BFAM - bioinformatics for the functional analysis of mammalian genomes

SPONSORED BY THE

Federal Ministry of Education and Research

A Fast Algorithm for the Iterative Calculation of Betweenness Centrality

J.Griebsch, T. Binder and A. Buchner
Institut für Informatik, Lehrstuhl für Effiziente Algorithmen
Technische Universität München, D-80290 München
{griebsch, buchner}@in.tum.de

Introduction

$$C_B = \sum_{m \in M} \sum_{m' \in M \setminus m} \frac{\sigma_{mm'}(r)}{\sigma_{mm'}}$$

with

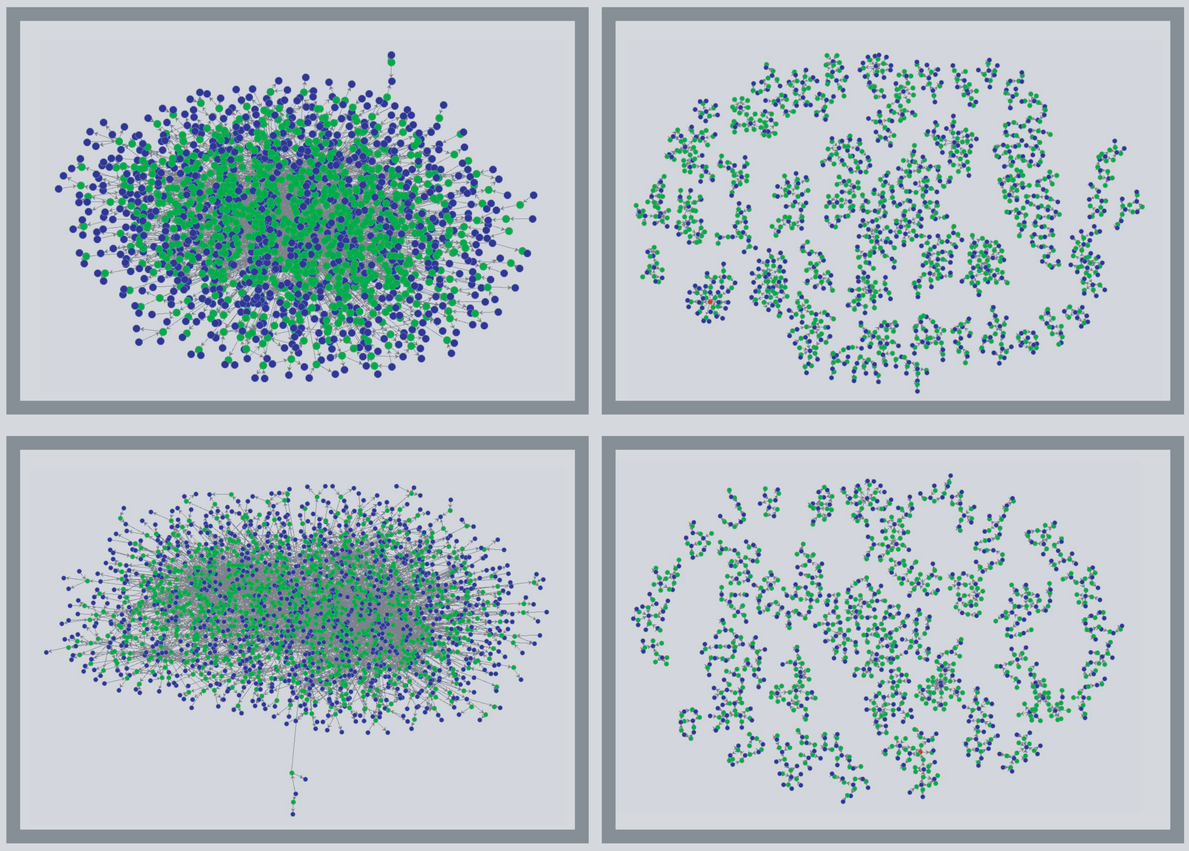
r	a reaction/metabolite
M	the set of substrates
$\sigma_{mm'}$	total number of shortest paths
$\sigma_{mm'}(r)$	number of shortest paths passing through r

Betweenness Centrality Index: Which fraction of all shortest paths runs through a node ?

Various centrality measures capture different notions of a vertex's importance in a network [KS04]. The computationally rather involved betweenness centrality index has also been employed to identify functionally related sub-units/-networks in biochemical networks: Based on the idea that biochemical networks consist of a set of smaller, (often co-regulated) sub-pathways that are connected through central nodes, such functionally related sub-networmay be identified by iteratively removing the most central node in a network and recomputing the centrality of the remaining nodes [HHJ03]. In our experiments such a decomposition of the metabolic network of E. coli (1993 nodes and 3997 edges) took 1.5hours on a 1GHz machine using the fastest known algorithm for the computation of betweenness centrality by Brandes [Br01].

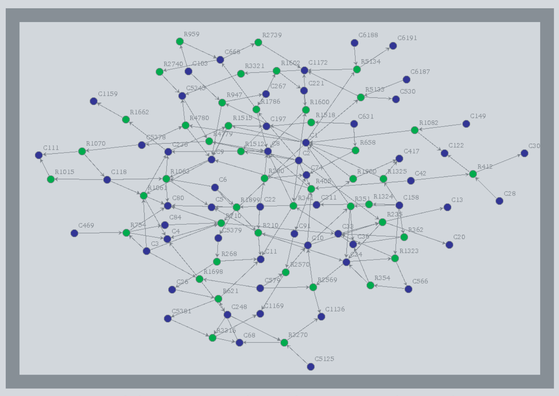
Methods

We use a dynamic variant for the iterative application of the Brandes Algorithm.
In each iteration only one node is removed and, on average, large parts of the shortest paths graphs do not need to be recomputed. Our algorithm uses this observation to restrict the necessary re-computations to the affected nodes - an approach that is employed by a whole class of algorithms, the so called "Dynamic Graph" algorithms [DPZ95,EEI99].
In detail our algorithm works as follows:
The first step is one complete iteration of the Brandes algorithm, but in order to reuse the intermediate results, we store for each start vertex the distances, the number of shortest paths, as well as the list of predecessors for the other nodes. After the identification of the most central vertex we perform a breadth-first-traversal, starting at that node, in order to mark all vertices in which the distances need to be recalculated (which is the case whenever the list of predecessors becomes empty). This step is followed by an update of the respective distance values, number of shortest paths, predecessor lists and the fraction of shortest paths that contain the vertex. In a final step the computation of the new centrality values and the deletion of the most central node complete the iteration of the dynamic algorithm. These steps are repeated until the graph breaks up into several components that might be regarded as functional subunits of the whole network.

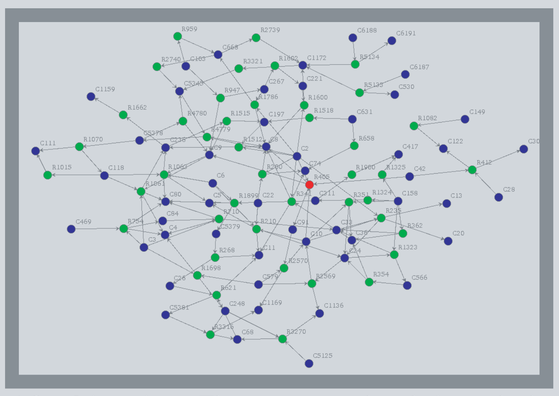


Decomposition of complete metabolic networks from E. coli (above) after 100 and H. sapiens (below) 110 iteration.

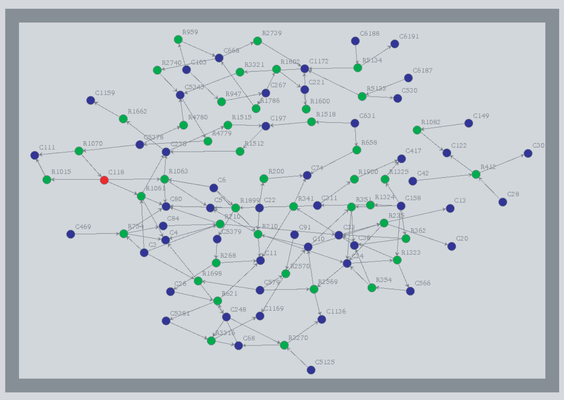
Results



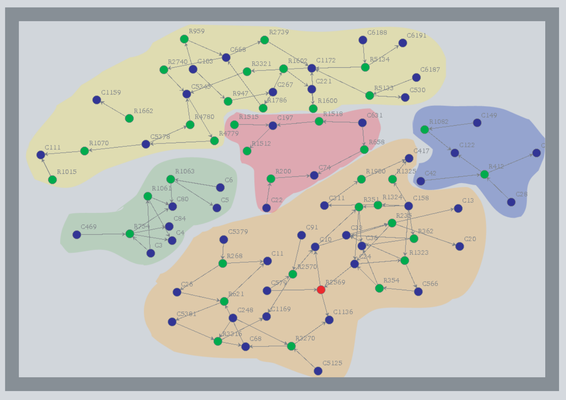
Decomposition example: Citrate Cycle and Glycolysis / Glyconeogenesis from *E.coli*



After 2 iterations: The most central node is marked red.



After 6 iterations



After 13 iterations: The orange component represents the citrate cycle.

Automated separation of two metabolic pathways using betweeness centrality

Conclusions

The worst-case time complexity of the dynamic algorithm is equal compared to the iterated Brandes Algorithm while the observed running time for the complete decomposition of metabolic networks of different species is approximately 3-5 times faster.
However, to achieve this speedup more memory usage is required.

For more details of the algorithm and the decomposition process of metabolic networks please contact one of the above mentioned email adresses.

References

[Br01] Brandes, U.: A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology. 25(2):163-177. 2001.*
[DPZ95] Djidjev, H. N., Pantziou, G. E., and Zaroliagis, C. D.: On-line and Dynamic Algorithms for Shortest Path Problems. *Proc. 12th Symp. Theoretical Aspects of Computer Science. 1995.*
[EGI99] Eppstein, D., Galil Z. and Italiano, G.: Dynamic Graph Algorithms. *Algorithms and Theoretical Computing Handbook .Chapter 8, M. J. Atallah, ed. CRC Press. 1999.*
[HHJ02] Holme, P., Huss, M. and Jeong, H.: Subnetwork Hierarchies of Biochemical Pathways. *Bioinformatics.19(4): 532-538. 2003.*
[KS04] Koschützki, D. and Schreiber, F.: Comparison of Centralities for Biological Networks. *GCB´04 to appear.*

NETWORK OF BIOINFORMATICS COMPETENCE CENTERS NBCC

This work was done within the context of the NGFN Nationales Genomforschungsnetz