Motif finding: under the hood

Counting mutual friends

```
abcDF = abDF.join(bcDF, "B").filter("A = 1")
abcDF.show()
```

```
+---+---+---+
|  B|  A|  C|
+---+---+---+
|  3|  1|  4|
|  3|  1|  5|
|  3|  1|  6|
|  3|  1|  7|
|  2|  1|  3|
|  2|  1|  5|
+---+---+---+
```

Counting mutual friends

```
abcDF = abDF.join(bcDF, "B").filter("A = 1")
abcDF.show()
```

```
+---+---+---+
|  B|  A|  C|
+---+---+---+
|  3|  1|  4|
|  3|  1|  5|
|  3|  1|  6|
|  3|  1|  7|
|  2|  1|  3|
|  2|  1|  5|
+---+---+---+
```

Motif finding

```
motifs = gf.find("(A)-[]->(B); (B)-[]->(C)")
```

Motif finding

1. Creating a collection of Patterns

Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

Example:

"(A)-[]->(B); (B)-[]->(C)"

[NamedVertex("A"),

AnonymousEdge(NamedVertex("A"), NamedVertex("B")),

NamedVertex("B"),

NamedVertex("B"),

AnonymousEdge(NamedVertex("B"), NamedVertex("C")),

NamedVertex("C")]

Motif finding

1. Creating a collection of Patterns
2. Iterate over it
   g - original GraphFrame

   currentResult - current result of iterations

Motif finding

How exactly does Motif finding handle every new pattern depends on 2 factors:

Motif finding

How exactly does Motif finding handle every new pattern depends on 2 factors:

1. Type of a pattern

Motif finding

How exactly does Motif finding handle every new pattern depends on 2 factors:

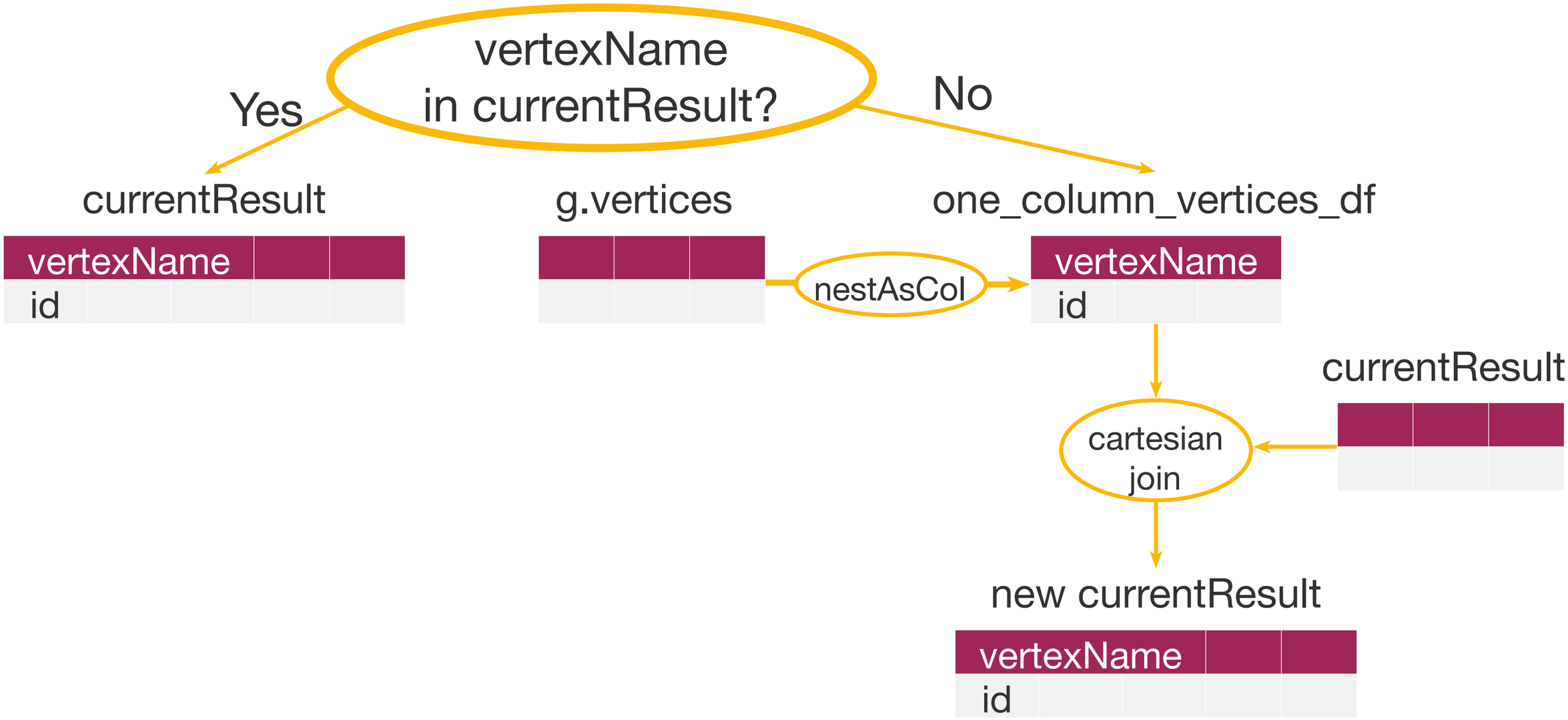
1. Type of a pattern

2. Whether the currentResult already contains the columns with the names mentioned in proceeding pattern

Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

NamedVertex(VertexName)



Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,

  src: NamedVertex(srcVertexName) | AnonymousVertex,

  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(

  src: NamedVertex(srcVertexName) | AnonymousVertex,

  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName

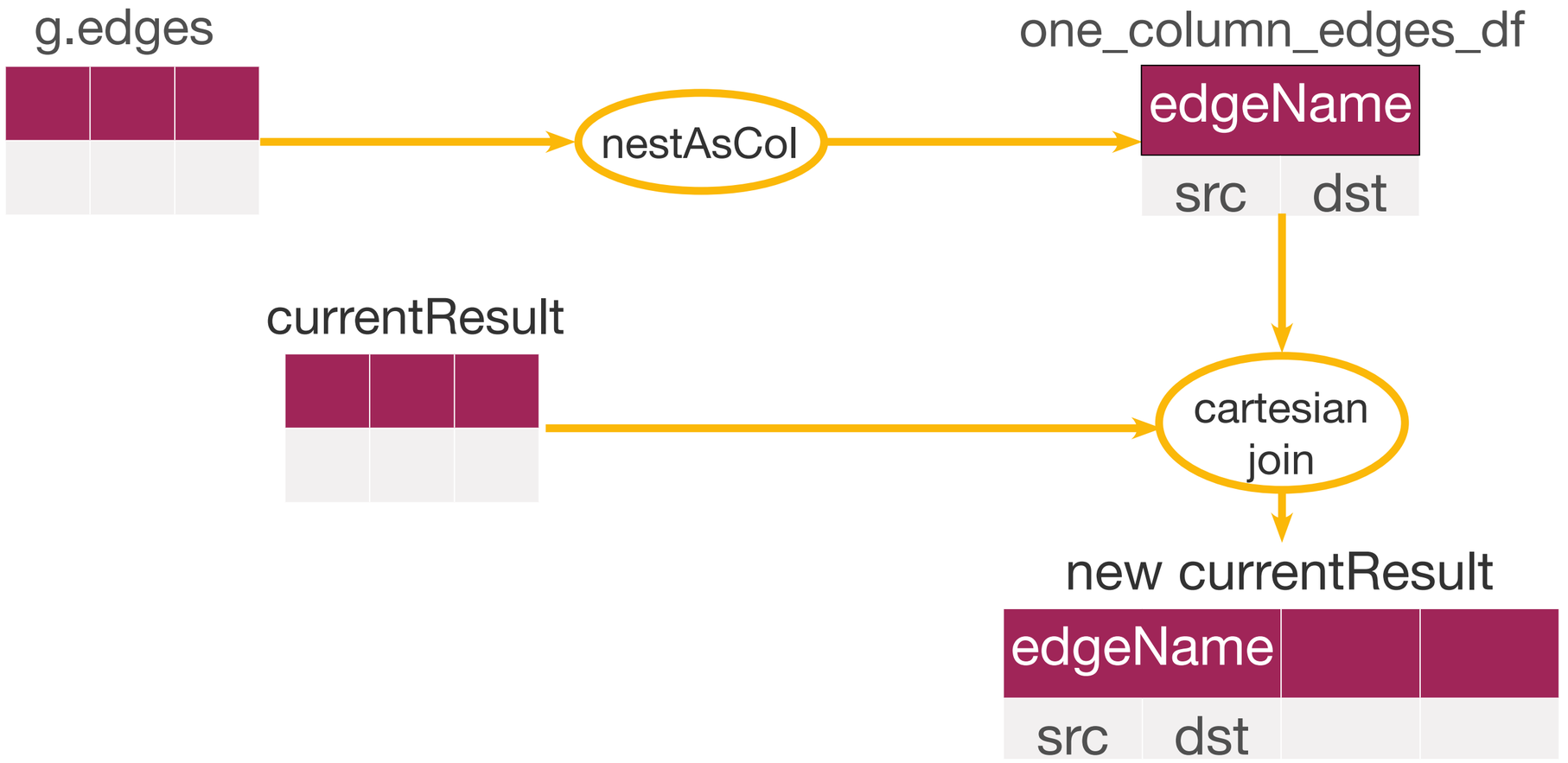
  src: NamedVertex(srcVertexName) | AnonymousVertex,

  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(

  src: NamedVertex(srcVertexName) | AnonymousVertex,

  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

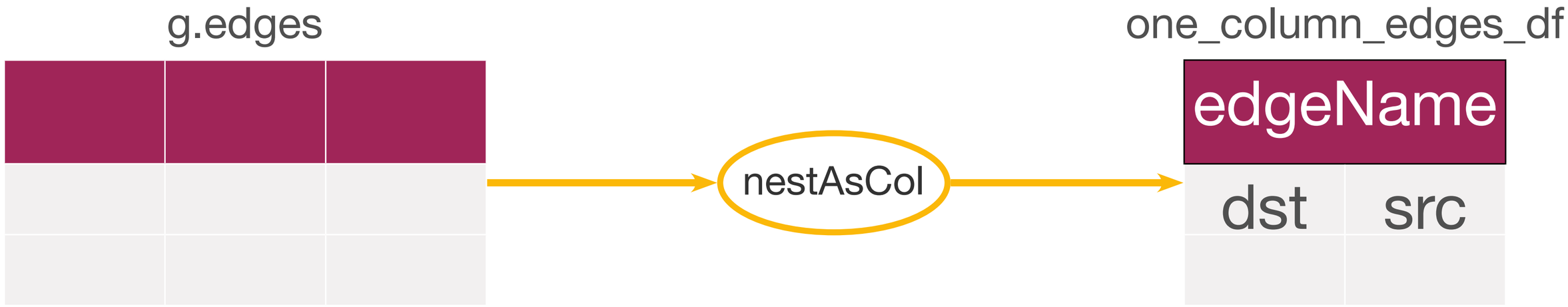
NamedEdge(edgeName,
src: AnonymousVertex, dst: AnonymousVertex)



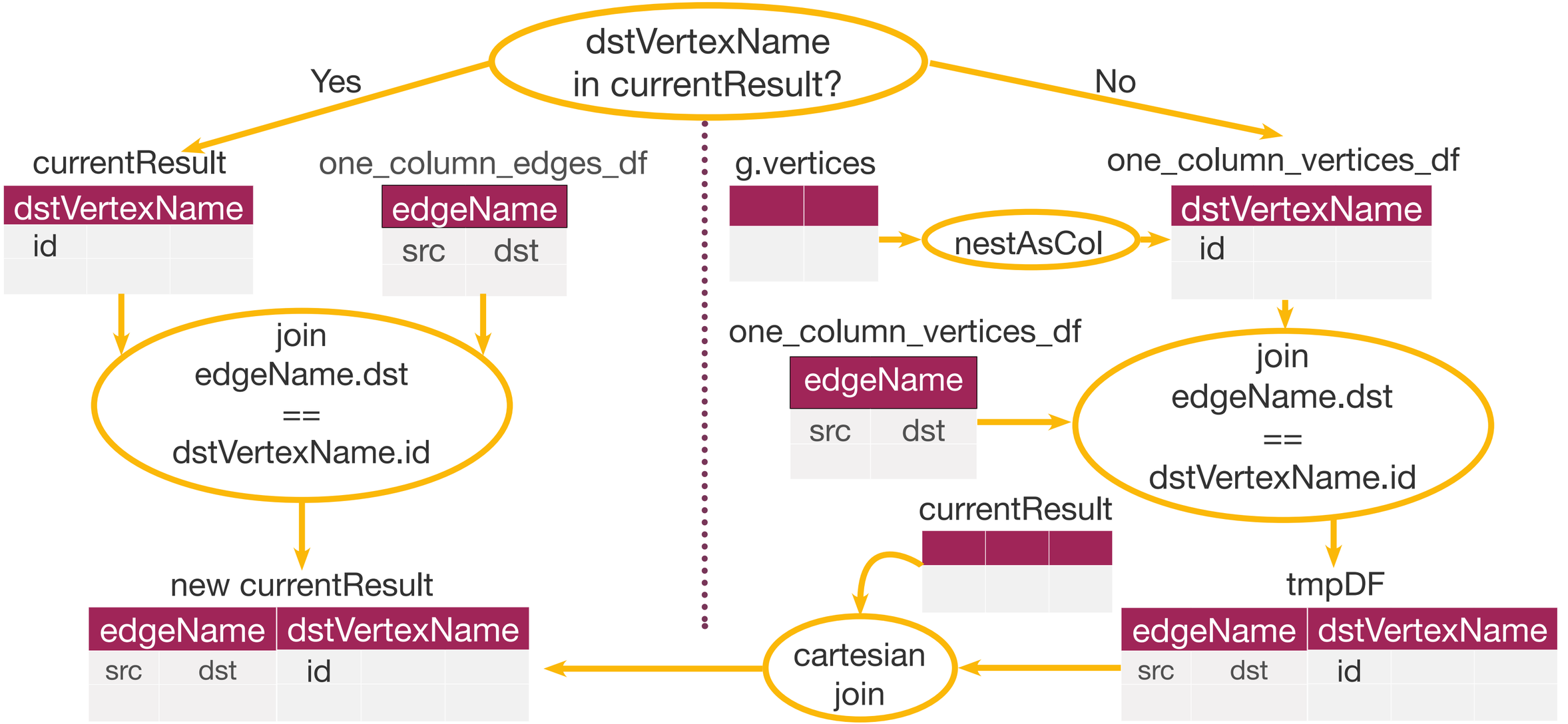
Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

NamedEdge(edgeName,
src: AnonymousVertex,
dst: NamedVertex(srcVertexName))



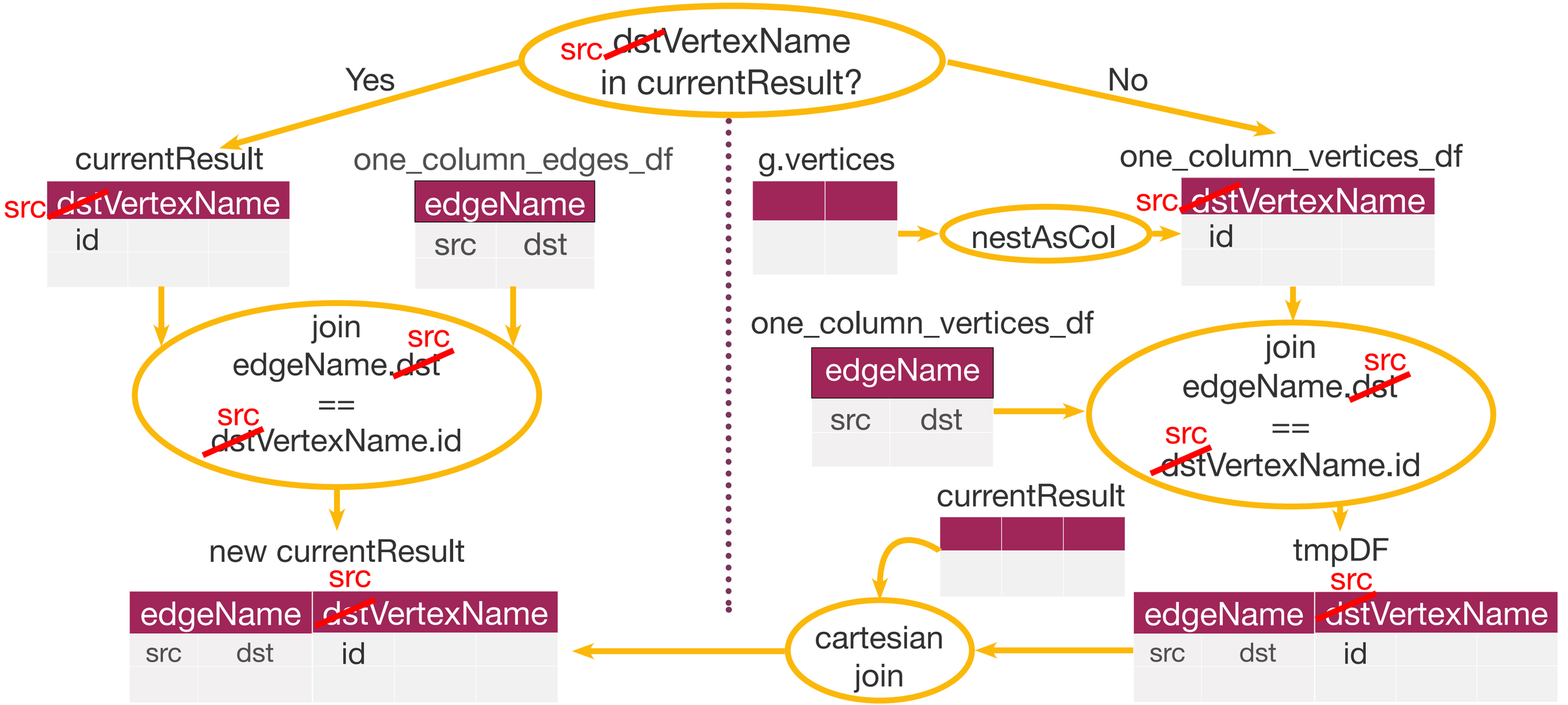
NamedEdge(edgeName,
src: AnonymousVertex,
dst: NamedVertex(dstVertexName))



Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: AnonymousVertex)



Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))

1 currentResult

srcVertexName	dstVertexName	

2 currentResult

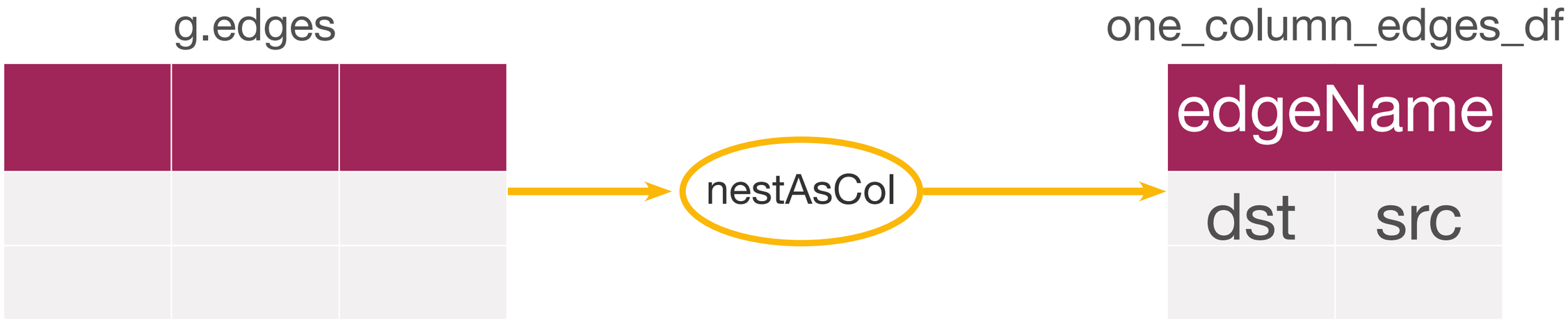
3 currentResult

	dstVertexName	

4 currentResult

srcVertexName		

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))



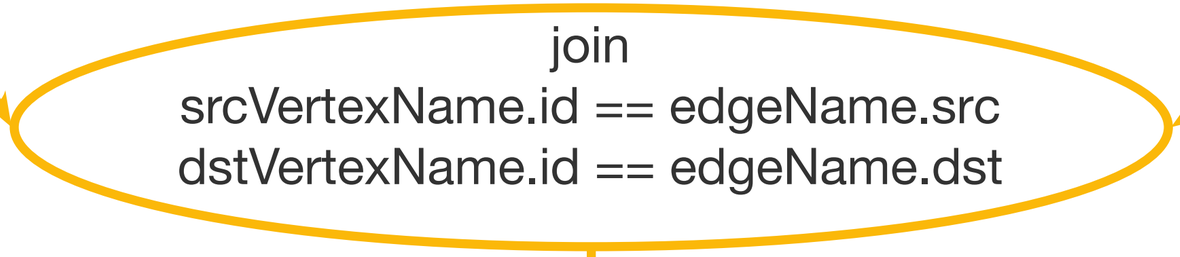
NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))

currentResult

srcVertexName	dstVertexName

one_column_edges_df

edgeName	
src	dst



new currentResult

srcVertexName			dstVertexName			edgeName	
id			id			src	dst

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))

1 currentResult

srcVertexName	dstVertexName	

2 currentResult

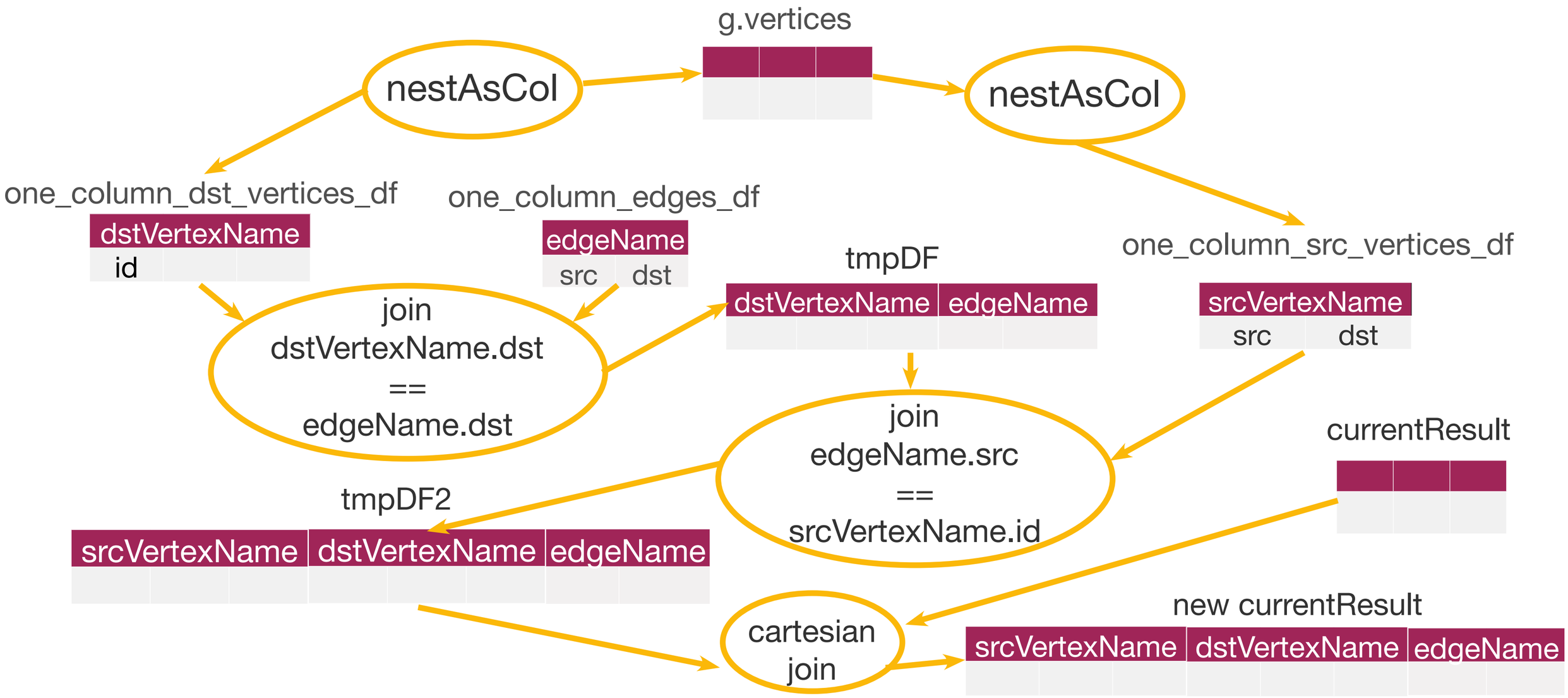
3 currentResult

	dstVertexName	

4 currentResult

srcVertexName		

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))



NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))

1 currentResult

srcVertexName	dstVertexName	

2 currentResult

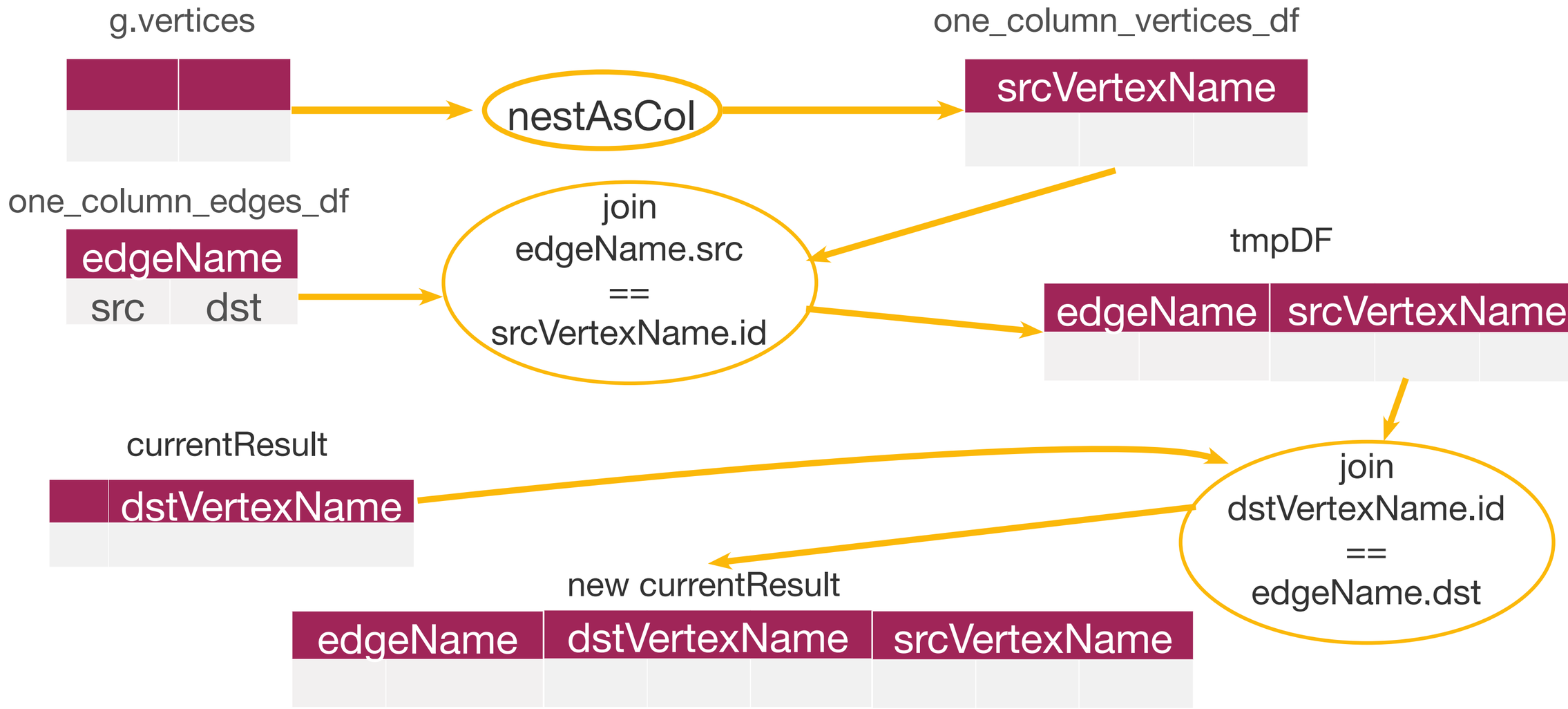
3 currentResult

	dstVertexName	

4 currentResult

srcVertexName		

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))



NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))

1 currentResult

srcVertexName	dstVertexName	

2 currentResult

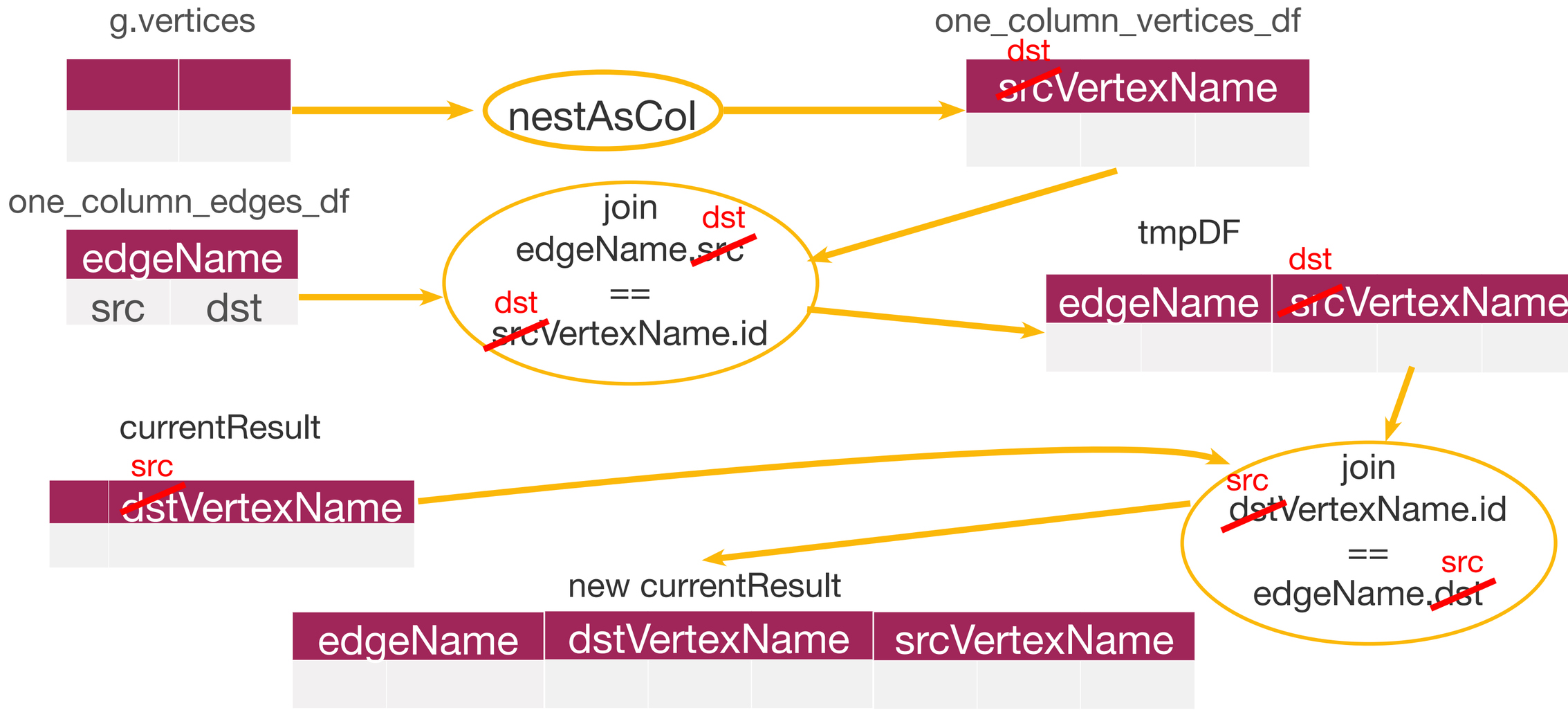
3 currentResult

	dstVertexName	

4 currentResult

srcVertexName		

NamedEdge(edgeName,
src: NamedVertex(srcVertexName),
dst: NamedVertex(dstVertexName))



Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)

AnonymousEdge (src: AnonymousVertex|NamedVertex,
dst: AnonymousVertex|NamedVertex)

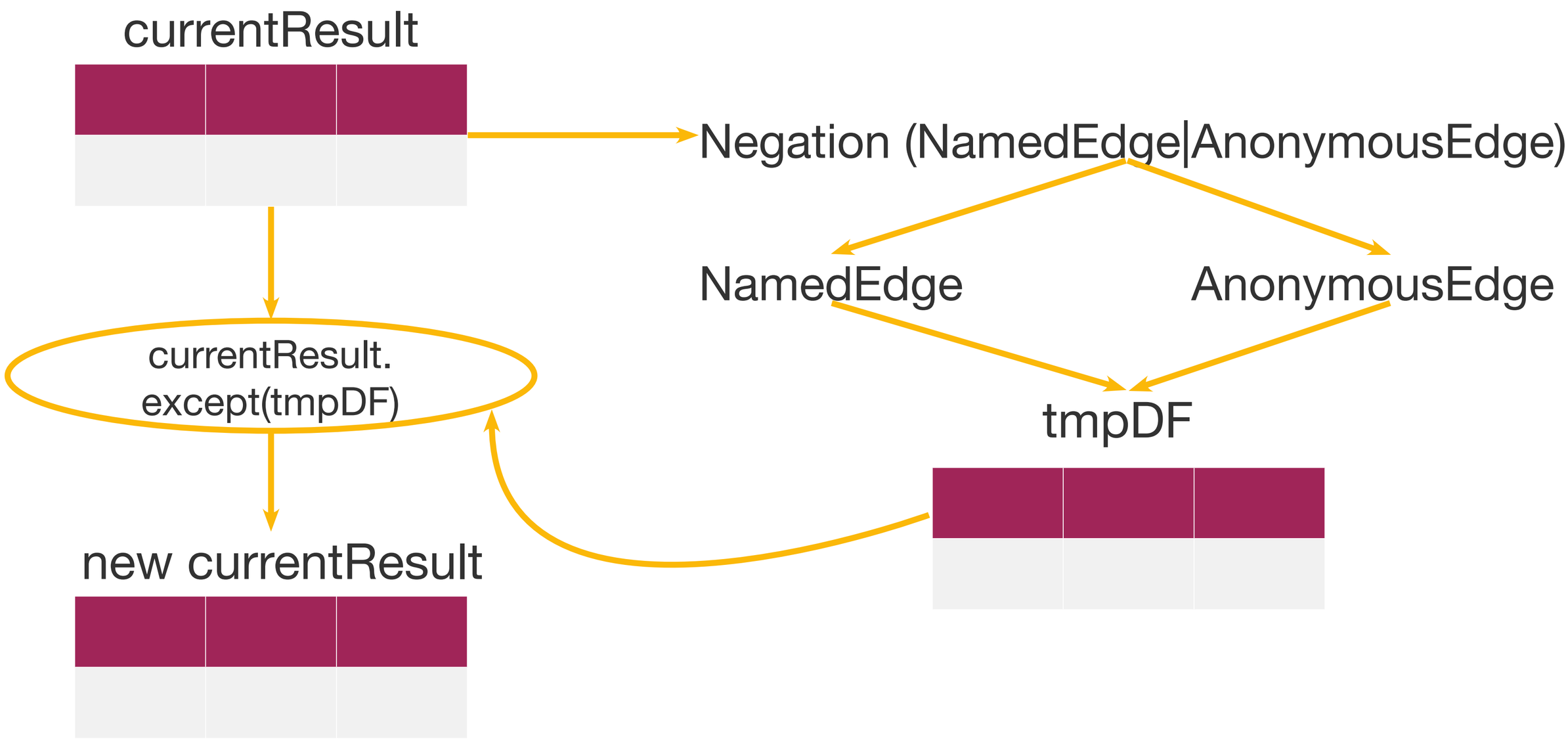
NamedEdge("__tmp", src: AnonymousVertex|NamedVertex,
dst: AnonymousVertex|NamedVertex)

new currentResult

		__tmp

Types of patterns:

- NamedVertex(vertexName)
- AnonymousVertex
- NamedEdge(edgeName,
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- AnonymousEdge(
  src: NamedVertex(srcVertexName) | AnonymousVertex,
  dst: NamedVertex(dstVertexName) | AnonymousVertex)
- Negation(NamedEdge | AnonymousEdge)


currentResult
Negation (NamedEdge|AnonymousEdge)
NamedEdge
AnonymousEdge
currentResult.
except(tmpDF)
tmpDF
new currentResult

Summary

- Now you know how motif finding algorithm works step by step