The 13th International Workshop on Structural and Syntactic Pattern Recognition

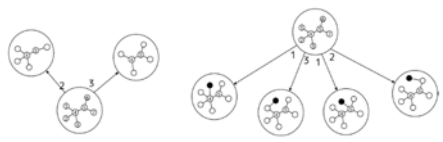
Indexing Tree and Subtree by using a Structure Network

Mingming ZHANG, Shinichiro OMACHI
Graduate School of Engineering, Tohoku University

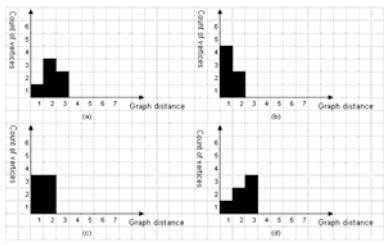
PURPOSE Describing the trees' super-sub relationships with a network, and using a numeric method to access it fast.

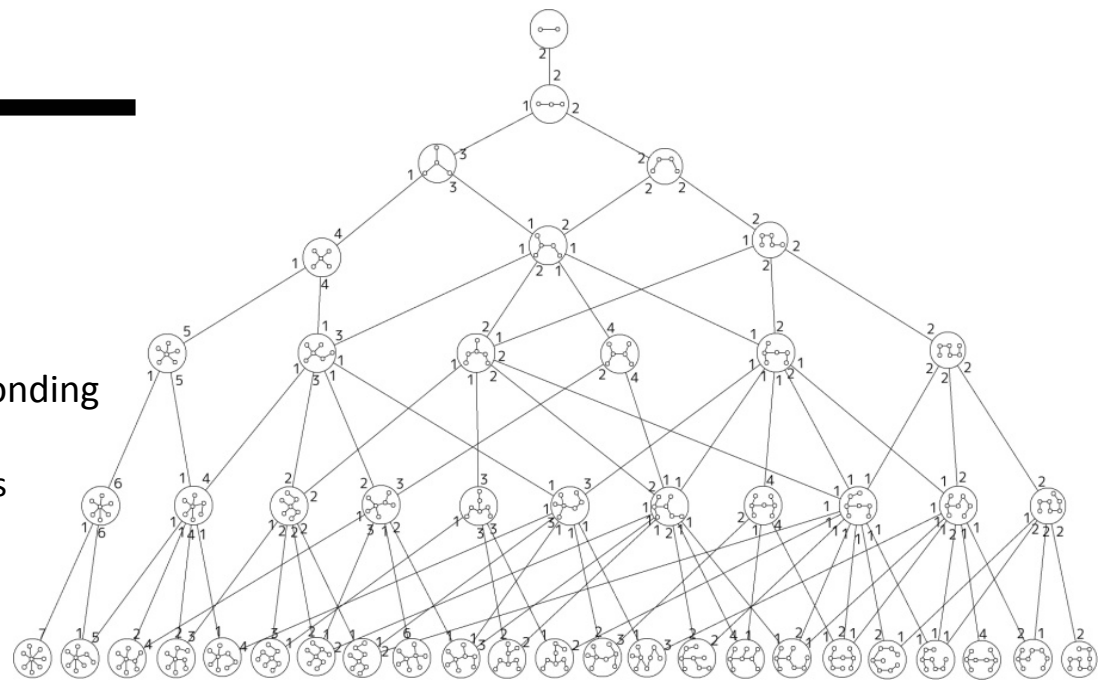
Structure Network

- Structure Network: A network describing the super-sub relationship of trees with the tree structures as the nodes.
- Find out super & sub trees for a given tree
  - Super-tree: Adding an edge
  - Sub-tree: Removing a leaf edge
- The structures of super & sub trees of a tree are corresponding to the results of clustering its vertices
  - Clustering vertices by using the histograms of graph distances



sub-structures super-structures

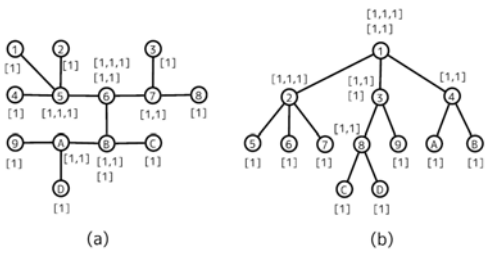




By using this network the trees' super-sub relationships can be represented.

Encoding a Structure

- Each vertex can be labeled with its graph distance histogram
- The labeled results are corresponding to the labeled result in a traditional matching method [1]



(a) (b)

Vertices in Tree (a)	Vertices in Tree (b)	Label	Histogram
1,2,4	5,6,7	[1]	[1,3,2,4,2]
9,D	C,D	[1]	[1,2,2,2,5]
3,8	A,B	[1]	[1,2,2,5,2]
C	9	[1]	[1,2,4,5]
5	2	[1,1,1]	[4,2,4,2]
A	8	[1,1]	[3,2,2,5]
7	4	[1,1]	[3,2,5,2]
B	3	[[1,1],[1]]	[3,4,5]
6	1	[[1,1,1],[1,1]]	[3,7,2]

The clustering result includes all the structure information of a tree

- Encoding the clustering result to represent the structure feature of a tree
  - Encoding each cluster as following field

$$
\begin{aligned}
h &:= \text{histogram sequence;}\\
l &:= \text{length of } h;\\
s &:= \text{size of cluster;}\\
field &:= (l+2), s, h;
\end{aligned}
$$

- Link all cluster fields to make a long numeric array that can represent the structure of a tree

$$
\begin{aligned}
(\ &\mathbf{5},\mathbf{1},\ 3,4,5,\mathbf{5},\mathbf{1},3,7,2,\mathbf{6},\mathbf{1},1,2,4,5,\mathbf{6},\mathbf{1},3,\\
&2,2,\ 5,\mathbf{6},\mathbf{1},3,2,5,2,\mathbf{6},\mathbf{1},4,2,4,2,\mathbf{7},\mathbf{2},1,2,\\
&2,2,\ 5,\mathbf{7},\mathbf{2},1,2,2,5,2,\mathbf{7},\mathbf{3},1,3,2,4,2)
\end{aligned}
$$

Experiments & Results

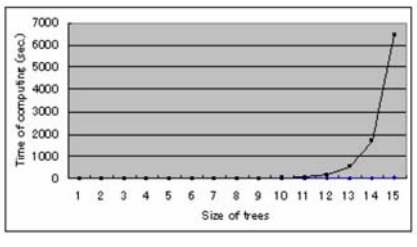
- Comparing the isomorphic ability with the traditional matching method[1]
  - 20,000 trees
  - Up to 20 vertices
  - Clustering them by their structures

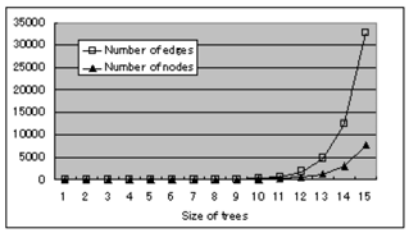
Correct Rate: 100%

Time complexity 9.5 days → 45 mins

Evaluation for constructing the Structure Network

- Either time and spatial complexities of constructing network are exponential increases
- The structure network need to be constructed **only once**





[1] Köbler, J., Schöning, U. and Torán, J.: "The graph isomorphism problem: its structural complexity", Springer, 1993