

Embedding Similarity aided Relationship Prediction in Heterogeneous Networks

Pramod Srinivasan and Vipul Venkataraman
Guided by Dr. Meng Jiang

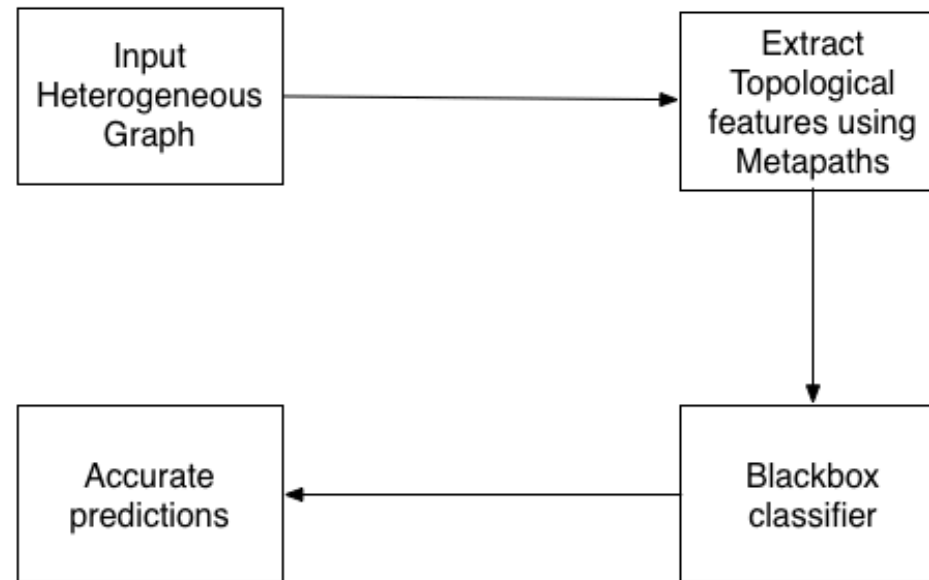
Outline

- Problem Statement
- PathPredict: Recap + Issues
- Using Embedding Similarity
- Workflow of our approach
- Experiments + Results
- Future work

Introduction

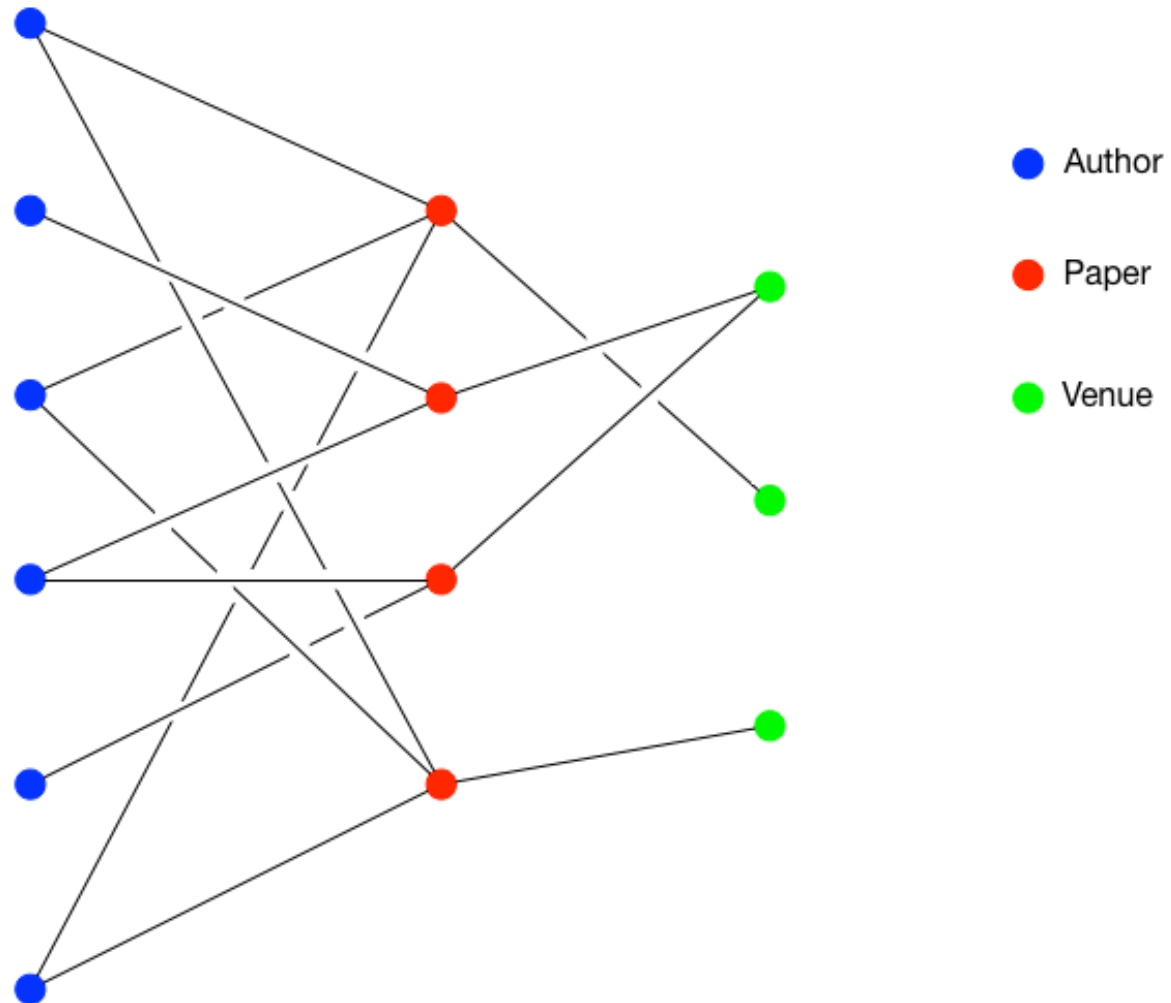
- Link prediction: Important problem with varied applications
- Our goal: Efficient relationship prediction in heterogeneous networks
 - DBLP
 - Task: Future acceptance (author, venue)

Path Predict: Workflow



- Facilitates testing significance of topological features

Issues with Path Predict



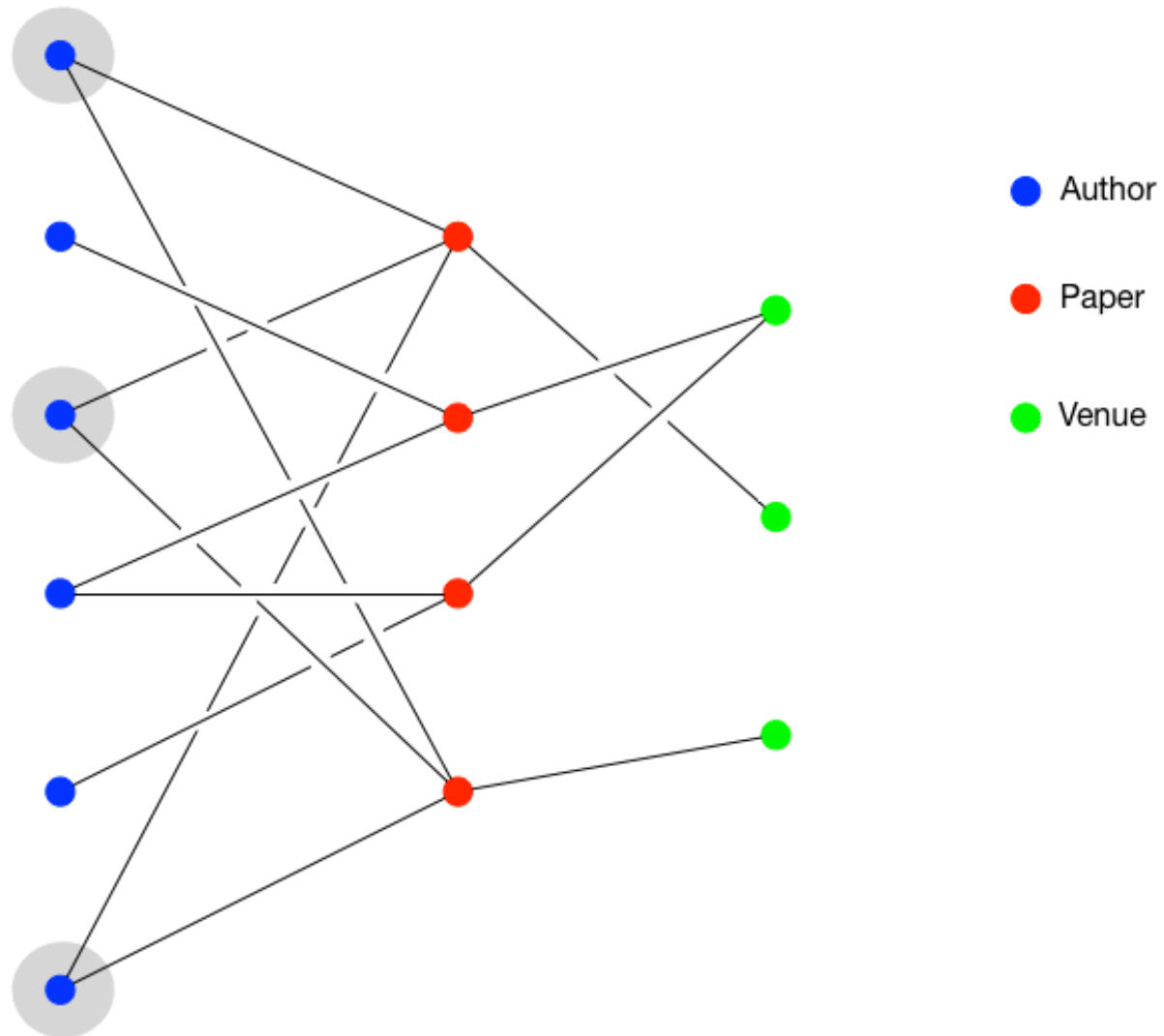
Intuition

- Challenge: Sparsity of the original network
 - 3-partite
 - Less informative meta-path features
- Key idea: Learn node similarity using network embedding
 - Construct richer meta-paths!
 - Can use any off-the-shelf embedding technique

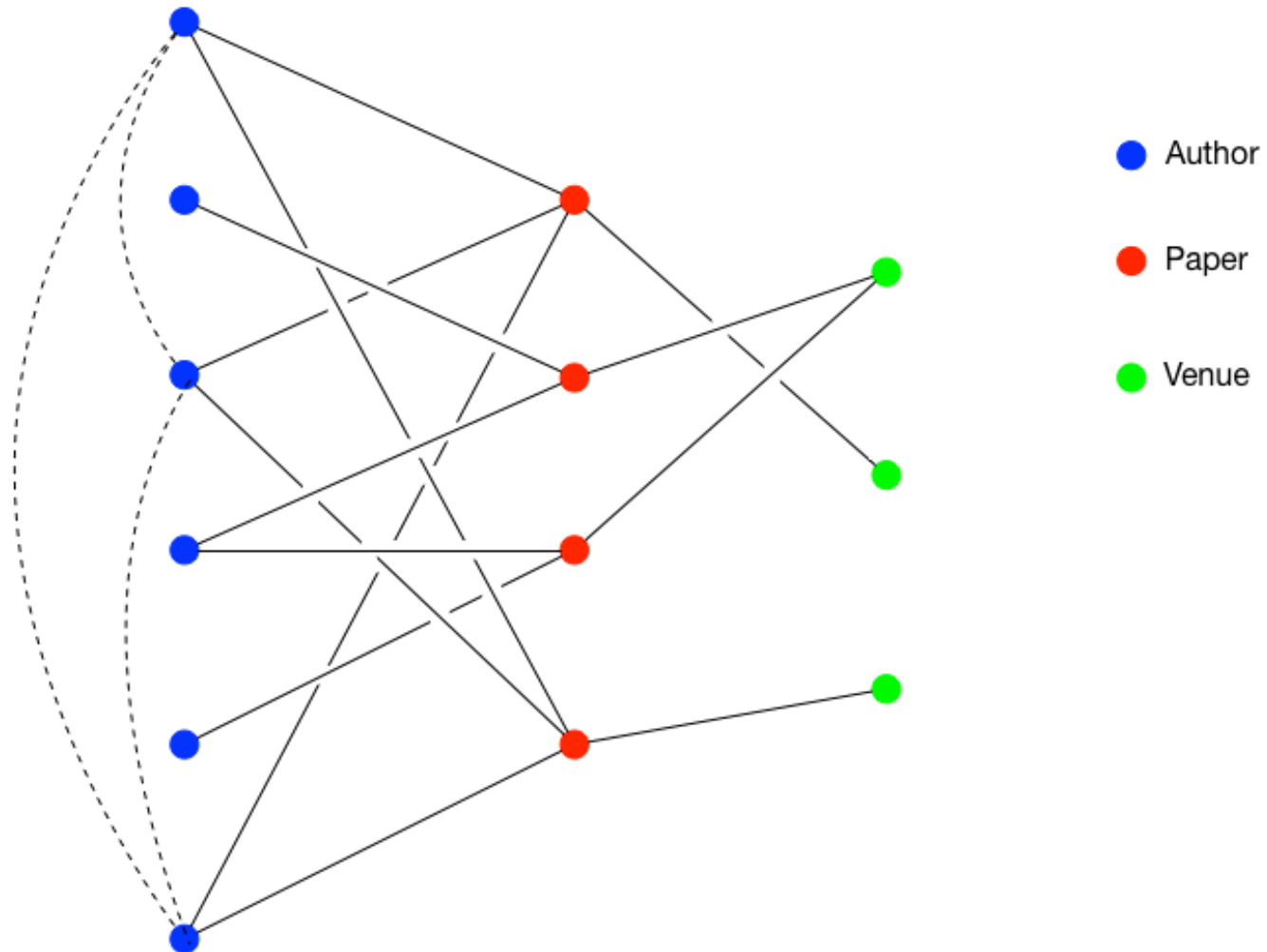
LINE: Quick Recap

- LINE: Large-scale Information Network Embedding
- Embeds large networks into low dimensional vector spaces
- Highly parallelizable!
- Preserves both local and global network structures
 - First order proximity
 - Second order proximity

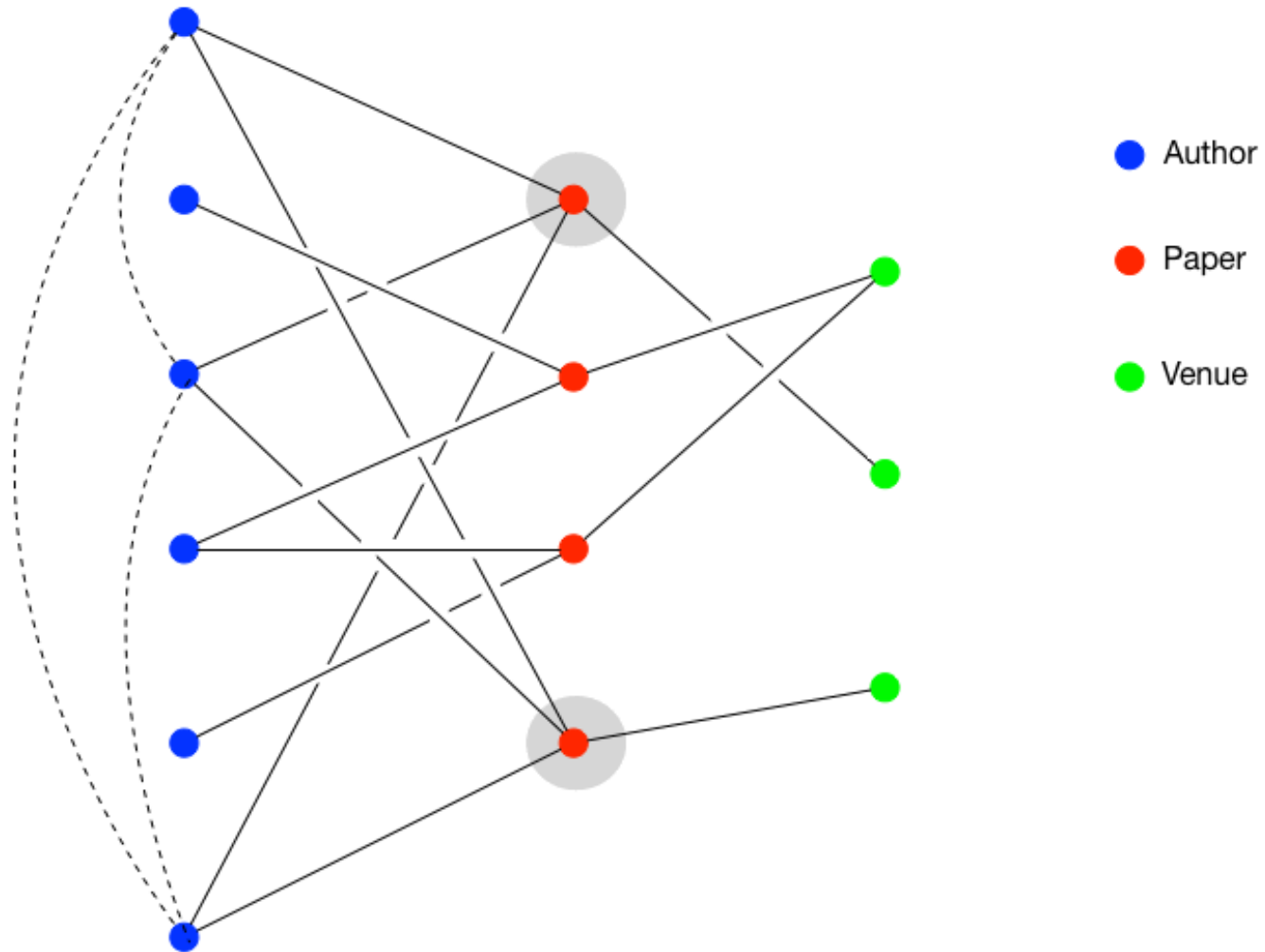
Similar Authors



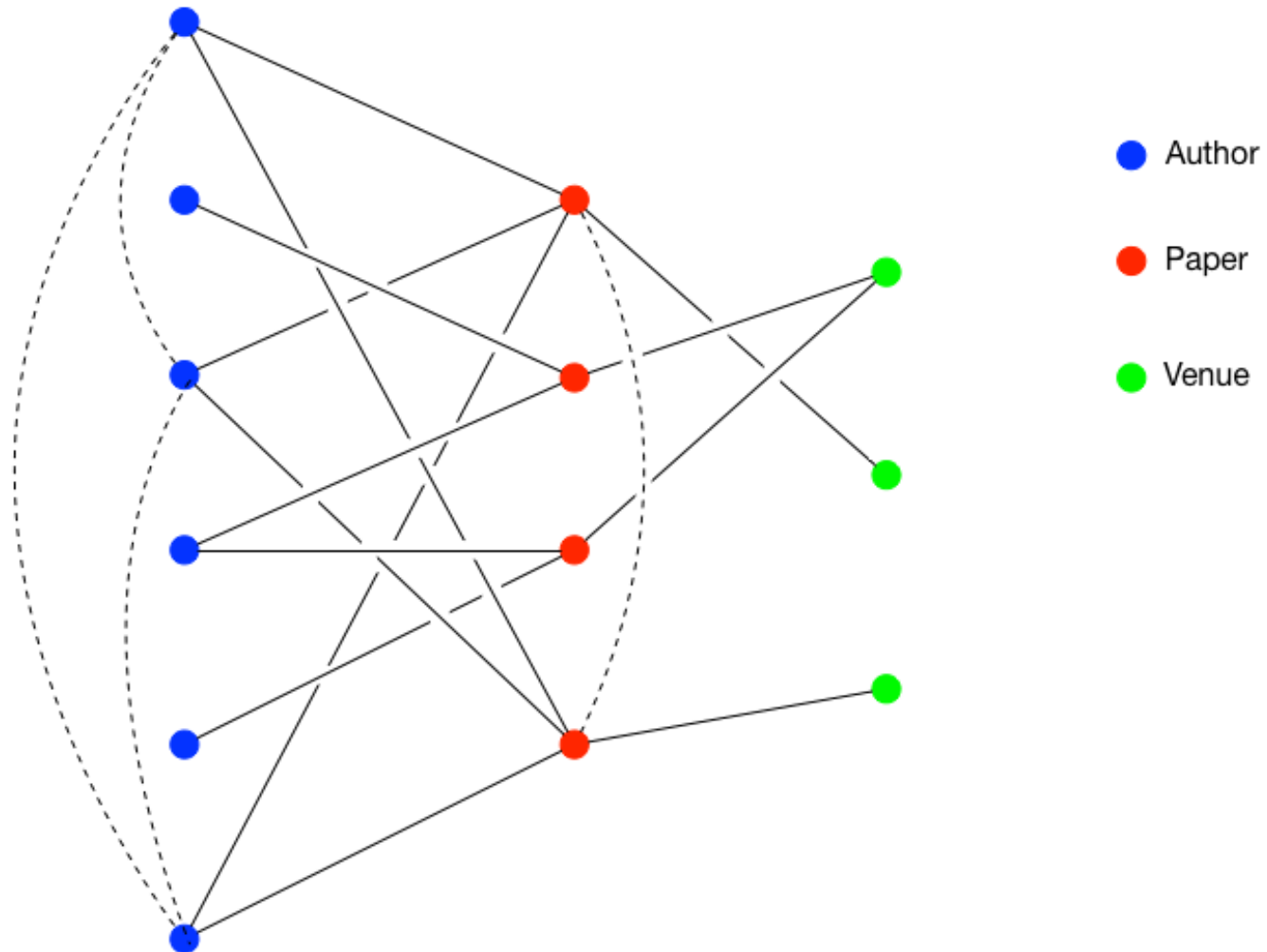
Similar Authors



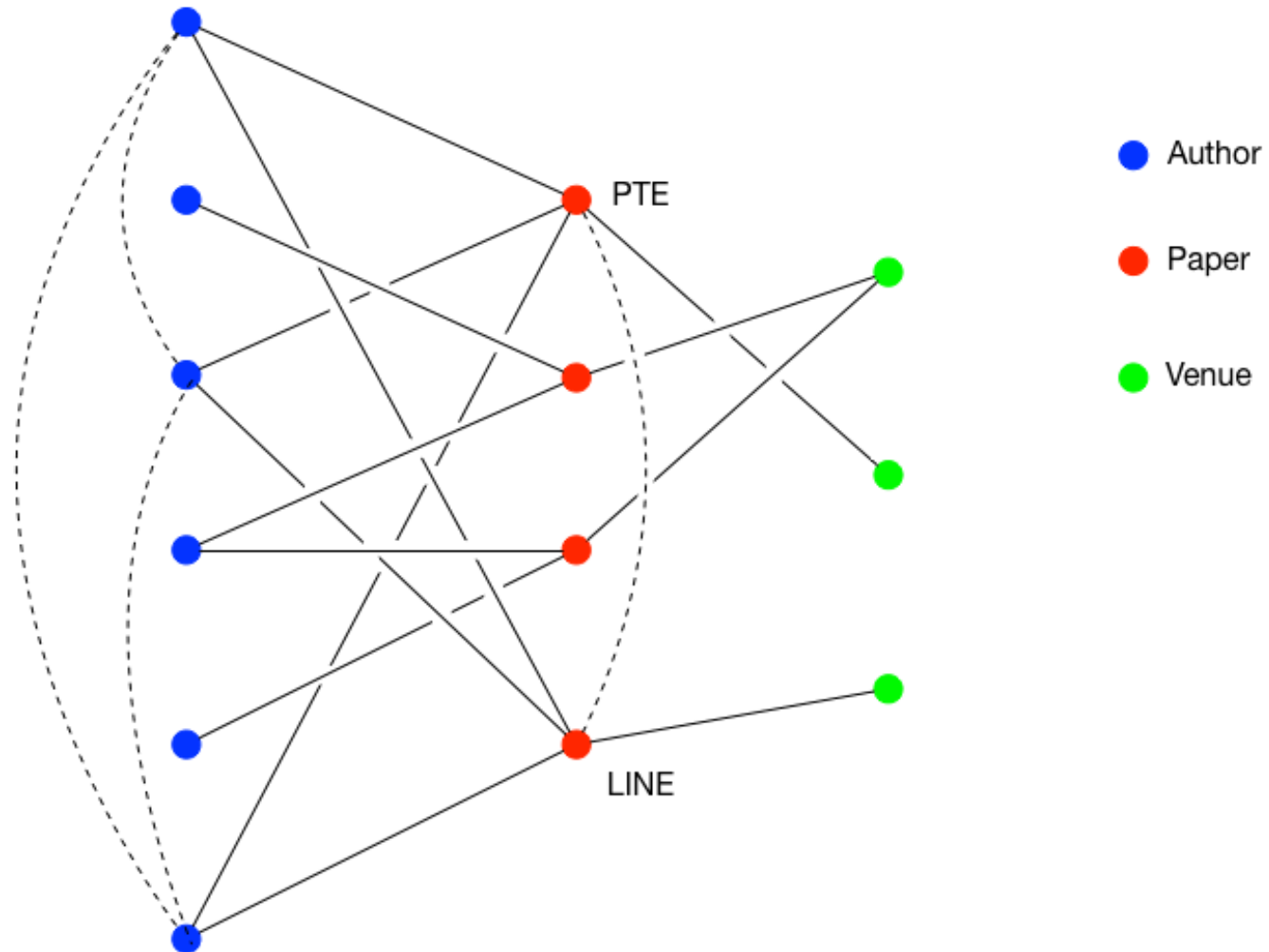
Similar Papers



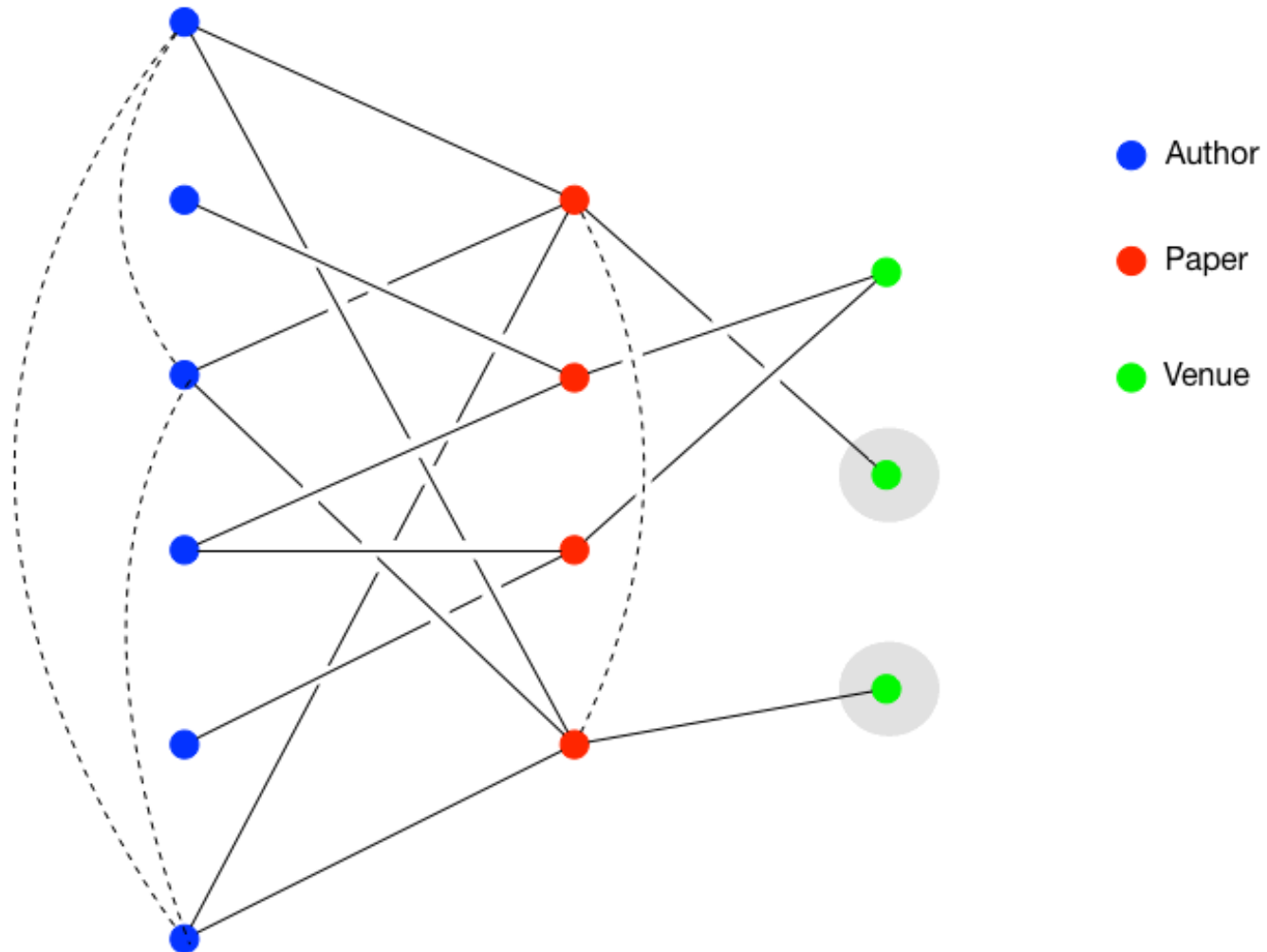
Similar Papers



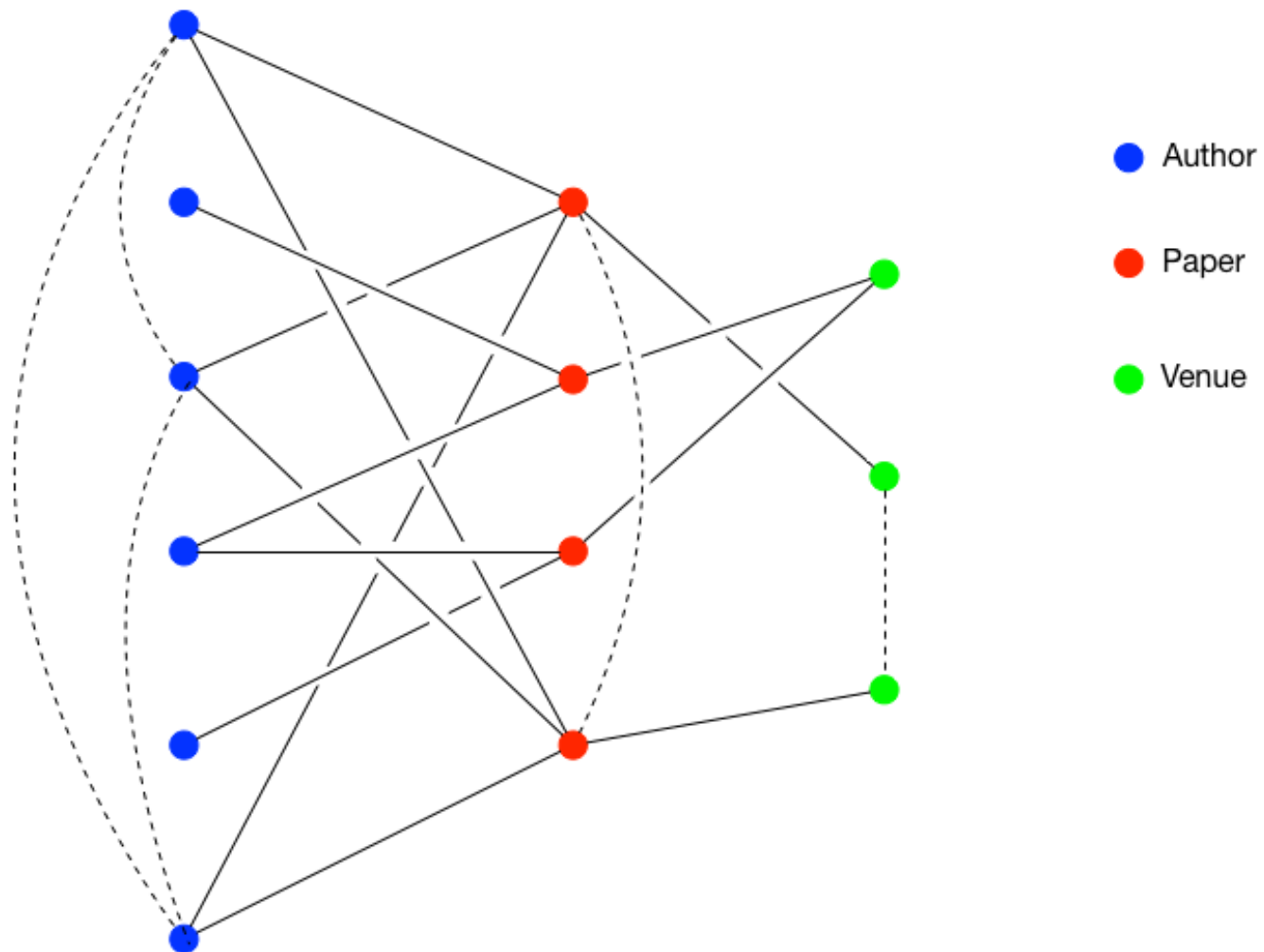
Similar Papers



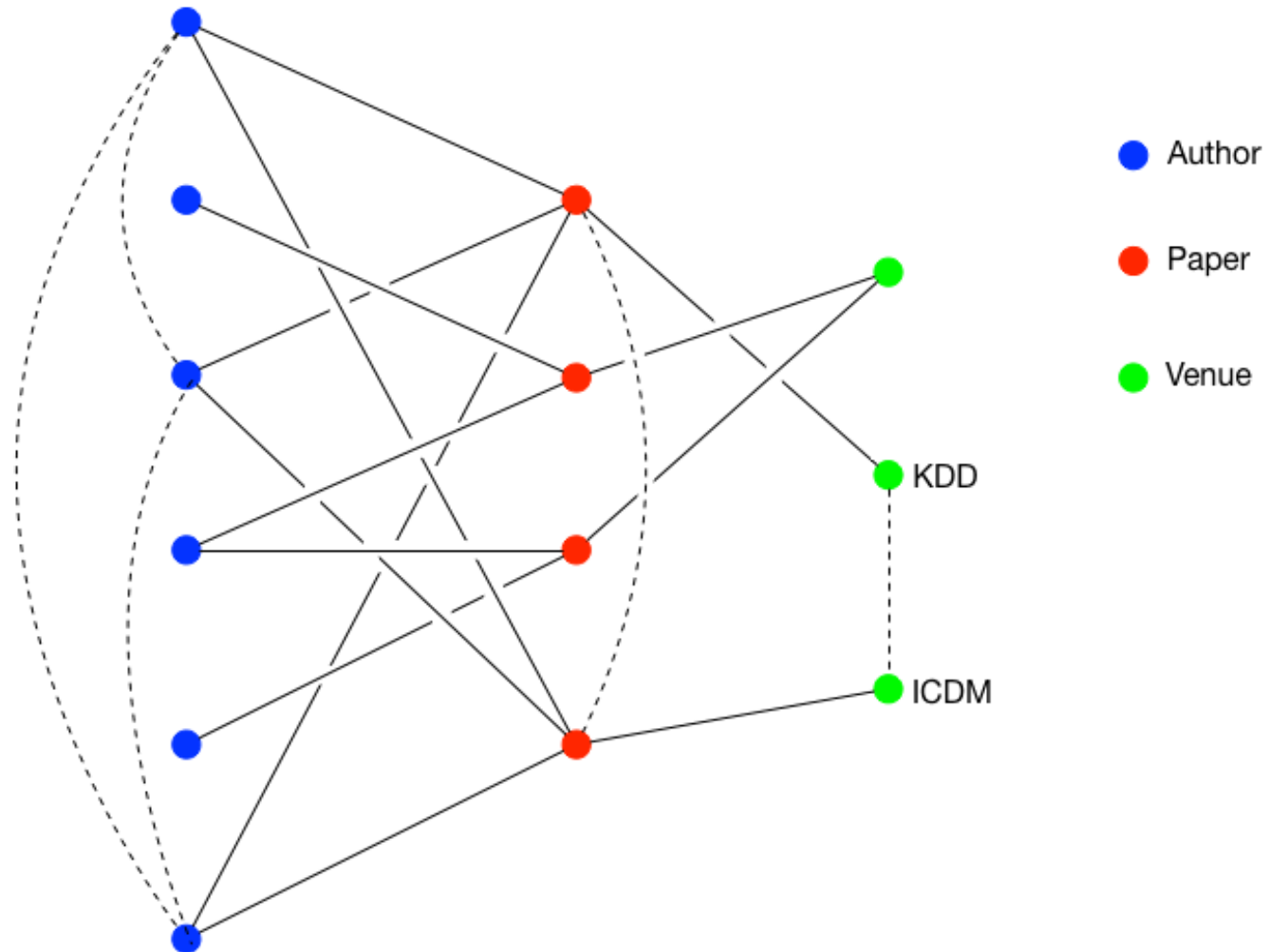
Similar Venues



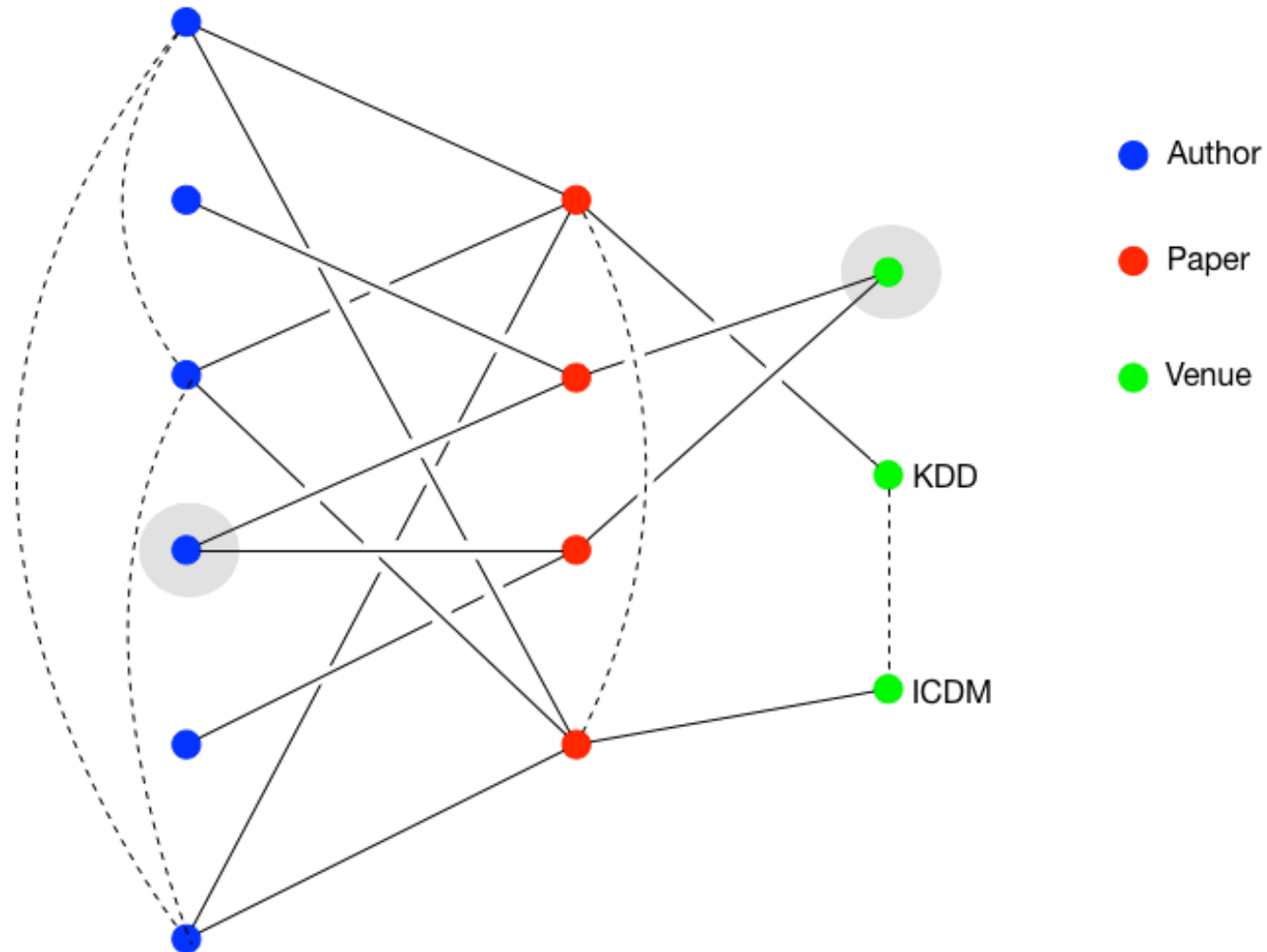
Similar Venues



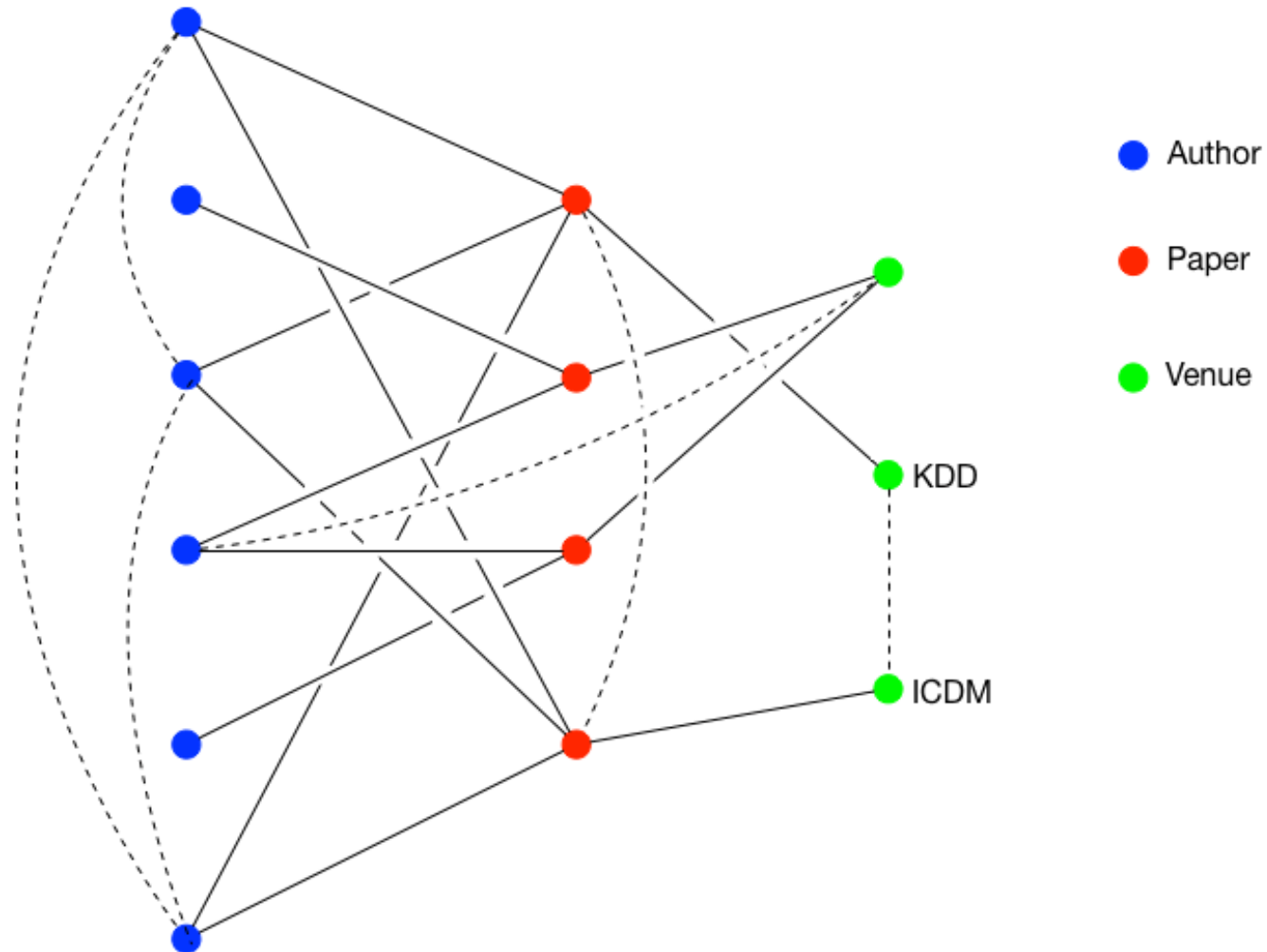
Similar Venues



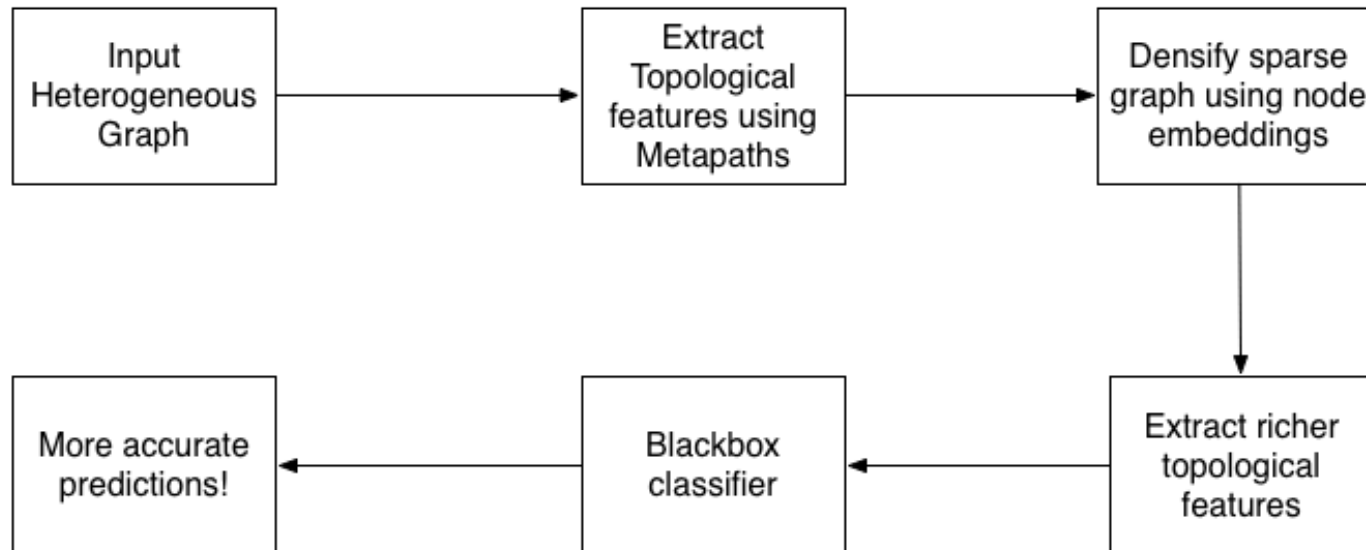
Heterogeneous Similarity



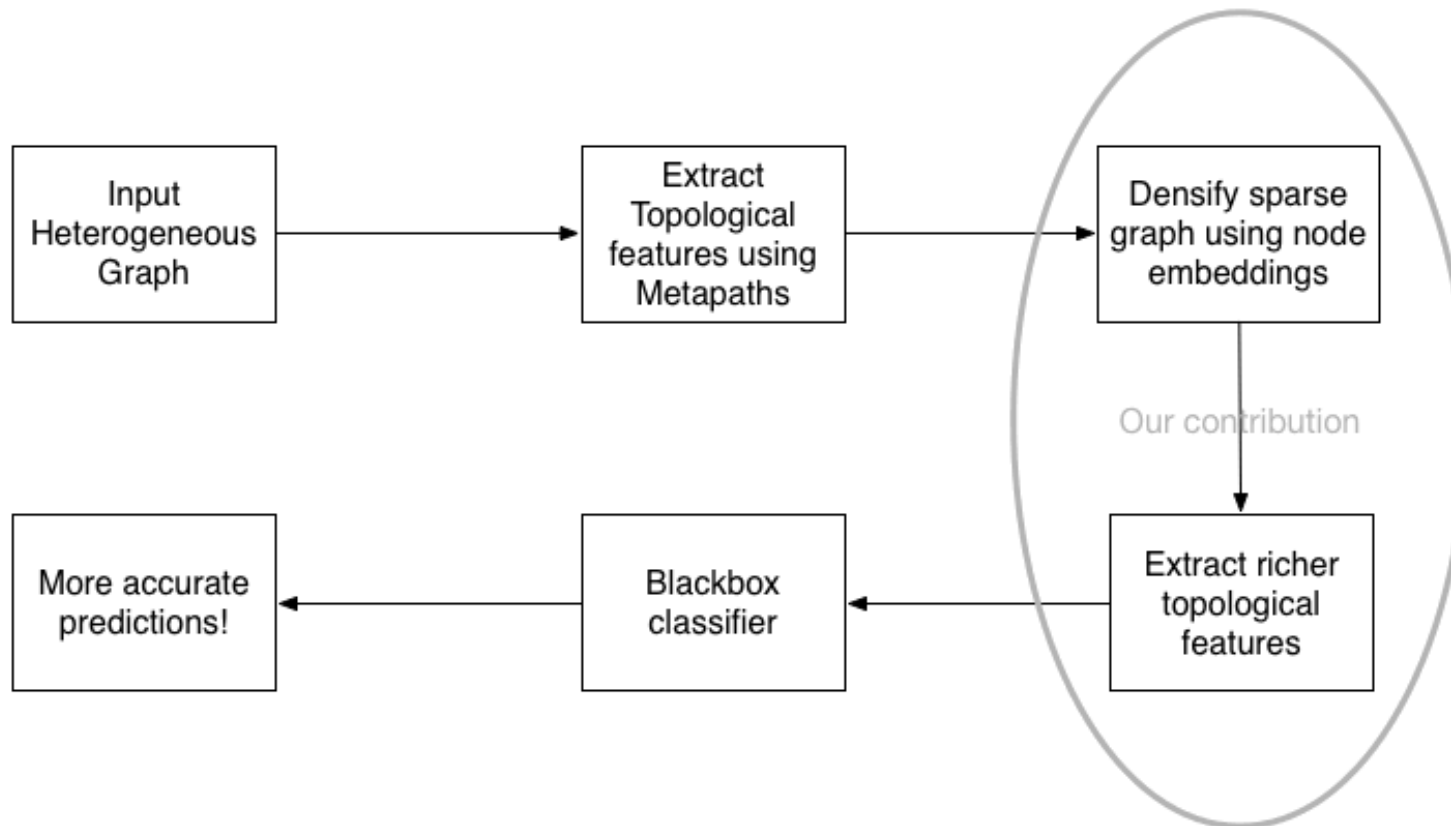
Densification of the original network



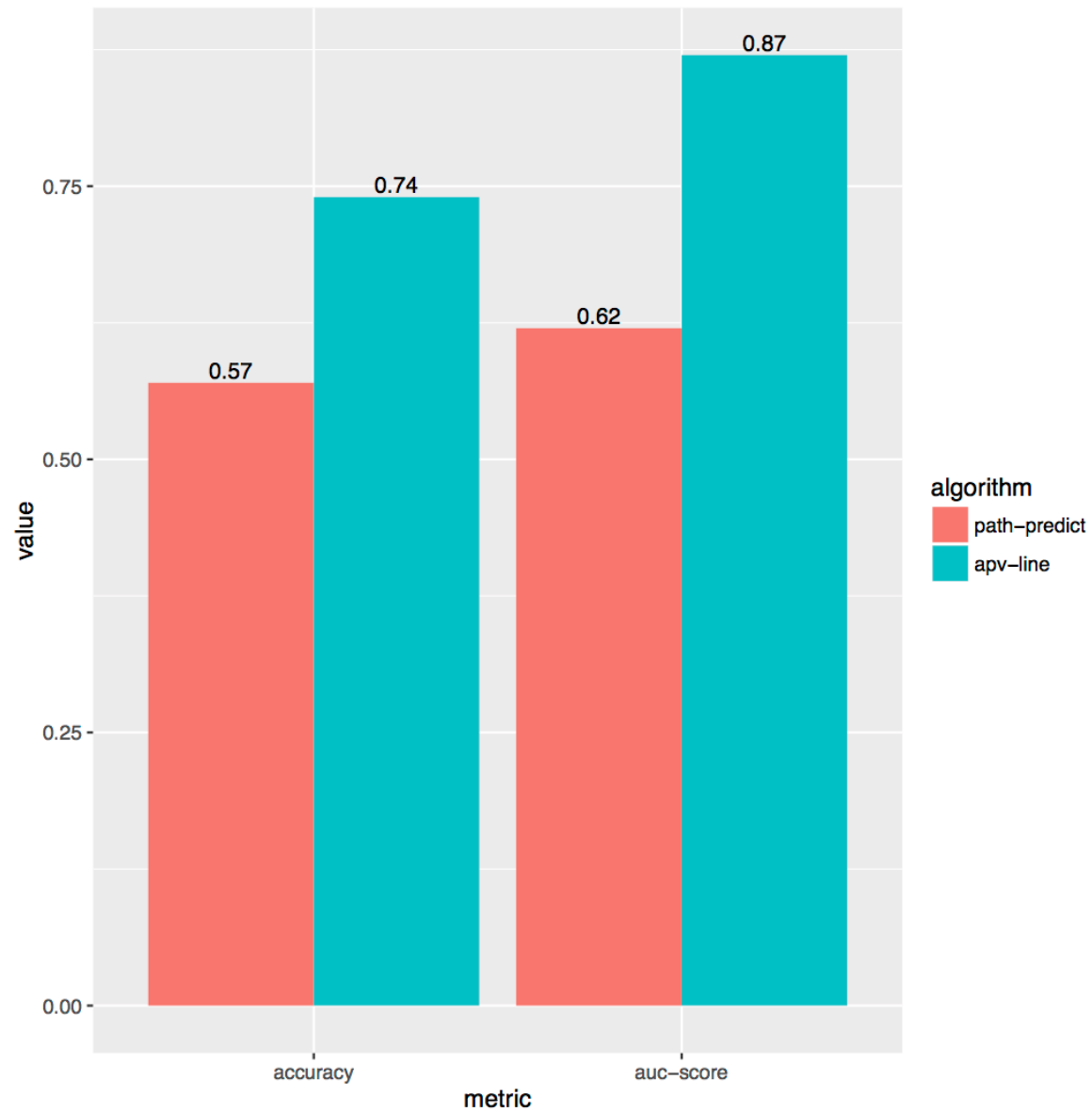
Workflow



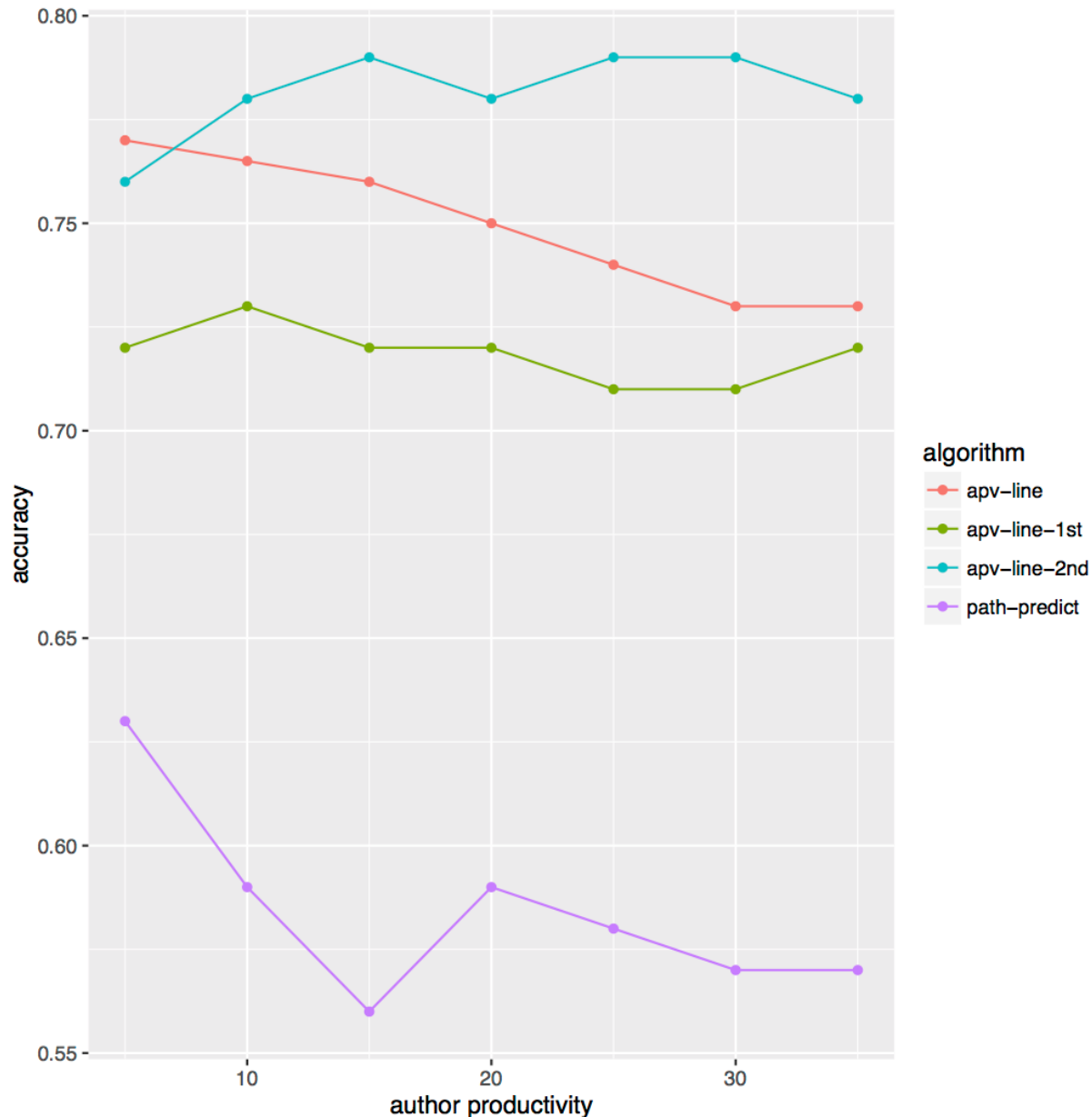
Workflow



Experiments



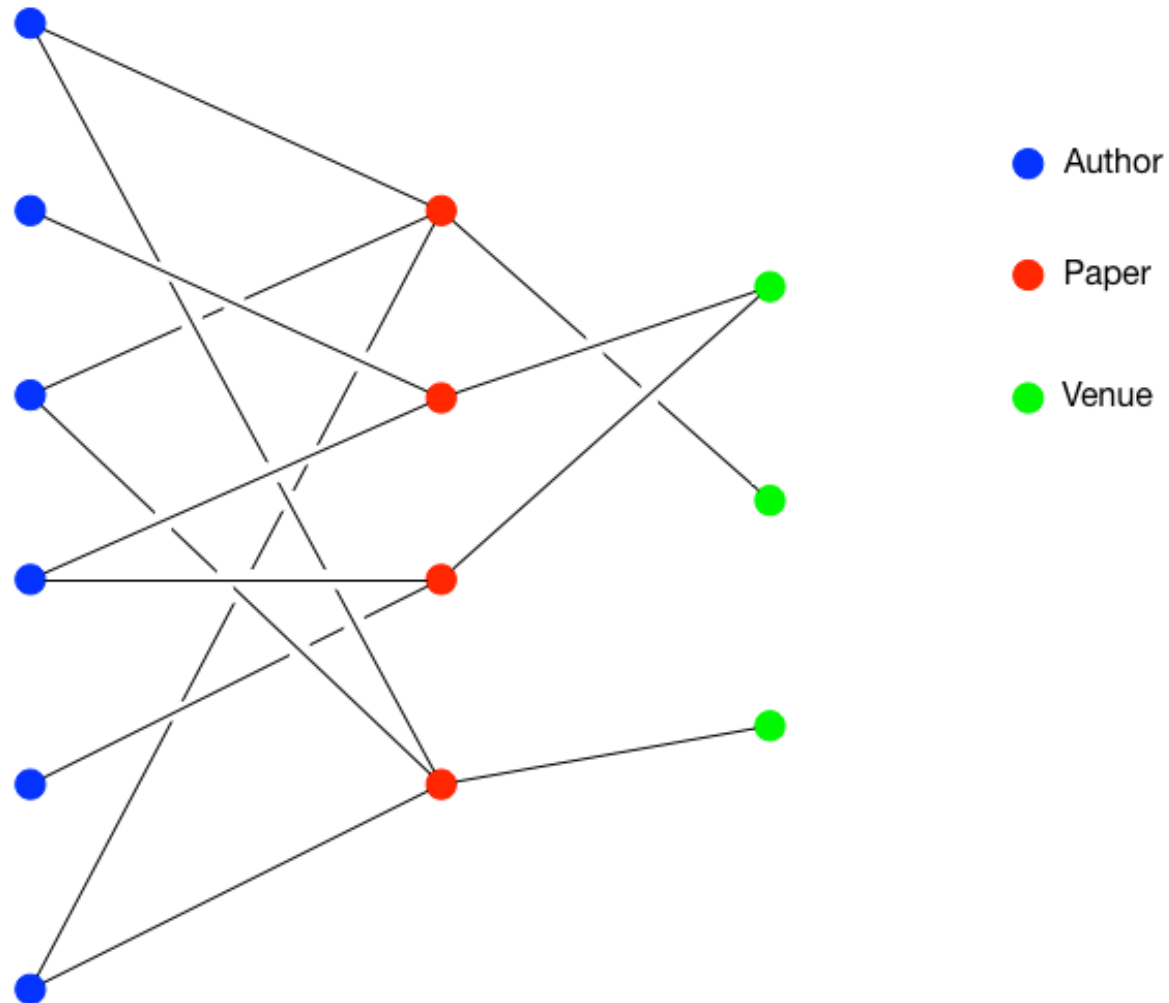
Experiments: Author Productivity



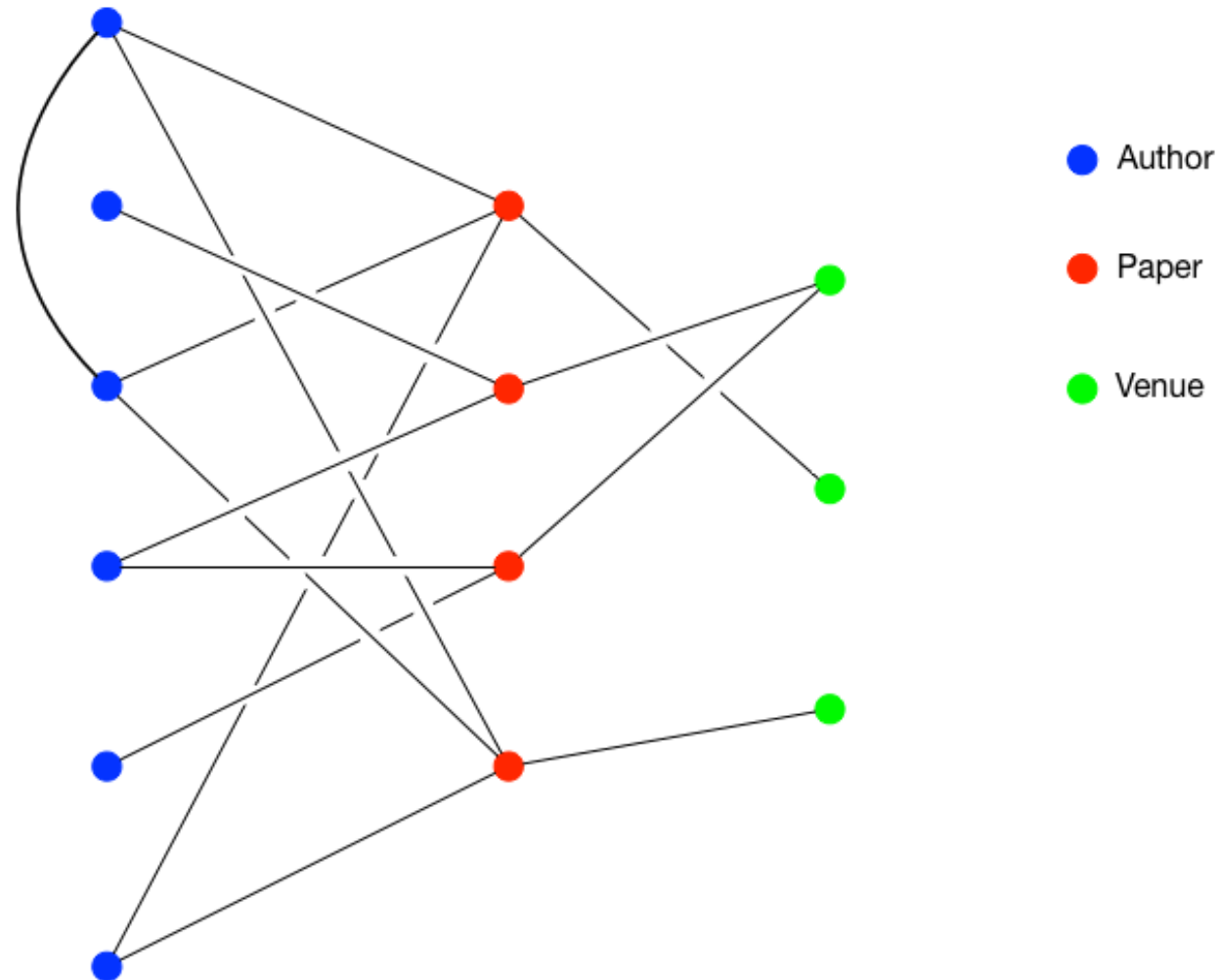
Insights:

- 1st order proximity not as good as 2nd order proximity
- This is possibly due to the lack of AV and AA edges in the original network
- Can we circumvent this, and reduce sparsity even further?

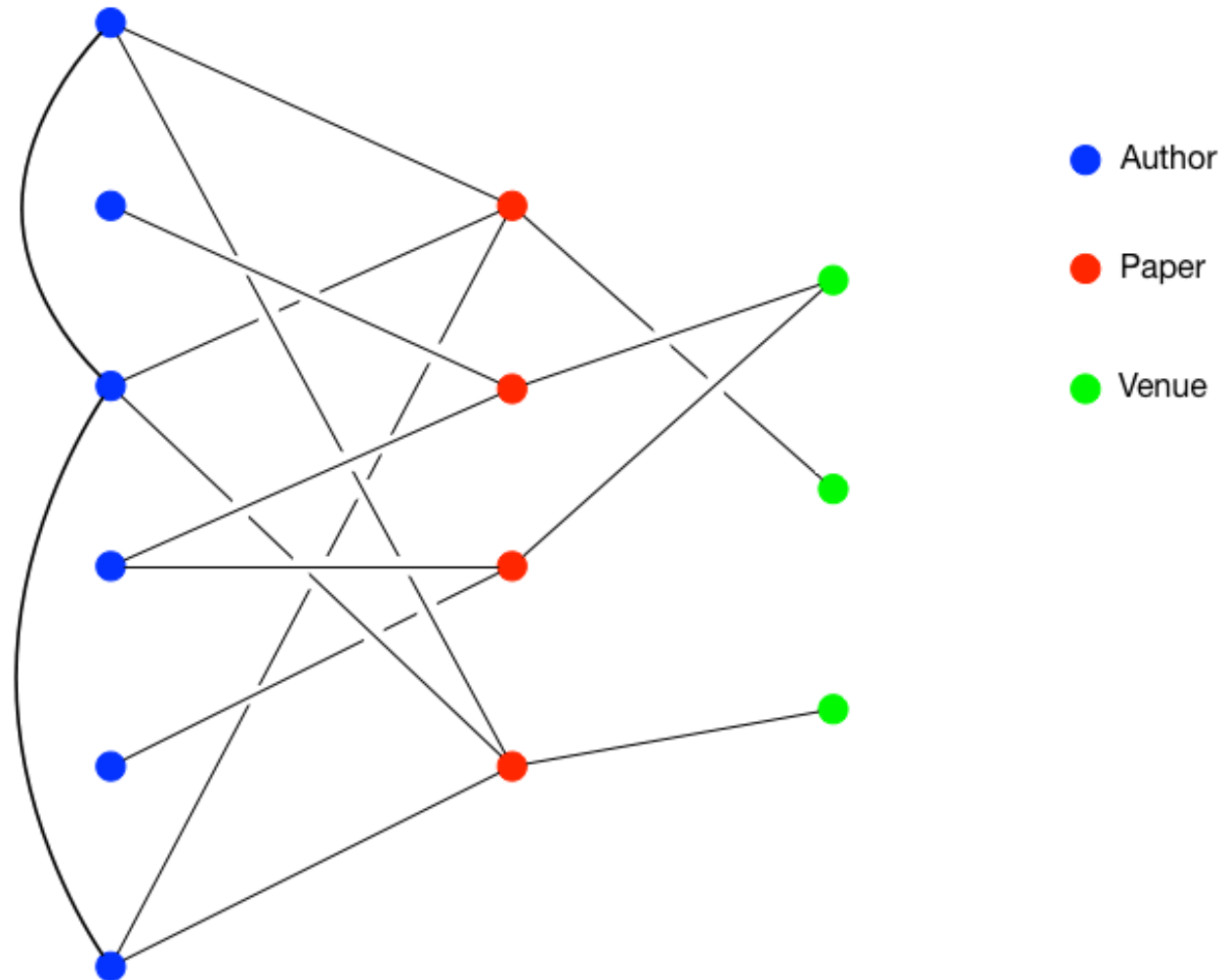
Enriching the APV network



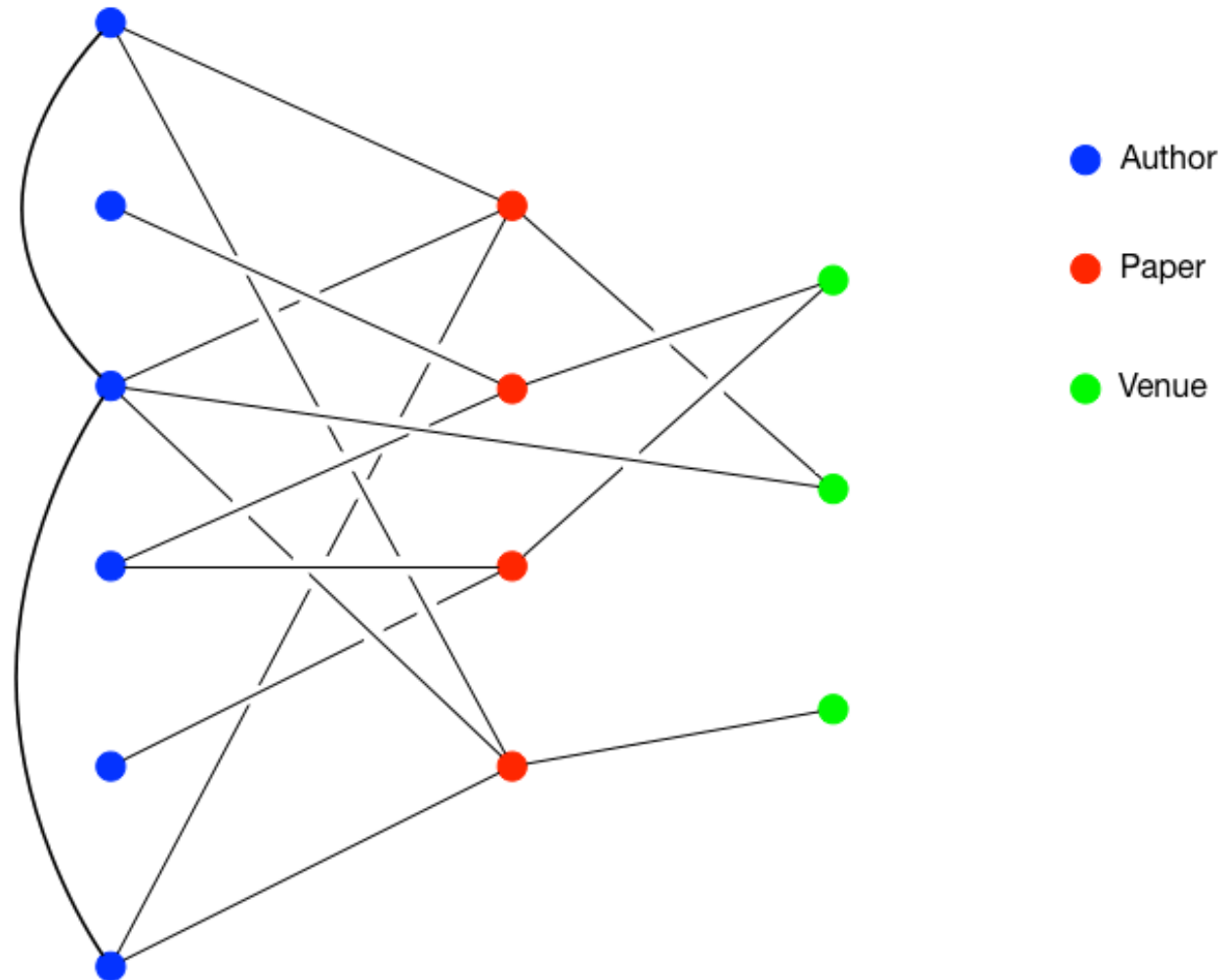
Enriching the APV network



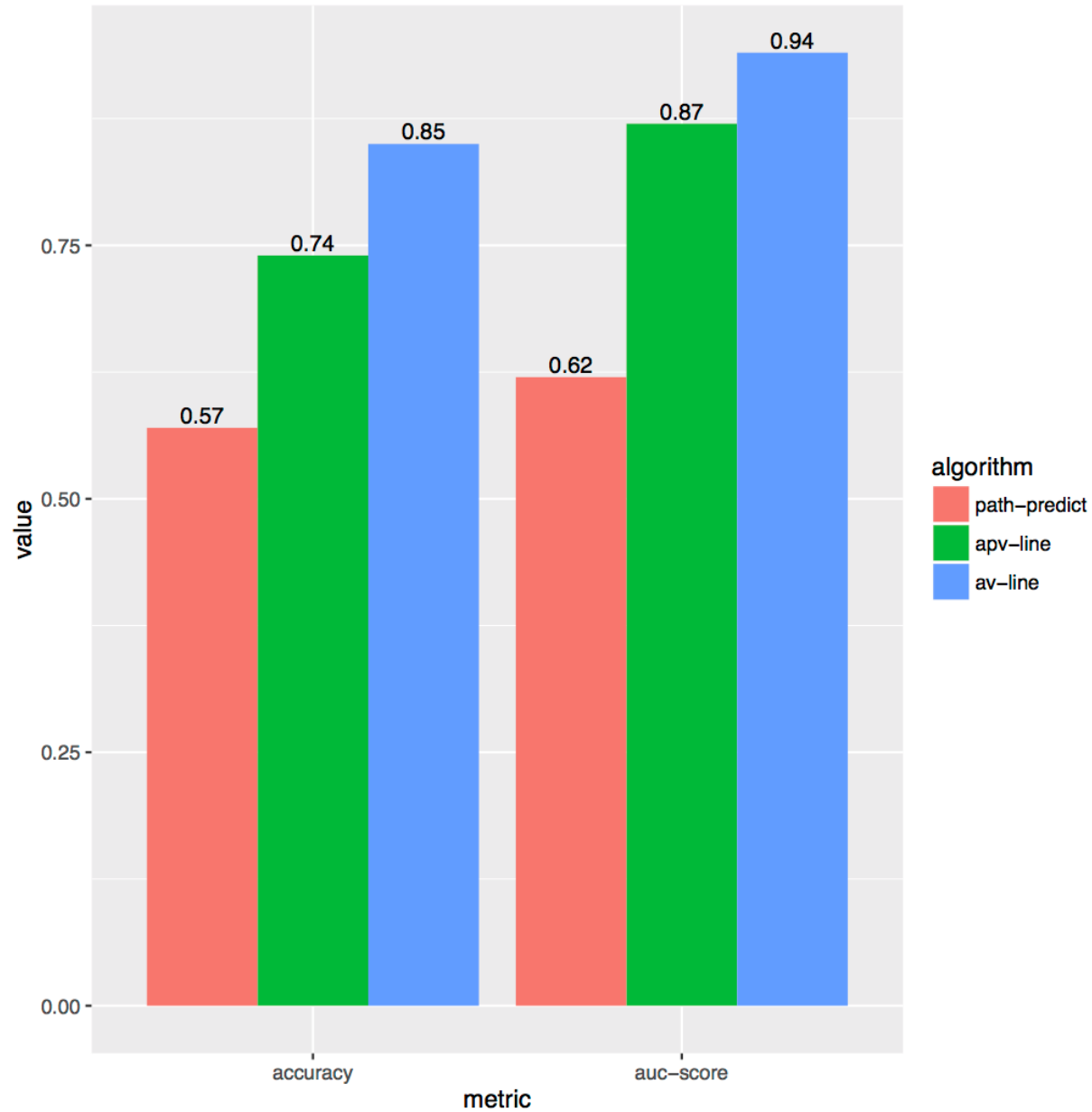
Enriching the APV network



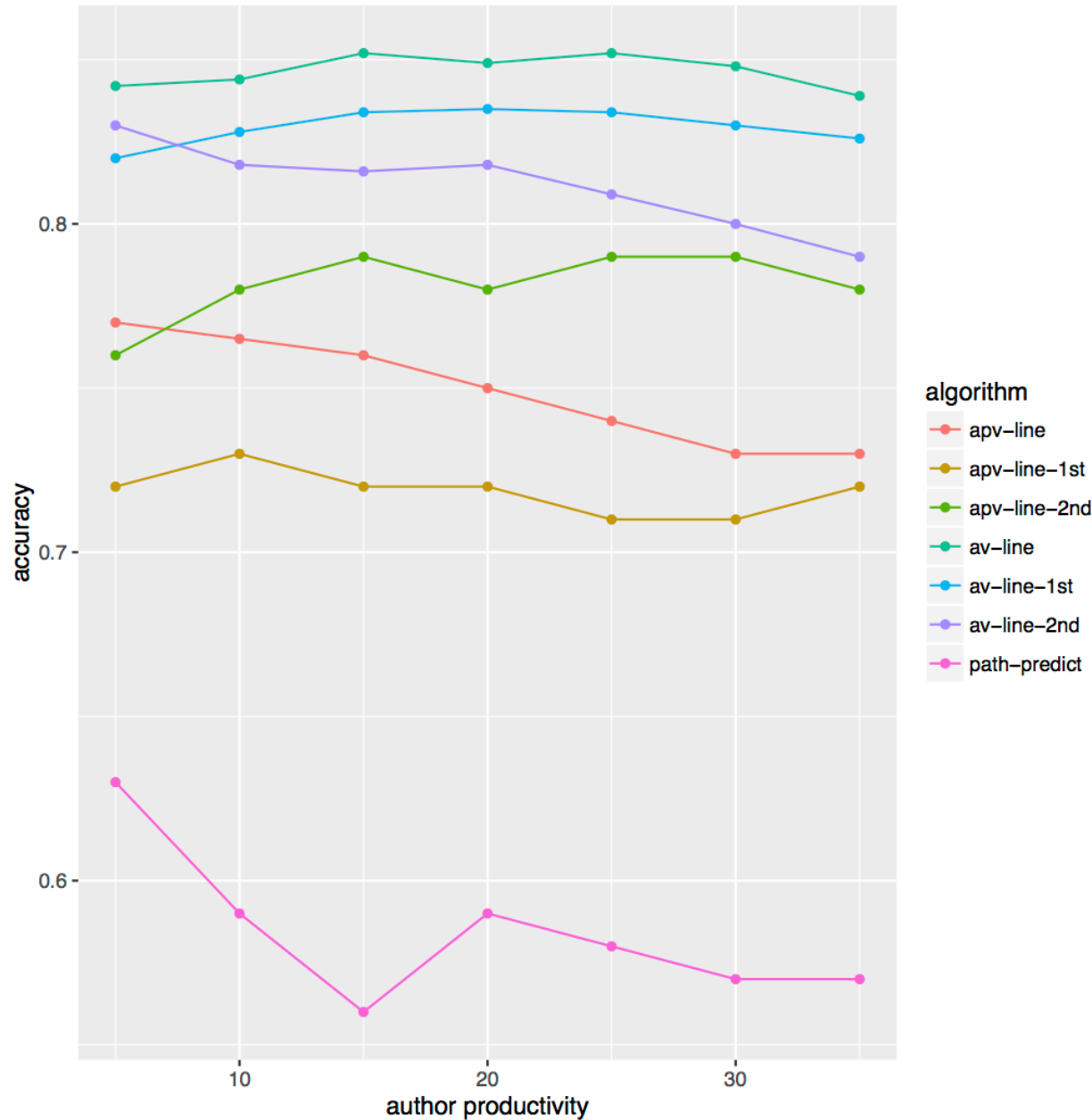
Enriching the APV network



Experiments: Head-to-head



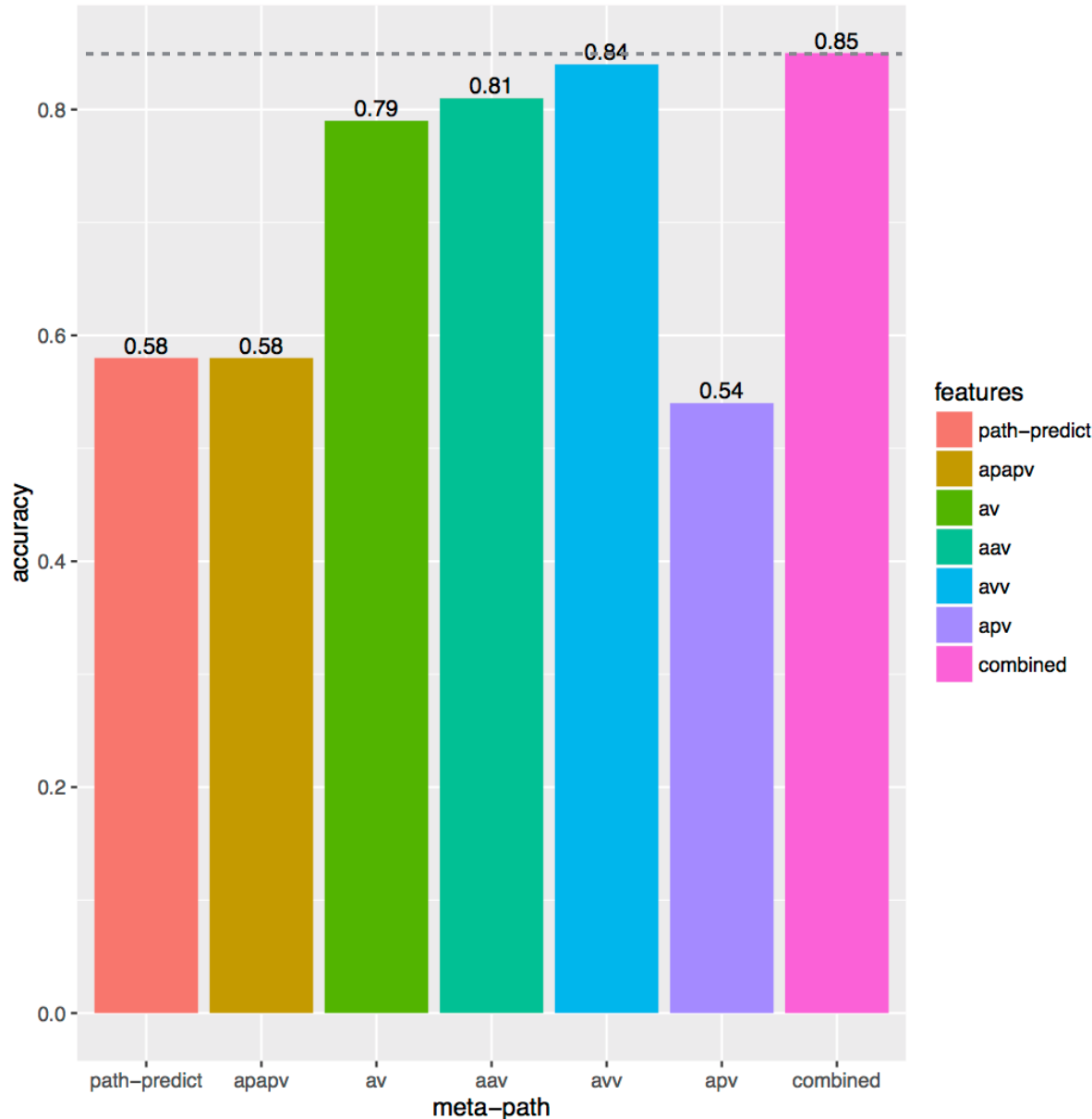
Experiments: Author Productivity



Insights:

- Introduction of AA and AV edges in the heterogeneous network improved accuracy even further!
- Gains from both 1st-order and 2nd-order embedding

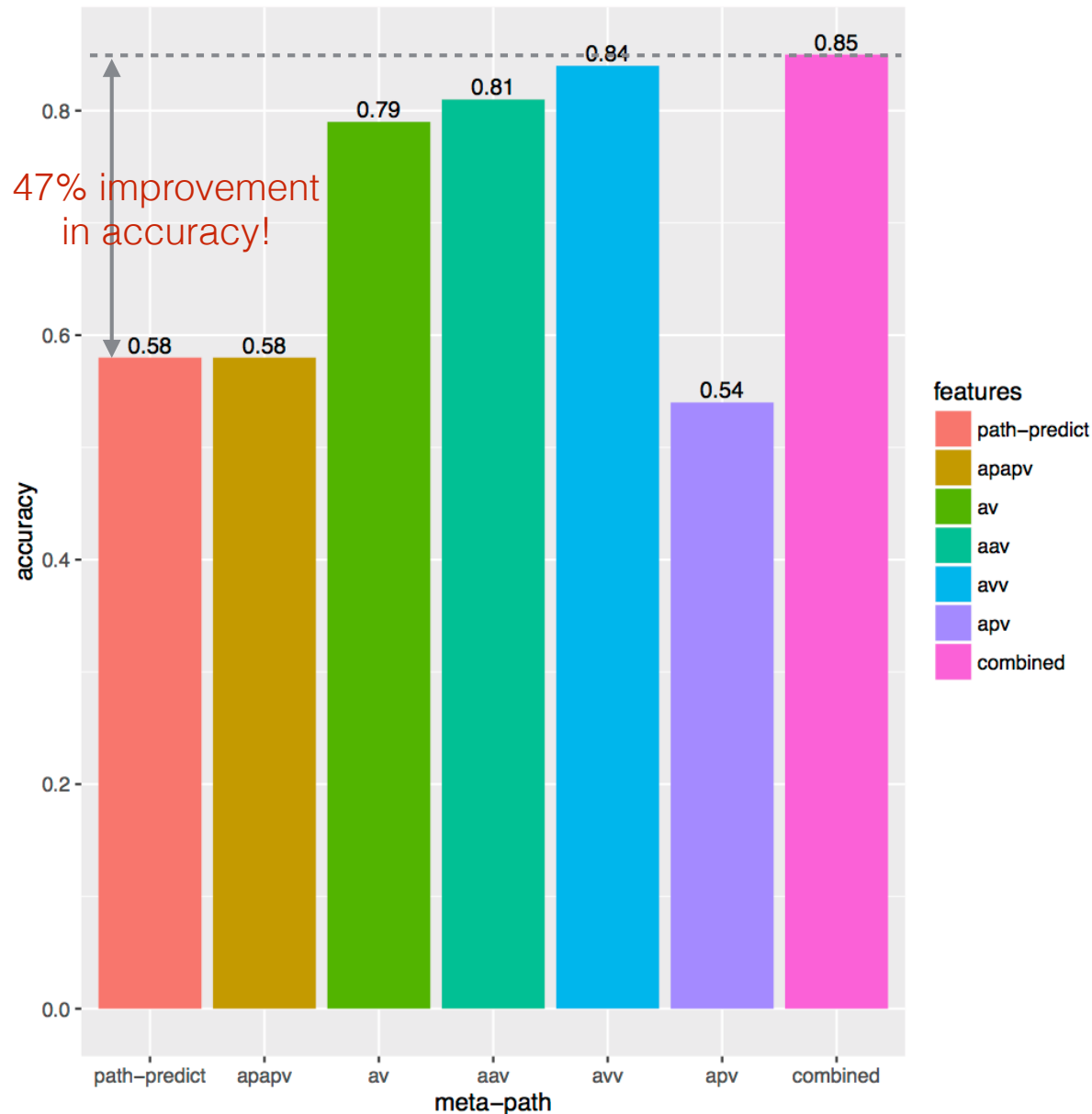
Experiments: Significant Features



Insights:

- The following metapaths singlehandedly outperform path-predict: AV, AAV, AVV
- These metapaths are not present in the original network
- Illustrates why embedding is useful

Experiments: Significant Features



Insights:

- The following metapaths singlehandedly outperform path-predict: AV, AAV, AVV
- These metapaths are not present in the original network
- Illustrates why embedding is useful

Future Work

- Use PTE instead of LINE
- Test performance with respect to change in network sparsity
- Extend method to other relationship prediction tasks:
 - Future co-authorship (a-p-a)
 - Future co-attendance (a-p-v-p-a)
- Testing model generalization
 - Different 5/10 years vs 5 years

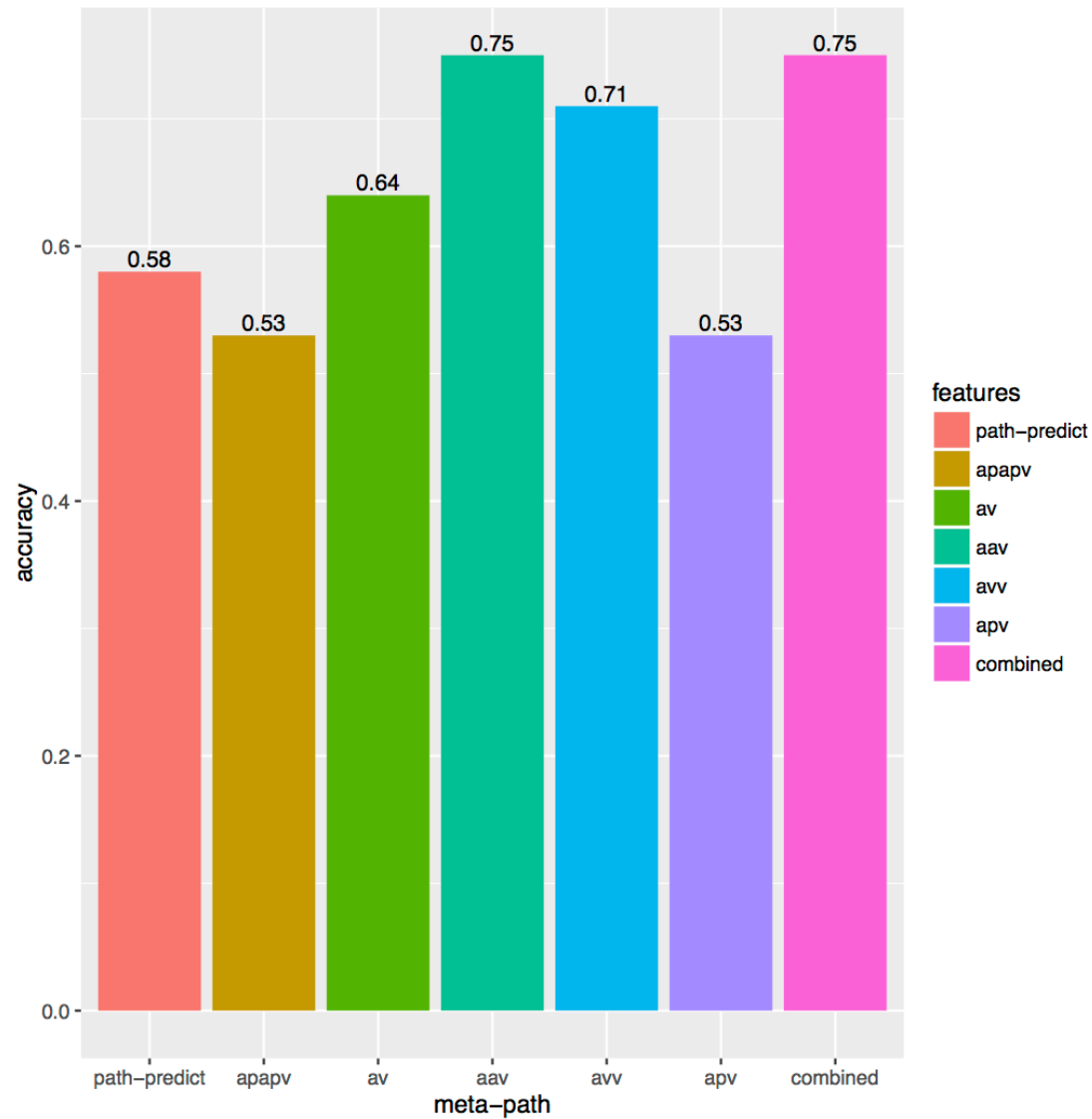
Conclusion

- Combating network sparsity with node similarities obtained from embedding techniques
- Enriching the APV network with links to capture high first order proximity
- Significant features that capture latent information
- Achieved 47% increase in accuracy from PathPredict

Questions!

Supplementary Slides

Experiments: Significant Features



Experiments

