Toposemantic Network Clustering

Leon Poutievski,
Profs. Ken Calvert, James Griffioen
{leon, calvert, griff}@netlab.uky.edu

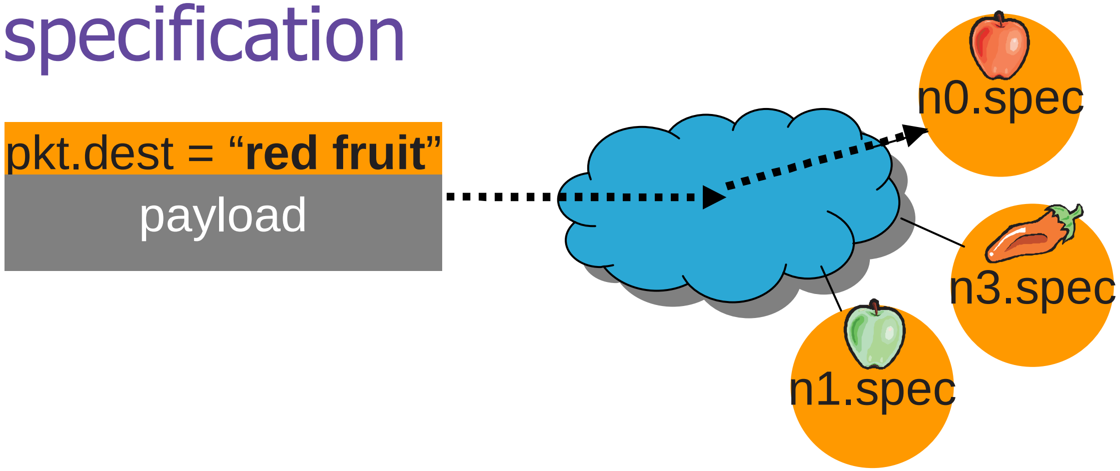
University of Kentucky
Laboratory for Advanced Networking

Thanks for NSF support

UK

Problem

- We explore new routing & forwarding architecture
  - Goal: scalable network-layer service with generalized addresses
- Given: graph, labels (specifications) on nodes
  - Node specifications might be topologically independent
- Network delivers a packet to all nodes that match the destination specification
  - E.g. Speccast

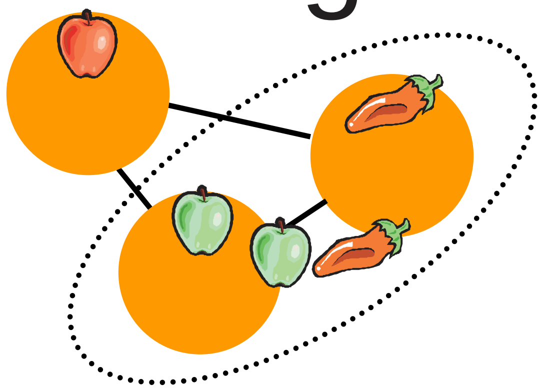


Naive Approach

- Modify traditional distance-vector or link-state protocol
  - Everyone must know about everyone
  - Conclusion: must reduce state

$\Rightarrow$ The only way is to use abstraction

Routing Information Hiding

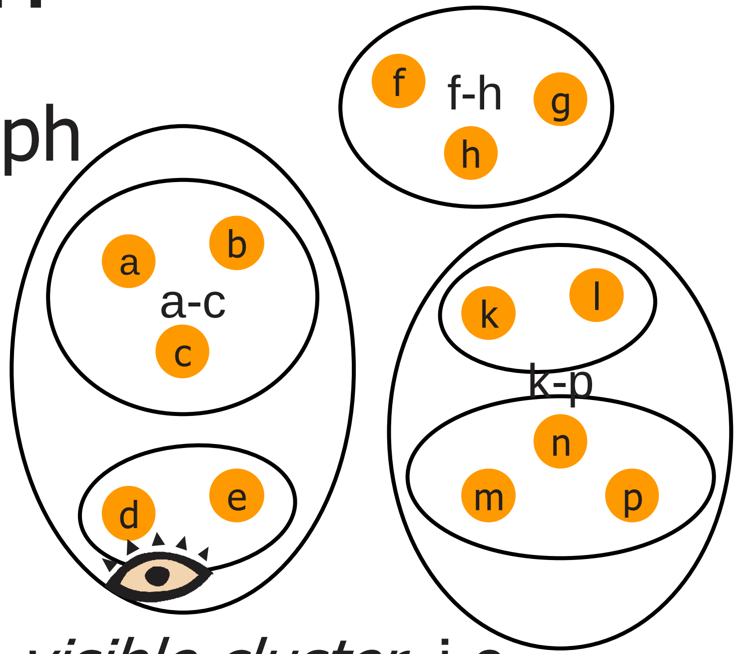


- Topological hiding
  - Abstract part of graph
  - Give up: delay
- Semantic hiding
  - Abstract destination descriptions
  - Give up: bandwidth

- Traditional approaches conflate topological and semantic hiding
  - E.g. IP prefixes assigned to AS's
- We want to explore relative importance of topological vs semantic information hiding

Approach

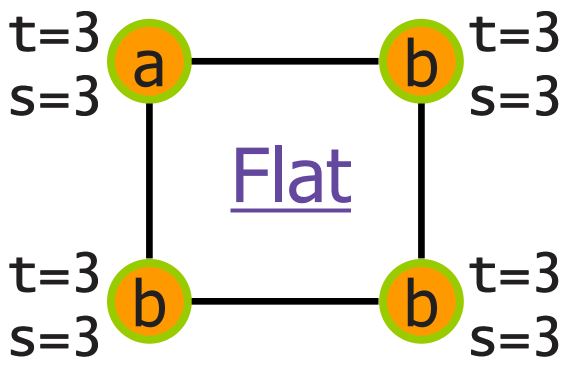
- **Clustering** – partitioning a graph into a hierarchy of connected subgraphs, assign a label to each created cluster



- Fish-eye view
  - One routing table entry (RTE) per *visible cluster*, i.e. sibling cluster or parent's sibling cluster
- Challenge: find clustering that minimizes

  state, delay, overdeliveries
- Problem is hard, usually greedy algorithms are used

Clustering

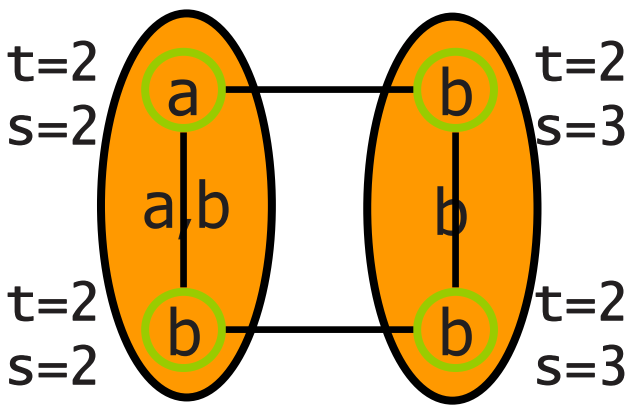


Total:
t=12, s=12

Topological Clustering

- Goal: minimize number of RTEs
- Solution: balanced hierarchy

t=8, s=10

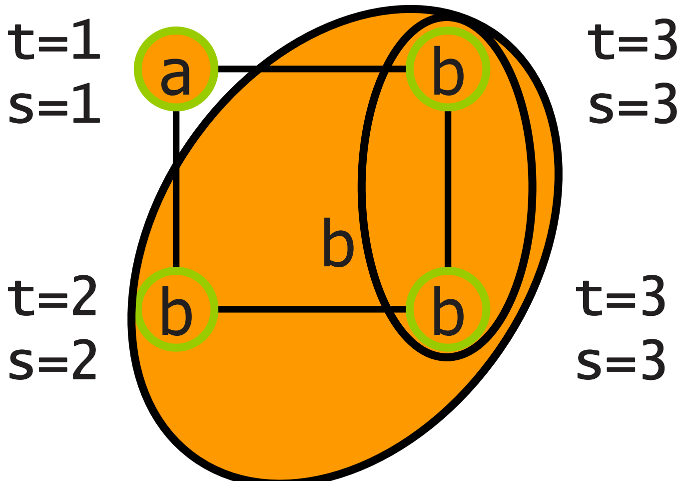


Semantic Clustering

- Data clustering: partition data into subsets, s.t. data in each subset is "similar"

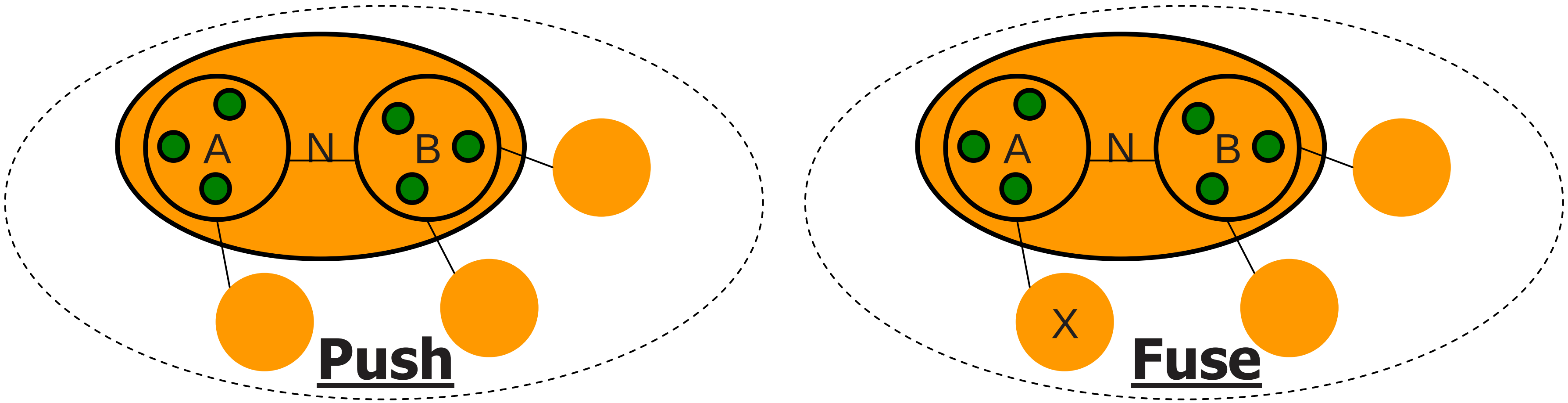
  with a constraint –topology

t=9, s=9



Toposemantic Clustering

- Idea: combine approaches to minimize state
- Two basic merging operations



- Estimating State Reduction: heuristic H(A, B)
- Knobs:

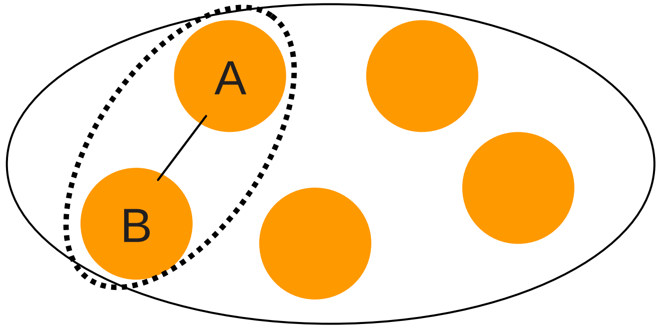
  σ – importance of semantic clustering

  τ – importance of topological clustering

 $\sigma = 0$ ($\tau \neq 0$) clustering based on topology

 $\tau = 0$ ($\sigma \neq 0$) clustering based on semantics

Algorithm

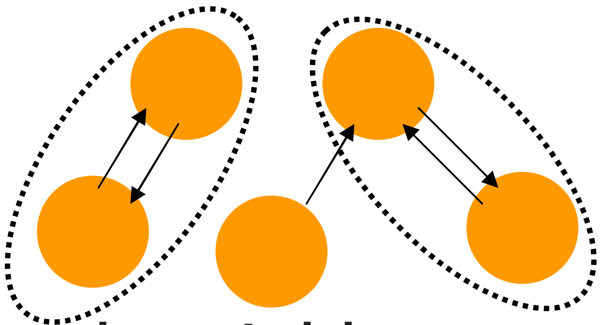


- Centralized
  1. Find a cluster with subclusters > threshold
  2. Evaluate H on every pair of neighboring clusters
  3. Pick a pair of clusters with max H(A,B)

     If (subclusters(A) + subclusters(B) ≤ threshold)

     then Fuse(A, B)

     else Push(A, B)

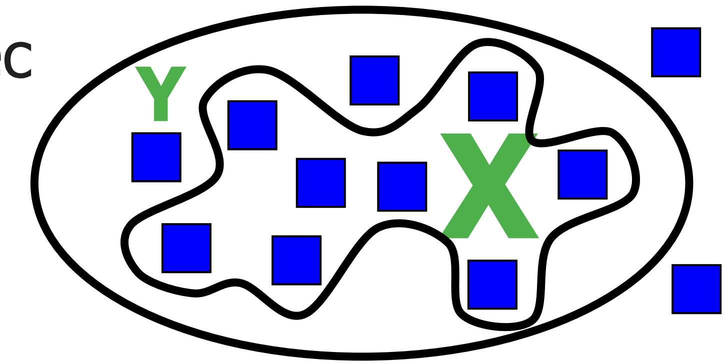


- Distributed/Random
  - Each cluster evaluates H to each neighbor and picks a neighbor with a maximum value
  - If two clusters pick each other, they merge

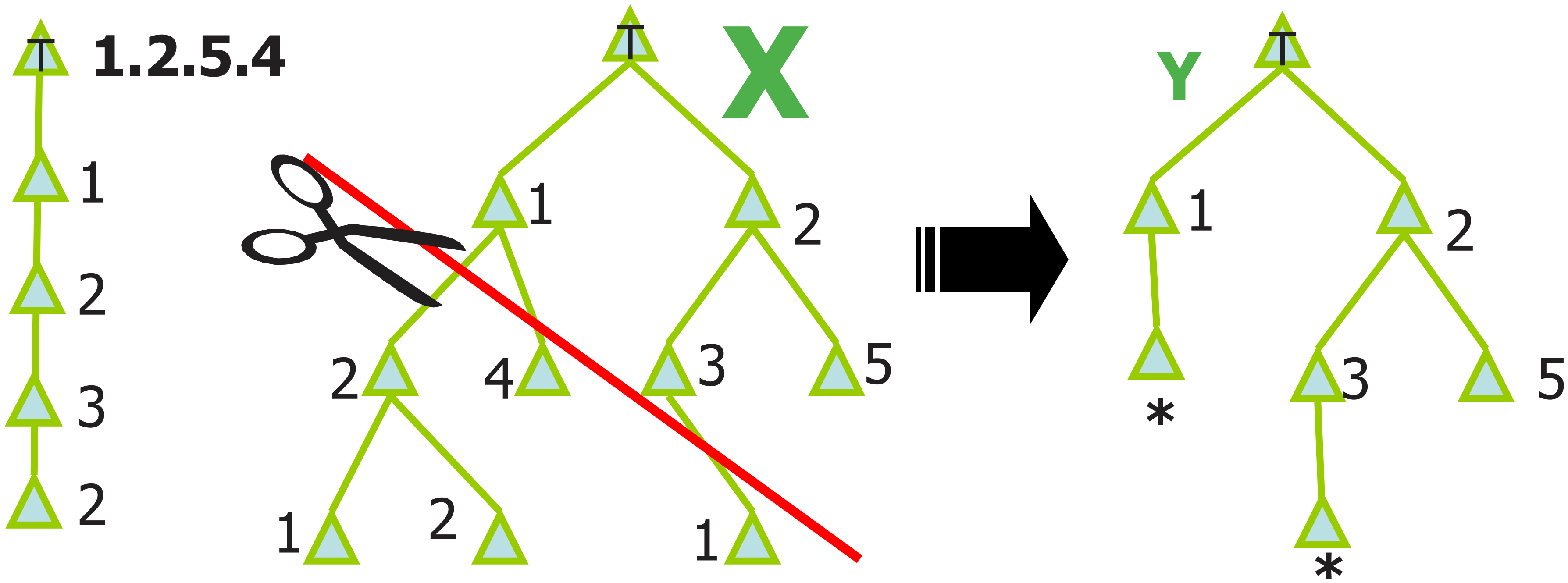
UK

Specification Abstraction

- The ability to replace any spec with a <u>more compact</u> but less specific one



- Our specification language



- New parameter: maximum cluster spec size

Evaluation

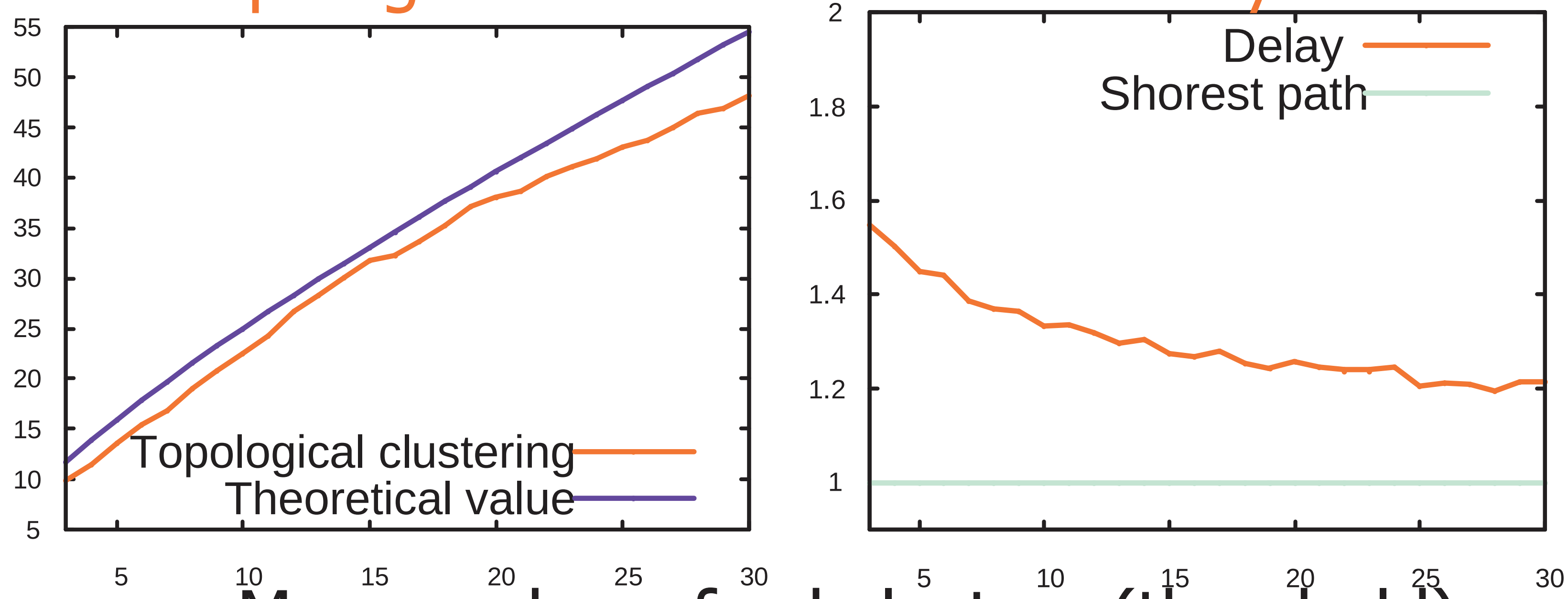
- Goal: compare topological, semantic and toposemantic clustering, explore parameter space
- Simulation parameters
  - Distributed clustering, Unicast traffic
  - Transit-stub (by GT-ITM), 600 nodes, 20 topologies
- Metrics
  - *Topological State* = average number of RTEs (visible clusters)
  - *Spec state ratio* = spec state with clustering/spec state without clustering
  - *Stretch* – delay in edges / shortest path delay
  - Load – number of links over which a pkt is forwarded
  - *Ratio of overdeliveries* = load with abstraction/load without abstraction

Topological Clustering

(σ = 0, semantics ignored)

- Trade-off #0

Topological state vs Delay



Max number of subclusters (threshold)

  - Hierarchy is balanced
  - Studied by Kamoun & Kleinrock

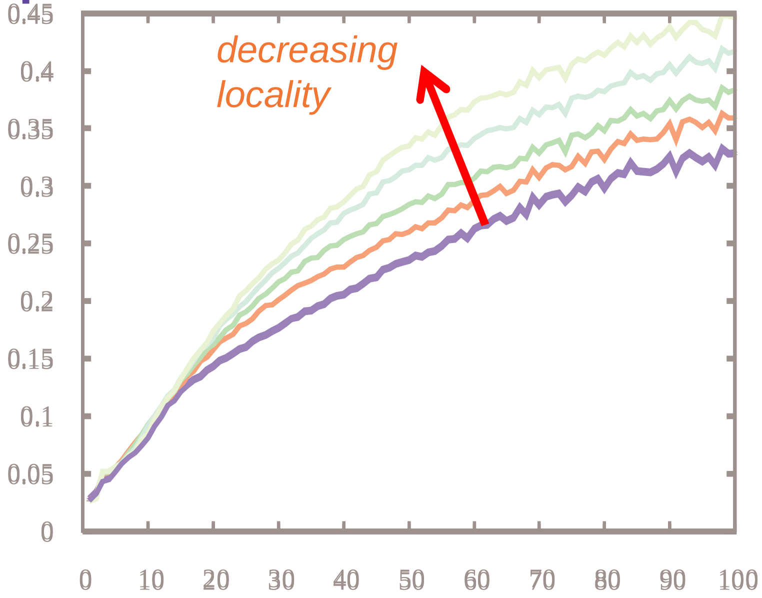
UK

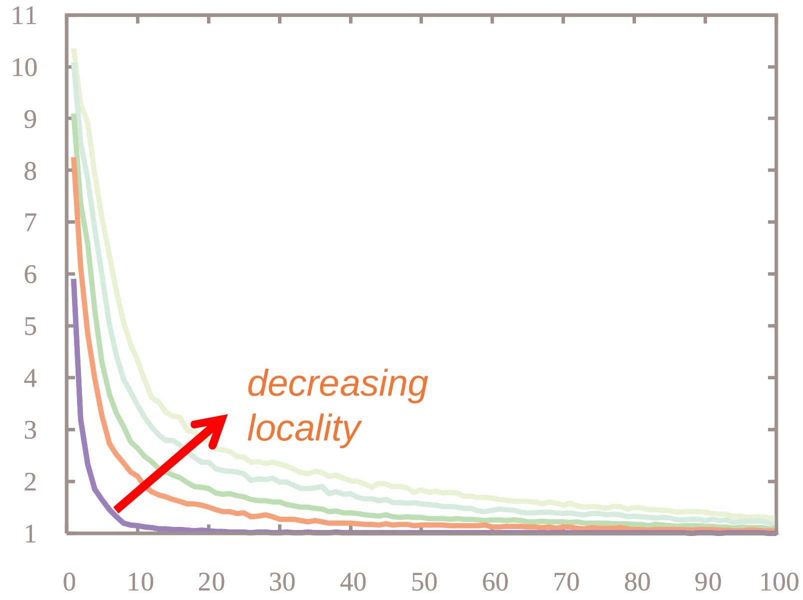
Topological Clustering

(σ = 0, semantics ignored)

- Trade-off #1

Specification state ratio vs Ratio of overdeliveries





Max cluster spec size

- ***Locality:*** correlation between node's location and spec
  - We start with high locality, then
  - Pick random pairs of nodes and swap their specs

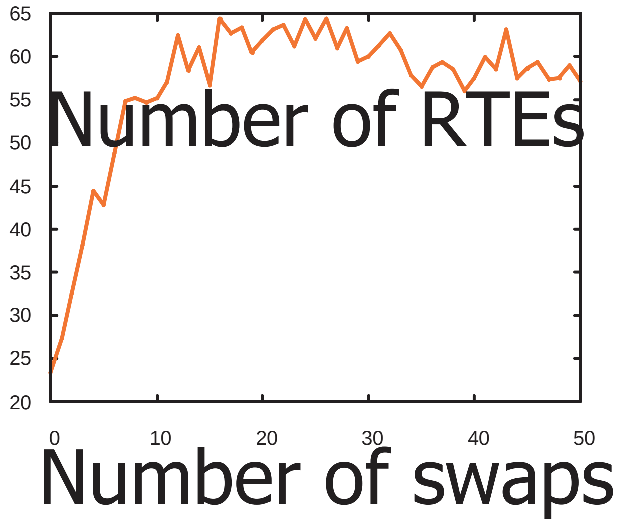
Semantic Clustering

(τ = 0, topology is ignored)

- For high locality (no swaps) semantic is better

- For lower locality confirmed
  - Number of RTEs: Topological < Semantic
  - Size of RTEs: Topological > Semantic

- Problem: even small amount of randomness leads to unbalanced hierarchy

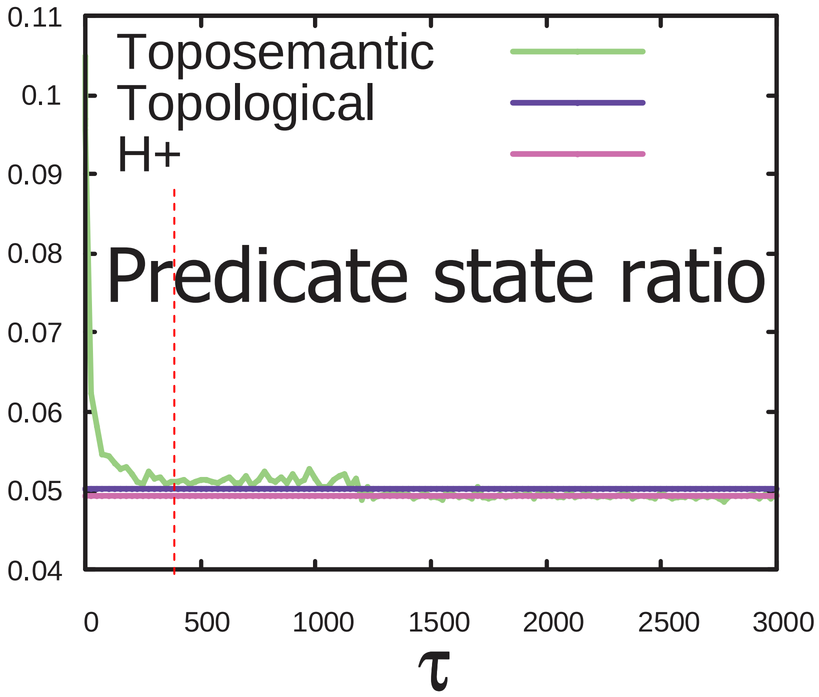


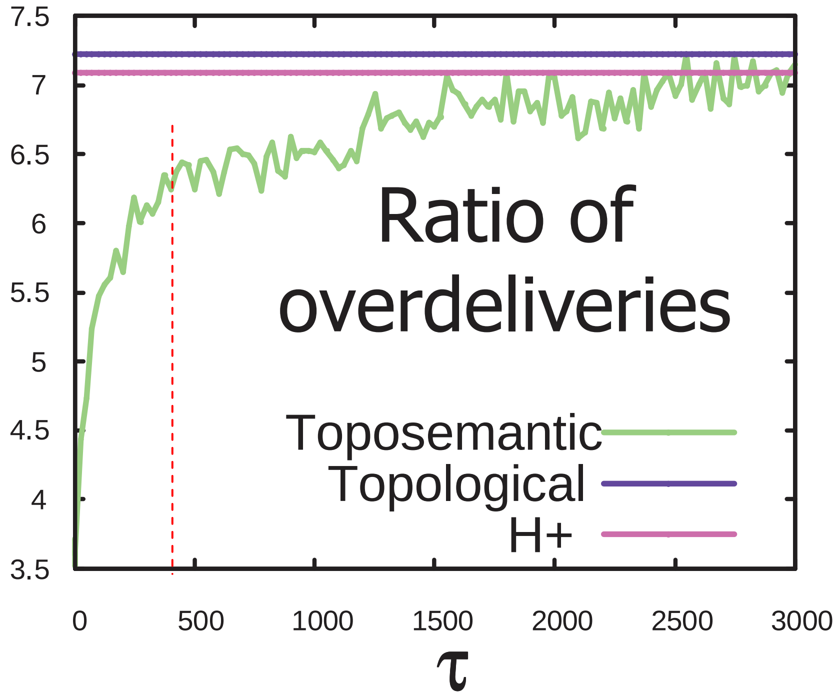
UK

Knob-Setting

Approach: set $\sigma = 1$, pick the best τ

(Up to 20% of nodes non-local)

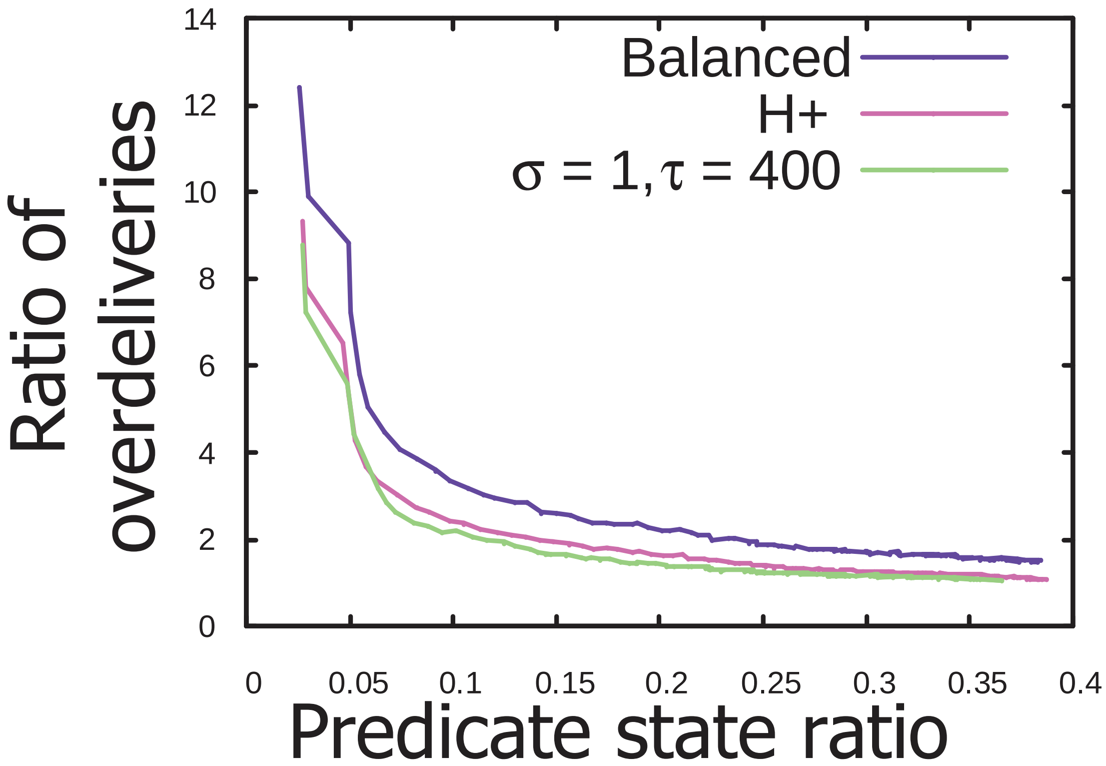




- Pick $\tau = 400$
- Approach "Toposemantic H^+": $\sigma = 1$, $\tau = 10^6$

Tradeoff Effects

- Measure spec. state ratio – overdeliveries tradeoff
- Toposemantic is the best but requires a parameter
- H^+ is the 2nd best, but no parameter is needed



UK

Conclusions

- Studied properties and compared
  - Topological clustering
  - Semantic clustering
- Defined a new clustering approach – Toposemantic that separates weight of topology and semantics
- Described centralized and distributed algorithms
- Analyzed tradeoffs (and knob settings)