LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

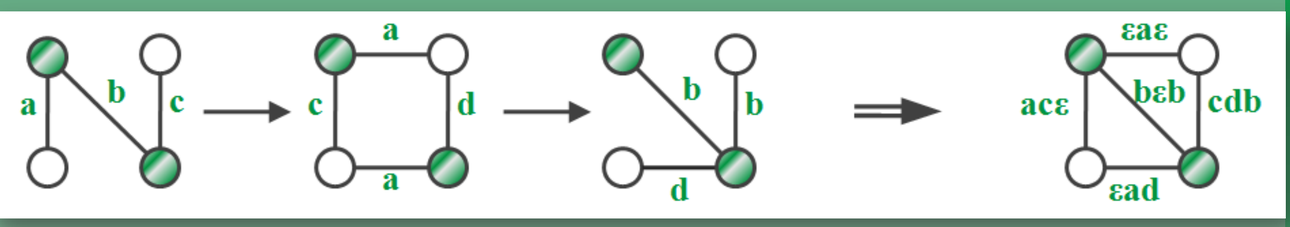
INSTITUTE FOR INFORMATICS

Frequent Subgraph Mining in Dynamic Networks

Contributions

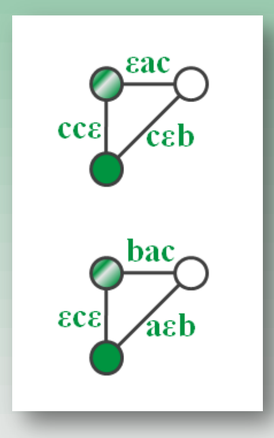
We present a new framework for performing data mining on dynamic networks in an on-top fashion. Existing subgraph mining algorithms on static graphs can be easily integrated into our framework. The efficient search for dynamic patterns inside static frequent sub-graphs is based on the idea of suffix trees.

Transformation into a Dynamic Graph



The three graphs on the left represent a time series of graphs with edge insertions and edge deletions over time. The graph on the right is a dynamic graph that summarizes all information represented by the time series by the use of so called *existence strings*.

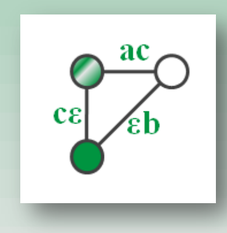
Efficient Dynamic Frequent Subgraph Discovery



Two embeddings of a frequent subgraph.

time point	1	2	3
edge	ε	a	c
edge	c	ε	b
edge	c	c	ε
concatenation	<1:εcc>	<2:aεc>	<3:cbε>
edge	b	a	c
edge	a	ε	b
edge	ε	c	ε
concatenation	<1:baε>	<2:aεc>	<3:cbε>

Matrix representation and common substring.



Dynamic pattern.

Algorithm 1 Dynamic frequent subgraph discovery

Input: All embeddings of one static frequent subgraph S

for each embedding S_i of S do

1) Order the edges of S_i according to canonical labeling
2) Store their existence strings in this order as a matrix
3) Translate each column of the matrix $M(S_i)$ into one character
4) Concatenate these characters into one string

end for

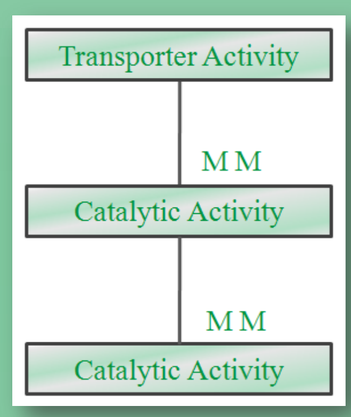
Perform frequent substring discovery on all the resulting strings

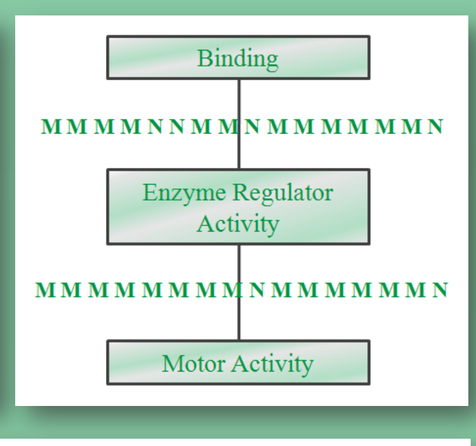
Output: Dynamic patterns within embeddings of S

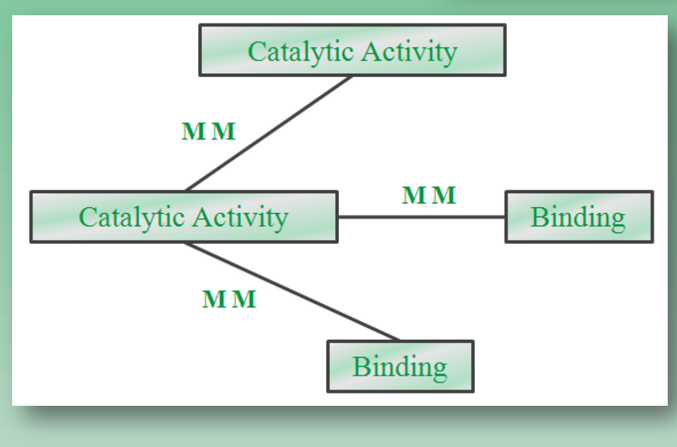
Results on Biological Yeast Data

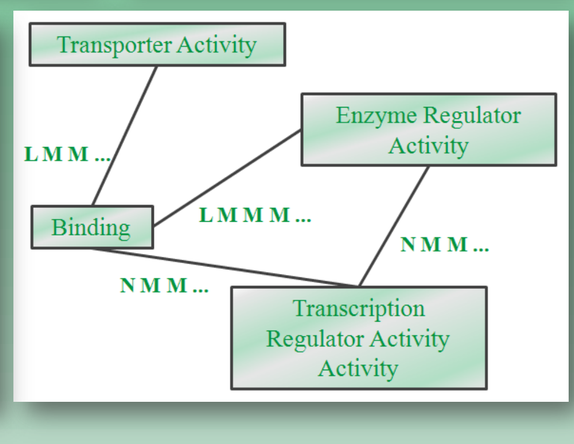
Dynamic network construction by integration of PPI data and a time series of yeast gene expression levels.
Enumeration of all static frequent subgraphs by the use of FANMOD [Wernicke, 2006].
The *p*-value of each static frequent subgraph is determined to asses the significance of the subgraph.
Evolutionary conservation rate indicates the quality of the result.

Most frequent and longest dynamic patterns in subgraphs with 3 and 4 vertices.









B. Wackersreuther[1], P. Wackersreuther[1], A. Oswald[1], C. Böhm[1], K.M. Borgwardt[2]

Database Systems LMU

[1] Department of Computer Science
Database Systems Group

[2] MPIs for Developmental Biology
and Biological Cybernetics