

id4812



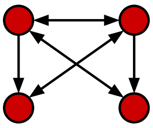
motif 130

id4814



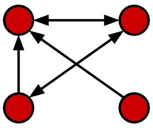
motif 131

id4830



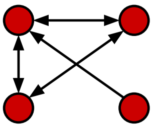
motif 132

id4946



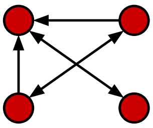
motif 133

id4950



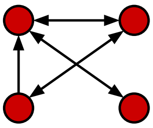
motif 134

id4952



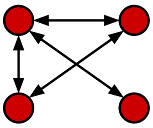
motif 135

id4954



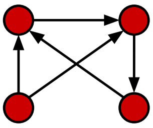
motif 136

id4958



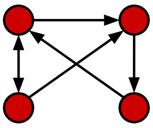
motif 137

id4994



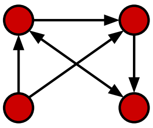
motif 138

id4998



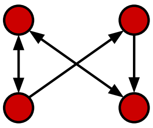
motif 139

id5002



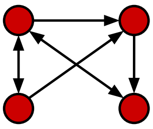
motif 140

id5004



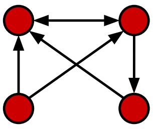
motif 141

id5006



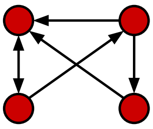
motif 142

id5010



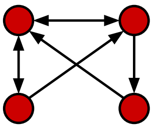
motif 143

id5012



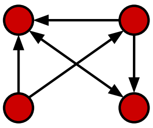
motif 144

id5014



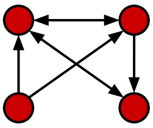
motif 145

id5016



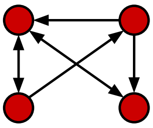
motif 146

id5018



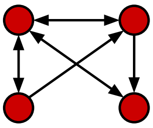
motif 147

id5020



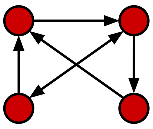
motif 148

id5022



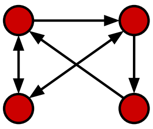
motif 149

id5058



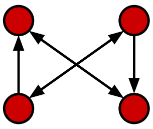
motif 150

id5062



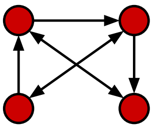
motif 151

id5064



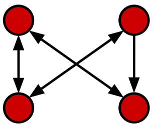
motif 152

id5066



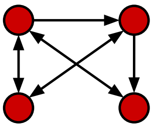
motif 153

id5068



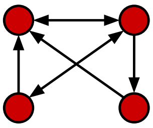
motif 154

id5070



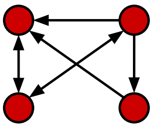
motif 155

id5074

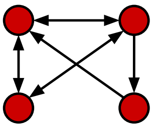


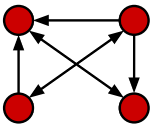
motif 156

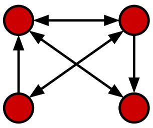
id5076

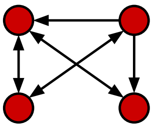


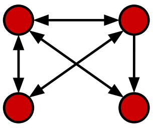
motif 157

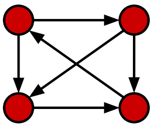
id5078

motif 158

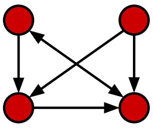
id5080

motif 159

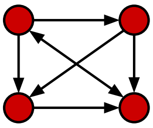
id5082

motif 160

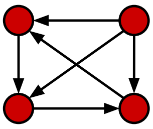
id5084

motif 161

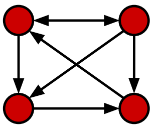
id5086

motif 162

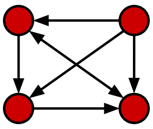
id6342

motif 163

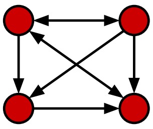
id6348

motif 164

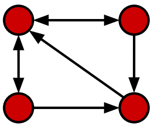
id6350

motif 165

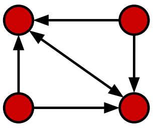
id6356

motif 166

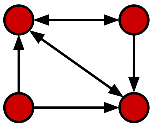
id6358

motif 167

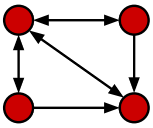
id6364

motif 168

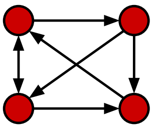
id6366

motif 169

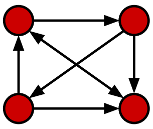
id6550

motif 170

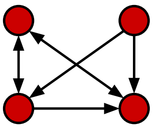
id6552

motif 171

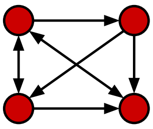
id6554

motif 172

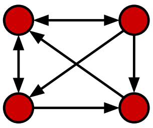
id6558

motif 173

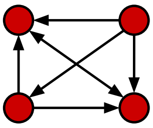
id6598

motif 174

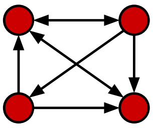
id6602

motif 175

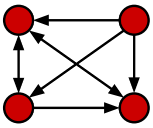
id6604

motif 176

id6606

motif 177

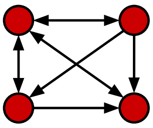
id6614

motif 178

id6616

motif 179

id6618

motif 180

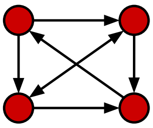
id6620

motif 181

id6622



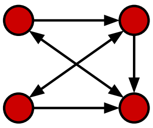
motif 182

id6854



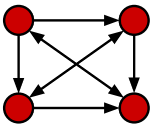
motif 183

id6858



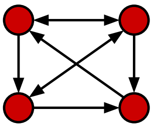
motif 184

id6862



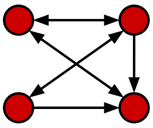
motif 185

id6870



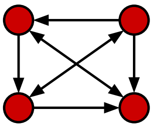
motif 186

id6874



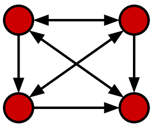
motif 187

id6876



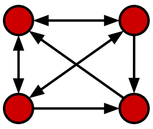
motif 188

id6878



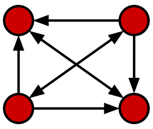
motif 189

id7126



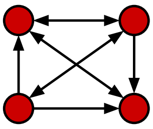
motif 190

id7128



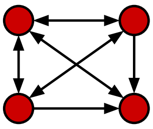
motif 191

id7130



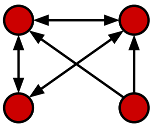
motif 192

id7134



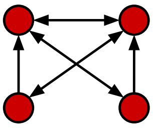
motif 193

id13142



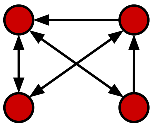
motif 194

id13146



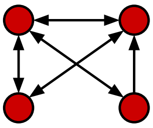
motif 195

id13148



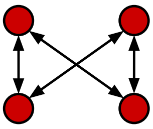
motif 196

id13150



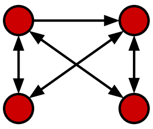
motif 197

id13260



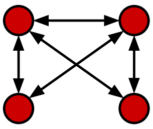
motif 198

id13262



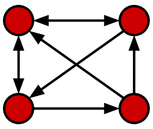
motif 199

id13278



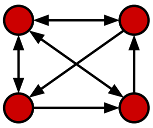
motif 200

id14678



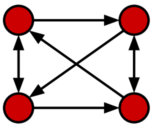
motif 201

id14686



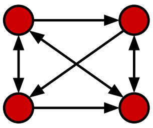
motif 202

id14790



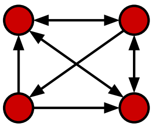
motif 203

id14798



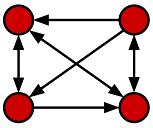
motif 204

id14810



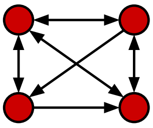
motif 205

id14812



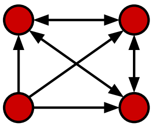
motif 206

id14814



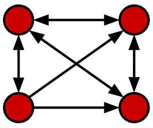
motif 207

id15258



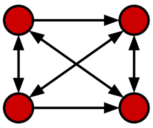
motif 208

id15262



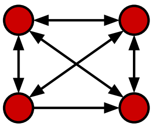
motif 209

id15310



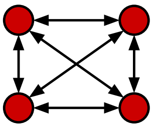
motif 210

id15326



motif 211

id31710



motif 212

# 3 Significant metabolites: Motif 9

## 3.1 Habitat variability

Figure S7 shows the mean frequency for metabolites occurring within motif 9 for the 115 metabolic networks, grouped into the specialised (blue bars) and varied (red bars) classes. Here metabolites are displayed in decreasing order according to the varied class. Figure S7 shows the 54 metabolites that were found at least once across the 115 metabolic networks. We find that the distribution of metabolites is slightly broader for the varied class, similar, but less prominent, to the results obtained for motif 5. Using Chi-square tests (Fisher's Exact test) we explored group differences for the individual metabolites. Figure S8 identifies only one metabolite, RNA, for which significant differences were found (Fisher's Exact test, $p < 0.001$).

## 3.2 Oxygen requirement

Figure S9 shows the mean frequency for metabolites occurring within motif 9 for the 383 metabolic networks that evolved in either the presence or absence of oxygen. Here metabolites are displayed in decreasing order according to the aerobic class (blue bars). Figure S9 shows the 65 metabolites that were found at least once across the two classes. Note that the distribution for the aerobic class and anaerobic class for motif 9 are a lot closer than was obtained for motif 5. Figure S10 shows that the metabolites with the most significant differences (Fisher's Exact test, p<0.001) included Glutathione, L-Arginine, L-Citrulline, N-(L-Arginino)succinate, Succinate, Succinyl-CoA and O-Succinyl-L-homoserine.

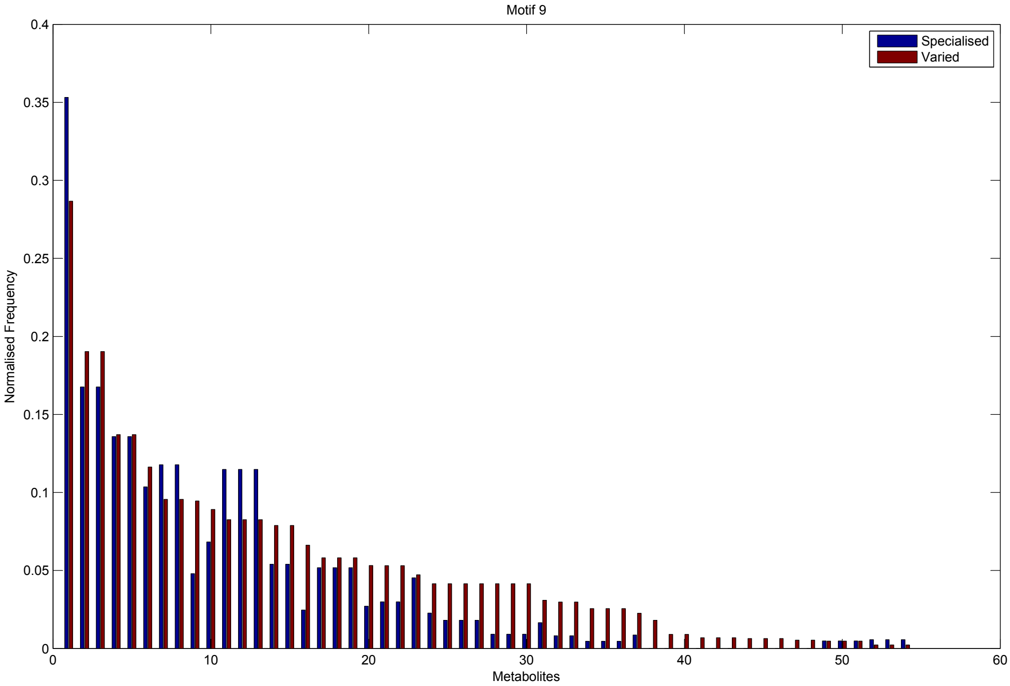


Figure S7: Mean normalised frequency for the 54 metabolites obtained for the 115 metabolic networks. Blue bars represent the specialised class and the red bars represent the varied class. Here, the metabolites are in descending order of the metabolite frequencies for the varied class.

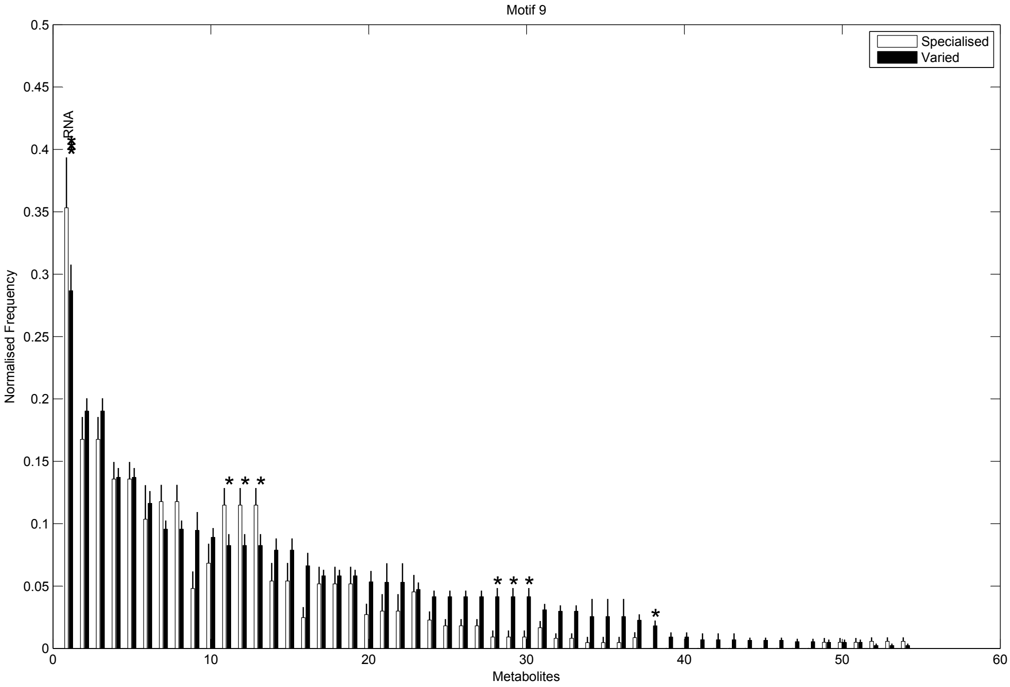


Figure S8: Mean normalised frequency for the 54 metabolites obtained for the 115 metabolic networks. Error bars are standard errors. Asterisks indicate levels of significance, with *, **, and *** corresponding to $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. Metabolite names are provided for the most significant metabolites.

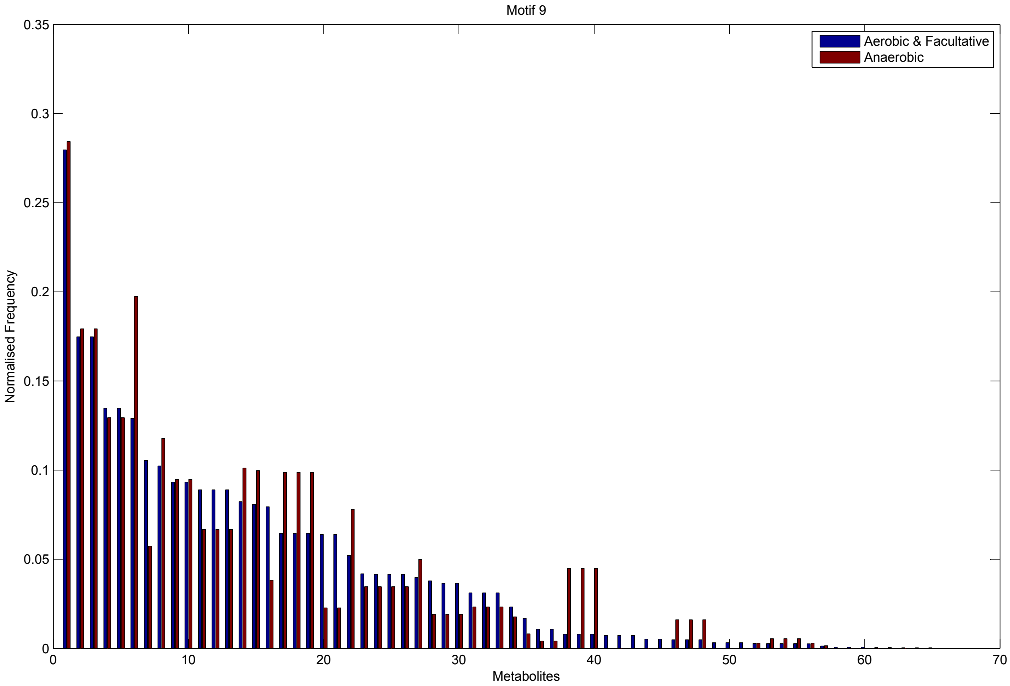


Figure S9: Mean normalised frequency for the 65 metabolites obtained for the 383 metabolic networks. Blue bars represent the aerobic-facultative class and the red bars represent the anaerobic class. Here, the metabolites are in descending order of the metabolite frequencies for the aerobic and facultative class.

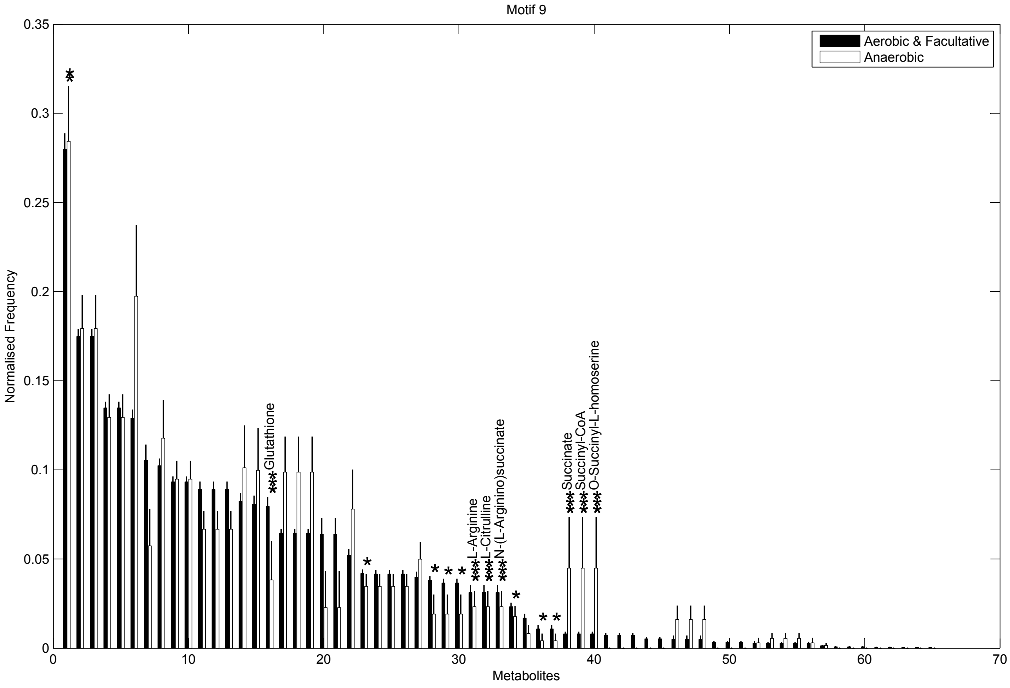


Figure S10: Mean normalised frequency for the 65 metabolites obtained for the 383 metabolic networks. Error bars are standard errors. Asterisks indicate levels of significance, with *, **, and *** corresponding to $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. Metabolite names are provided for the most significant metabolites.

# References

[1] Cramer, D, *A cautionary tale of two statistics: Partial correlation and standardised partial regression*, Journal of Psychology **137**(5):507–511, (2003).

[2] Mark Newman, *Networks: An Introduction*, Oxford University Press, 2010.