Provenance-*Only* Integration

Ashish Gehani Dawood Tariq

SRI



Integration Challenges

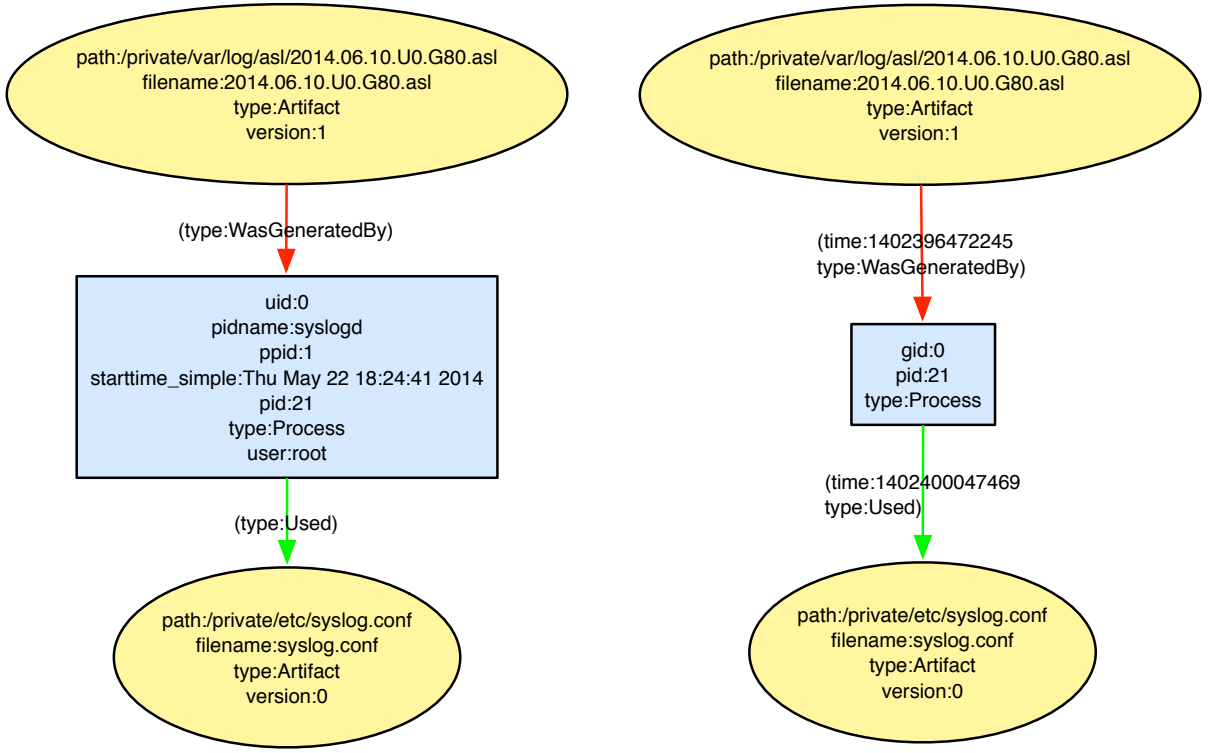
- Metadata variation:
  - Abstraction levels
  - Completeness
  - Identifiers
  - Semantics
- Querying requires:
  - Record assembly
  - Reconciling syntax
  - Mapping semantics

Related Work

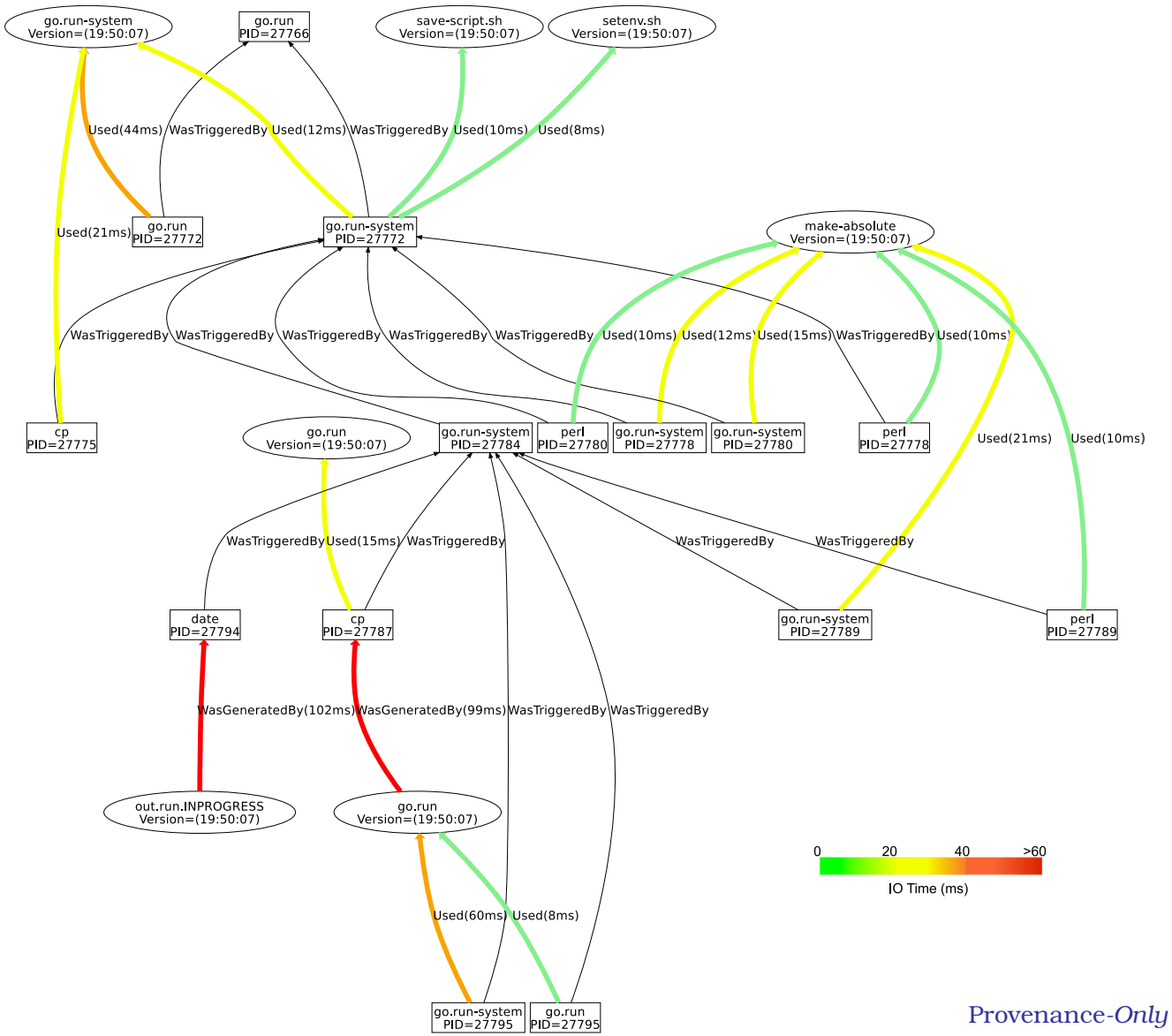
- Data integration
  - AnHai Doan, Alon Halevy, and Zachary Ives, Principles of data integration, Elsevier, 2012.
- Provenance integration
  - Semantic web (Umuhoza 2012)
  - Grid computing (Zhao 2008)
  - System interoperability (Angelino 2011)
  - Cross-organization sharing (Allen 2011)

Provenance-*Only* Integration

- Single underlying activity
- Multiple views of it
- Partial overlap in metadata



Speech Processing Hot Spots



Basic Provenance-Only Integration

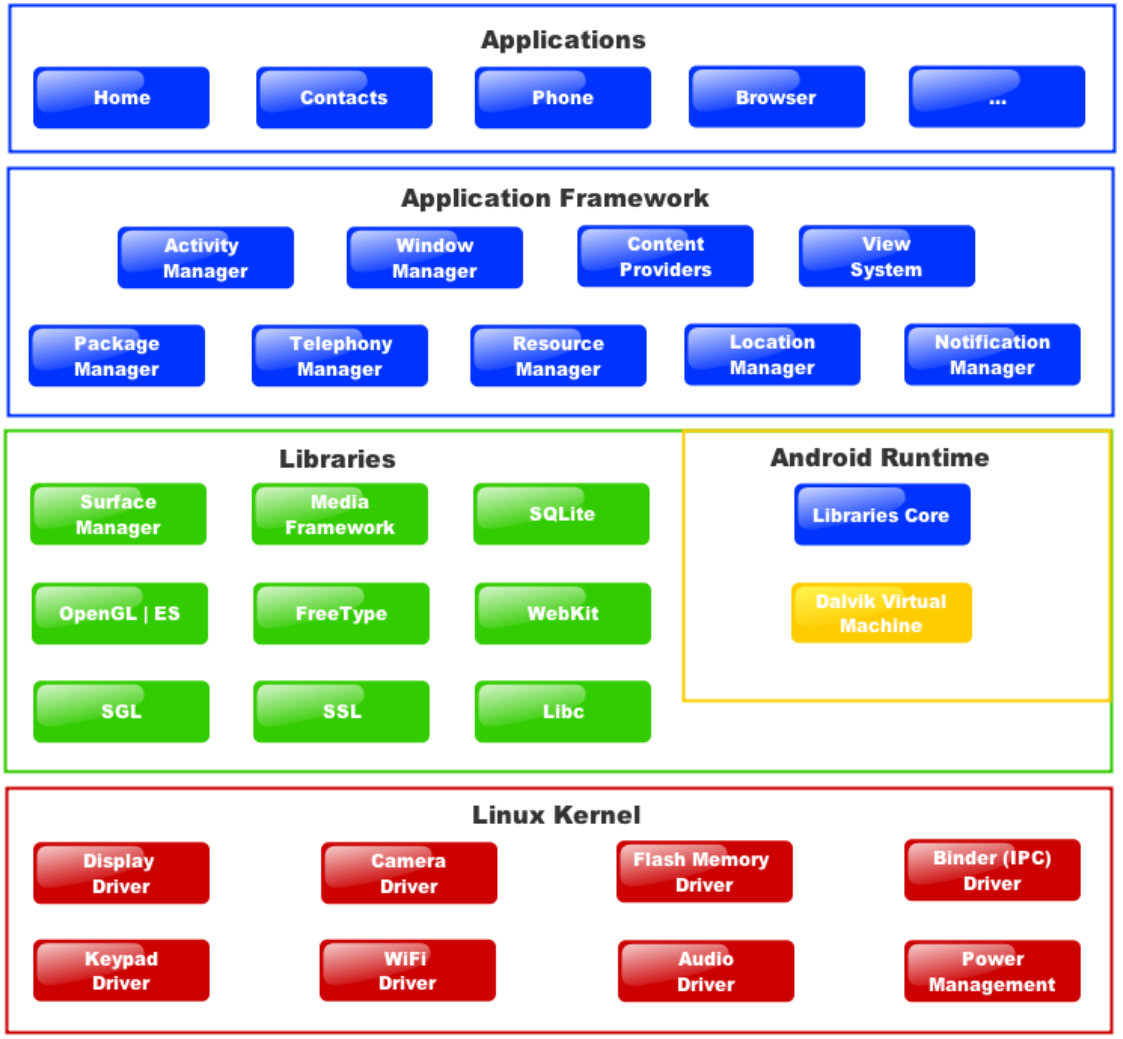
- Provenance from two vantage points
- Need to integrate the two
- Approach:
  - Define *matching threshold* τ
  - Merge vertex pair if τ-similar
  - Merge edge pair if τ-similar
  - Cost from conflating owners
- Goal:
  - Minimize τ
  - Keep cost $<$ tolerance Υ

Android Provenance

- Security analysis
  - System-wide monitoring
- Resource-constrained
  - Disrupts power management
  - Blinded by garbage collection
- *Multiple abstraction levels*
  - *Kernel interface*
  - *Inter-application (Binder)*
  - *Provenance-only integration*



Android

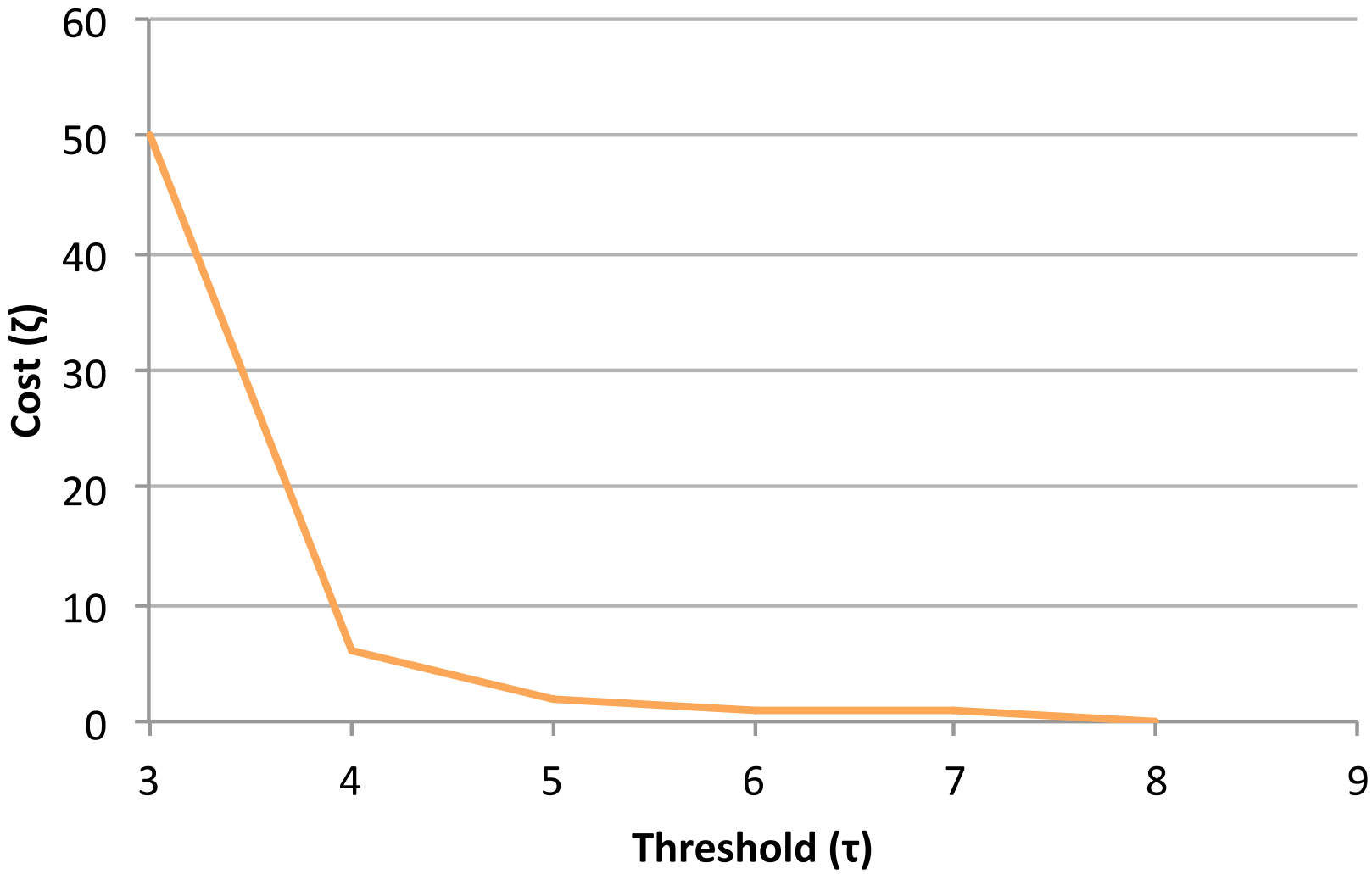


by Alvaro Fuentes Vasquez via Wikimedia Commons (CC-BY-SA-3.0-2.5-2.0-1.0)

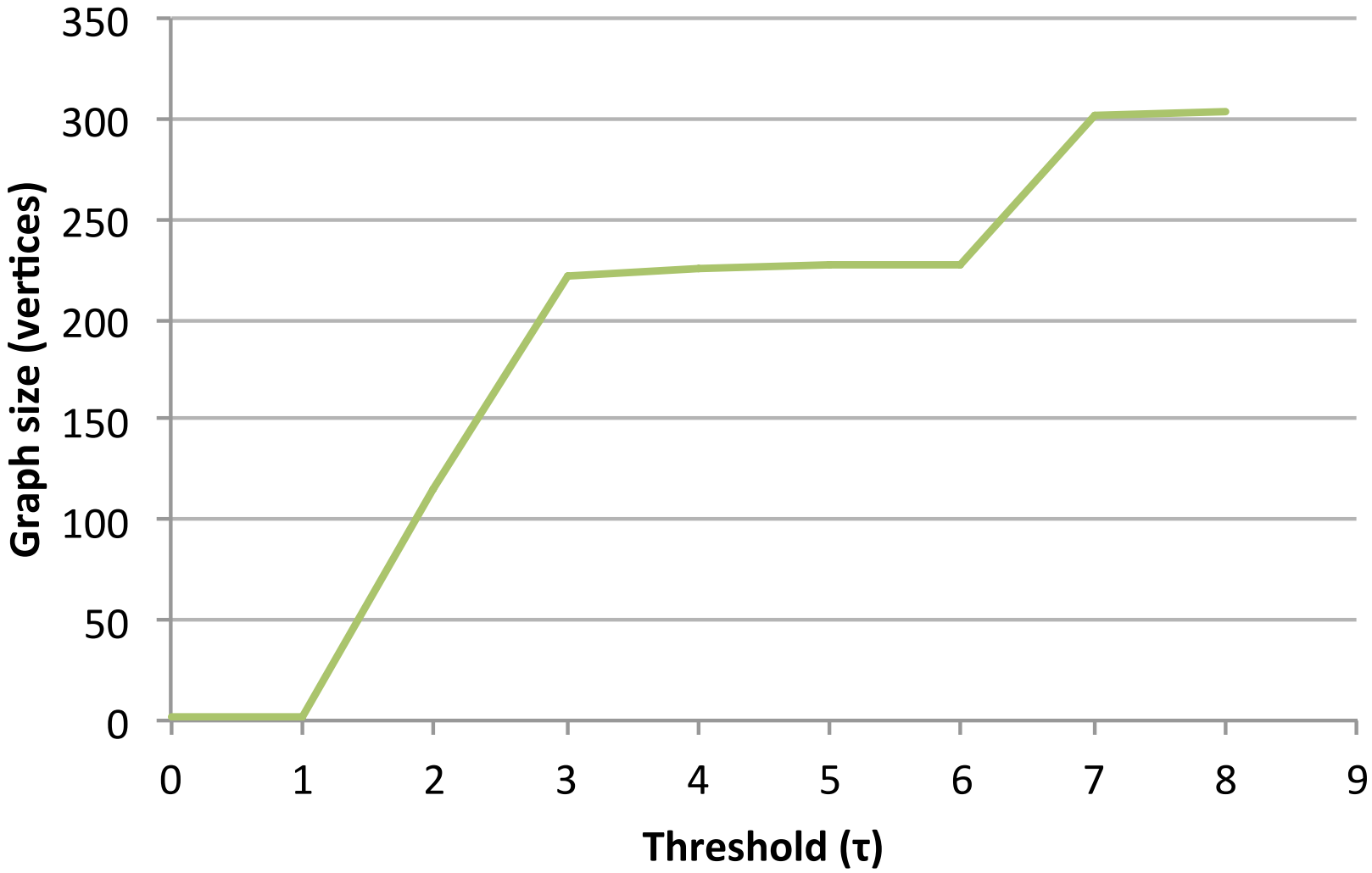
Fast Integration

- Integrate *all* τ-similar elements
- Don't have to find matching pairs
  - Avoids subgraph isomorphism problem
- Separate vertex, edge matching thresholds
- Thresholds are input now
- Cost is per match now
- Approach:
  - Merge τ_v-similar vertices, if cost < Υ
  - Merge τ_e-similar edges, if cost < Υ

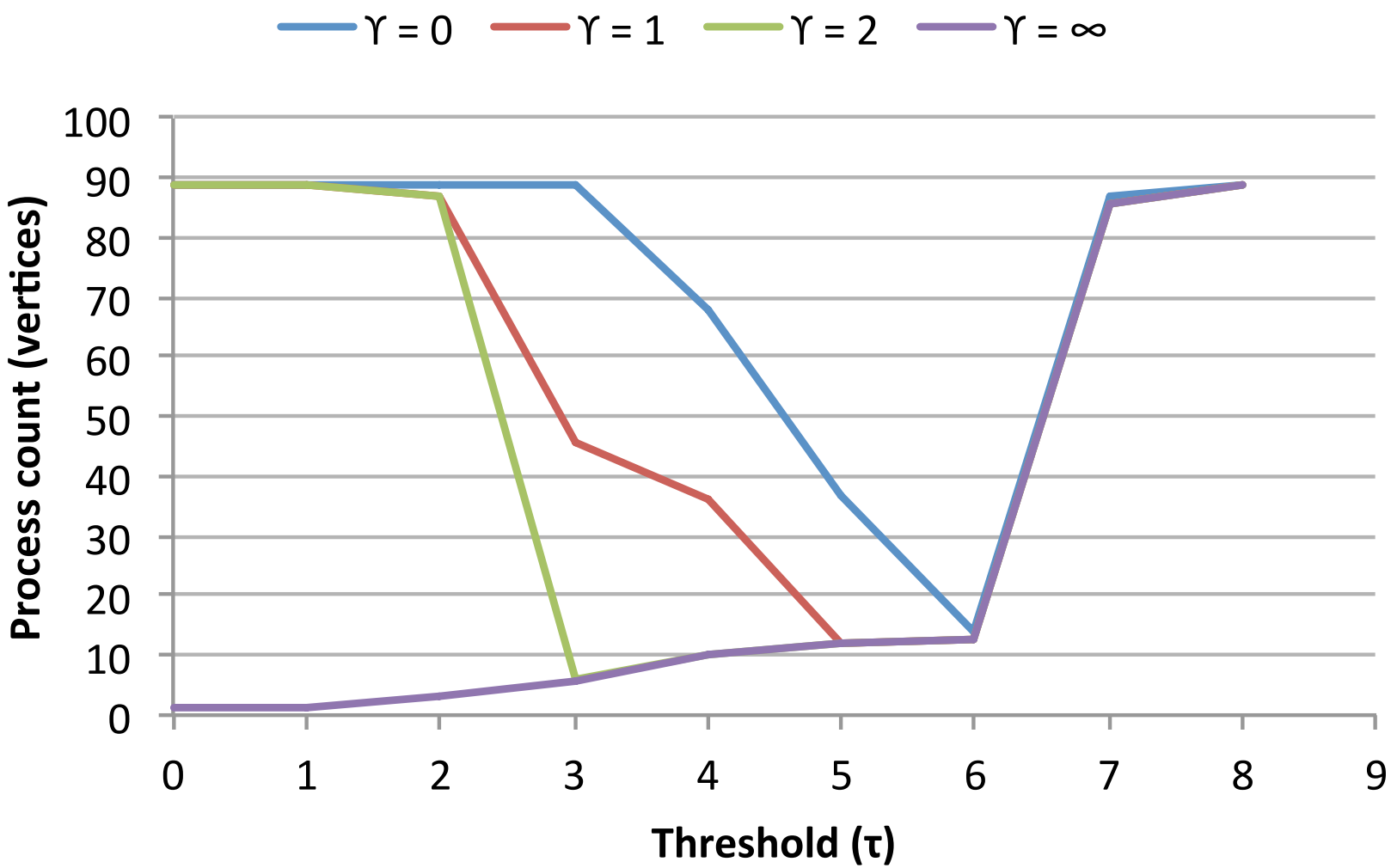
False Integration → High Cost



Integration as Abstraction



Fidelity of Attribution



Conclusion

- Provenance-only integration
- Basic form as constrained optimization
- Fast version $\rightarrow$ automated abstraction

- Acknowledgement
  - TaPP '14 organizers, reviewers
  - US NSF Grant IIS-1116414

- URL: http://data-provenance.googlecode.com
- Email: ashish.gehani@sri.com

- Questions?