Subgraph Isomorphism and Related Problems

Subgraph isomorphism is an important and very general form of pattern matching that finds practical application in areas such as pattern recognition and computer vision, computer-aided design, image processing, graph grammars, graph transformation, and biocomputing.

In this talk, several problems related to subgraph isomorphism will be discussed and recent results relating subgraph isomorphism, maximum common subgraph, minimum common supergraph, and graph distance will be reviewed.

- Read, R. C. and Corneil, D. G. (1977). The graph isomorphism disease. Journal of Graph Theory 1, 339–363.

- Gati, G. (1979). Further annotated bibliography on the isomorphism disease. Journal of Graph Theory 3, 95–109.

Subgraph Isomorphism and Related Problems

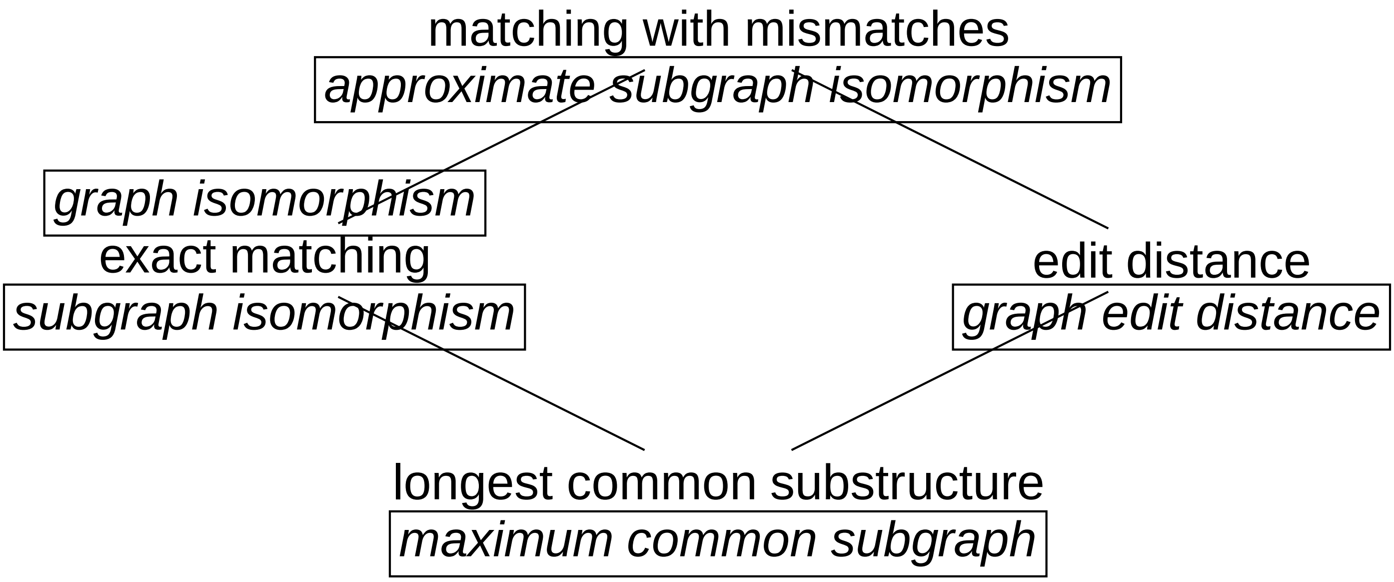
- Mathematical motivation
  - NP-complete problems are a challenge to theoretical computer science

- Non-mathematical motivation
  - Pattern recognition and computer vision
  - Computer-aided design
  - Image processing
  - Graph grammars and graph transformation
  - Biocomputing

- Subgraph isomorphism is an important and very general form of *exact* pattern matching
  - String searching
  - Sequence alignment
  - Tree comparison
  - Pattern matching on graphs

Subgraph Isomorphism and Related Problems

- A hierarchy of pattern matching problems
  - Graph isomorphism
  - *Subgraph isomorphism*
  - Maximum common subgraph
  - Approximate subgraph isomorphism
  - Graph edit distance



Subgraph Isomorphism and Related Problems

- Given a *pattern* G and a *text* H
  - Decision problem
    Answer whether H contains a subgraph isomorphic to G
  - Search problem
    Return an occurrence of G as a subgraph of H
  - Counting problem
    Return a count of the number of subgraphs of H that are isomorphic to G
  - Enumeration problem
    Return all occurrences of G as a subgraph of H

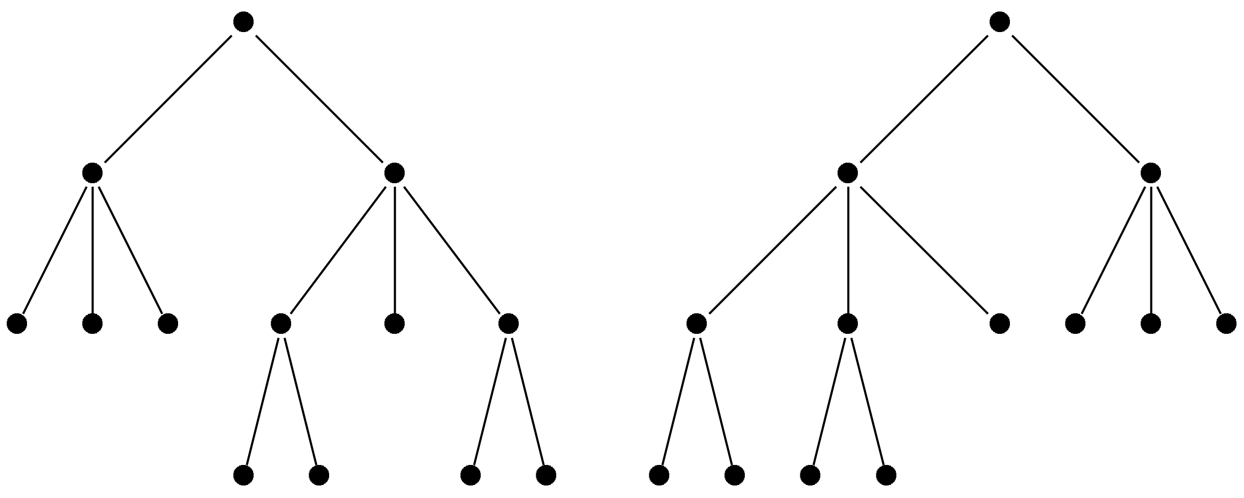
- Given a *pattern* G and a *text* H
  - General problem
    Both G and H are input graphs
  - Restricted problem
    Both G and H are input graphs belonging to a particular class, such as trees or planar graphs
  - Fixed problem
    G is an input graph but H is a fixed graph, or vice versa

Subgraph Isomorphism and Related Problems

Definition. *Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are* isomorphic*, denoted by $G_1 \cong G_2$, if there is a bijection $\varphi : V_1 \to V_2$ such that, for every pair of vertices $v_i, v_j \in V_1$, $(v_i, v_j) \in E_1$ if and only if $(\varphi(v_i), \varphi(v_j)) \in E_2$*

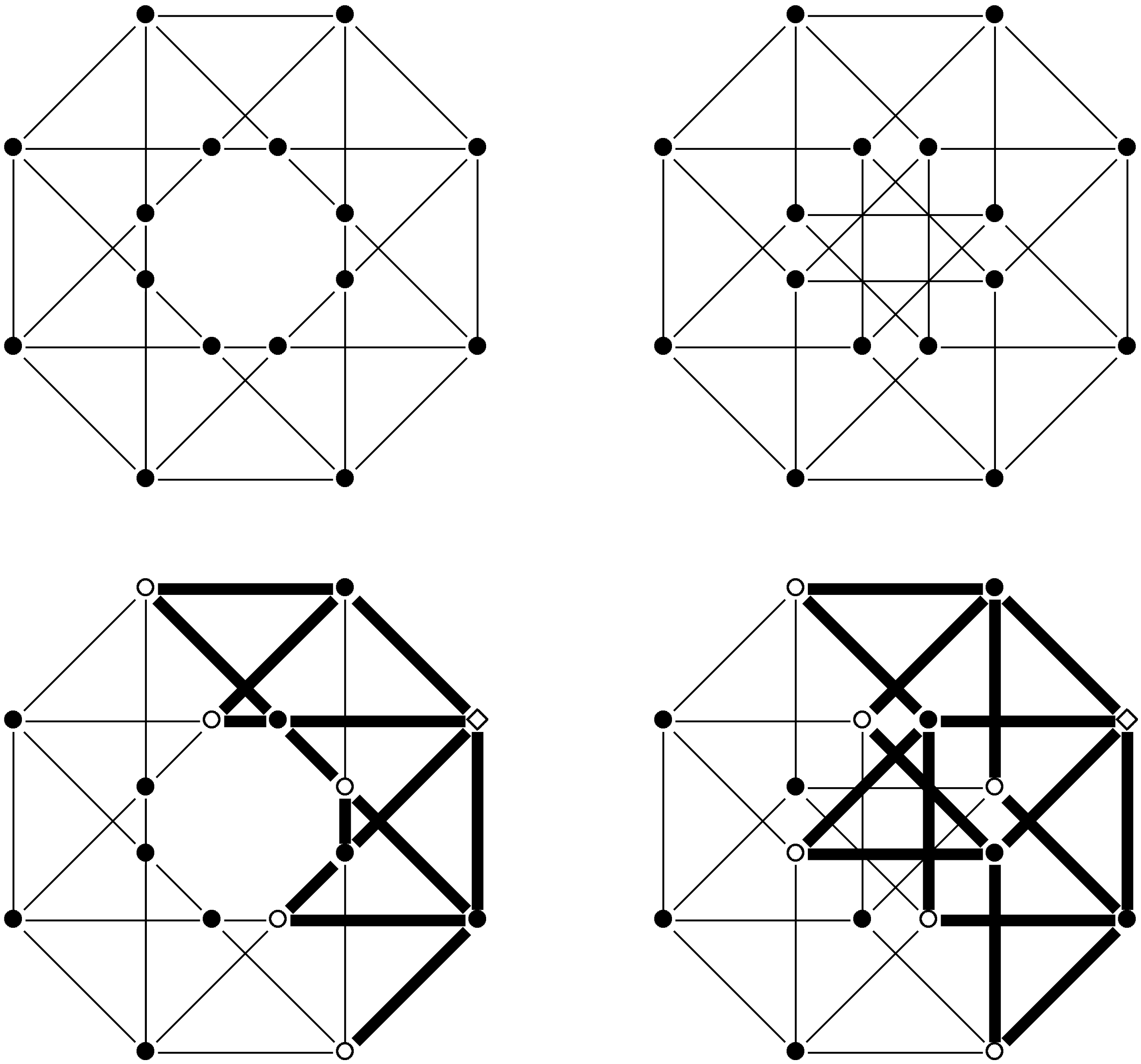
- For input graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $V_1 = \{u_1, \ldots, u_n\}$ and $V_2 = \{v_1, \ldots, v_n\}$, a necessary condition for $G_1 \cong G_2$ is that the multisets $\{\Gamma(u_i) \mid 1 \leqslant i \leqslant n\}$ and $\{\Gamma(v_i) \mid 1 \leqslant i \leqslant n\}$ be equal

Example. *Tree isomorphism*



The Subgraph Isomorphism Problem

Example. *Graph isomorphism*



Subgraph Isomorphism and Related Problems

Definition. *A graph* $G_1 = (V_1, E_1)$ *is* isomorphic to a subgraph *of a graph* $G_2 = (V_2, E_2)$*, denoted by* $G_1 \cong S_2 \subseteq G_2$*, if there is an injection* $\varphi : V_1 \to V_2$ *such that, for every pair of vertices* $v_i, v_j \in V_1$*, if* $(v_i, v_j) \in E_1$ *then* $(\varphi(v_i), \varphi(v_j)) \in E_2$

- For input graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, vertex $u_i \in V_1$ cannot be mapped by a subgraph isomorphism to vertex $v_j \in V_2$ unless $\deg(u_i) \leqslant \deg(v_j)$, for all $1 \leqslant i \leqslant n_1$ and $1 \leqslant j \leqslant n_2$

SUBGRAPH ISOMORPHISM

INSTANCE Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$

QUESTION Does G_1 contain a subgraph isomorphic to G_2?

Reference Transformation from CLIQUE

Comment Contains CLIQUE, COMPLETE BIPARTITE SUBGRAPH, HAMILTONIAN CIRCUIT as special cases

Subgraph Isomorphism and Related Problems

Definition. *A* common subgraph *of two graphs G_1 and G_2 consists of a subgraph H_1 of G_1 and a subgraph H_2 of G_2 such that $H_1 \cong H_2$. The* maximum common subgraph *of two graphs is a common subgraph that is not a proper subgraph of another common subgraph*

- The *maximum* common subgraph is the largest possible common subgraph, while a common subgraph is *maximal* if it cannot be extended to another common subgraph by the addition of vertices or edges

MAXIMUM COMMON SUBGRAPH

INSTANCE Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, positive integer K

QUESTION Do there exist subsets $E_1' \subseteq E_1$ and $E_2' \subseteq E_2$ with $|E_1'| = |E_2'| \geqslant K$ such that the two subgraphs $G_1' = (V_1, E_1')$ and $G_2' = (V_2, E_2')$ are isomorphic?

Reference Transformation from CLIQUE

Subgraph Isomorphism and Related Problems

Definition. *A* common subgraph *of two graphs G_1 and G_2 consists of a subgraph H_1 of G_1 and a subgraph H_2 of G_2 such that $H_1 \cong H_2$. The* maximum common subgraph *of two graphs is a common subgraph that is not a proper subgraph of another common subgraph*

- The *maximum* common subgraph is the largest possible common subgraph, while a common subgraph is *maximal* if it cannot be extended to another common subgraph by the addition of vertices or edges

MAXIMUM COMMON SUBGRAPH

INSTANCE Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$

SOLUTION A common subgraph: graphs $G_1' \subseteq G_1$ and $G_2' \subseteq G_2$ such that G_1' and G_2' are isomorphic

MEASURE Size of the common subgraph

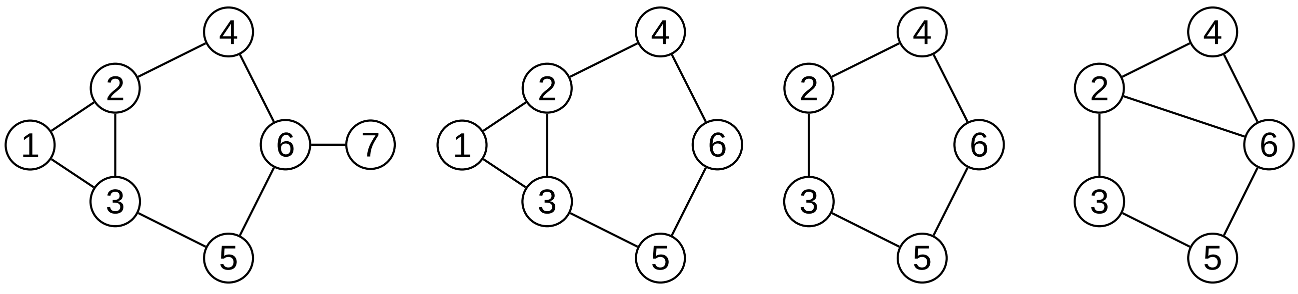
Subgraph Isomorphism and Related Problems

Definition. *The* edit distance *between two graphs is the shortest or the least cost sequence of elementary graph edit operations that transform one graph into the other*

- Elementary edit operations include
  - rotation
  - substitution
  - deletion
  - insertion

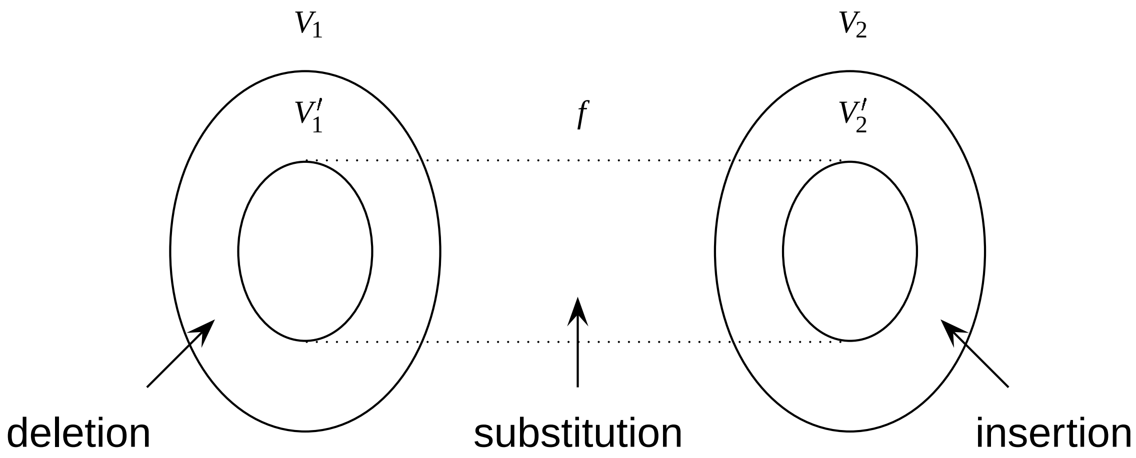
  of vertices and edges

Example. *Computing edit distance by deletion and insertion*



Subgraph Isomorphism and Related Problems

Definition. *An* approximate graph matching *from a graph G_1 to a graph G_2 is a bijective function $f : V_1' \to V_2'$, where $V_1' \subseteq V_1$ and $V_2' \subseteq V_2$*



Definition. *A* cost function *is a tuple $C = (c_{vd}, c_{vi}, c_{vs}, c_{es})$ of nonnegative real numbers*

- c_{vd}, c_{vi}, c_{vs} model the cost of vertex deletion, insertion, substitution
- c_{es} models the cost of edge substitution

Edge deletion cost c_{ed} and edge insertion cost c_{ei} are assumed to be included in the costs of the corresponding vertex deletions and insertions

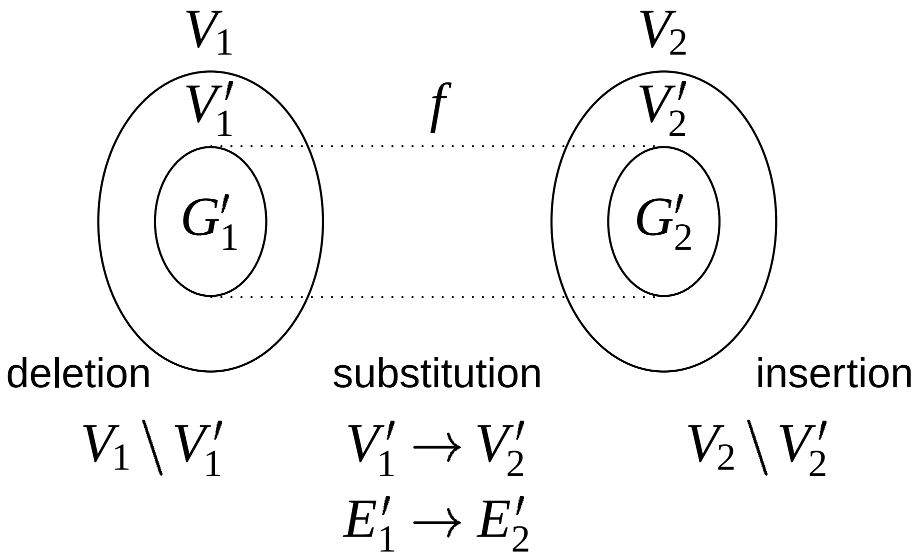
Subgraph Isomorphism and Related Problems

Definition. *The* cost *of an approximate graph matching $f : V_1' \to V_2'$ from a graph G_1 to a graph G_2 is given by*

$$
\begin{aligned}
\gamma_C(f) \quad &= \quad \sum_{v \in V_1 \setminus V_1'} c_{vd}(v) + \sum_{v \in V_2 \setminus V_2'} c_{vi}(v) \\
&\quad + \sum_{v \in V_1'} c_{vs}(v) + \sum_{e \in E_1'} c_{es}(e)
\end{aligned}
$$

where $C = (c_{vd}, c_{vi}, c_{vs}, c_{es})$ is a cost function

Subgraph Isomorphism and Related Problems



The costs $c_{vd}(v), c_{vi}(v), c_{vs}(v), c_{es}(v)$ correspond to

- deleting a vertex $v \in V_1 \setminus V_1'$ from G_1
- inserting a vertex $v \in V_2 \setminus V_2'$ into G_2
- substituting a vertex $v \in V_1'$ by $f(v) \in V_2'$
- substituting an arc $e_1 = (u, v) \in E_1'$ by $e_2 = (f(u), f(v)) \in E_2'$

where $G_1' \sqsubseteq G_1$ and $G_2' \sqsubseteq G_2$

Subgraph Isomorphism and Related Problems

Definition. *The* edit distance *between two graphs G_1 and G_2 is the (cost of the) least cost* approximate graph matching *from G_1 to G_2*

$$\delta(G_1, G_2) = \min\{\gamma_C(f) \mid f : G_1 \to G_2\}$$

Definition. *A distance function δ over graphs is a* metric *if it satisfies*

- δ *is positive definite:*
  - $\delta(G_1, G_2) \geqslant 0$
  - $\delta(G_1, G_2) = 0$ *if and only if* $G_1 \cong G_2$
- δ *is symmetric:* $\delta(G_1, G_2) = \delta(G_2, G_1)$
- δ *is triangular:* $\delta(G_1, G_2) \leqslant \delta(G_1, G_3) + \delta(G_3, G_2)$

A distance metric is useful for searching in a metric space

Subgraph Isomorphism and Related Problems

GRAPH EDIT DISTANCE

INSTANCE Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, positive integer K

QUESTION Is $\delta(G_1, G_2) \leqslant K$?

MINIMUM GRAPH TRANSFORMATION

INSTANCE Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$

SOLUTION A transformation that makes G_1 isomorphic to G_2

MEASURE Number of edges removed from E_1 and added to E_2

Subgraph Isomorphism and Related Problems

- Subgraph isomorphism is NP-complete [GT48]
  - Restriction to planar graphs remains NP-complete
  - Fixed planar subgraph isomorphism is in P
- Maximum common subgraph is NP-complete [GT49]
  - Approximation is APX-hard [GT46]
  - Restriction to graphs of bounded degree is in APX
- Graph edit distance is NP-complete
  - Approximation is APX-hard [GT49]

Subgraph Isomorphism and Related Problems

- Bunke, H. (1997). On a relation between graph edit distance and maximum common subgraph. Pattern Recogn. Lett. 18, 8, 689–694.

  – The graph edit distance coincides with

  $$\delta(G_1, G_2) = |V_1| + |V_2| - 2|\hat{V}_{12}|$$

  if the cost function is such that

  $$\begin{aligned} c_{vd} &= c_{vi} = 1 \\ c_{vs} &= c_{es} = \infty \end{aligned}$$

- Bunke, H. and Schearer, K. (1998). A graph distance metric based on the maximal common subgraph. Pattern Recogn. Lett. 19, 3–4, 255–259.

  – The graph distance measure given by

  $$\delta(G_1, G_2) = 1 - \frac{|\hat{V}_{12}|}{\max(|V_1|, |V_2|)}$$

  is a metric

Subgraph Isomorphism and Related Problems

- Bunke, H. (1999). Error-correcting graph matching: On the influence of the underlying cost function. IEEE T. Pattern Anal. 21, 9, 917–922.

  – The graph edit distance is a metric if the cost function is such that

$$c_{vd} + c_{vi} \leqslant c_{vs}$$
$$c_{vd} + c_{vi} \leqslant c_{es}$$

- Messmer, B. T. and Bunke, H. (1999). A decision tree approach to graph and subgraph isomorphism detection. Pattern Recogn. 32, 12, 1979–1998.

  – Fixed graph and subgraph isomorphism is dealt with by storing all permutation matrices of the fixed graph in a decision tree

  – Computational complexity is time $\Theta(n_1^3)$ and space $\Theta(3^{n_2})$ after preprocessing time $\Theta(n_2^{n_2})$

Subgraph Isomorphism and Related Problems

- Eppstein, D. (1999). Subgraph isomorphism in planar graphs and related problems. Journal of Graph Algorithms and Applications 3, 3, 1–27.

  - Fixed planar subgraph isomorphism is dealt with by partitioning the planar graph into pieces of small tree width, and applying dynamic programming within each piece
  - Computational complexity is $\Theta(n_2)$

- Cortadella, J. and Valiente, G. (2000). A relational view of subgraph isomorphism. In Proc. 5th Int. Seminar on Relational Methods in Computer Science (Québec, Canada, 2000), pp. 45–54.

  - An explicit representation of the relation containing all and only all subgraph isomorphisms is built by intersection of binary relations
  - Space efficiency is achieved by using symbolic techniques

Subgraph Isomorphism and Related Problems

- Larrosa, J. and Valiente, G. (2001). Graph pattern matching using constraint satisfaction. To appear in Math. Structures Comput. Sci.
  - Neighborhood constraints are exploited for domain filtering
  - The new algorithm never visits more nodes than *really full look-ahead* and than *forward checking* using degree constraints and structure constraints
  - A benchmark for subgraph isomorphism is proposed

- Fernández, M.-L. and Valiente, G. (2001). A graph distance measure combining maximum common subgraph and minimum common supergraph. To appear in Pattern Recogn. Lett.
  - The graph distance measure given by

    $$\delta(G_1, G_2) = |\check{G}_{12}| - |\hat{G}_{12}|$$

    is a metric, where $|G| = |V| + |E|$

Subgraph Isomorphism and Related Problems

- Wilson, R. C., Evans, A. N., and Hancock, E. R. (1995). Relational matching by discrete relaxation. Image and Vision Computing 13, 5, 411–421.

  - Approximate subgraph isomorphism is dealt with as a nonlinear optimization problem for a global measure of relational consistency
  - A Bayesian measure of relational consistency is based on the sum of the matching probabilities over $\Gamma(v)$ for all $v \in V_1$

- Gold, S. and Rangarajan, A. (1996). A graduated assignment algorithm for graph matching. IEEE T. Pattern Anal. 18, 4, 377–388.

  - Approximate subgraph isomorphism is dealt with as a nonlinear optimization problem
  - The algorithm uses a "continuation method" to transform the discrete assignment problem into a continuous problem, in order to avoid poor local minima
  - Computational complexity is $O(m_1 m_2)$

Subgraph Isomorphism and Related Problems

- Cross, A. D. J., Wilson, R. C., and Hancock, E. R. (1997). Inexact graph matching using genetic search. Pattern Recogn. 30, 6, 953–970.

  - Approximate subgraph isomorphism is dealt with as a nonlinear optimization problem for a global Bayesian measure of relational consistency
  - The crossover process is realized at the level of subgraphs, rather than using string-based or random crossover
  - Empirical results show
    - Polynomial convergence time
    - Convergence rate more rapid than simulated annealing

Subgraph Isomorphism and Related Problems

- Messmer, B. T. and Bunke, H. (1998). A new algorithm for error-tolerant subgraph isomorphism detection. IEEE T. Pattern Anal. 20, 5, 493–504.

  - Fixed approximate subgraph isomorphism is dealt with by storing a recursive decomposition of a set of fixed graphs
  - Common subgraphs of different fixed graphs are represented only once
  - The method is only sublinearly dependent on the number of fixed graphs

- El-Sonbaty, Y. and Ismail, M. A. (1998). A new algorithm for subgraph optimal isomorphism. Pattern Recogn. 31, 2, 205–218.

  - Approximate subgraph isomorphism is dealt with as minimum weighted bipartite matching of decomposed subgraphs
  - Graph G_1 is decomposed into n_1 subgraphs
  - Graph G_2 is decomposed into n_2 subgraphs
  - Computational complexity is average case $\Theta(n_1^2 n_2^2)$, worst case $\Theta(n_1^2 n_2^2 \min(n_1, n_2))$
  - Hidden weight of structure preservation

Subgraph Isomorphism and Related Problems

- Special cases
  - Restriction to planar graphs remains NP-complete
    - Planar clique is in P
    - Planar Hamiltonian circuit is NP-complete
    - Fixed planar subgraph isomorphism is in P
  - Restriction to chordal graphs
    - Chordal clique is in P
    - Bounded degree clique is in P
  - Restriction to interval graphs
  - Restriction to graphs of bounded degree
    - Bounded degree clique is in P

Subgraph Isomorphism and Related Problems

- Approximation algorithms
  - Most algorithms for approximate subgraph isomorphism and related problems are *not* approximation algorithms
    - Approximate solutions are empirically shown to be close to the optimum, only for particular problem instances
  - Theoretical analysis of existing algorithms for approximate subgraph isomorphism and related problems
  - Polynomial-time approximation algorithms with bounded absolute or relative error (for special cases)
  - Polynomial-time approximation algorithms with bounded input-dependent relative error