Graph Matching

CSE 373
Data Structures

Input: 2 digraphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$

Questions to ask:
1. Are $G_1$ and $G_2$ isomorphic?
2. Is $G_1$ isomorphic to a subgraph of $G_2$?
3. How similar is $G_1$ to $G_2$?
4. How similar is $G_1$ to the most similar subgraph of $G_2$?
Isomorphism for Digraphs

$G_1$ is isomorphic to $G_2$ if there is a 1-1, onto mapping $h: V_1 \rightarrow V_2$ such that

$(v_i,v_j) \in E_1 \iff (h(v_i), h(v_j)) \in E_2$

Find an isomorphism $h: \{1,2,3,4,5\} \rightarrow \{a,b,c,d,e\}$. Check that the condition holds for every edge.

Subgraph Isomorphism for Digraphs

$G_1$ is isomorphic to a subgraph of $G_2$ if there is a 1-1 mapping $h: V_1 \rightarrow V_2$ such that

$(v_i,v_j) \in E_1 \Rightarrow (h(v_i), h(v_j)) \in E_2$

Isomorphism and subgraph isomorphism are defined similarly for undirected graphs. In this case, when $(v_i,v_j) \in E_1$, either $(v_i,v_j)$ or $(v_j,v_i)$ can be listed in $E_2$, since they are equivalent and both mean $\{v_i,v_j\}$. 
**Similar Digraphs**

Sometimes two graphs are close to isomorphic, but have a few “errors.”

Let \( h(1)=b, h(2)=e, h(3)=c, h(4)=a, h(5)=d \).

\[
\begin{align*}
(1,2) & \quad (b,e) \\
(2,1) & \quad (e,b) \\
(4,5) & \quad (a,d) \\
(2,5) & \quad (e,d) \\
(3,2) & \quad \times \\
(3,4) & \quad (c,a)
\end{align*}
\]

The mapping \( h \) has 2 errors.

- \((c,b) \in G_2\), but \((3,1) \notin G_1\)
- \((3,2) \in G_1\), but \((c,e) \notin G_2\)

**Error of a Mapping**

Intuitively, the error of mapping \( h \) tells us
- how many edges of \( G_1 \) have no corresponding edge in \( G_2 \) and
- how many edges of \( G_2 \) have no corresponding edge in \( G_1 \).

Let \( G_1=(V_1,E_1) \) and \( G_2=(V_2,E_2) \), and let \( h:V_1 \rightarrow V_2 \) be a 1-1, onto mapping.

- **forward error** \( EF(h) = |\{(v_i,v_j) \in E_1 \mid (h(v_i),h(v_j)) \notin E_2\}| \)
- **backward error** \( EB(h) = |\{(v_i,v_j) \in E_2 \mid (h^{-1}(v_i),h^{-1}(v_j)) \notin E_1\}| \)
- **total error** \( Error(h) = EF(h) + EB(h) \)

**relational distance** \( GD(G_1,G_2) = \min \ Error(h) \)

for all 1-1, onto \( h:V_1 \rightarrow V_2 \)
Variations of Relational Distance

1. normalized relational distance:
   Divide by the sum of the number of edges in E1 and those in E2.

2. undirected graphs:
   Just modify the definitions of EF and EB to accommodate.

3. one way mappings:
   h is 1-1, but need not be onto
   Only the forward error EF is used.

4. labeled graphs:
   When nodes and edges can have labels, each node should be mapped to a node with the same label, and each edge should be mapped to an edge with the same label.

Graph Matching Algorithms

1. graph isomorphism
2. subgraph isomorphism
3. relational distance
4. attributed relational distance (uses labels)

Subgraph Isomorphism

Given model graph $M = (VM, EM)$
   data graph $D = (VD, ED)$

Find 1-1 mapping $h : VM \rightarrow VD$

satisfying $(vi, vj) \in EM \Rightarrow (h(vi), h(vj)) \in ED.$
Method: Backtracking Tree Search

![Diagram of a tree search algorithm]

procedure Treesearch(VM, VD, EM, ED, h)
{
  v = first(VM);
  for each w ∈ VD
    {
      h' = h ∪ {(v,w)}; //add to set
      OK = true;
      for each edge (vi,vj) in EM
      {
        if one of vi or vj is v and the other has been assigned a value in h'
        if ( h'(vi),h'(vj) ) is NOT in ED )
        { OK = false; break; };
      }
      if OK
      {
        VM' = VM – v; //remove from set
        VD' = VD – w'
        if isempty(VM') output(h');
        else Treesearch(VM',VD',EM,ED,h')
      }
    }
}
Branch-and-Bound Tree Search

Keep track of the least-error mapping.

M

D

map_err = 0
bound_err = 99999

map_err = 0
map_err = 0
map_err = 0
map_err = 0

map_err = 1
map_err = 1
map_err = 1
map_err = 0; bound_err = 1

mapping = {(1,a),(2,b),(3,c)}
mapping = {(1,b),(2,d),(3,c)}