Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam

Graph Twiddling in a MapReduce World

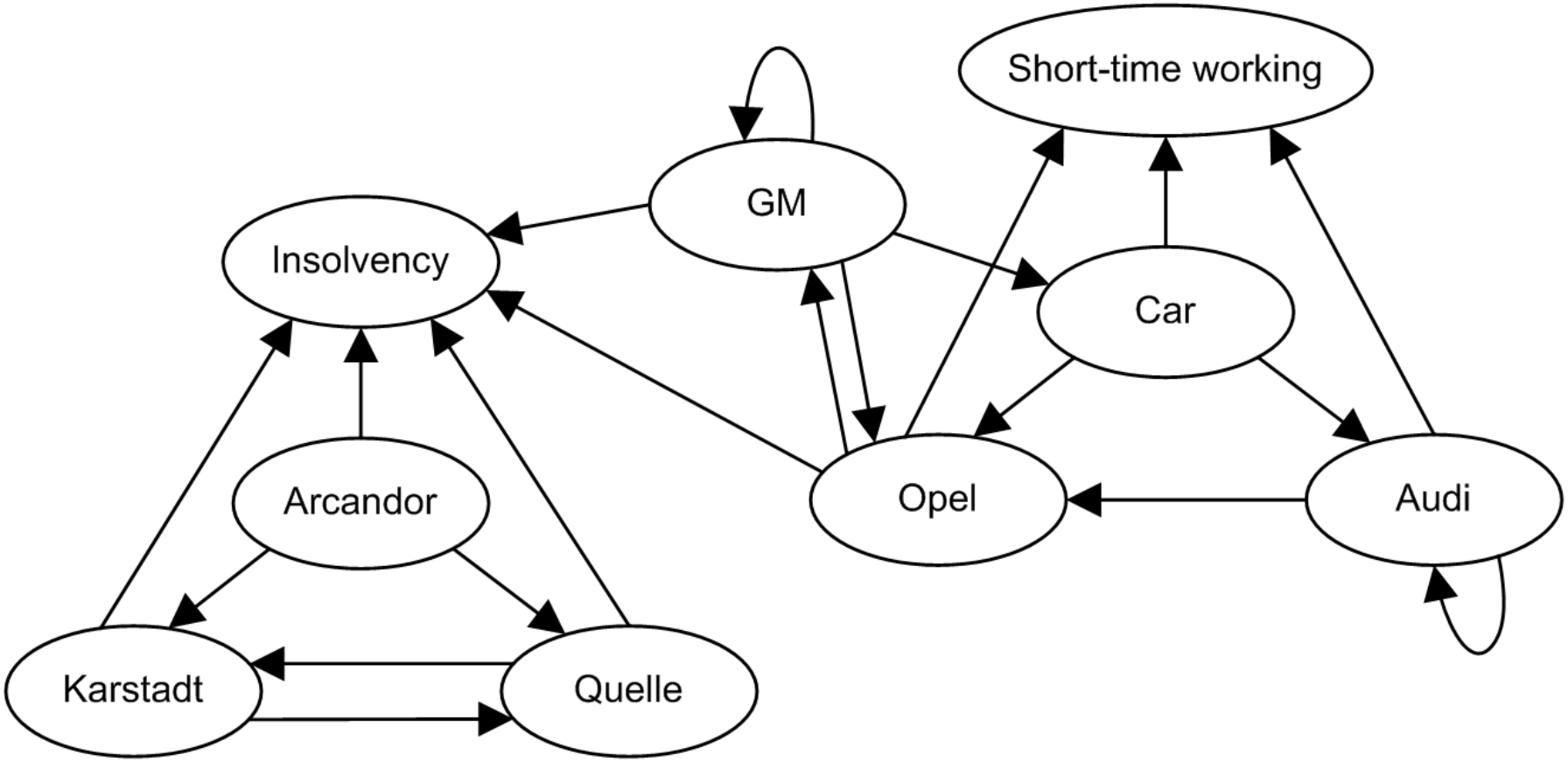
Jonathan Cohen

Adv. MapReduce Algorithms winter term 09/10
HPI

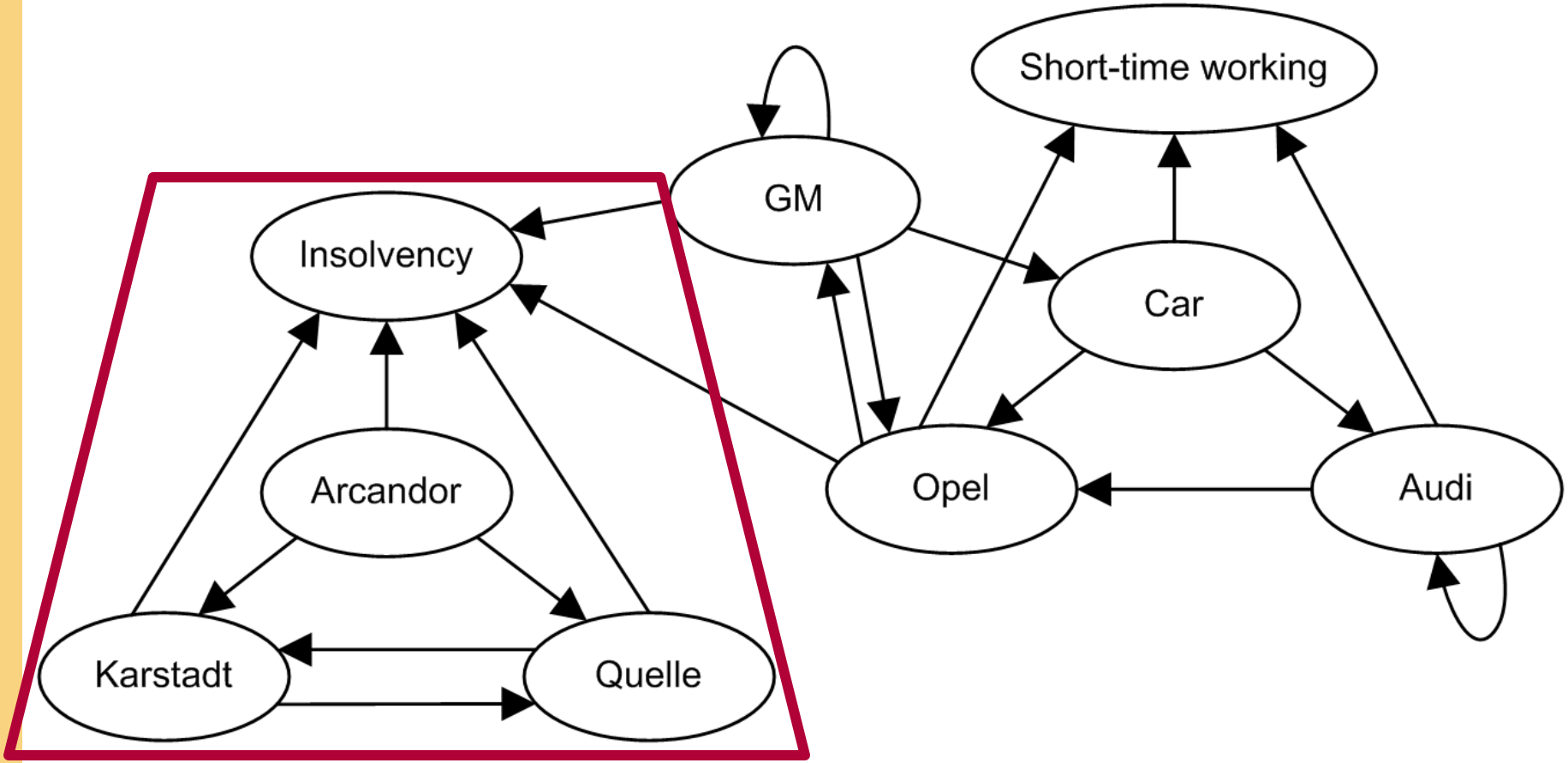
Winter presentation – Problem statement

Arvid Heise, Michael Leben

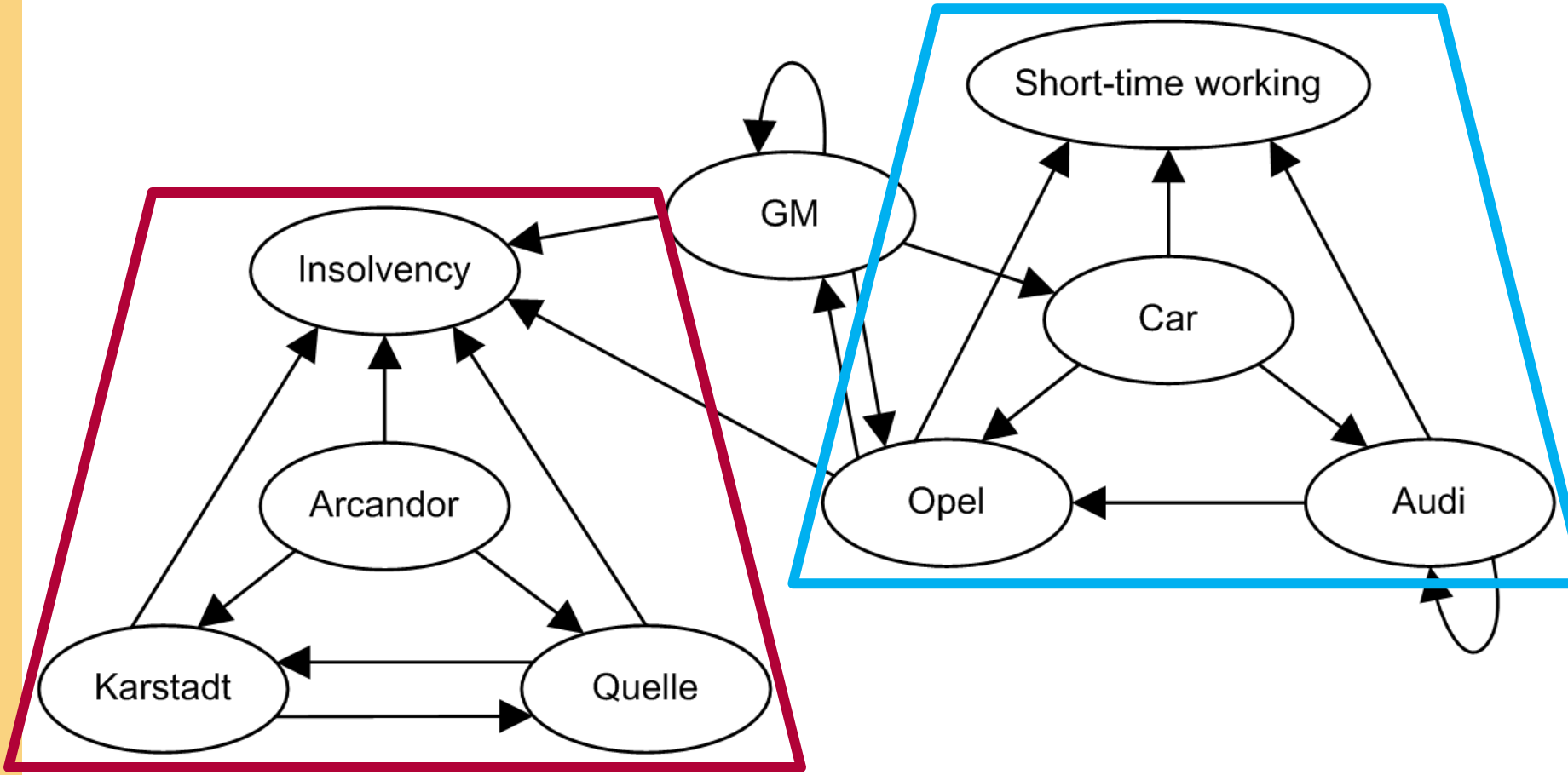
Use Case: Find Domains in Wikipedia



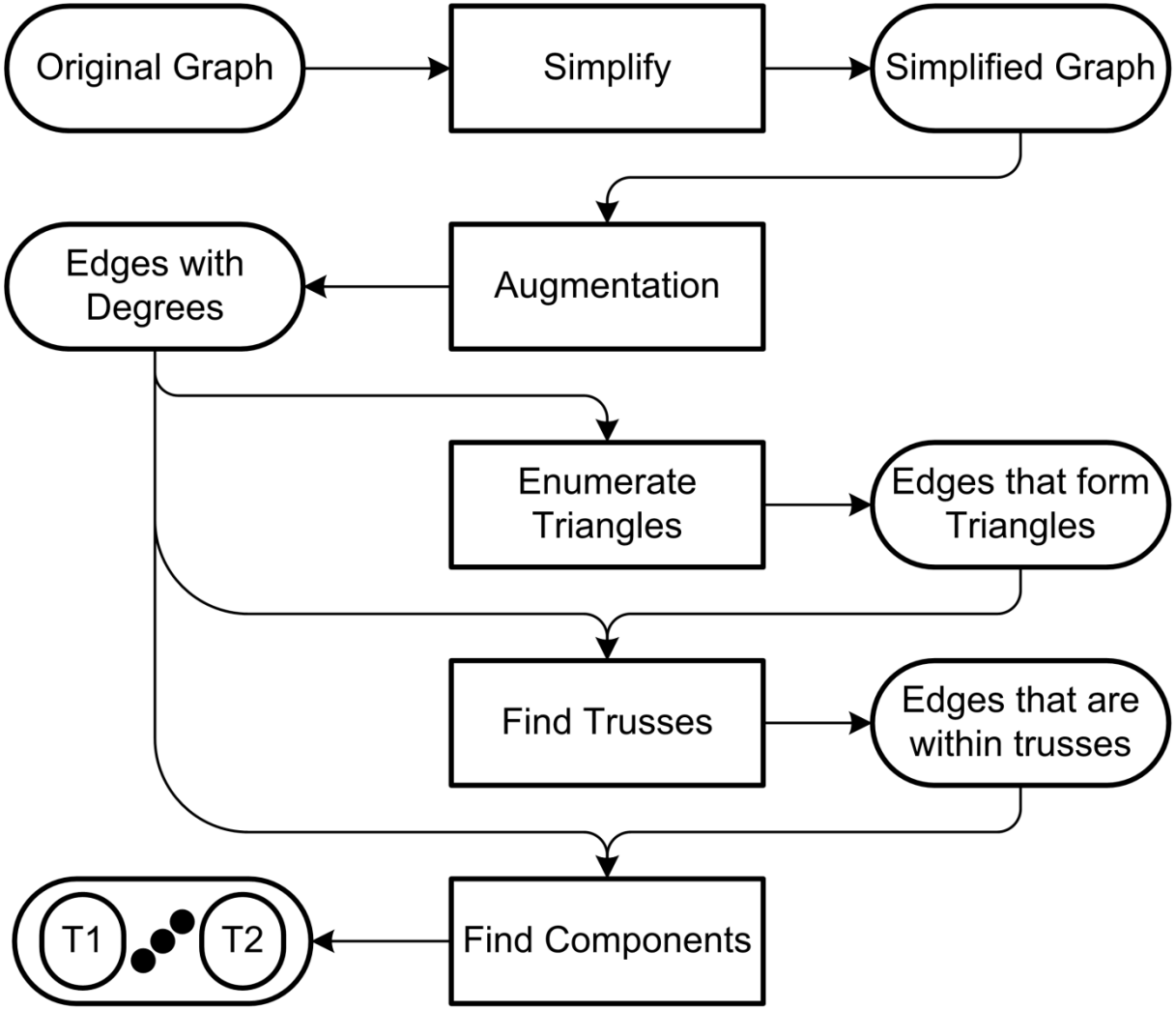
Use Case: Find Domains in Wikipedia



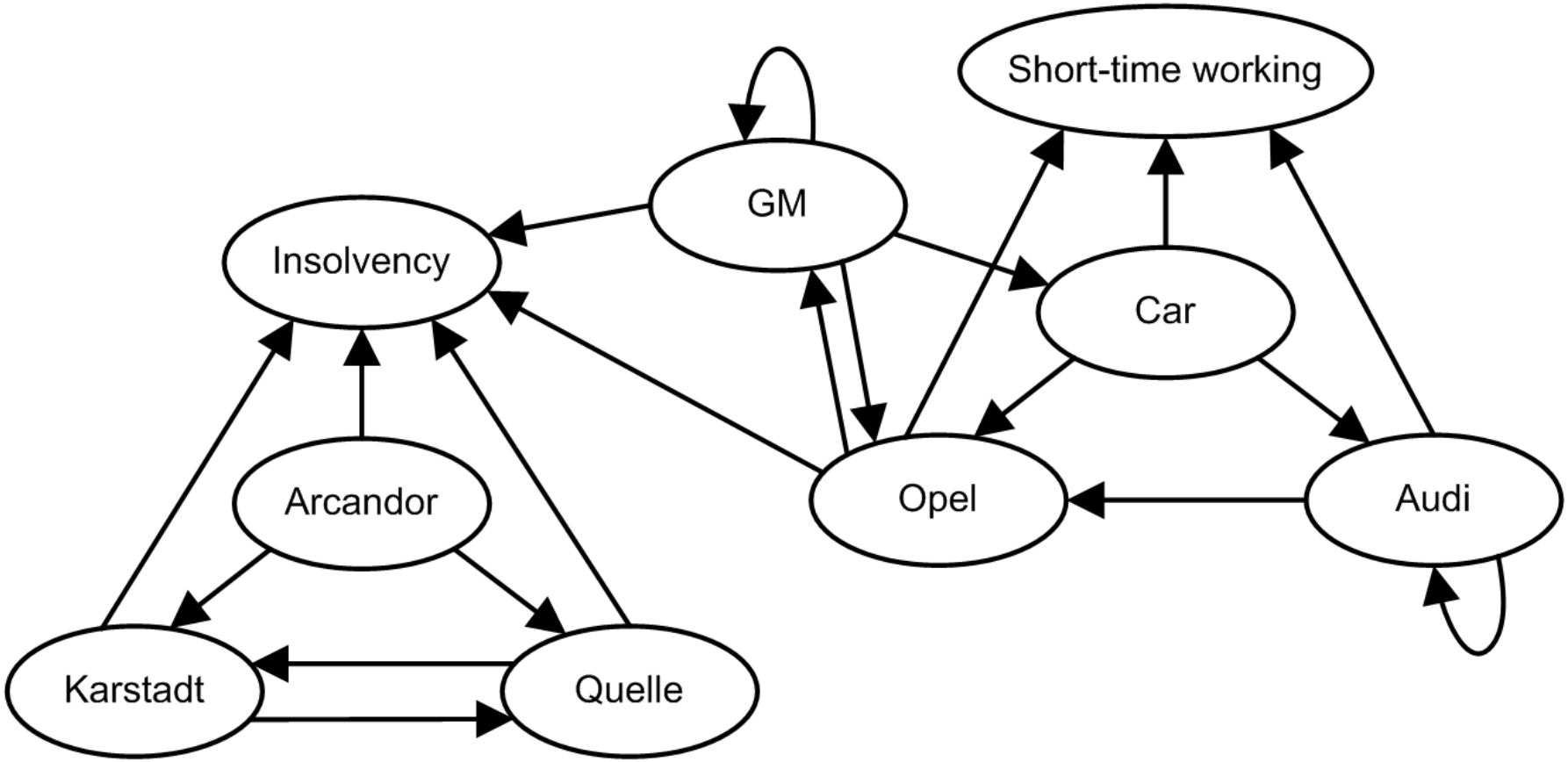
Use Case: Find Domains in Wikipedia



Map/Reduce Process for finding trusses

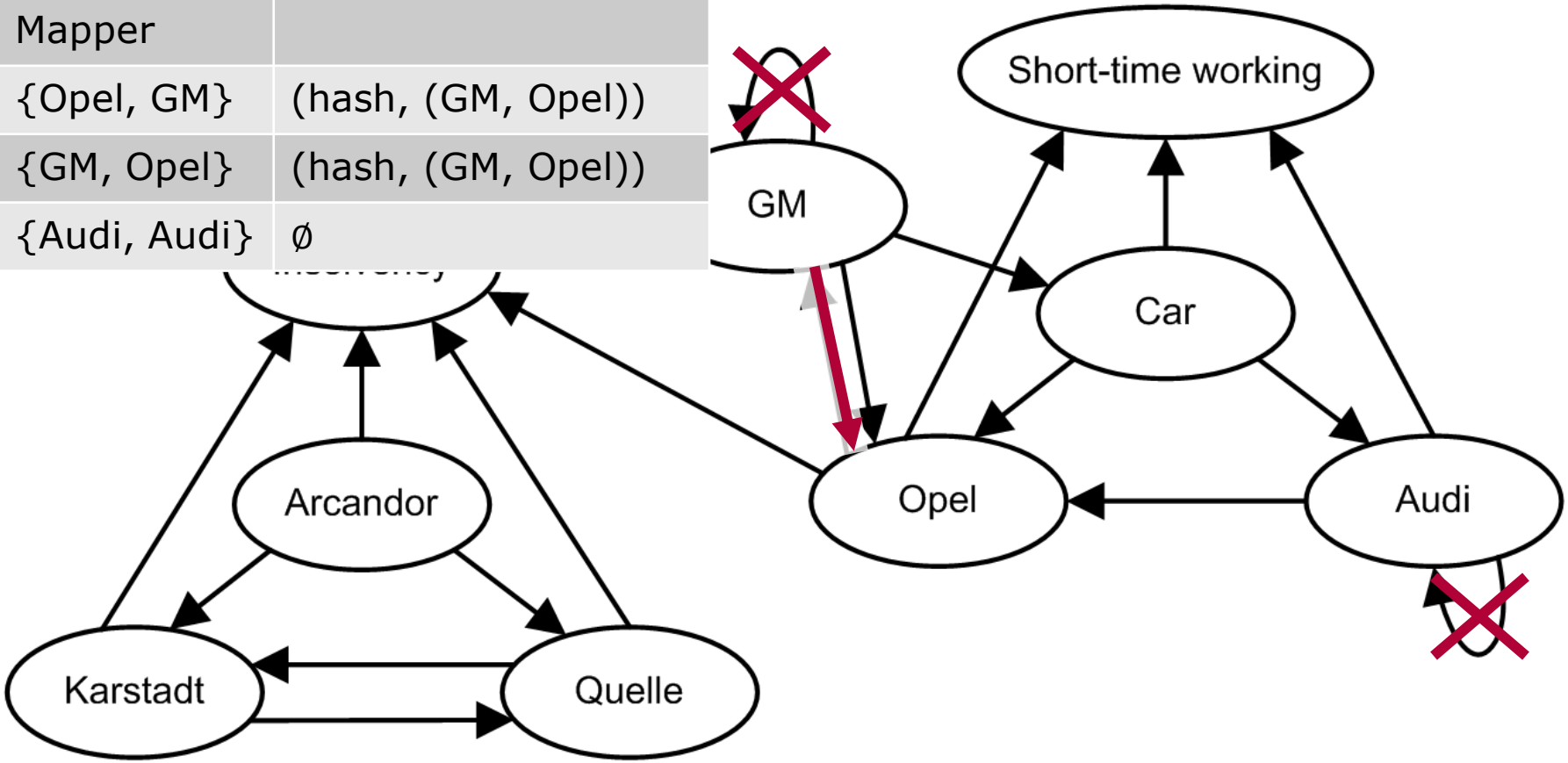


Simplification of Graph



Simplification of Graph

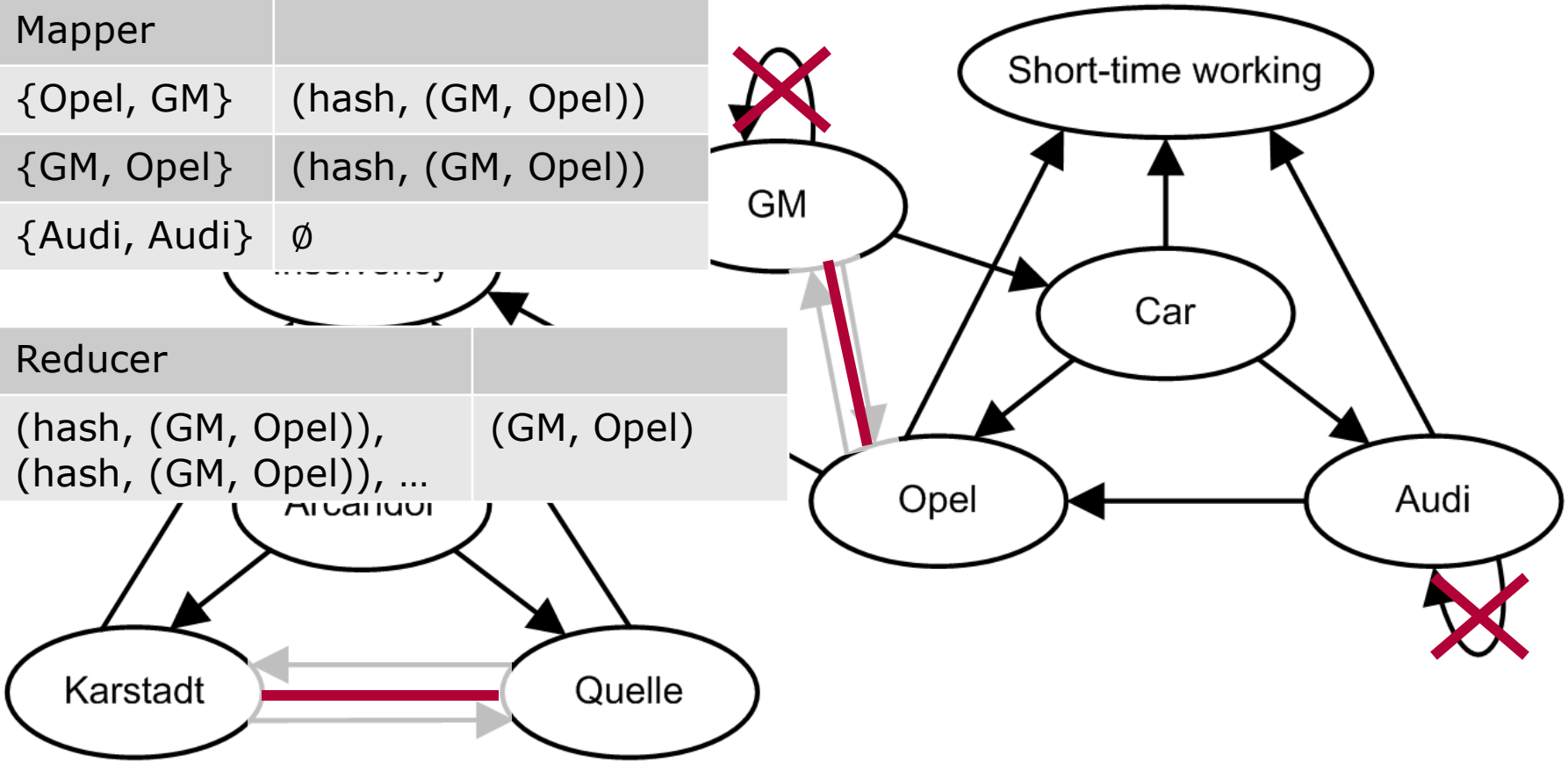
Mapper	
{Opel, GM}	(hash, (GM, Opel))
{GM, Opel}	(hash, (GM, Opel))
{Audi, Audi}	∅



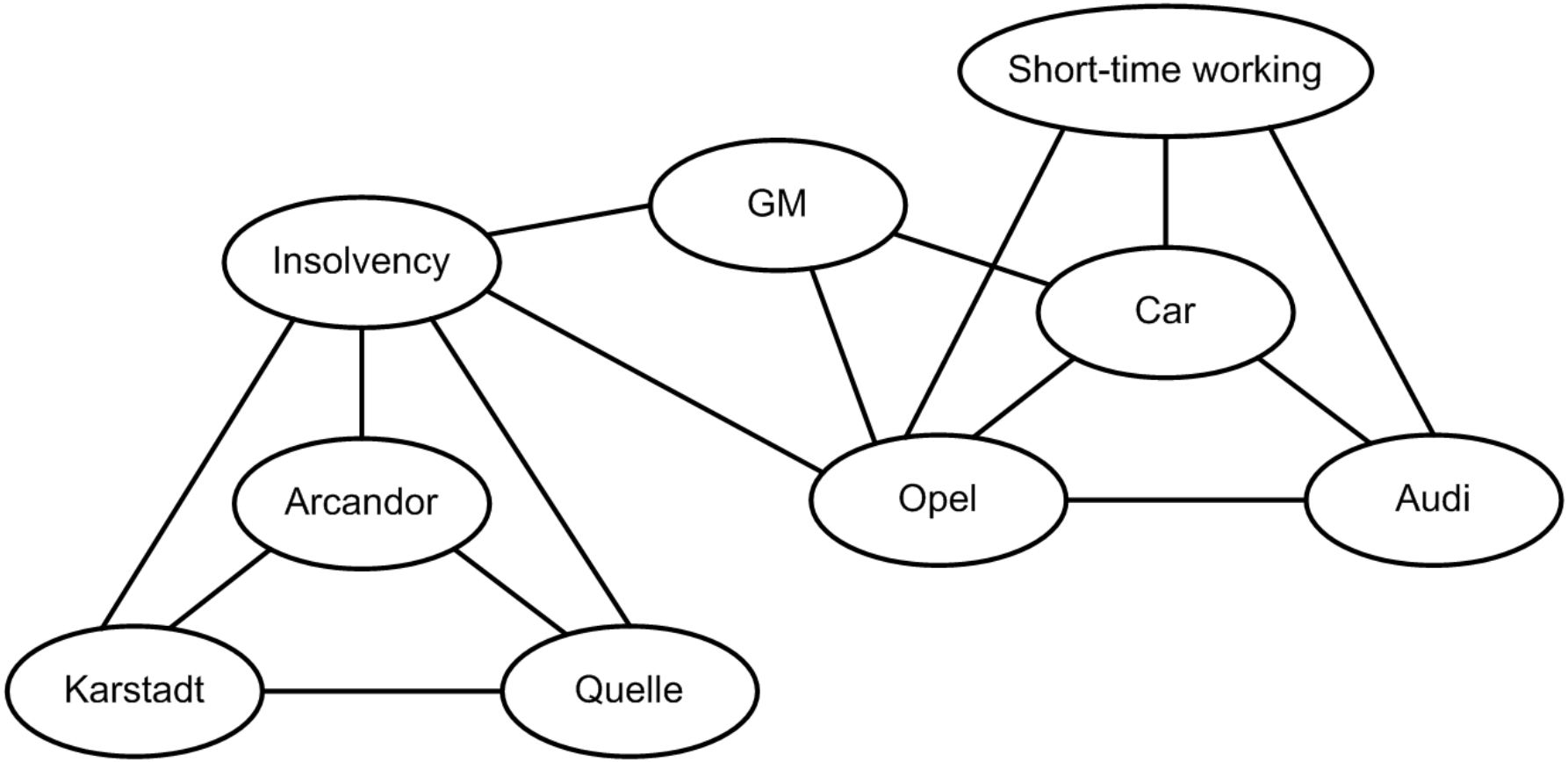
Simplification of Graph

Mapper	
{Opel, GM}	(hash, (GM, Opel))
{GM, Opel}	(hash, (GM, Opel))
{Audi, Audi}	∅

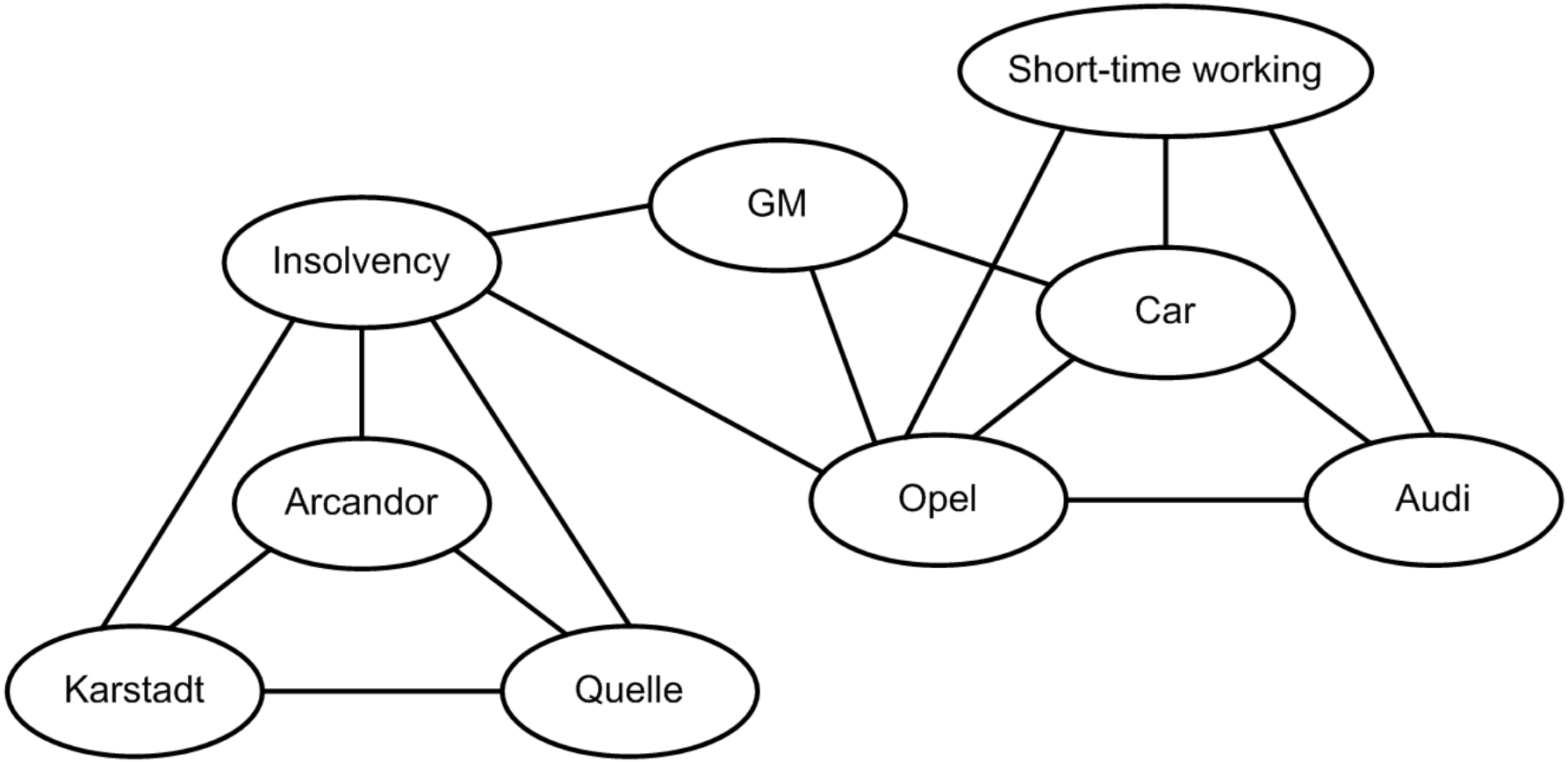
Reducer	
(hash, (GM, Opel)), (hash, (GM, Opel)), …	(GM, Opel)



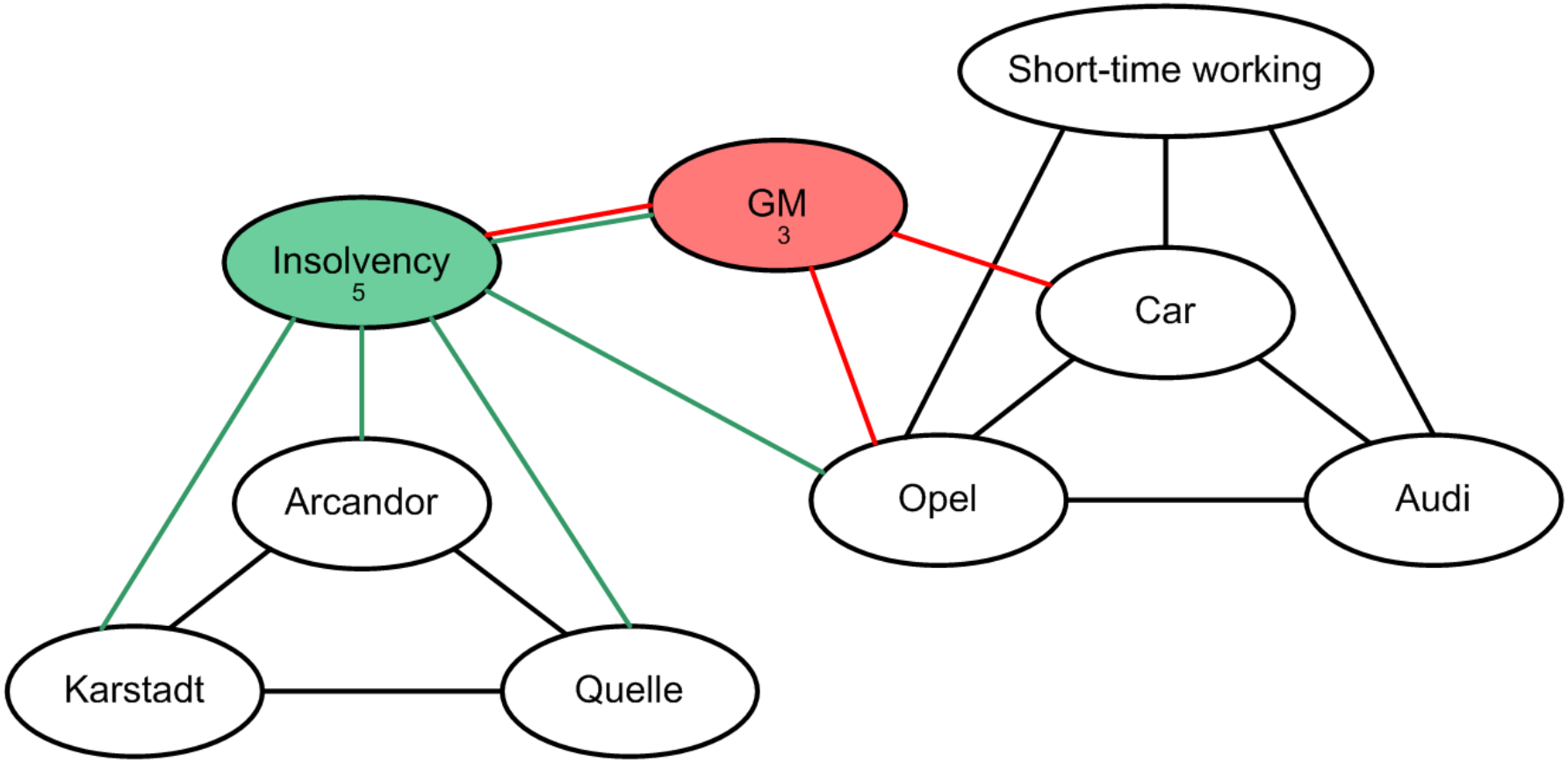
Simplification of Graph



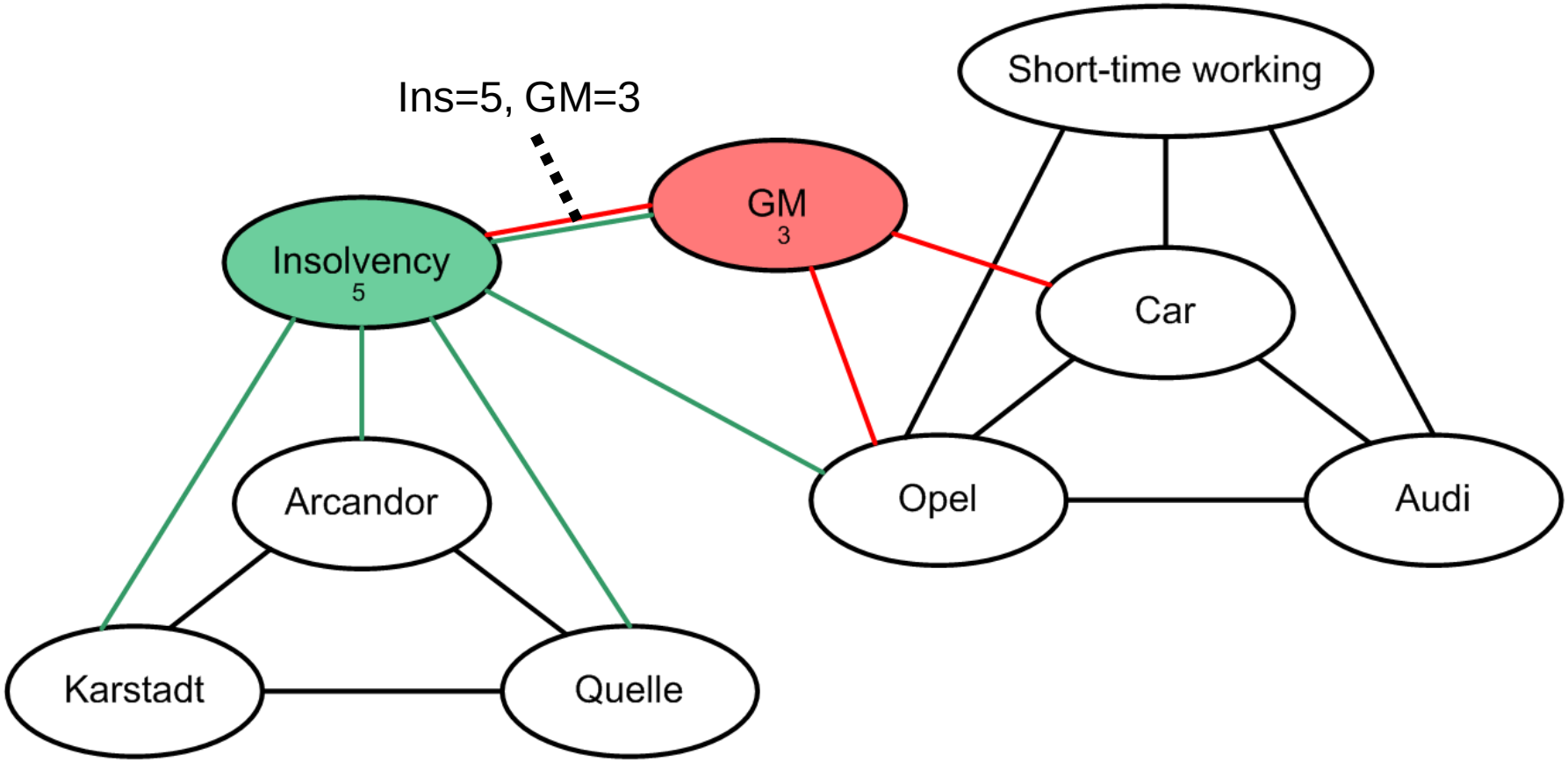
Augmenting Edges with Degrees



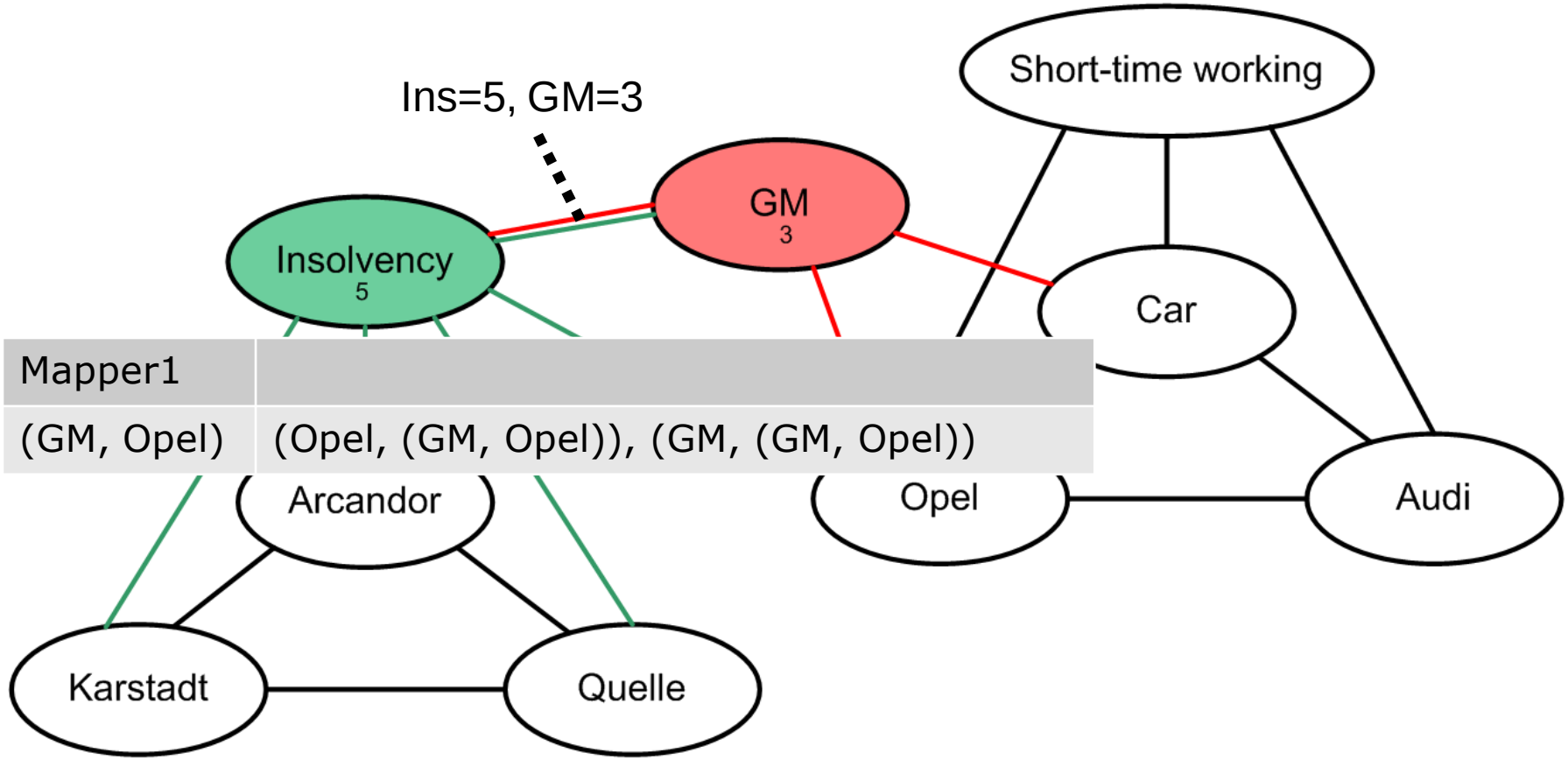
Augmenting Edges with Degrees



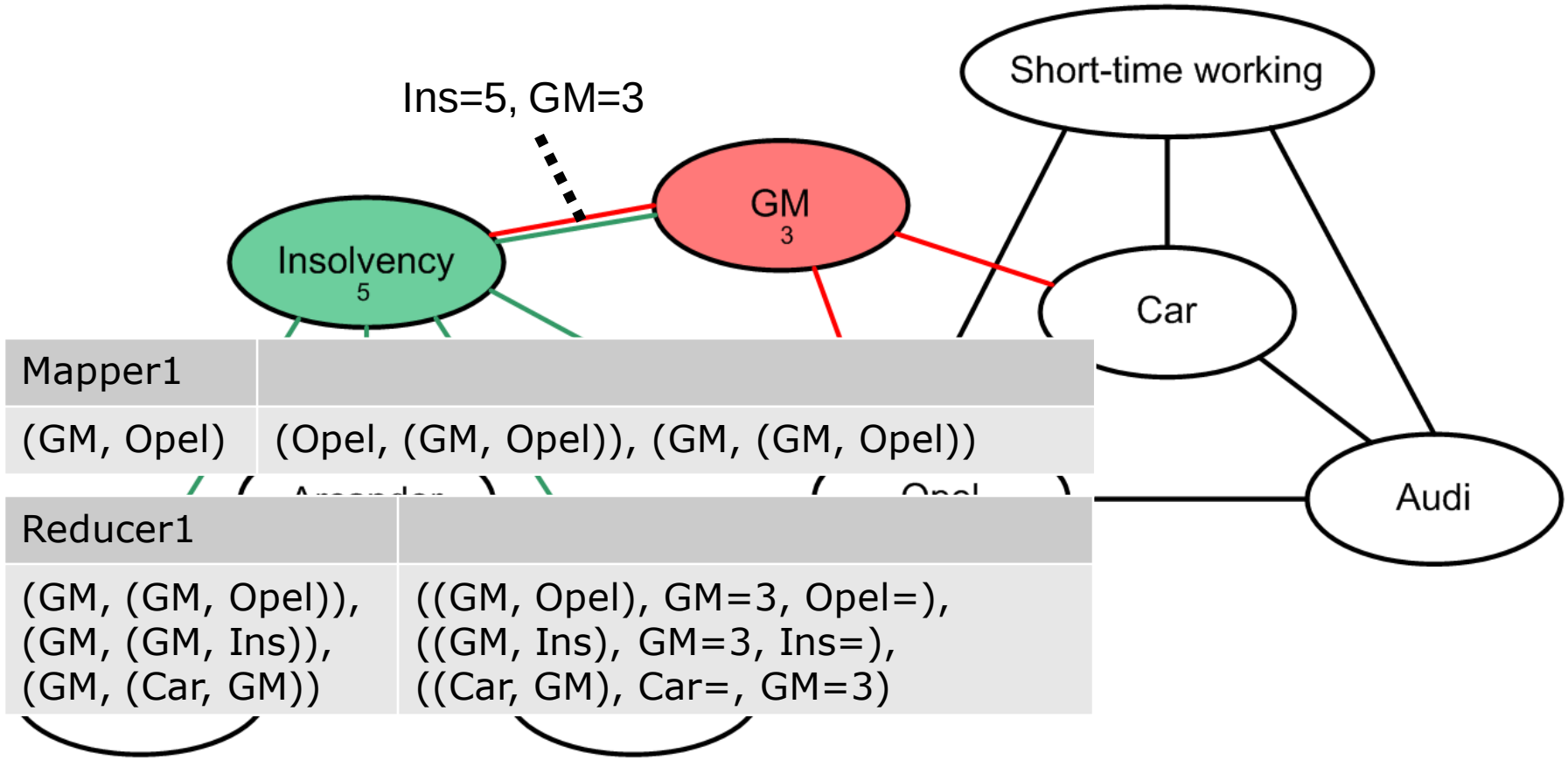
Augmenting Edges with Degrees



Augmenting Edges with Degrees



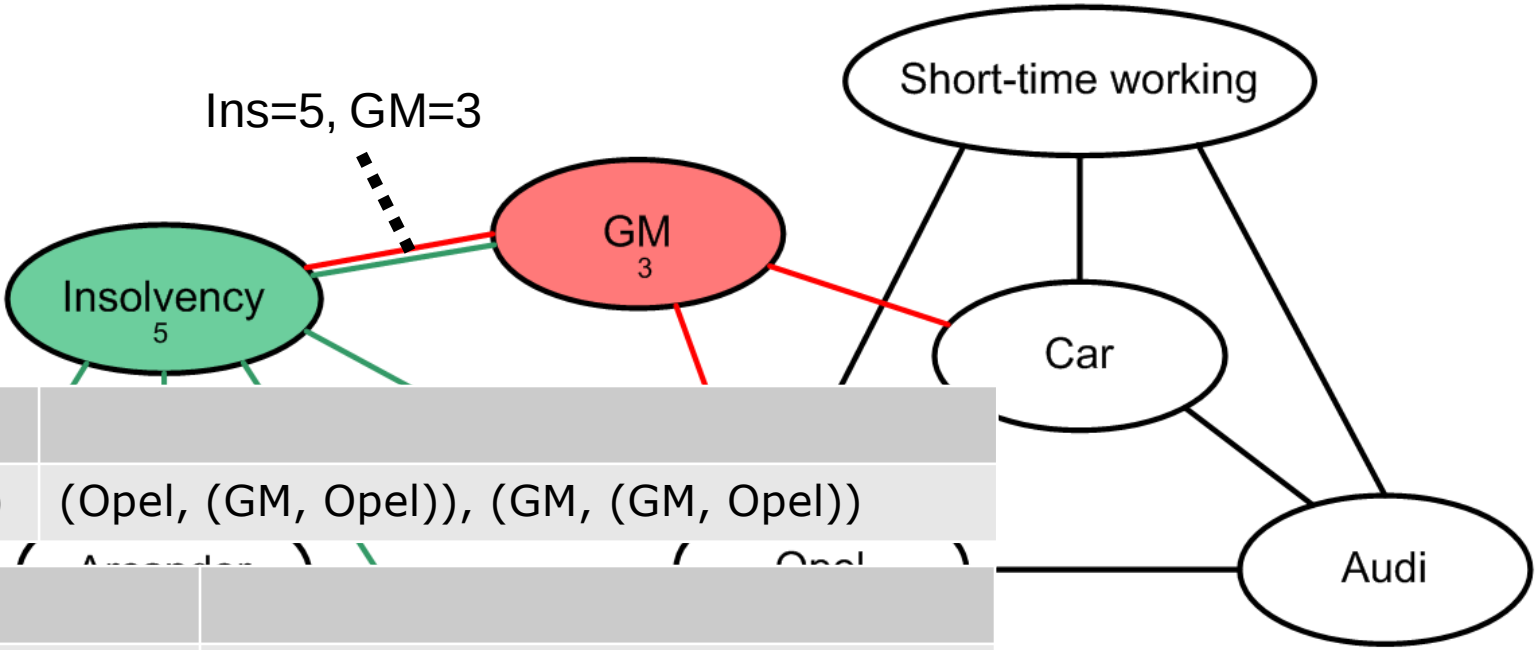
Augmenting Edges with Degrees



Mapper1	
(GM, Opel)	(Opel, (GM, Opel)), (GM, (GM, Opel))

Reducer1	
(GM, (GM, Opel)), (GM, (GM, Ins)), (GM, (Car, GM))	((GM, Opel), GM=3, Opel=), ((GM, Ins), GM=3, Ins=), ((Car, GM), Car=, GM=3)

Augmenting Edges with Degrees

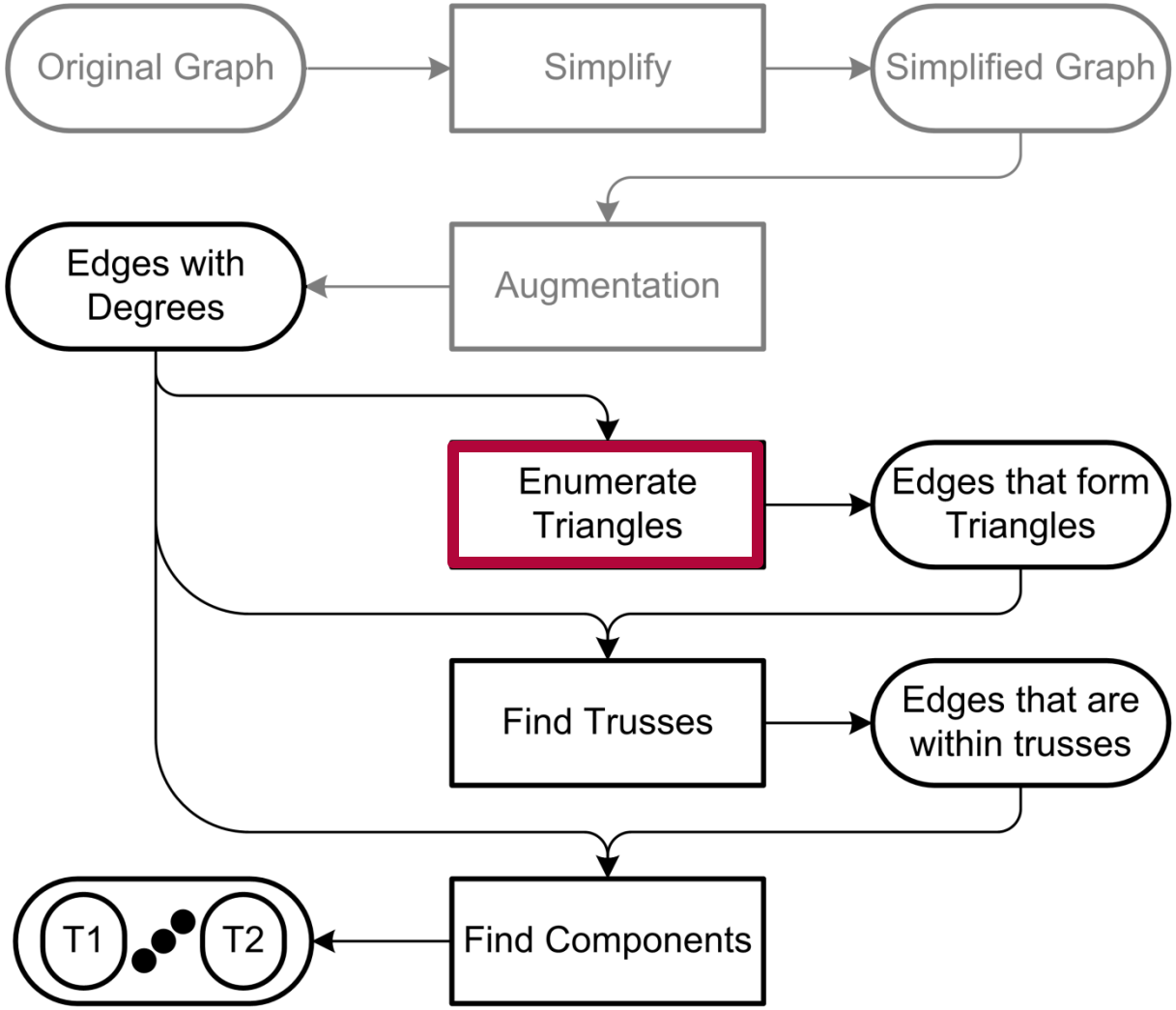


Mapper1	
(GM, Opel)	(Opel, (GM, Opel)), (GM, (GM, Opel))

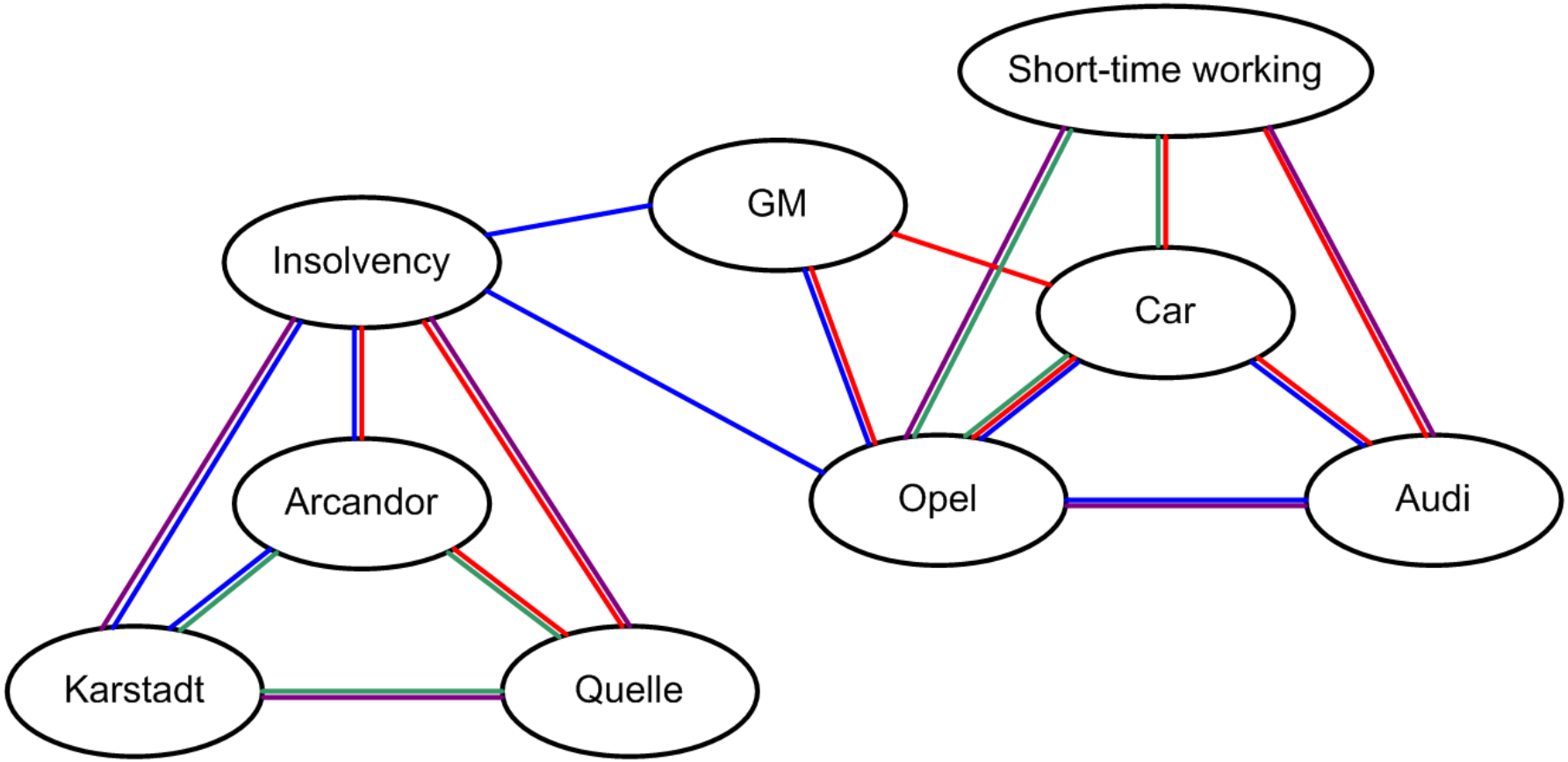
Reducer1	
(GM, (GM, Opel)), (GM, (GM, Ins)), (GM, (Car, GM))	((GM, Opel), GM=3, Opel=), ((GM, Ins), GM=3, Ins=), ((Car, GM), Car=, GM=3)

Reducer2	
((GM, Ins), GM=3, Ins=), ((GM, Ins), GM=, Ins=5)	((GM, Ins), GM=3, Ins=5)

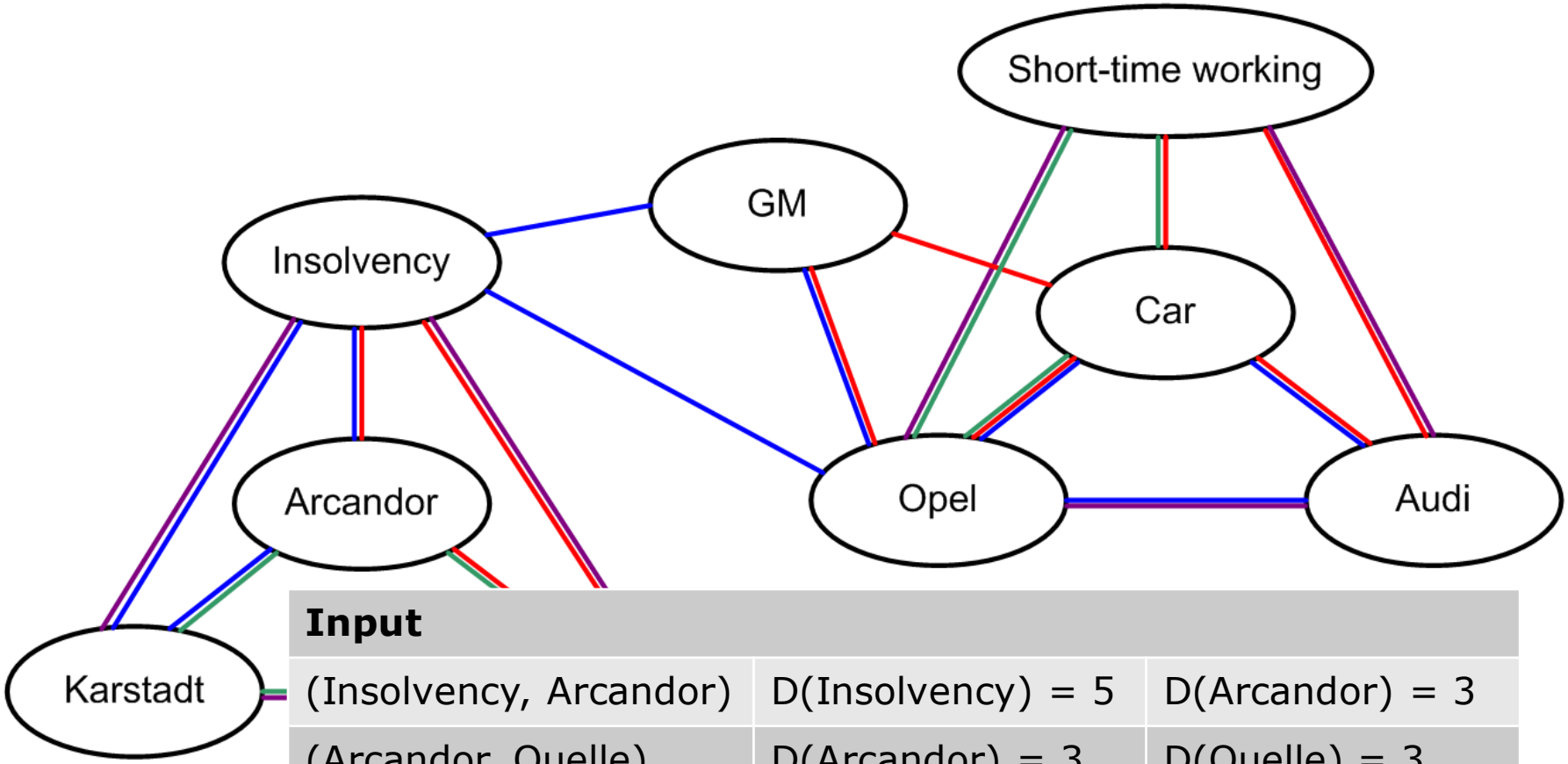
Process Overview



In Detail: Finding Triangles 1|5
Input

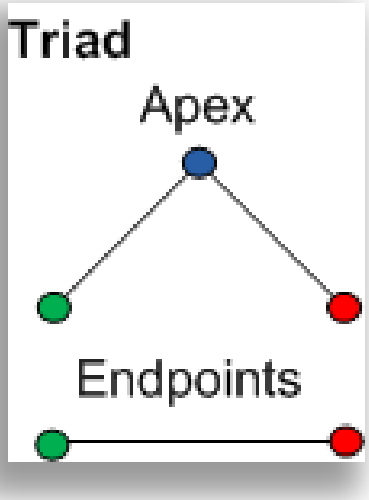


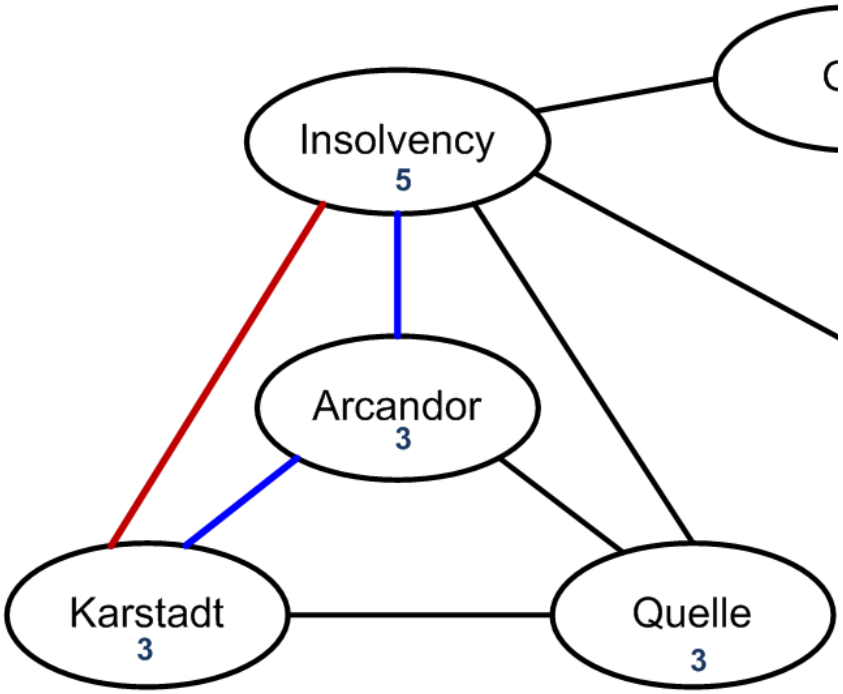
In Detail: Finding Triangles 1|5
Input



Input		
(Insolvency, Arcandor)	D(Insolvency) = 5	D(Arcandor) = 3
(Arcandor, Quelle)	D(Arcandor) = 3	D(Quelle) = 3
(Insolvency, GM)	D(Insolvency) = 5	D(GM) = 3
…	…	…

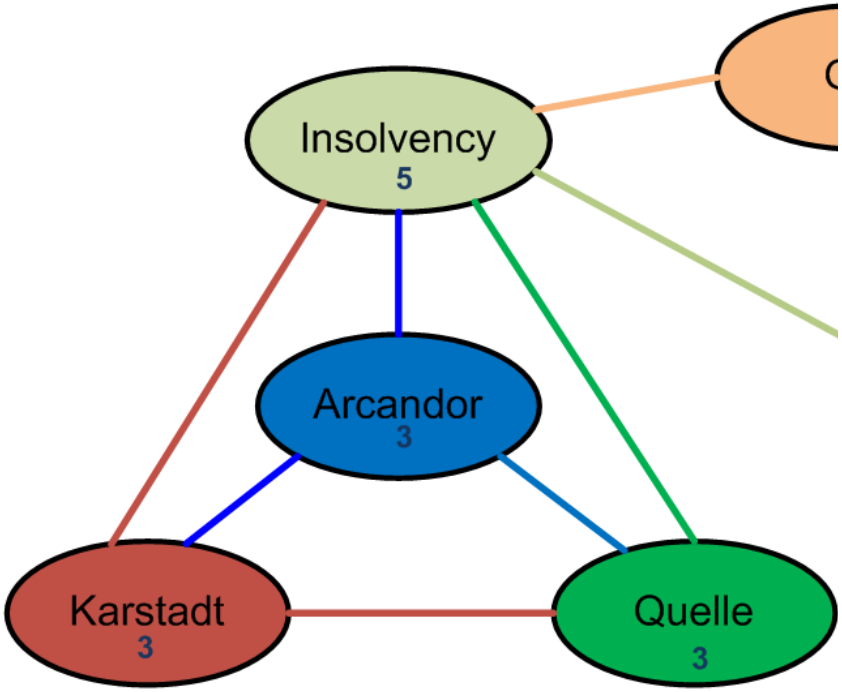
In Detail: Finding Triangles 2|5 Idea

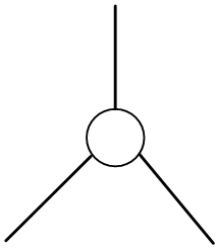




In Detail: Finding Triangles 3|5
1. Mapper: Key by Apex

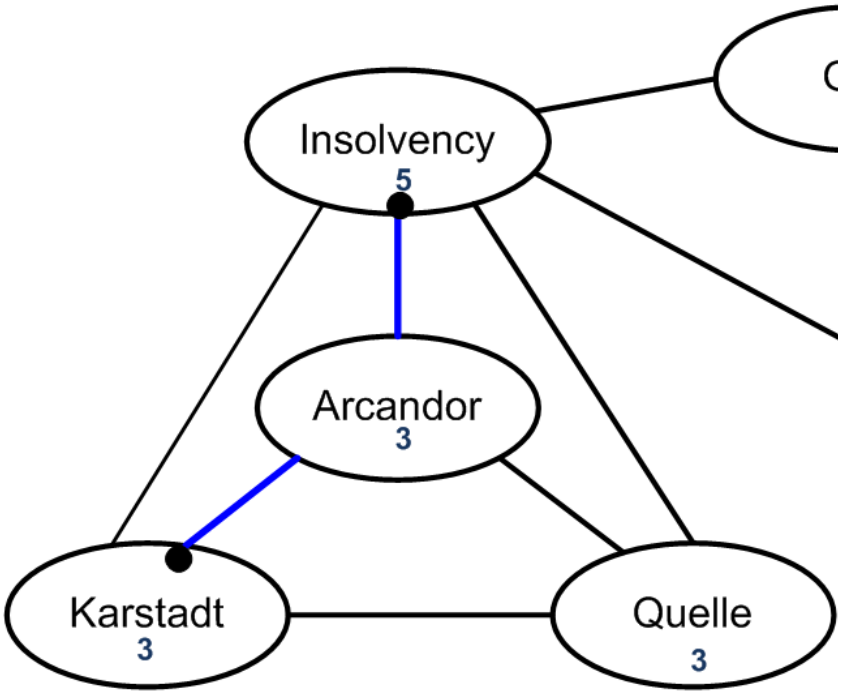
Output of first Mapper: Edges to Apex	
Key (Apex)	**Edges**
Arcandor	(**Arcandor**, Insolvency) (**Arcandor,** Karstadt) (**Arcandor**, Quelle)
Quelle	(Insolvency, **Quelle**)
Karstadt	(Insolvency, **Karstadt**) (**Karstadt**, Quelle)

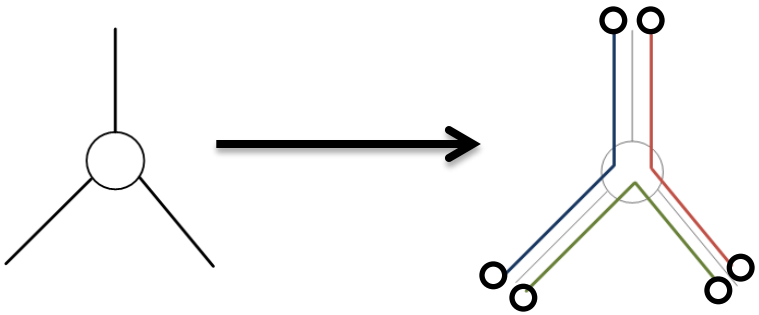




In Detail: Finding Triangles 4|5
1. Reduce: Form Triads

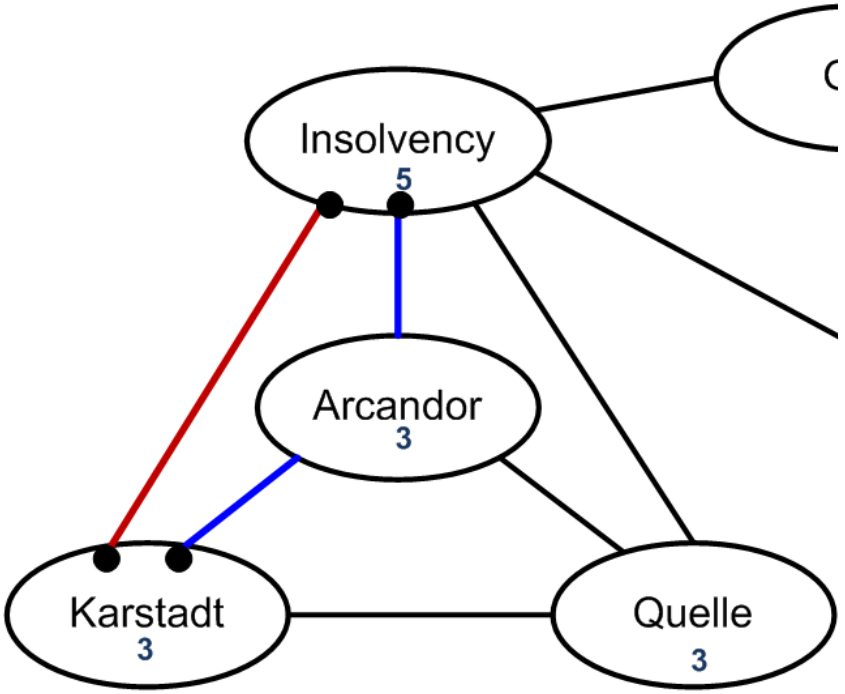
Output of first Reducer: Emit Triads, keyed by endpoints	
Key	**Triad**
Insolvency, Karstadt	(Arcandor, Insolvency) (Arcandor, Karstadt)
Karstadt, Quelle	(Arcandor, Karstadt) (Arcandor, Quelle)
…	…

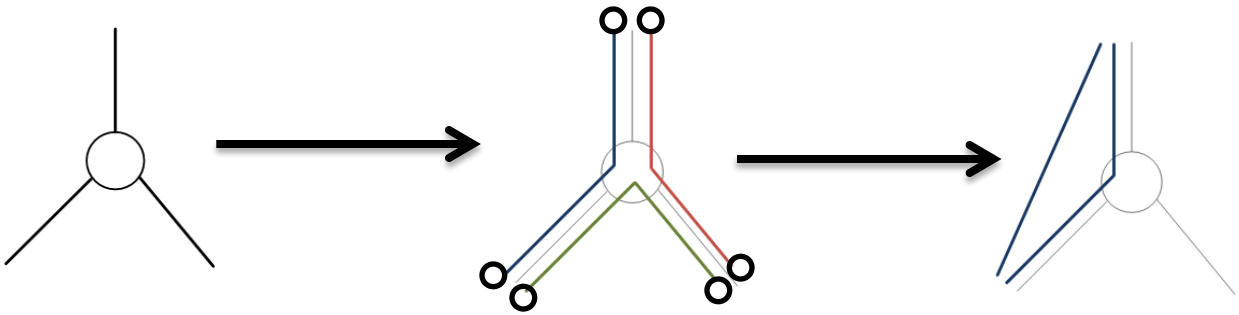




In Detail: Finding Triangles 5|5
2. Phase: Close Triads with Edges

Output of second Mapper: Map edges and triads to their endpoints	
Key	**Triad or Edge**
Insolvency, Karstadt	Arcandor, Insolvency Arcandor, Karstadt
Insolvency, Karstadt	Insolvency, Karstadt
	…



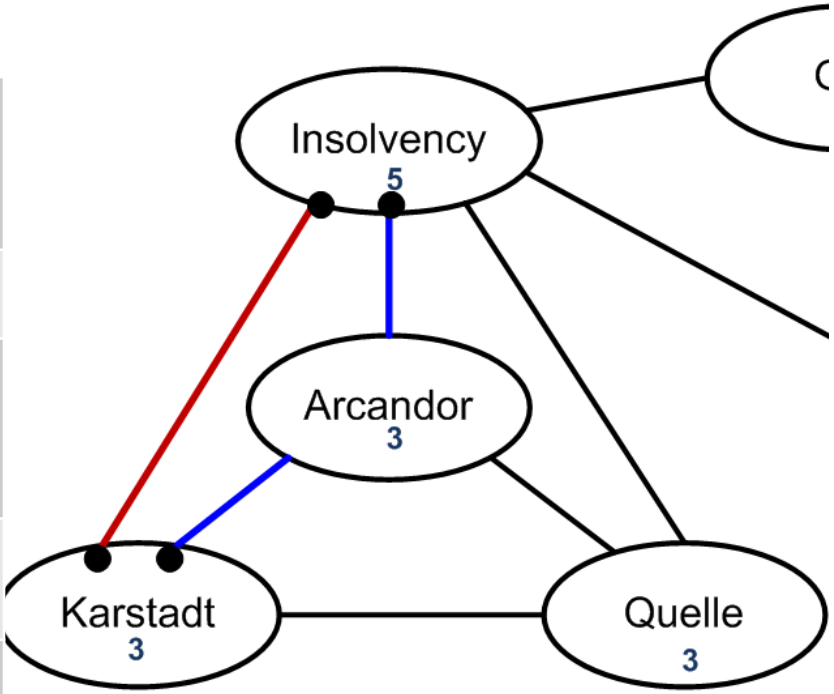


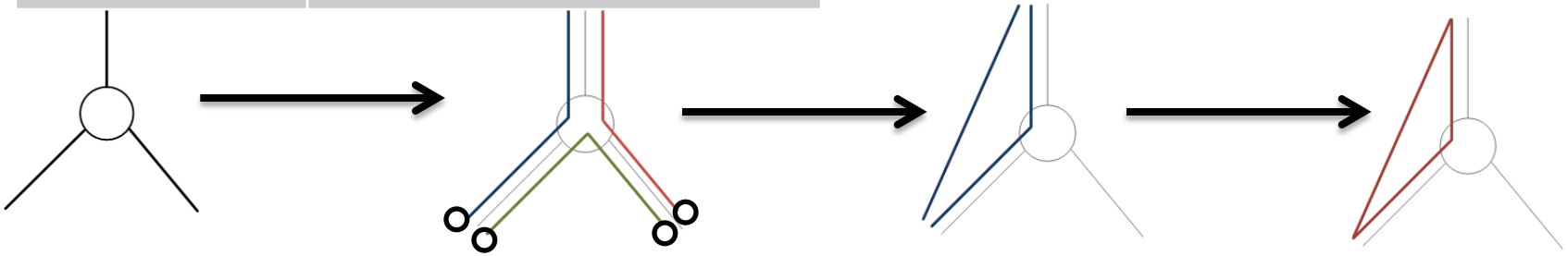
In Detail: Finding Triangles 5|5
2. Phase: Close Triads with Edges

Output of second Mapper:

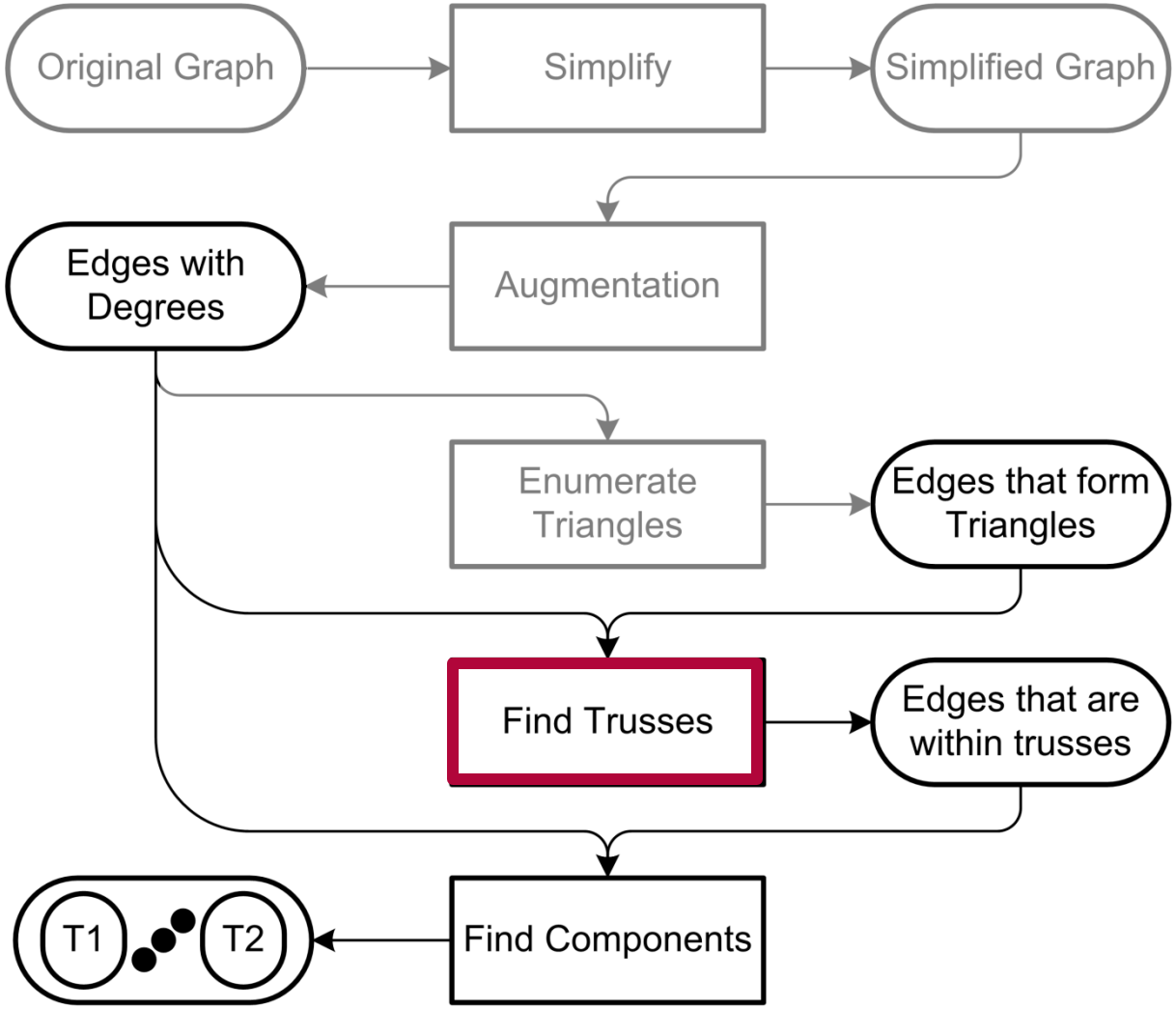
Output of second Reducer:
Emit Triangles
Merge Edges and Triads to Triangles

Key	Triangle
…	Arcandor, Insolvency Arcandor, Karstadt Insolvency, Karstadt
	…
	…

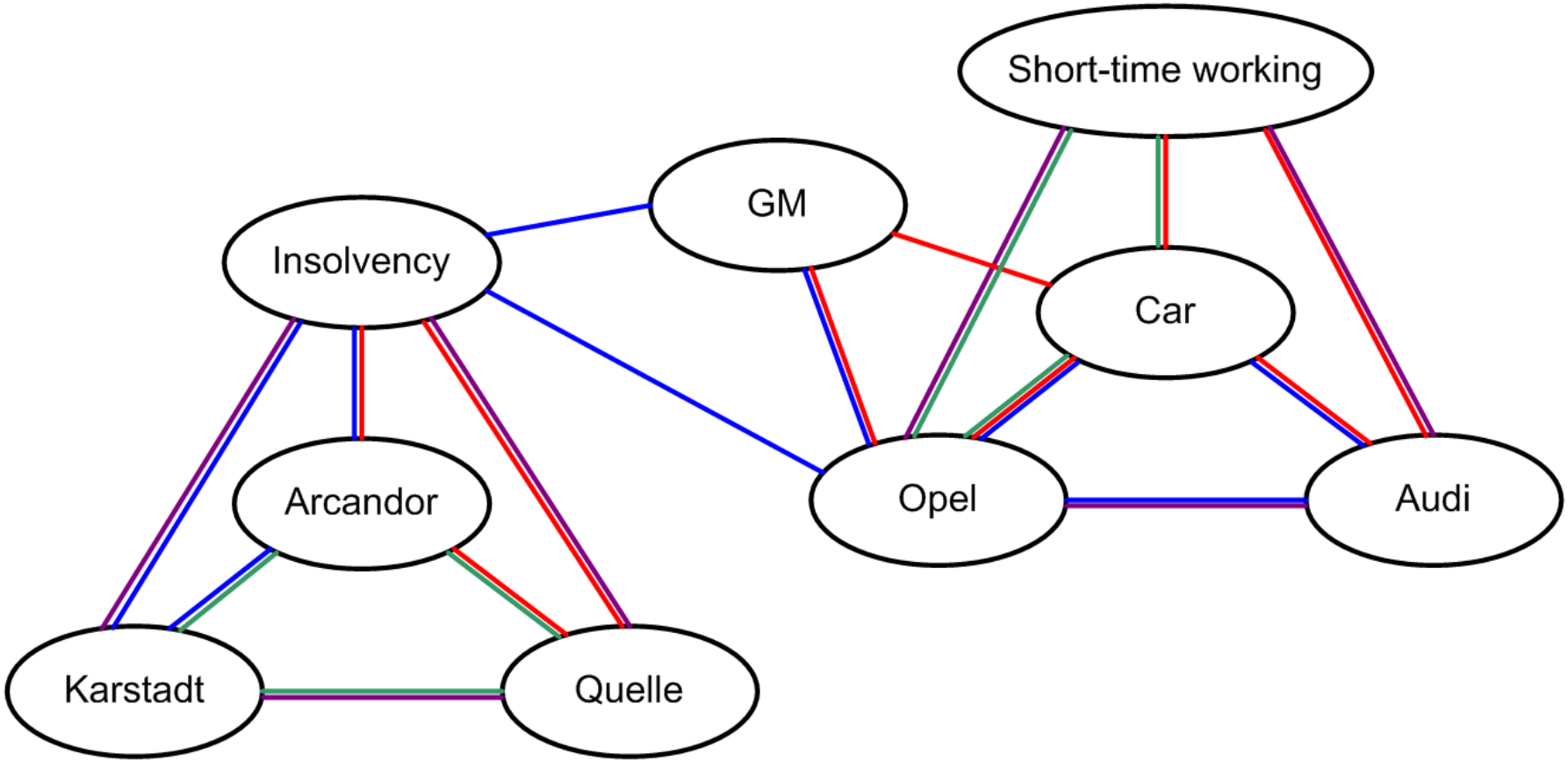




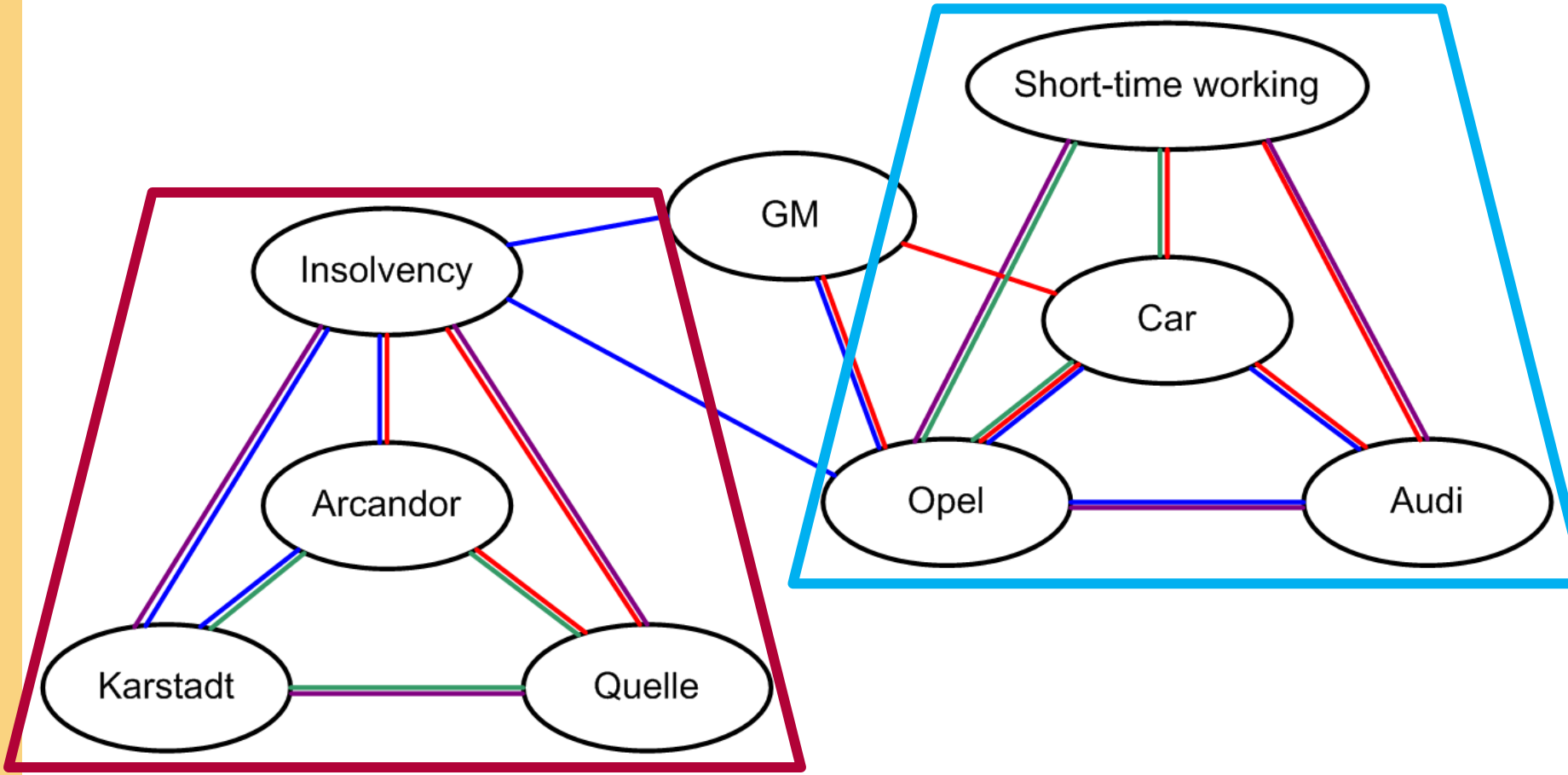
Process Overview



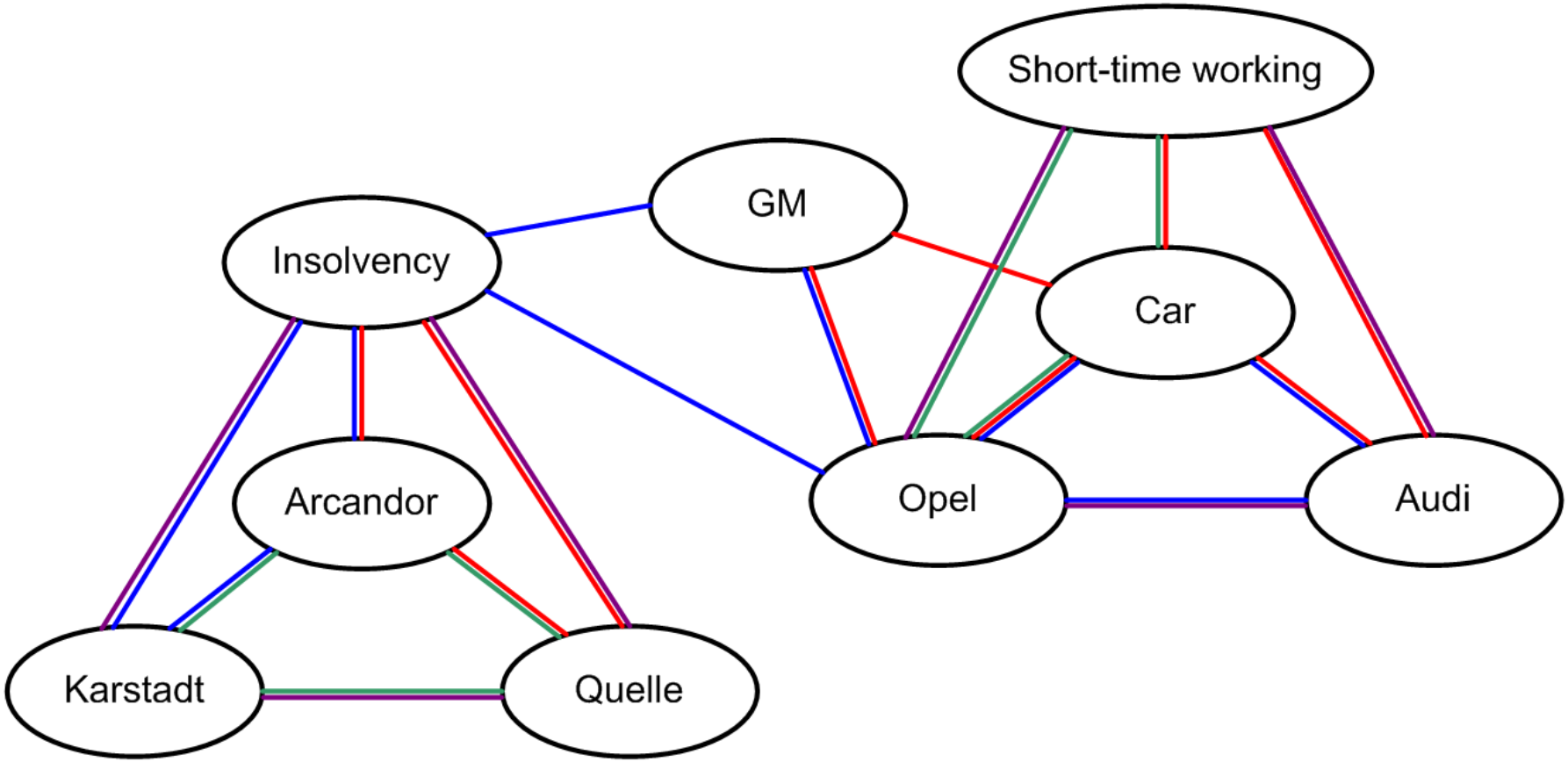
Finding Trusses



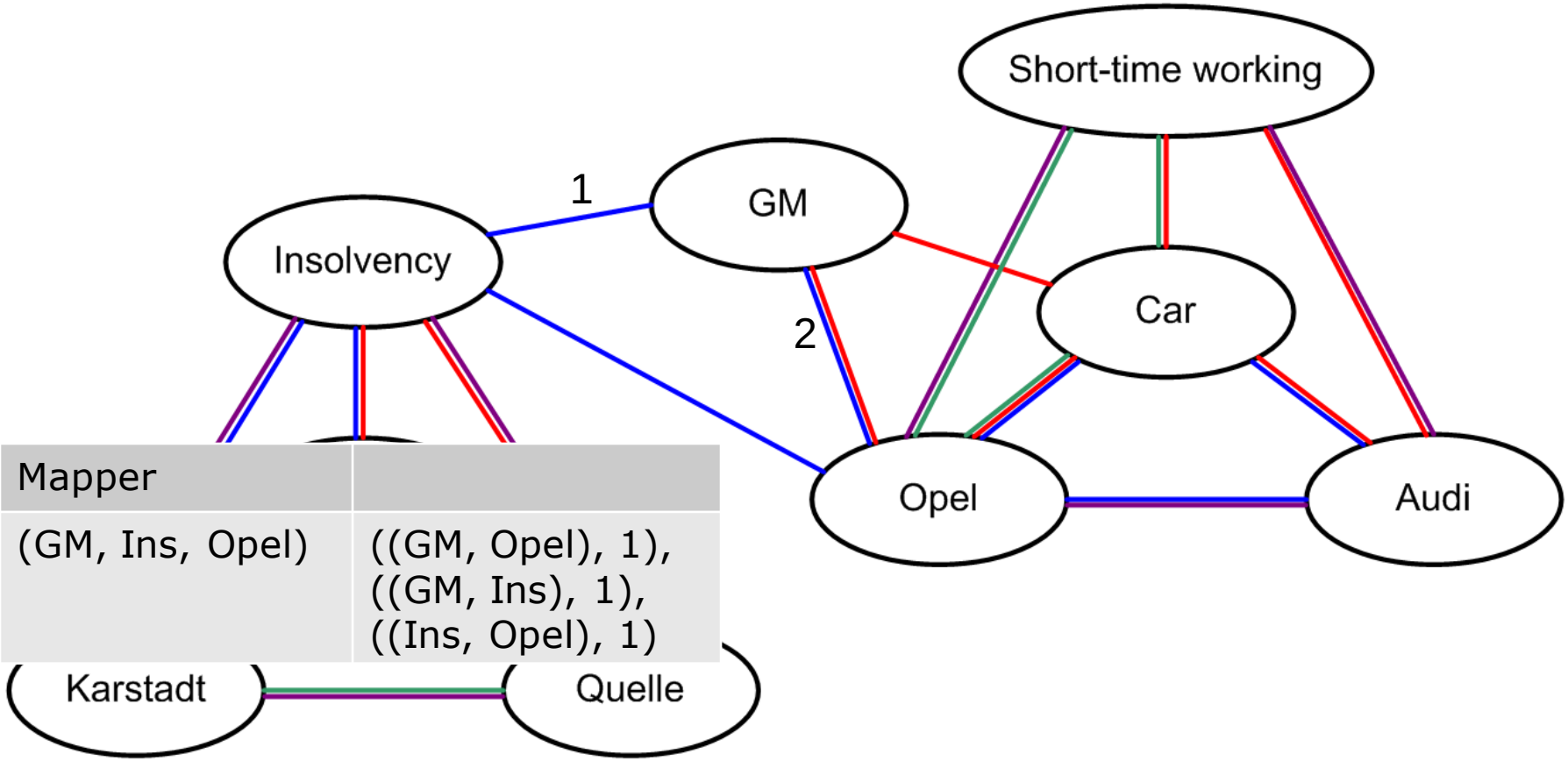
Finding Trusses



Finding Trusses

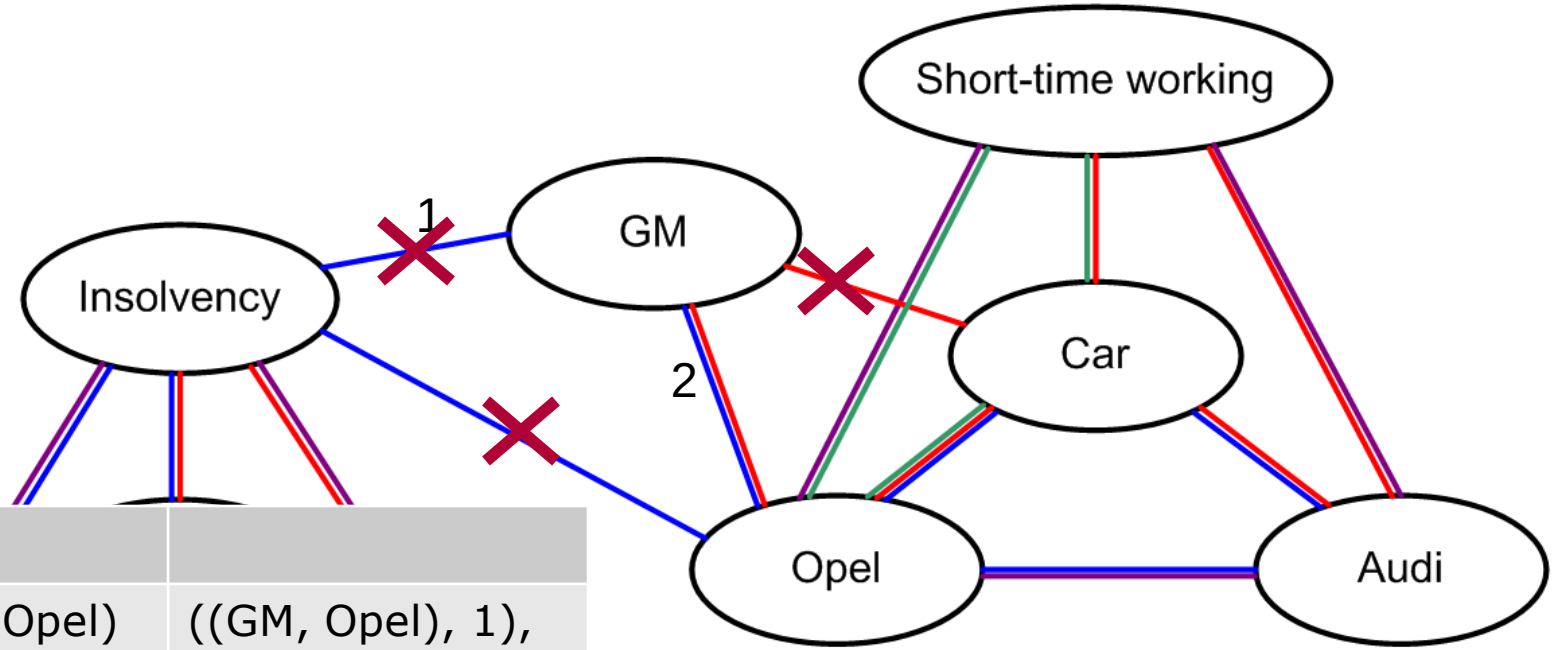


Finding Trusses



Mapper	
(GM, Ins, Opel)	((GM, Opel), 1), ((GM, Ins), 1), ((Ins, Opel), 1)

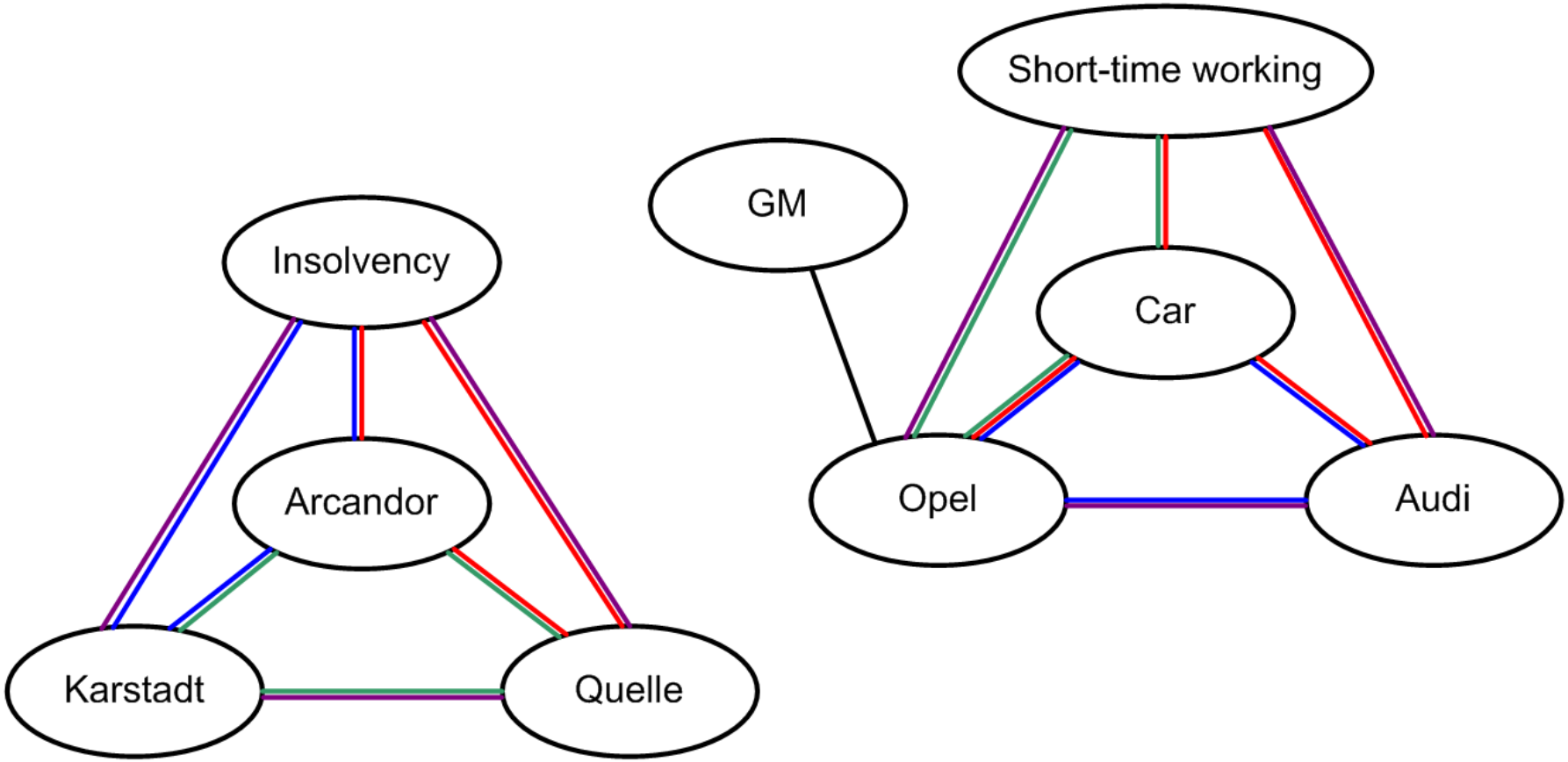
Finding Trusses



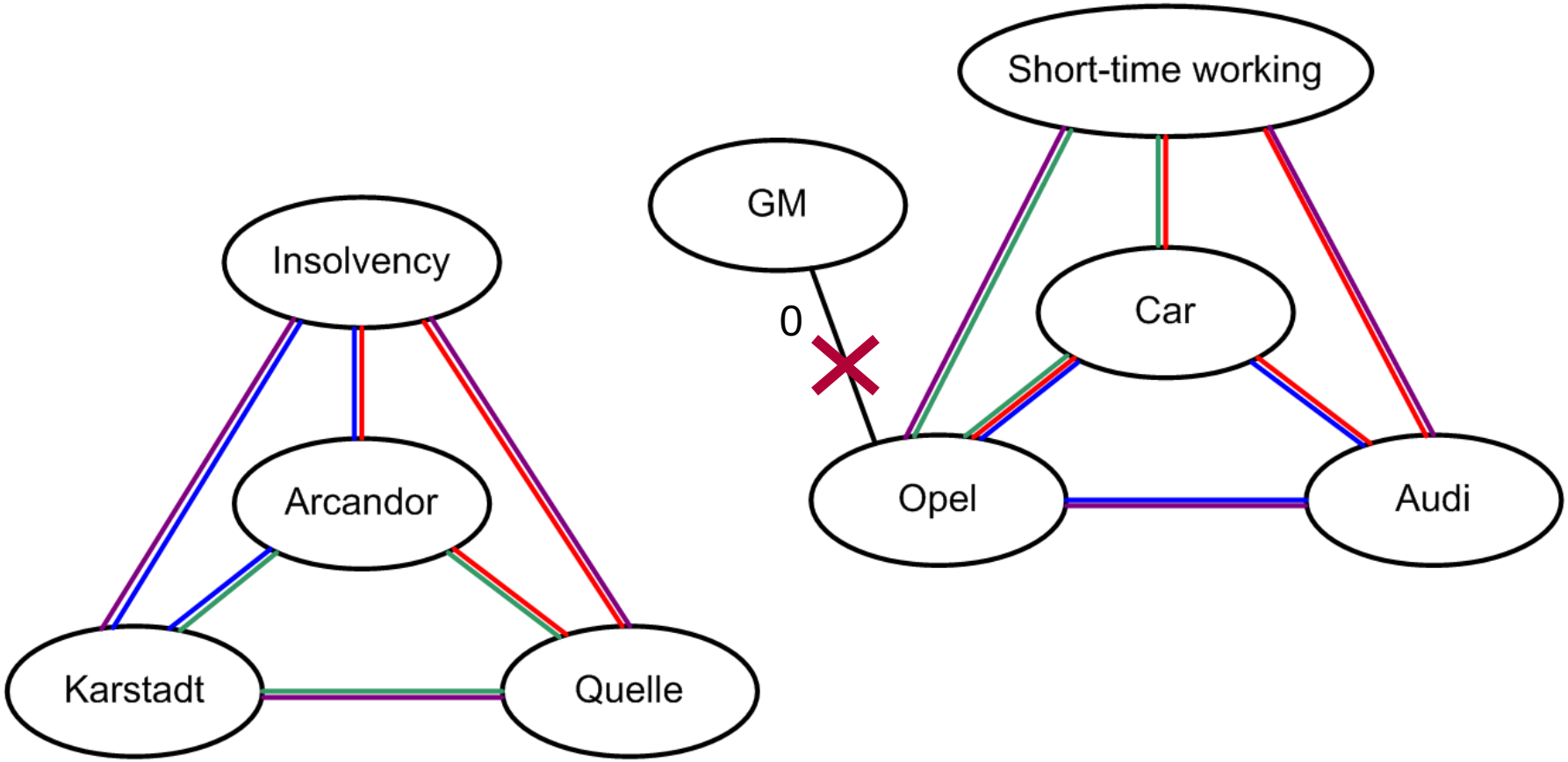
Mapper	
(GM, Ins, Opel)	((GM, Opel), 1), ((GM, Ins), 1), ((Ins, Opel), 1)

Reducer	
((GM, Ins), 1)	∅
((GM, Opel), 1), ((GM, Opel), 1)	(GM, Opel)

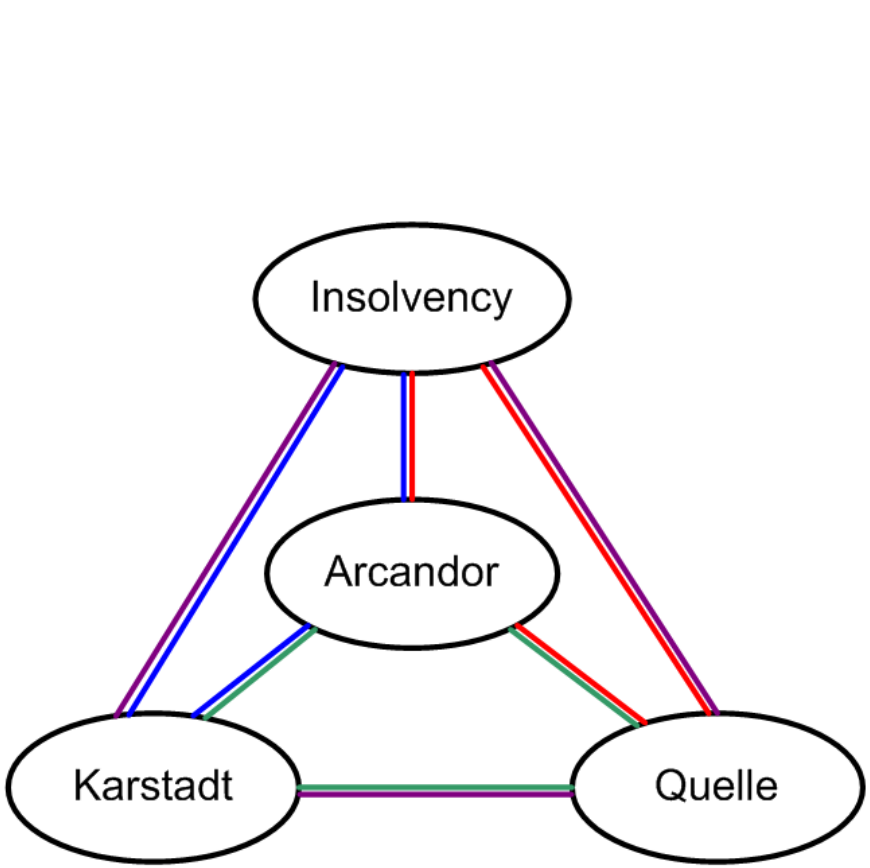
Finding Trusses #2

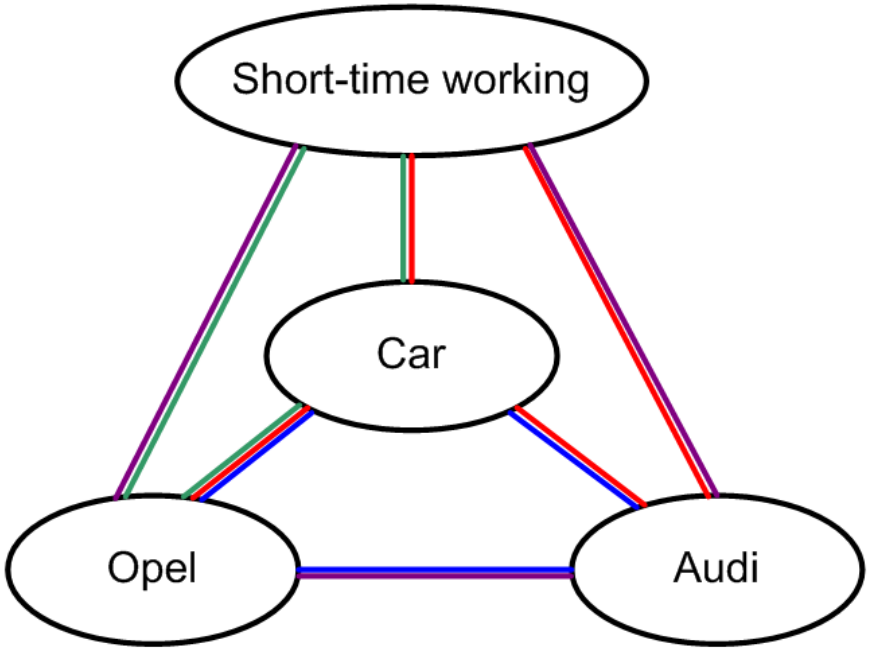


Finding Trusses #2

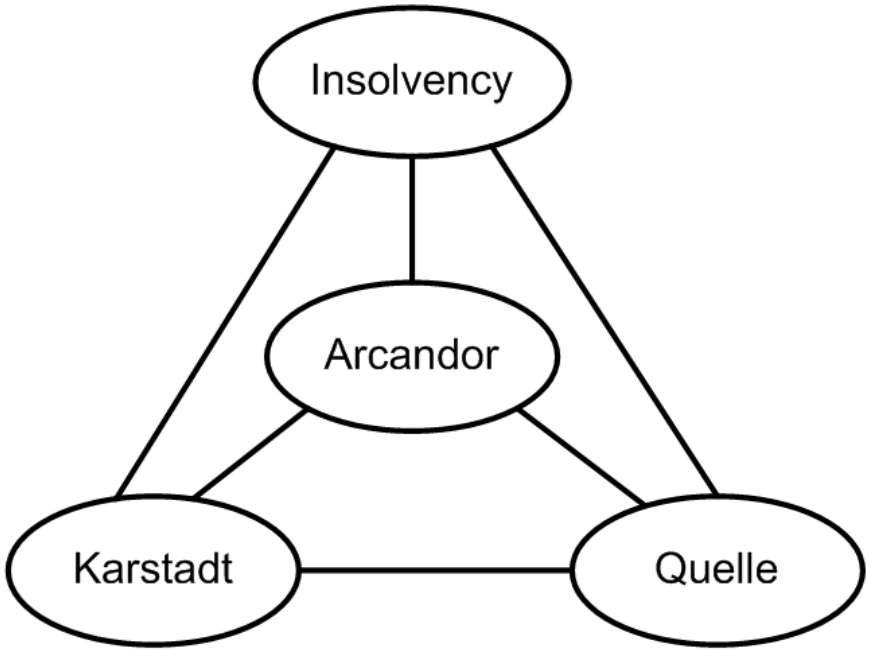


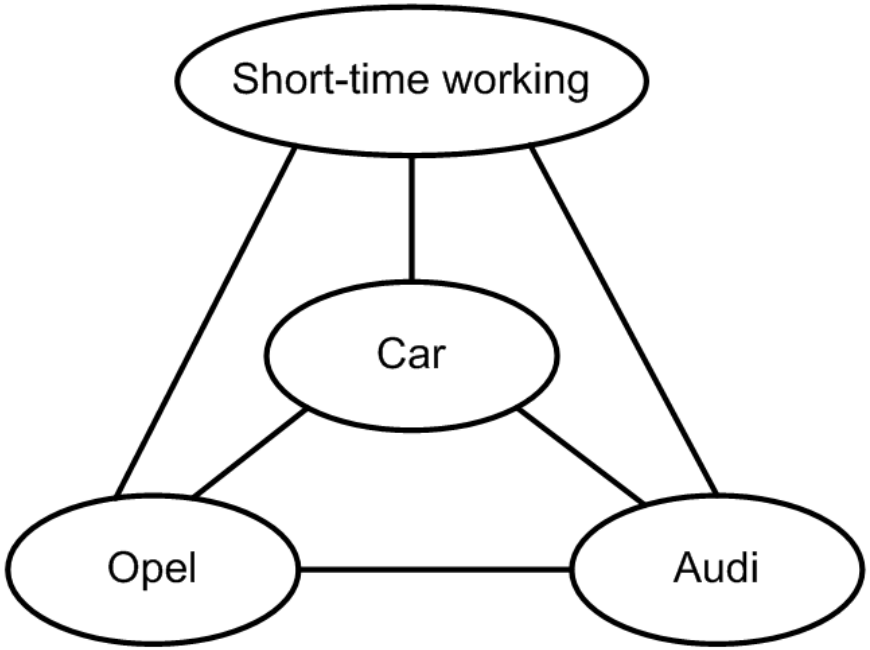
Finding Trusses #2





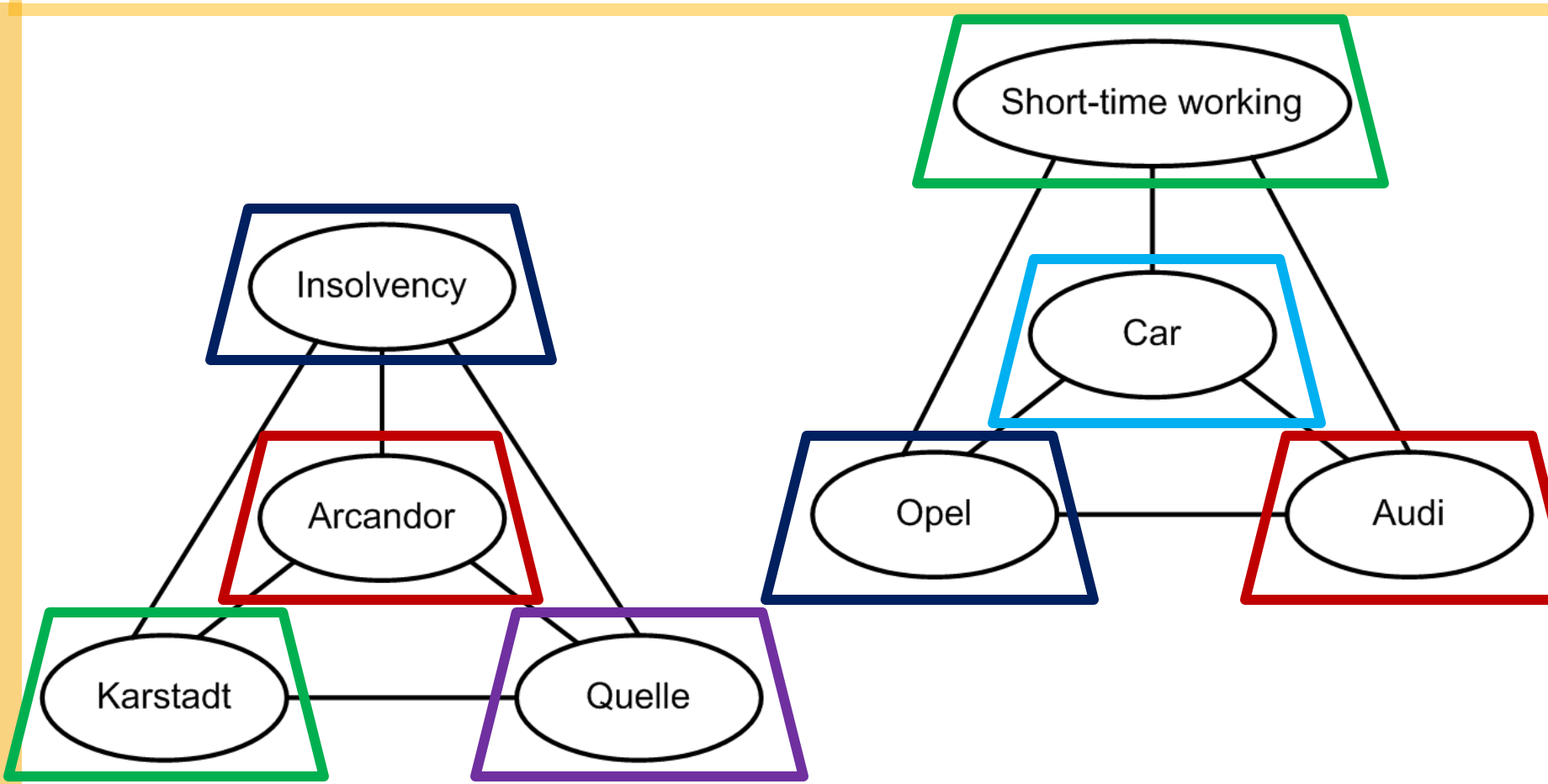
Finding Components





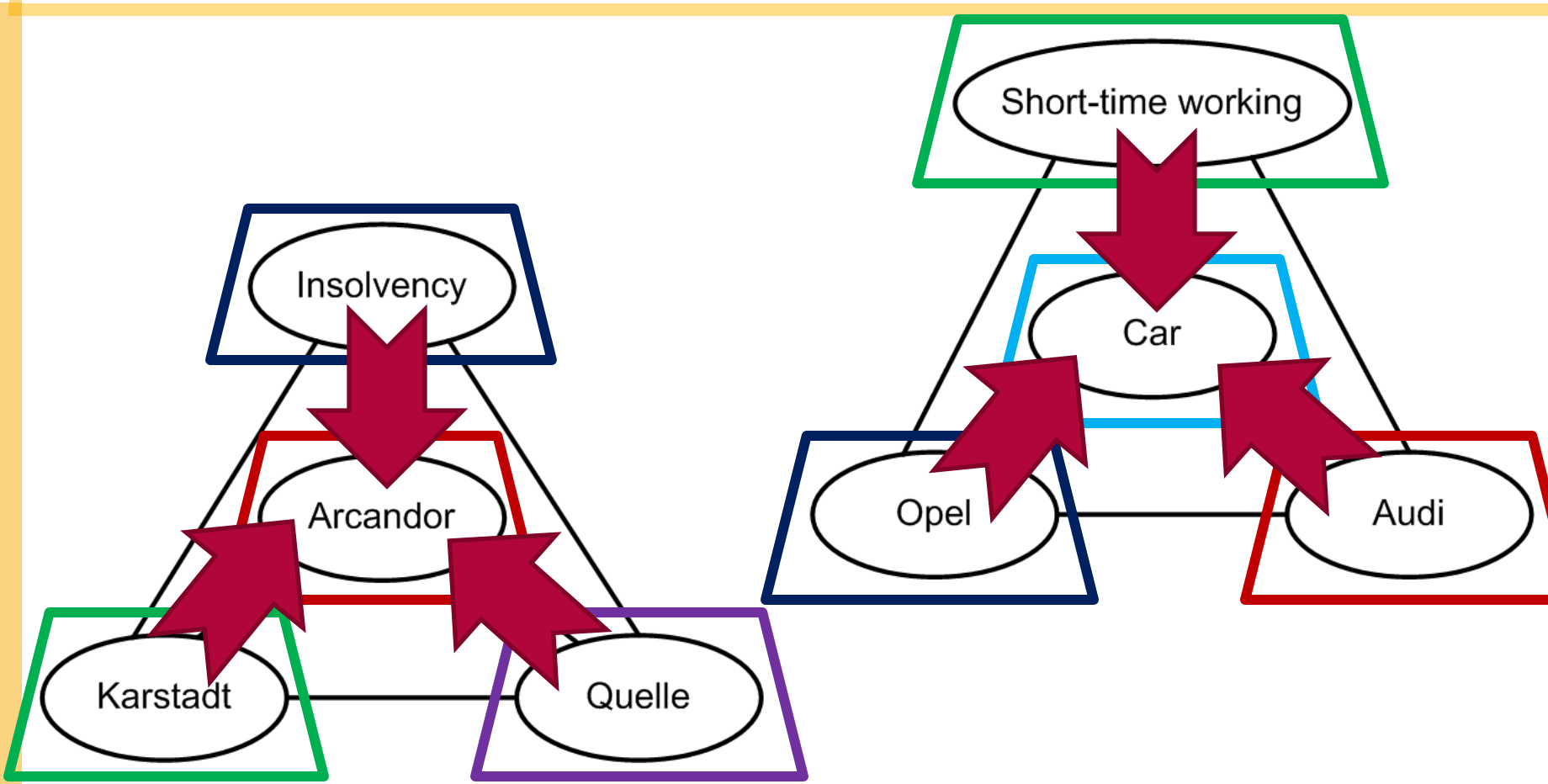
- Expands zones from all nodes simultaneously
- Merges incrementally zones in two steps (3 MapReduce Jobs)

Finding Components



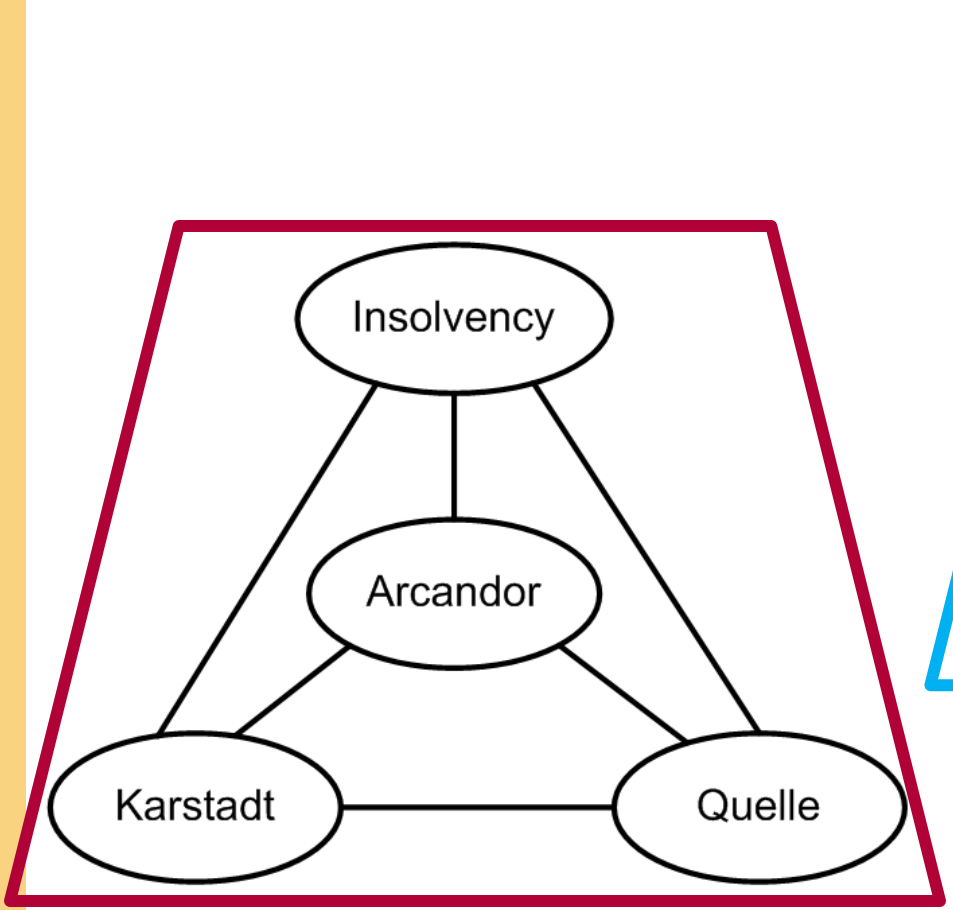
- Expands zones from all nodes simultaneously
- Merges incrementally zones in two steps (3 MapReduce Jobs)

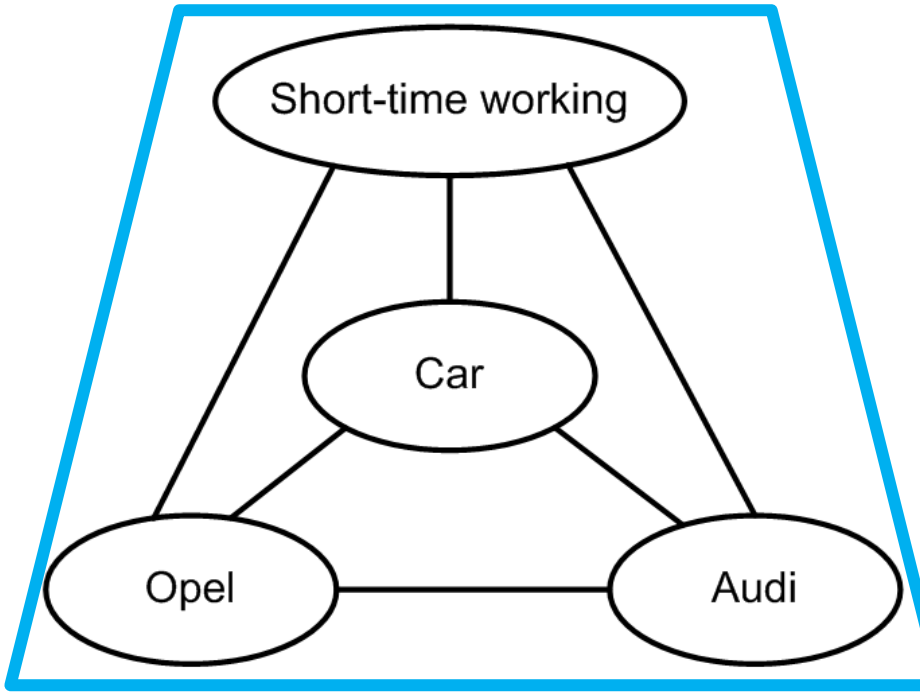
Finding Components



- Expands zones from all nodes simultaneously
- Merges incrementally zones in two steps (3 MapReduce Jobs)

Finding Components





- Expands zones from all nodes simultaneously
- Merges incrementally zones in two steps (3 MapReduce Jobs)

Outlook

- Implement 5 algorithms
  - Simplification, augmentation of edges with degrees ✓
  - Enumeration of triangles, trusses, and components

- Evaluation
  - Resource usage / scalability
    - Amount of data might be bottle neck
  - Is it applicable for all graphs?
    - DBpedia and sub graphs as test data
  - Description of algorithms