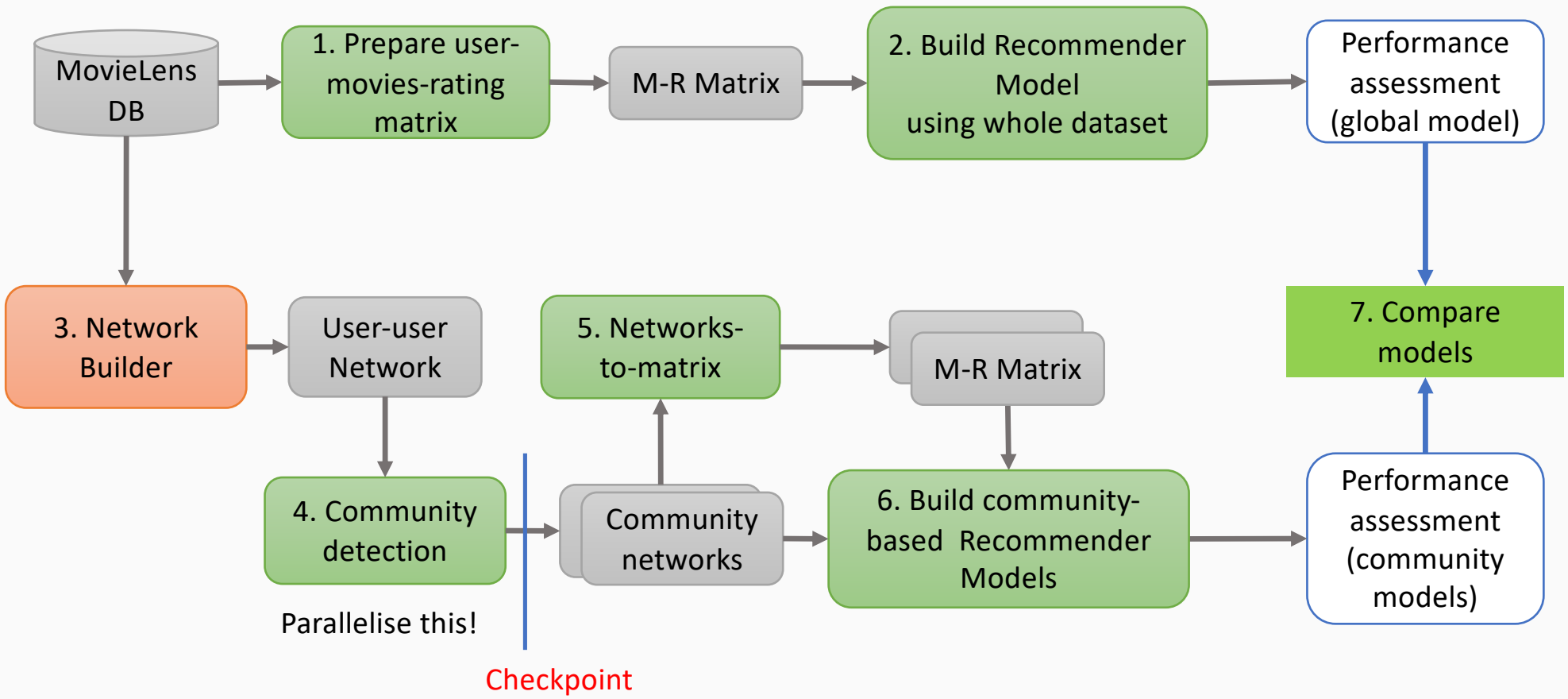
MovieLens DB
1. Prepare user-movies-rating matrix
M-R Matrix
2. Build Recommender Model using whole dataset
Performance assessment (global model)
3. Network Builder
User-user Network
5. Networks-to-matrix
M-R Matrix
7. Compare models
4. Community detection
Community networks
6. Build community-based Recommender Models
Performance assessment (community models)
Parallelise this!
Checkpoint

Girvan-Newman Parallel Implementation

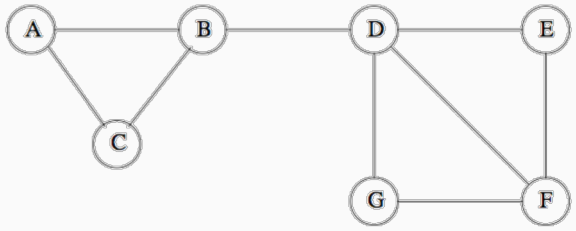

Figure 10.1: Example of a small social network

The edge betweenness for edge e:

$$\sum_{s,t \neq v} \frac{\sigma_{st}(e)}{\sigma_{st}},$$

where

- σ_{st} is total number of shortest paths from node s to node t and $\sigma_{st}(e)$ is the number of those paths that pass through e.

Implementation in Spark:

1. Parse graph and broadcast to workers
2. Partition graph nodes in RDD
3. Single source shortest path sequential in each partition (worker). *Map*
4. Compute betweenness in parallel. *Reduce*
5. Remove edges of highest betweenness in graph objects in workers