BMI206: Paper 3 Individual Project

Laura Shub

December 2019

	bridging_centrality	betweenness_centrality	clustering_coefficient
MG1655	9.155e-04	9.203e-02	3.567e-01
W3110	9.233e-04	9.243e-02	3.575e-01
EDL933	1.115e-03	1.020e-01	2.465e-01
SAKAI	1.144e-03	1.033e-01	2.487e-01
CFTO73	2.528e-03	5.370e-02	3.508e-01
UTI89	2.410e-03	5.209e-02	3.437e-01
core	9.237e-04	2.706e-02	3.503e-01

Table 1: Enrichment p-values for essential reactions in the top 5% of reactions, ranked by each graph metric for the 7 metabolic models examined.

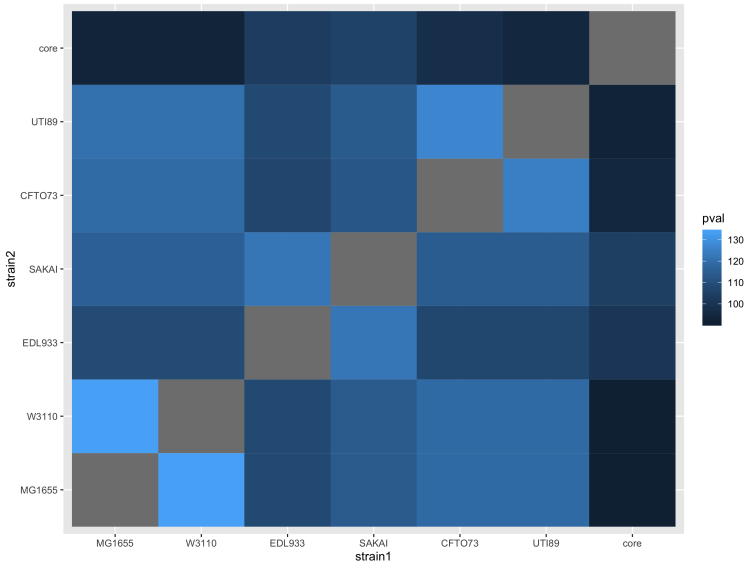

Figure 1: -log(p) for the enrichment of overlapping reactions in the top 5% of reactions by bridging centrality.

1 Methods

The metabolic networks were provided by [2] in XML format. The network reconstruction and graph metric calculations were performed using a Python script. Metabolites and reactions were stripped of their prefixes and suffixes, and transport reactions were pruned. No self-loops were allowed. Enrichment p-values were calculated using a hypergeometric distribution in R. For Figure 1, essential and non-essential reactions were classified using the results of flux balance analysis calculated under aerobic conditions, provided in the supplement of [2]. In calculating the overlap enrichment displayed in Figure 1, reactions in strain 1 were considered successful if they also appeared in the top 5% of strain 2. Both the Python and R code used to generate these figures along with the ranked reactions for each strain for the three graph metrics tested (bridging centrality, betweenness centrality, and clustering coefficient) can be found at `https://github.com/laurashub/BMI206IndividualProject`.

2 Results

The relation between high bridging centrality and essential reactions held for all seven of the *E. coli* strains, with each displaying significant enrichment ($p < 0.05$). The only strain showing significant enrichment of essential reactions in the top 5% by betweenness centrality was the 'Core' model, at $p = 2.706 \times 10^{-2}$. Other essential enrichment p-values for betweenness centrality ranged from 5.21×10^{-2} to 1.03×10^{-2}, through no other strains met the threshold for significance. Clustering coefficient did not show significant enrichment for essential reactions in any of the strains.

In looking at the overlap in the top 5% reactions for bridging centrality, each pair of strains had significant enrichment (p-val $<<<$ 0.05) in overlapping reactions. The 'Core' strain showed the lowest overlap with the remaining six strains. MG1655 and W3110 showed the greatest similarity in the top reactions by bridging centrality ($p = 2.63 \times 10^{-134}$).

3 Discussion

The significant enrichment of essential reactions in the top reactions by bridging centrality held for all seven strains studied. In comparison, betweenness centrality did not show the same significant enrichment as observed in [1] for six of the seven strains. This discrepancy is likely due to the difference in the size of the networks. The pruned iJO133 contained 1251 reactions and 9099 arcs, while the seven reactions tested here had an average of approximately 1535 reactions and 237,818 arcs. This large difference in the relative number of arcs to nodes likely changed the available shortest paths between nodes, impacting the betweenness centrality of those reactions considered "essential" and causing them to appear lower in the ranking. The core model had the least of these additional edges, and as such was the only strain to show a significant enrichment. Similarly,

these additional edges would influence the clustering coefficient by making the neighborhood of essential reactions more connected, raising their clustering coefficient. Removal of the additional edges as discussed in [1] might be able to recover significant enrichment.

Both in Table 1 and Figure 1, the bacterial strains appear to be paired, showing the most significant overlap with the strain that shares the given pathogenic phenotype. MG1655 and W3110 are both variations of non-pathogenic k-12 E. coli, EDL933 and SAKAI are both enterohemorragic, and CFT073 and UTI89 are both uropathogenic strains. This is likely because these models include reactions common to the individual pathogenesis that are unique to these pairs, impacting the graph metrics similarly and causing the same reactions to appear in the top 5%.

The reactions in the top 5% by bridging centrality were almost identical in all seven strains and largely essential, as indicated by the very high enrichment observed for essential reactions (Table 1) as well as overlap (Figure 1). This makes sense from a biological standpoint, as essential metabolic reactions would most likely be conserved across different strains. These data suggest that bridging centrality is robust to relatively small changes in graph topology such as those seen between different strains of the same species, providing additional support for the use of the correlation between essential reactions and high bridging centrality described in [1].

[1] Kim, E., Ashlock, D. & Yoon, S. (2019). Identification of critical connectors in the directed reaction-centric graphs of microbial metabolic networks. *BMC Bioinformatics* 20, 328 doi:10.1186/s12859-019-2897-z

[2] Baumler, D. J., Peplinski, R. G., Reed, J. L., Glasner, J. D., & Perna, N. T. (2011). The evolution of metabolic networks of E. coli. *BMC systems biology*, 5, 182. doi:10.1186/1752-0509-5-182