Min, Max and PTIME Anti-Monotonic Overlap Graph Measures

Toon Calders[1] Jan Ramon[2] Dries Van Dyck[3]

[1]Eindhoven University of Technology,
Department of Mathematics and Computer Science,
[2]Katholieke Universiteit Leuven, Department of Computer Science,
[3]Hasselt University & Transnational University of Limburg
Department of Mathematics, Physics & Computer Science,
Theoretical Computer Science Group,

6th International Workshop on Mining and Learning with Graphs (MLG-2008)
July 4–5, Helsinki, Finland

Introduction

Single graph mining

Mine frequent patterns in single, typically large graph, e.g.,

- worldwide web graph
- social networks
- biological networks
- citation graphs

Patterns are graphs themselves

Introduction

Single graph mining

Mine frequent patterns in single, typically large graph, e.g.,

- worldwide web graph
- social networks
- biological networks
- citation graphs

Patterns are graphs themselves

Main question

How define frequency of a graph in a graph?

Not as easy as it seems!

Anti-monotonicity

Anti-monotonicity

Support measure f is anti-monotonic if frequency of superpattern P is never higher than frequency of subpattern p:

$$\forall p, P, G \mid p \subseteq P : f(p, G) \geq f(P, G)$$

$\Rightarrow$ Algorithms rely critically on anti-monotonicity to prune search space

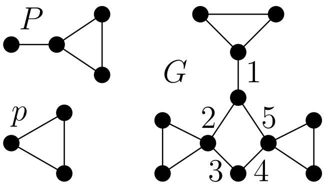
Anti-monotonicity

Anti-monotonicity

Support measure f is anti-monotonic if frequency of superpattern P is never higher than frequency of subpattern p:

$$\forall p, P, G \mid p \subseteq P : f(p, G) \geq f(P, G)$$

$\Rightarrow$ Algorithms rely critically on anti-monotonicity to prune search space



Counting instances does not work!

- $f(p, G) = 3 < f(P, G) = 5$

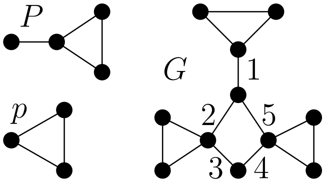
Anti-monotonicity

Anti-monotonicity

Support measure f is anti-monotonic if frequency of superpattern P is never higher than frequency of subpattern p:

$$\forall p, P, G \mid p \subseteq P : f(p, G) \geq f(P, G)$$

$\Rightarrow$ Algorithms rely critically on anti-monotonicity to prune search space



Counting instances does not work!

- $f(p, G) = 3 < f(P, G) = 5$
- frequency P higher than frequency $p \subseteq P$!
- not anti-monotonic

Overlap graphs

Most anti-monotonic support measures in single graph mining are based on overlap graph:

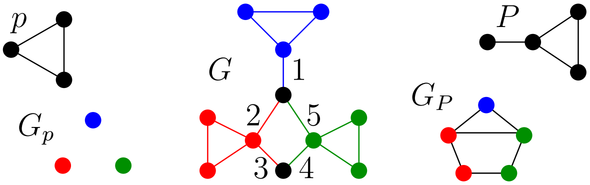
- Undirected graph G_P with instances of pattern P as vertices
- Edge between vertices if correspondig instances overlap in G
- Two types of overlap: edge or vertex overlap

Overlap graphs

Most anti-monotonic support measures in single graph mining are based on overlap graph:

- Undirected graph G_P with instances of pattern P as vertices
- Edge between vertices if correspondig instances overlap in G
- Two types of overlap: edge or vertex overlap

Example for vertex overlap:



Extension to labeled or directed graphs straightforward

Relation between overlap graphs

N. Vanetik, S. E. Shimony, E. Gudes, "Support measures for graph data", Data Min. Knowl. Discov., Vol. 13 (2), 2006, 243–260:

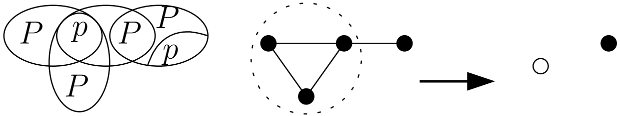
Observations

- Each instance of P contains instance of p
- G_P can be transformed into G_p by three operations:
  1. Clique Contraction (CC)
  2. Vertex Addition (VA)
  3. Edge Removal (ER)

Clique Contraction

Clique Contraction (CC)

contracts clique formed by multiple instances of P sharing single instance of p to single vertex, only overlap with all P-instances of clique remains

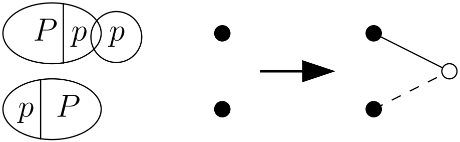


N. Vanetik, S. E. Shimony, E. Gudes, "Support measures for graph data", Data Min. Knowl. Discov., Vol. 13 (2), 2006, 243–260

Vertex Addition

Vertex Addition (VA)

adds instance of p not part of instance of P, assumes overlap with all other instances



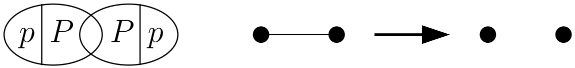
N. Vanetik, S. E. Shimony, E. Gudes, "Support measures for graph data", Data Min. Knowl. Discov., Vol. 13 (2), 2006, 243–260

Edge Removal

Edge Removal (ER)

removes redundant edges

- from overlap between instances of P for which contained instances of p do not overlap
- redundant edges introduced by Vertex Addition



N. Vanetik, S. E. Shimony, E. Gudes, "Support measures for graph data", Data Min. Knowl. Discov., Vol. 13 (2), 2006, 243–260

Anti-monotonic overlap graph measures

Overlap graph measure

Support measure f is overlap graph measure
$\iff \forall P, G : f(P, G) = \hat{f}(G_P)$

- Overlap graph measure uses only topology overlap graph
- We call $\hat{f}$ graph measure associated with f

Anti-monotonic overlap graph measures

Overlap graph measure

Support measure f is overlap graph measure
$\iff \forall P, G : f(P, G) = \hat{f}(G_P)$

- Overlap graph measure uses only topology overlap graph
- We call $\hat{f}$ graph measure associated with f

N. Vanetik, S. E. Shimony, E. Gudes, "Support measures for graph data", Data Min. Knowl. Discov., Vol. 13 (2), 2006, 243–260:

Theorem (Vanetik, Gudes, Shimony)

Overlap graph measure f is anti-monotonic
$\iff$ *$\hat{f}$ is increasing under CC, VA, ER*

Proof for labeled graphs and edge overlap graph measures

Meaningful measures

Overlap measure f is meaningful if

(1) f anti-monotonic

(2) frequency of k non-overlapping matches is k

$\Rightarrow \hat{f}(\overline{K_k}) = k$ ($\overline{K_k} = k$ isolated vertices)

Meaningful measures

Overlap measure f is meaningful if
(1) f anti-monotonic
(2) frequency of k non-overlapping matches is k
$\Rightarrow \hat{f}(\overline{K_k}) = k$ ($\overline{K_k} = k$ isolated vertices)

1. Maximum Independent Set Measure: $\hat{f}(G_P) = MIS(G_P) =$ maximal size of set of pairwise unconnected vertices
   - introduced and proven to be antimonotonic by Vanetik, Gudes and Shimony

Meaningful measures

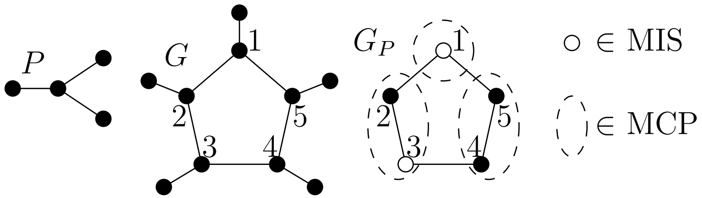
Overlap measure f is meaningful if
(1) f anti-monotonic
(2) frequency of k non-overlapping matches is k
$\Rightarrow \hat{f}(\overline{K_k}) = k$ ($\overline{K_k} = k$ isolated vertices)

1. Maximum Independent Set Measure: $\hat{f}(G_P) = MIS(G_P) =$ maximal size of set of pairwise unconnected vertices
   - introduced and proven to be antimonotonic by Vanetik, Gudes and Shimony
2. Minimum Clique Partition: $\hat{f}(G_P) = MCP(G_P) =$ minimum number of cliques in which graph can be partitioned

Meaningful measures

Overlap measure f is meaningful if
(1) f anti-monotonic
(2) frequency of k non-overlapping matches is k
$\Rightarrow \hat{f}(\overline{K_k}) = k$ ($\overline{K_k} = k$ isolated vertices)

1. Maximum Independent Set Measure: $\hat{f}(G_P) = MIS(G_P) =$ maximal size of set of pairwise unconnected vertices
   - introduced and proven to be antimonotonic by Vanetik, Gudes and Shimony
2. Minimum Clique Partition: $\hat{f}(G_P) = MCP(G_P) =$ minimum number of cliques in which graph can be partitioned



Bounding theorem for meaningful measures

Theorem

For every meaningful measure f it holds that:

$$\forall P, G : MIS(G_P) \leq f(P, G) = \hat{f}(G_P) \leq MCP(G_P)$$

Proof: use $\hat{f}$ increasing under CC, VA, ER:

Bounding theorem for meaningful measures

Theorem

For every meaningful measure f it holds that:

$$\forall P, G : MIS(G_P) \leq f(P, G) = \hat{f}(G_P) \leq MCP(G_P)$$

Proof: use $\hat{f}$ increasing under CC, VA, ER:

1. MIS: let G_P be any graph and $I \subseteq V(G_P)$ MIS of size k
   - we can construct G_P starting from $\overline{K_k} = k$ isolated vertices

Bounding theorem for meaningful measures

Theorem

For every meaningful measure f it holds that:

$$\forall P, G : MIS(G_P) \leq f(P, G) = \hat{f}(G_P) \leq MCP(G_P)$$

Proof: use $\hat{f}$ increasing under CC, VA, ER:

1. MIS: let G_P be any graph and $I \subseteq V(G_P)$ MIS of size k
   - we can construct G_P starting from $\overline{K_k} = k$ isolated vertices
   - add all other vertices with VA, result is graph with all edges except between vertices $\in I$

Bounding theorem for meaningful measures

Theorem

For every meaningful measure f it holds that:

$$\forall P, G : MIS(G_P) \leq f(P, G) = \hat{f}(G_P) \leq MCP(G_P)$$

Proof: use $\hat{f}$ increasing under CC, VA, ER:

1. MIS: let G_P be any graph and $I \subseteq V(G_P)$ MIS of size k
   - we can construct G_P starting from $\overline{K_k} = k$ isolated vertices
   - add all other vertices with *VA*, result is graph with all edges except between vertices $\in I$
   - remove all redundant edges with *ER* to get G_P

Bounding theorem for meaningful measures

Theorem

For every meaningful measure f it holds that:

$$\forall P, G : MIS(G_P) \leq f(P, G) = \hat{f}(G_P) \leq MCP(G_P)$$

Proof: use $\hat{f}$ increasing under CC, VA, ER:

1. MIS: let G_P be any graph and $I \subseteq V(G_P)$ MIS of size k
   - we can construct G_P starting from $\overline{K_k} = k$ isolated vertices
   - add all other vertices with *VA*, result is graph with all edges except between vertices $\in I$
   - remove all redundant edges with *ER* to get G_P
   - $\hat{f}(\overline{K_k}) = k = MIS(G_P)$ and in each step $\hat{f}$ can only increase $\Rightarrow \hat{f}(G_P) \geq MIS(G_P)$

2. MCP: let G_P be any graph and let $\{V_1, \ldots, V_k\}$ be MCP
    - each V_i is clique and can be contracted by CC

2. MCP: let G_P be any graph and let $\{V_1, \ldots, V_k\}$ be MCP
   - each V_i is clique and can be contracted by CC
   - result is $\overline{K_k} = k$ isolated vertices

2. MCP: let G_P be any graph and let $\{V_1, \ldots, V_k\}$ be MCP
   - each V_i is clique and can be contracted by CC
   - result is $\overline{K_k} = k$ isolated vertices
   - $\hat{f}(\overline{K_k}) = k$ and in each step $\hat{f}$ can only increase $\Rightarrow$ $\hat{f}(G_P) \leq k = MCP(G_P)$

2. MCP: let G_P be any graph and let $\{V_1, \ldots, V_k\}$ be MCP
   - each V_i is clique and can be contracted by CC
   - result is $\overline{K_k} = k$ isolated vertices
   - $\hat{f}(\overline{K_k}) = k$ and in each step $\hat{f}$ can only increase $\Rightarrow$ $\hat{f}(G_P) \leq k = MCP(G_P)$

Bad News

MIS and MCP are both NP-hard problems in size overlap graph

PTIME computable graph measure

A known PTIME graph measure is the Lovász number θ:

Lovász number

$\theta(G) = \min_A \lambda_{\max}(A)$, where $\lambda_{\max}(A)$ denotes largest eigenvalue of matrix A and minimum is taken over all feasible matrices A such that $A^\top = A$, $A_{ii} = 1$ and $A_{ij} = 1$ if $(i, j) \notin E(G)$.

Theorem (Sandwich Theorem, Lovász)

$$MIS(G) \leq \theta(G) \leq MCP(G)$$

PTIME computable graph measure

A known PTIME graph measure is the Lovász number θ:

Lovász number

$\theta(G) = \min_A \lambda_{\max}(A)$, where $\lambda_{\max}(A)$ denotes largest eigenvalue of matrix A and minimum is taken over all feasible matrices A such that $A^\top = A$, $A_{ii} = 1$ and $A_{ij} = 1$ if $(i, j) \notin E(G)$.

Theorem (Sandwich Theorem, Lovász)

$$MIS(G) \leq \theta(G) \leq MCP(G)$$

Theorem

θ is meaningful overlap graph measure

Proof: linear algebra

Generalization of results of Vanetik, Gudes and Shimony

Theorem (Vanetik, Gudes, Shimony)

Overlap graph measure f is anti-monotonic
$\iff \hat{f}$ *is increasing under CC, VA, ER*

- Only holds for labeled graphs and edge overlap graph measures

Generalization of results of Vanetik, Gudes and Shimony

Theorem (Vanetik, Gudes, Shimony)

Overlap graph measure f is anti-monotonic
$\iff \hat{f}$ *is increasing under CC, VA, ER*

- Only holds for labeled graphs and edge overlap graph measures
- What about
  1. unlabeled graphs?
  2. vertex overlap graphs?
  3. homomorphic or homeomorphic matches?

Generalization of results of Vanetik, Gudes and Shimony

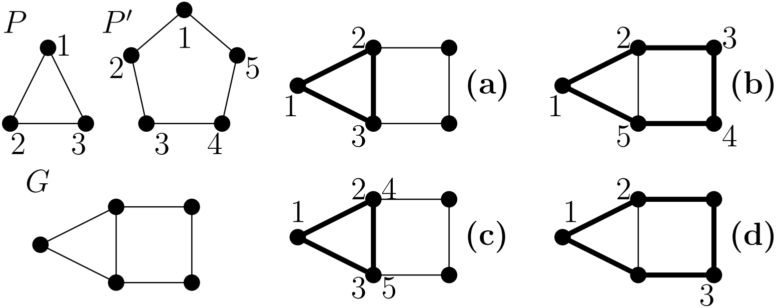
Theorem (Vanetik, Gudes, Shimony)

Overlap graph measure f is anti-monotonic
$\iff$ *$\hat{f}$ is increasing under CC, VA, ER*

- Only holds for labeled graphs and edge overlap graph measures
- What about
  1. unlabeled graphs?
  2. vertex overlap graphs?
  3. homomorphic or homeomorphic matches?
  4. any combination of the above?

Other types of matches

1. subgraph homomorphism $\cong$ subgraph isomorphism without injectivity
2. subgraph homeomorphism $\cong$ subgraph isomorphism in which edges can be mapped on vertex-disjoint paths



(a) iso of P **(b)** iso of P' **(c)** homo of P' **(d)** homeo of P in G

Extension to other settings

All results in this presentation:

1. f anti-monotonic overlap graph measure
   $\iff$ $\hat{f}$ increasing under CC, VA and ER
2. MIS, MCP, θ are meaningful (anti-monotonic + $\hat{f}(\overline{K_k}) = k$)
3. all meaningful overlap measures f:
   $MIS(G) \leq \hat{f}(G) \leq MCP(G)$

can be extended to all 24 combinations of

1. labeled or unlabeled graphs
2. directed or undirected graphs
3. isomorphic, homomorphic or homeomorphic matches
4. vertex or edge overlap graphs

Proofs are either direct or reductions from one setting to another