

## Enquadramento

O presente relatório descreve de forma sucinta a abordagem e as funcionalidades adotadas para a implementação sobre a biblioteca *NetworkX* da possibilidade de criar redes metabólicas, a sua visualização e realizar a sua análise topológica.

Para obter estas representações, foi criada a classe **MetabolicNetworkX** que utiliza métodos das bibliotecas *NetworkX*, *ElementTree XML*, *Pyvis* e *Matplotlib* que, em conjunto, permitem importar redes de vários formatos, a sua manipulação e visualização, assim como o seu estudo topológico com 4 diferentes métricas, *Degree Distribution*, *Shortest Path Analysis*, *Clustering Coefficient* e *Hubs and Centrality Measures*.

O objectivo desta ferramenta foi possibilitar a manipulação e visualização de redes metabólicas provenientes de várias fontes, a visualização destas redes com representação da sua análise topológica bem como comparar alguma destas métricas com dados teóricos, gerados através da biblioteca *NetworkX*.

## Funcionalidades

Esta secção descreve brevemente as categorias de funcionalidades incorporadas no *MetabolicNetworkX*.

1. Leitura de Redes: A ferramenta permite importação de redes metabólicas em formato txt e XML-SBML provenientes do EBI-BioModels (ficheiros exemplo presentes em anexo);
2. Visualização de Redes: Na tentativa de replicar a visualização do Cytoscape, com recurso à biblioteca *Pyvis* é possível criar uma representação de qualquer rede importada ou gerada manualmente em formato html, a esta visualização foi também incorporada elementos da análise topológica. A natureza deste método, no entanto não permite modificações dinâmicas, representando a rede de forma estática na estrutura em que foi gerada;
3. Análise Topológica: Todas as redes podem ser sujeitas a análise com as métricas já descritas através de incorporação de funcionalidades presentes no *networkX* e incorporação/adaptação de código fornecido.

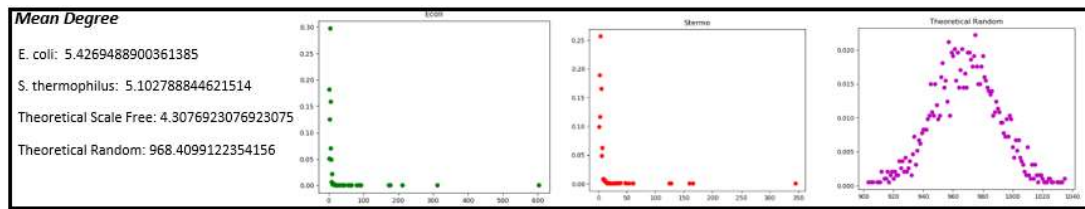
Uma versão mais completa das funcionalidades e guia de utilização da ferramenta estão presentes no ficheiro Jupyter *MetaNetXShowcase* (e respetivo .html para fácil visualização) anexado com o presente relatório, bem como o código, *MetabolicNetworkX.py*, para uma leitura completa dos métodos e respetivos comentários.

## Metodologia e Resultados

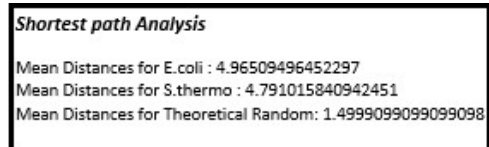
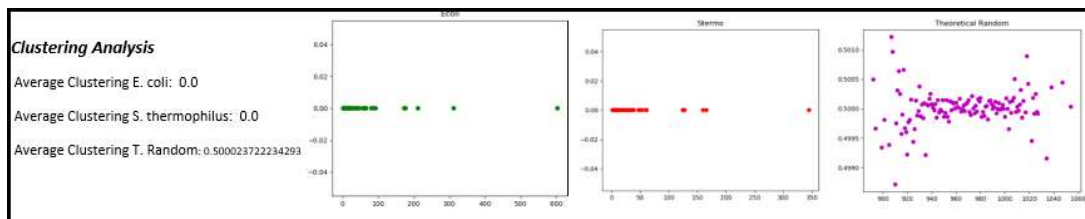
Para melhor demonstrar as funcionalidades da ferramenta, uma metodologia prática foi desenvolvida para estudar a rede metabólica da *Escherichia coli*, fornecida em formato txt, *Streptococcus thermophilus*, obtida da base de dados BioModels, uma rede *scale-free* e uma rede aleatória Erdős–Rényi geradas através de métodos teóricos presentes no *networkX* com o objectivo de replicar parte do trabalho de Barabási, 2004. A rede *E. Coli* foi ainda utilizada como exemplo da visualização com prévia análise topológica.

### Comparação entre Theoretical Scale free e Random Networks com dados biológicos reais:

O cálculo das métricas das redes, relativas à sua topologia, pode se revelar importante para inferir propriedades biológicas no modelo.



No esquema podemos observar o *Mean Degree*, que representa a média de edges que entram ou saem de um node, onde os resultados dos dados biológicos são muito próximos dos *scale free*. Nas redes *scale free*, a probabilidade de um node ter elevado número de nós com muitas ligações é estatisticamente inferior a de uma *random network*, sendo que as propriedades destas redes podem ser determinadas por alguns nós altamente conectados, conhecidos por *hubs*. Na *random network* os *degrees* dos nodes seguem uma distribuição de *Poisson*, o que indica que a maioria dos nodes têm aproximadamente o mesmo número de ligações, próximo da media (*mean degree*).



Na análise do coeficiente de *clustering*, observa-se a representação gráfica do coeficiente de agrupamento médio, este coeficiente testa a hipótese de uma rede, ser um grafo completo (*small-world*), em que todos os nodes estão ligados, e o caminho a percorrer entre eles, nunca toma valores elevados. Os grafos correspondentes aos dados biológicos são considerados completos, se apresentarem um valor do coeficiente significativamente maior que o coeficiente de um grafo gerado aleatoriamente, e para além disso, os resultados do *mean shortest path analysis* teriam de ser menores ou mais próximos do que a rede gerada aleatoriamente. Os nossos resultados rejeitam a hipótese de as redes biológicas testadas serem redes completas.

Uma propriedade comum das redes complexas é que qualquer par nós está à distância de um número reduzido de ligações, *small-world effect*. No entanto nas redes *scale-free* este efeito é muito pequeno. Nas células este é referido como *ultra small world effect*, em que nos nossos resultados dão paths de 4-5 passos entre pares de nodos. Esta distância indica que perturbações nas concentrações de metabolitos podem atingir o resto da rede rapidamente. Nos resultados verifica-se que *E.coli* tem aproximadamente o mesmo shortest path que o organismo *S.thermophilus*, o que pode indicar que existem mecanismos evolucionários que mantêm a average path length.

### Representação da rede metabólica

A construção da representação visual foi possível graças à biblioteca Pyvis e suas funcionalidades. Para a representação base da rede, metabolitos são representados como pontos amarelos e reações como triângulos verdes com ligações entre eles. Para além da visualização é possível ainda, nos ficheiros gerados com a representação, alterar a física da rede para melhor visualizar certos pontos ou melhorar a sua organização, o algoritmo de aproximação Barnes Hut é utilizado como gerador da posição inicial devido à sua eficiência num processo de *render* que pode ser relativamente demorado mediante o tamanho da rede. Para uma melhor visualização recomendamos no entanto o algoritmo ForceAtlas2 para representações como presentes na figura 1.

As medidas de centralidade foram incorporadas nesta representação. Para a representação dos nodos com maior grau de entradas e saídas (*highest degree*) uma mudança de cor foi implementada. Os metabolitos identificados são convertidos de amarelo para vermelho e reações de verde para azul. Esta representação é visível nas figuras 2 a 4 (nenhuma reação foi identificada como pertencente ao topo da lista de *highest degree*, no entanto esta modificação é possível ainda que improvável teoricamente).

*Closeness centrality*, identificando os nodos com menor distância média entre os mesmo e os restantes nodos da rede, são representados como pontos de reunião(*hubs*) aumentando o seu tamanho de acordo com a sua posição da lista. Esta representação pode ser observada nas figuras 3 e 4 com ambos metabolitos e reações aumentados.

*Betweenness centrality*, identificando nodos cujas porporções de caminhos mais curtos são menores, são identificados pelo aumento de espessura das suas edges bem como a herança de cor da origem, demonstrando as edges que algoritmos de caminho mais curto utilizarão com elevada probabilidade como podemos ver na figura 4.

Vários exemplos da representação de redes metabólicas de múltiplos microrganismos com e sem análise topológica estão presentes nos anexos deste relatório na pasta *miscnetworks*.

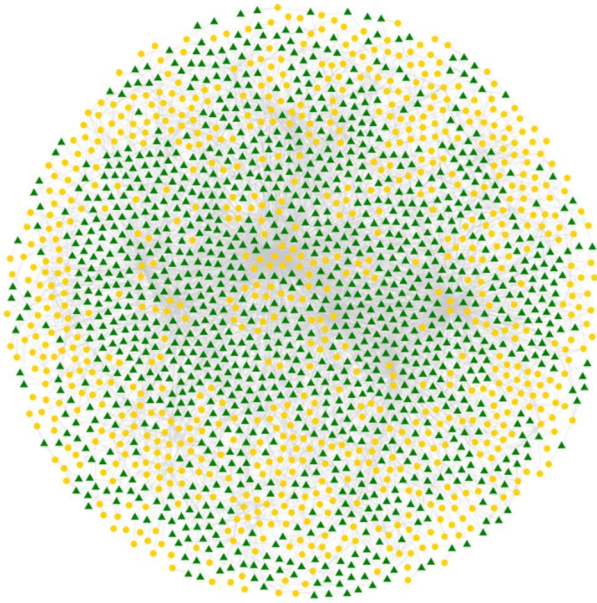


Fig.1 – Visualização da rede Metabolitos-Reações base (amarelo: metabolitos; verde: reações).

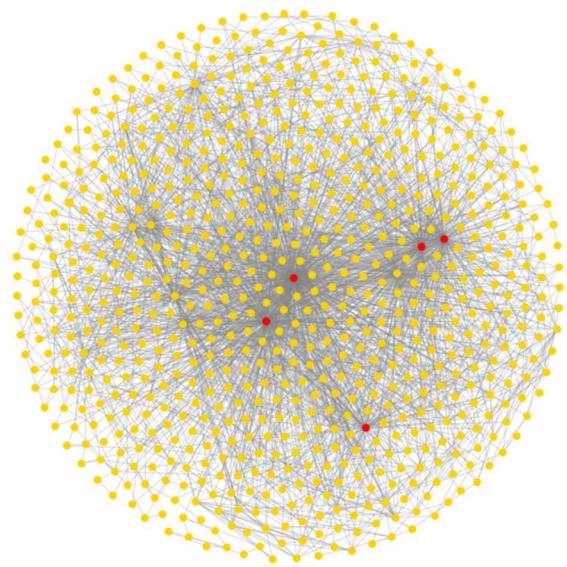


Fig.2 – Visualização da rede Metabolito-Metabolico (reações removidas) com representação do top 5 de nodes com *highest degrees* (amarelo: metabolitos; vermelho: top5 *highest degrees*).

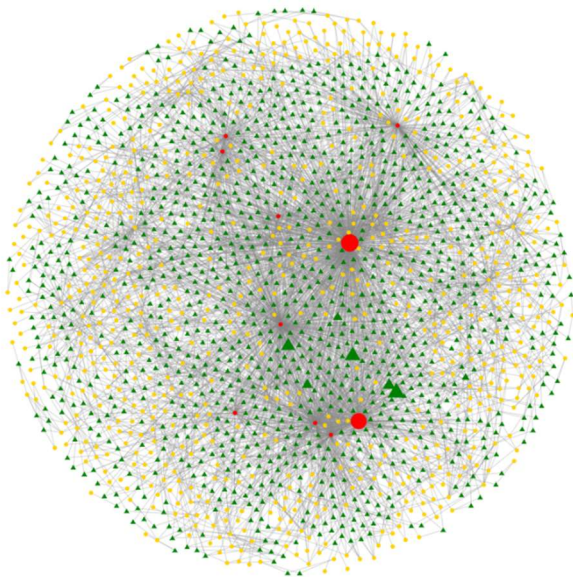


Fig.3 – Visualização da rede metabólica com representação do top 10 de nodes com *highest degrees* e *closeness centrality* (verde: reações; amarelo: metabolitos; vermelho: top10 *highest degrees*; tamanho maior: top10 *closeness centrality*).

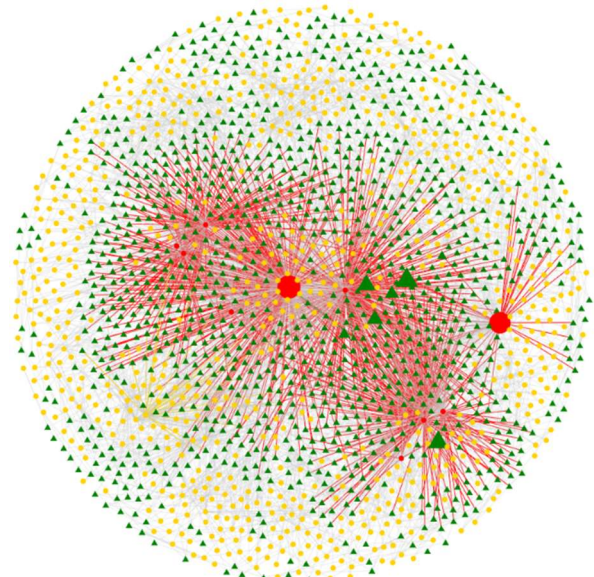


Fig.4 – Visualização da rede com presença visual todas as medidas de centralidade (verde: reações; amarelo: metabolitos; vermelho: top10 *highest degrees*; tamanho maior: top10 *closeness centrality*; edges coloridas: nodes de origem, top10 *betweenness centrality*)