

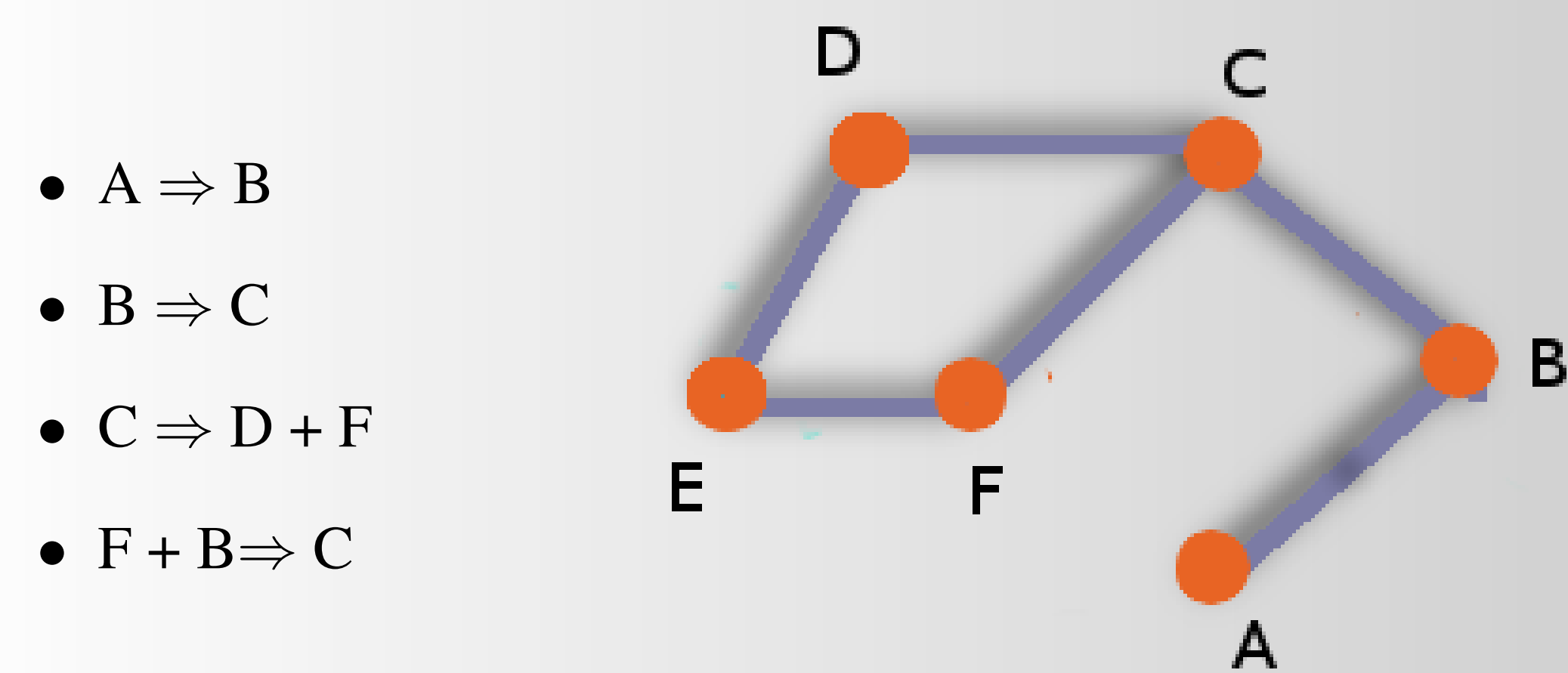
INTRODUÇÃO

A teoria de grafos é usada para descrever vários sistemas complexos como as redes sociais, redes de transmissão de energia, internet, etc. Nesse contexto, estudamos dois tipos de redes biológicas: redes metabólicas e redes de interações de proteínas (PPI - *Protein Protein Interaction*). O que buscamos indentificar são proteínas e moléculas que participem da mesma função biológica nos organismos. Esse problema, pode ser resumido a encontrar os *clusters*, ou agrupamentos nas redes. Eles são conjuntos de nós muito conectados entre si, mas, ao mesmo tempo, fracamente ligados com o resto da rede. Em resumo, metabólitos e proteínas que fazem parte de uma mesma função devem ser membros de um mesmo *cluster*.

REDES METABÓLICAS E REDES PPI

Muito embora as redes metabólicas e PPI sejam representadas da mesma maneira matematicamente, a sua interpretação física é completamente diferente. Nas redes metabólicas, cada nó (ou vértice) simboliza um metabólito, já uma ligação entre esses nós significa que os metabólitos conectados formam um par produto-reagente numa mesma reação química. Por outro lado, numa rede PPI um vértice representa uma proteína e uma ligação entre os vértices nos diz que essas proteínas interagem entre si em algum processo.

Abaixo, temos um exemplo de rede metabólica que contém os metabólitos A,B,C,D,E e F:

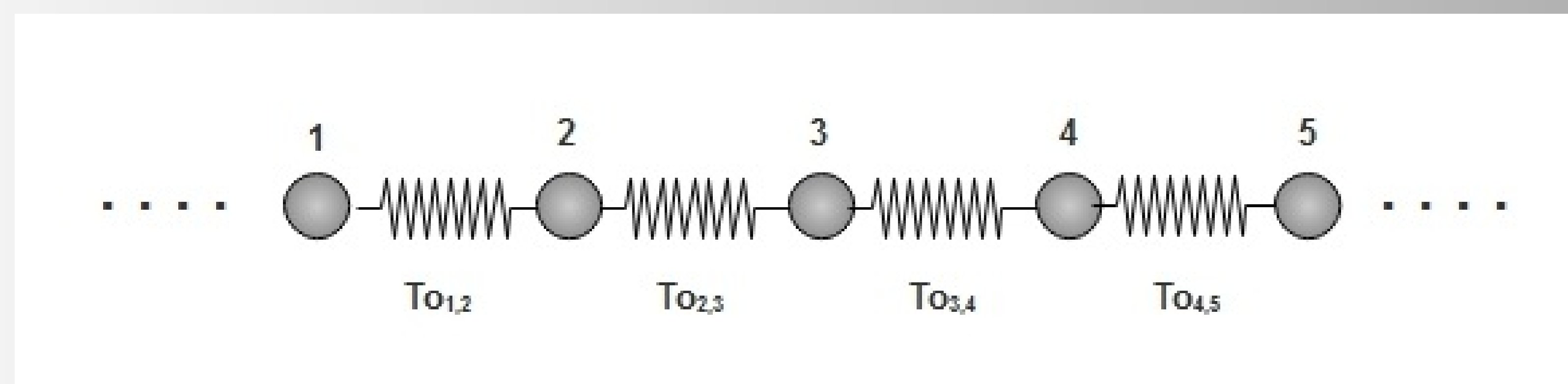


Matematicamente essa rede é descrita por sua matriz de adjacência:

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \quad (1)$$

METODOLOGIA

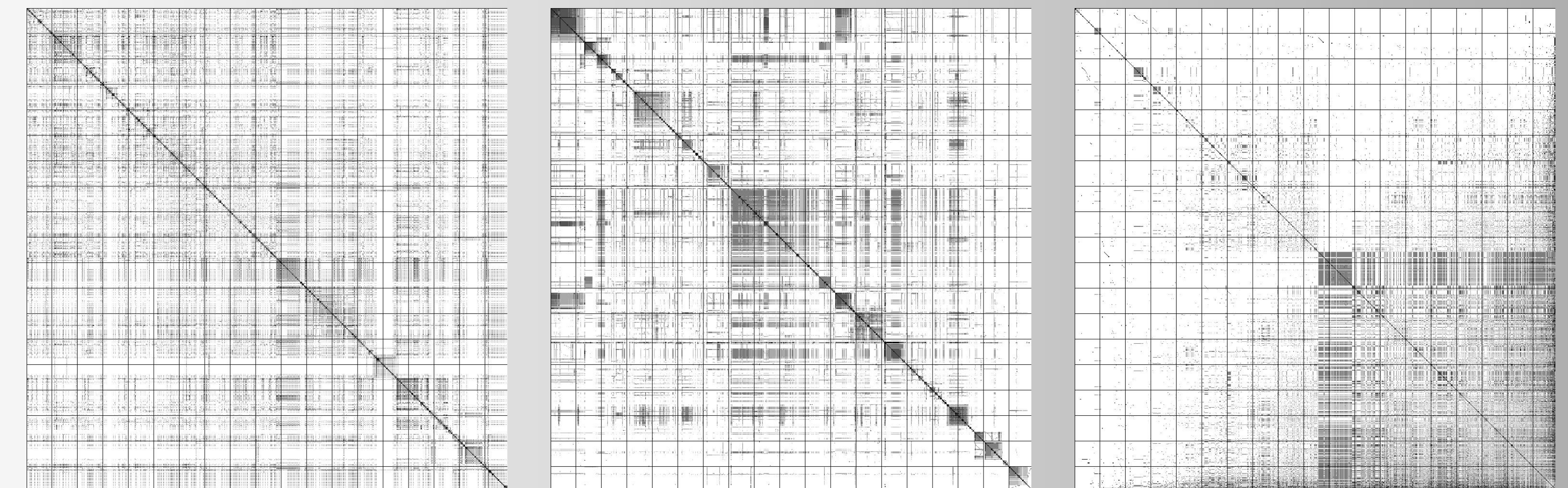
Fizemos um algoritmo para reordenar a rede de modo que os clusters fiquem em evidência. Imaginamos que os nós estão dispostos num espaço unidimensional discreto e que existe uma força atuante entre cada par de nós i e j (para $i \neq j$) da rede. Essa força pode ser de dois tipos: repulsiva, se os nós não compartilharem nenhum vizinho, ou atrativa do tipo massa-mola se os nós tiverem vizinhos em comum. Nesse caso, o *topological overlap* $T_{0,i,j}$ entre os nós atua como a constante elástica da mola.



Analisando a força resultante sobre cada nó, vemos se o nó tem tendência a se deslocar para a esquerda ou para a direita na rede. O algoritmo, então, faz permutações entre os nós - trocando nós à esquerda com tendência de ir para a direita por nós à direita que tenham tendência de ir para a esquerda. Essas permutações visam minimizar a energia potencial associada a essas forças e, além disso, aproximar os nós com maior topological overlap entre si - aproximando os *clusters*. É importante salientar que, o algoritmo não faz mudanças na posição dos nós na rede real, quando falamos em reordenar a rede ou permutações entre os nós - estamos falando de trocas na posição dos índices dos nós nas matrizes que representam as redes.

RESULTADOS

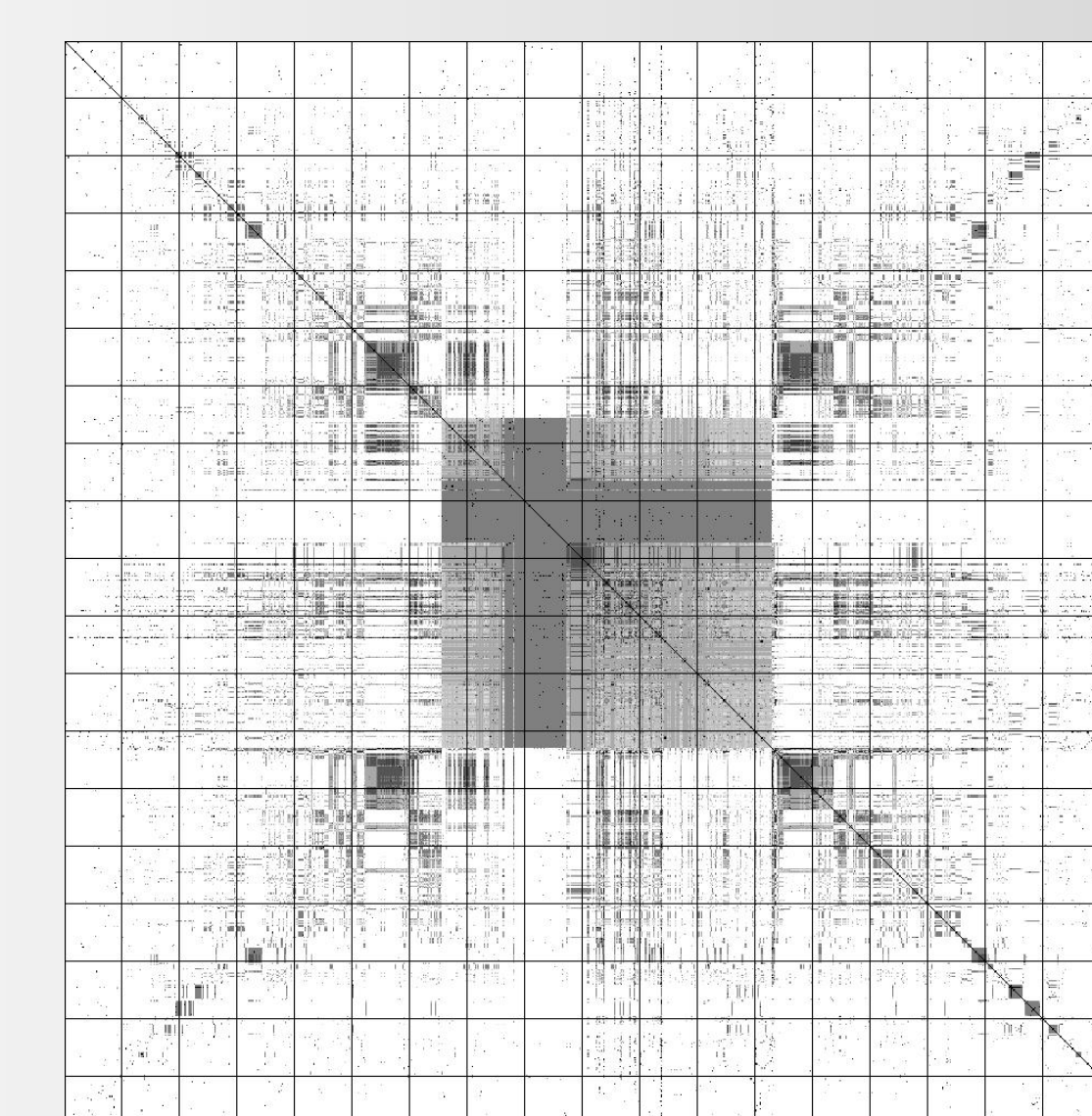
Comparamos o resultado do nosso algoritmo (denotado por MC) com os obtidos pelo PageRank [2] e o Kruskal. O PageRank é o algoritmo que o google usa para ordenar os resultados das pesquisas em seu site. Ele visa obter a relevância de cada nó (no caso, os sites) e reordena a rede de acordo com ela. O Kruskal, por sua vez, calcula a árvore geradora mínima do grafo para ordená-lo.



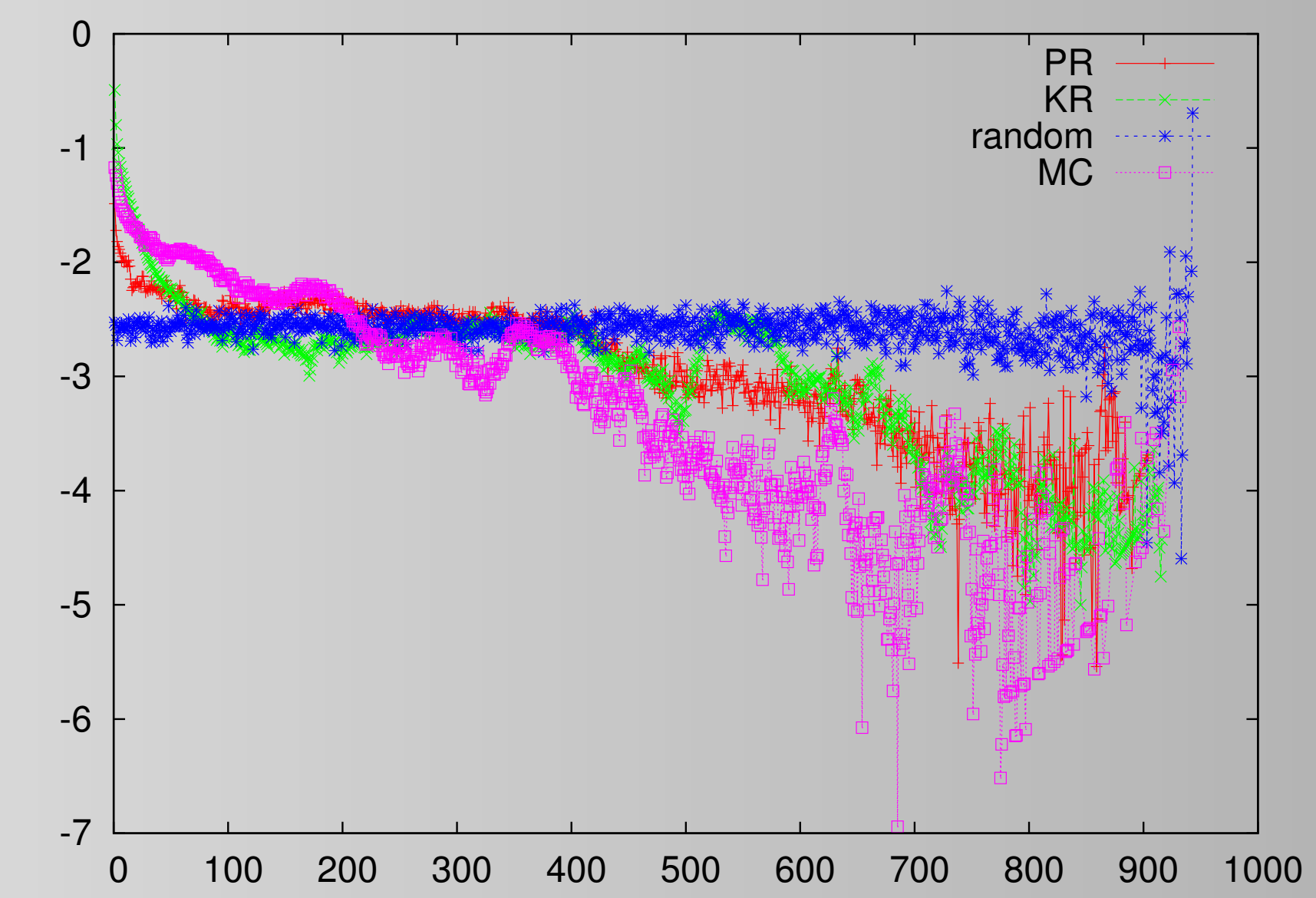
I - Aleatório

II - Kruskal

III -PageRank



IV - MC



V -Comparação

As figuras I,II,III e IV são gráficos do *topological overlap* com ordenamento aleatório (I) e ordenamentos obtidos pelo Kruskal (II), PageRank (III) e pelo nosso algoritmo (IV). O *topological overlap* é uma medida que varia entre 0 e 1 com base na quantidade de vizinho em comum entre um determinado par de nós, é 0, se os nós não tem nenhum vizinho em comum, e 1, se compartilharem todos os vizinhos. No gráfico quanto mais escura uma célula mais perto de 1 o *topological overlap* entre os nós. O gráfico de comparação indica a probabilidade de que uma célula a uma distância n da diagonal esteja preenchida, como representamos em escala logarítmica, vemos que o nosso algoritmo minimiza a chance de haver um nó com grande *topological overlap* longe da diagonal [1].

CONCLUSÃO

O algoritmo que construímos mostrou resultados melhores que o Kruskal e o PageRank para a reorganização da rede, conforme gráfico acima. No entanto, outros estudos são necessários para analisar a relação entre os *clusters* encontrados e as respectivas funções metabólicas.

REFERÊNCIAS

[1] Rybarczyk-Filho, José Luiz et al. *Towards a genome-wide transcriptogram: the Saccharomyces cerevisiae case*. Nucleic Acids Research. 2011;39(8):3005-3016.
 [2] Gallardo, P.F.2005. *Google's secret and linear algebra*.