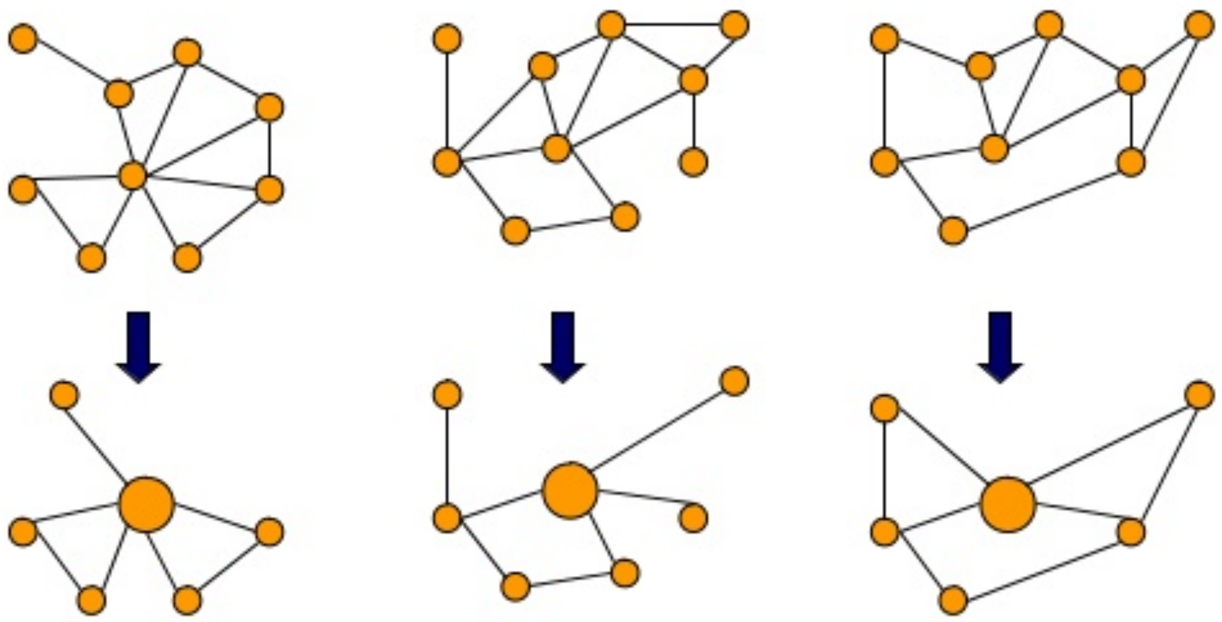
Solving graph compression via optimal transport

Vikas Garg, Tommi Jaakkola

CSAIL, MIT

Graph compression



- Improved performance
  - reduced storage requirements
  - faster algorithms
  - removal of spurious features for downstream tasks
- summarization/better visualization

Image source: Houw Liong The, Lecture 12 on graph mining.

Standard approach 1: Compression via coarsening

Find a matching, merge the matched vertices, and repeat.

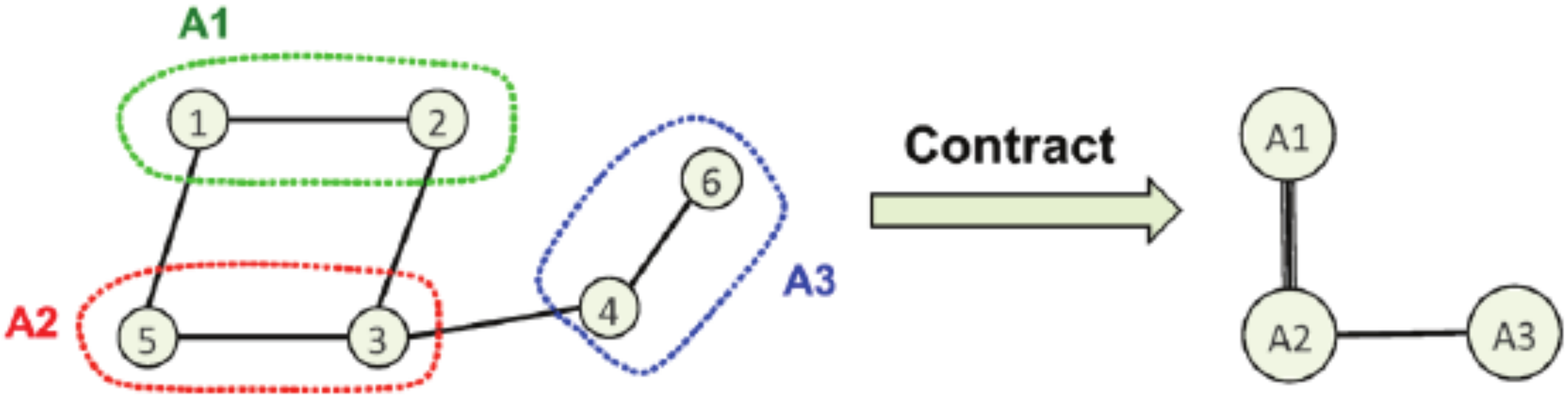


Image source: Aydin Buluc, SIAM Journal on Scientific Computing, 2011.

Standard approach 2: Compression via sparsification

Keep the vertices intact, and delete edges instead.

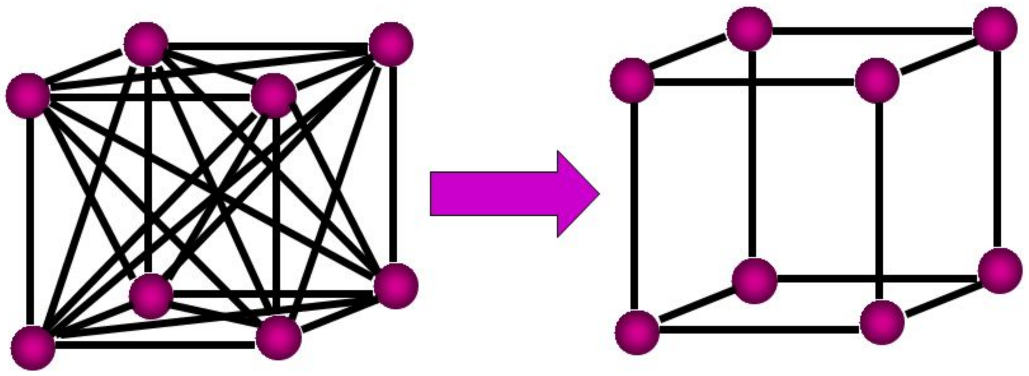


Image source: Daniel Spielman, Workshop on Algorithms for Modern Massive Datasets.

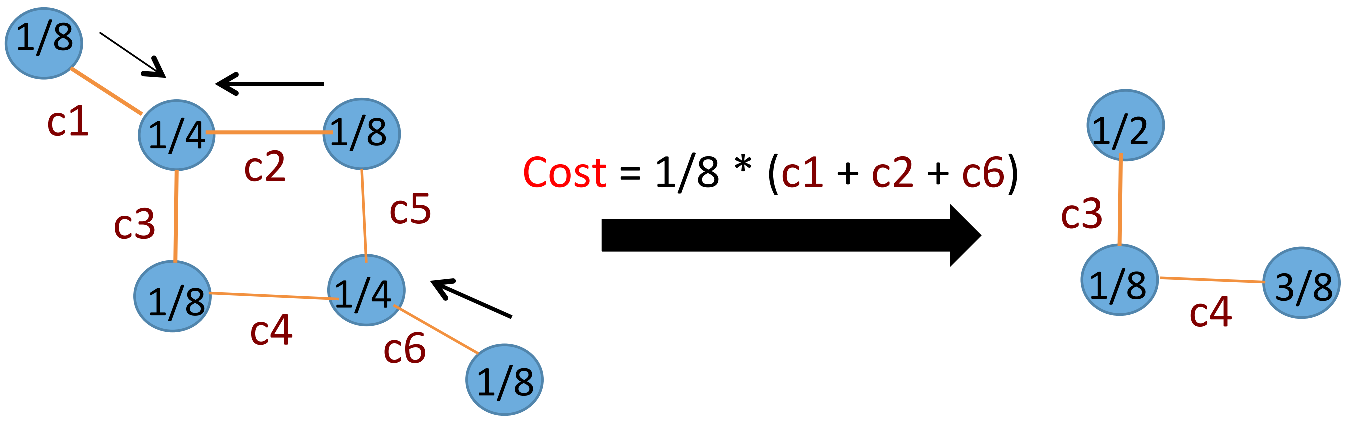
What's missing?

- Existing approaches try to preserve the graph spectrum, cuts etc.
  - oblivious to attribute information, e.g., graph labels and node features
- Compression criteria not given as an optimization problem
  - less suitable for robustly linking with downstream tasks

Optimal transport cost on a graph

Suppose we fix an initial distribution and a target distribution on the nodes.

What’s the minimum cost to transfer mass if we only allow flow along the edges?



Note that here the target distribution is over a smaller support (only 3 nodes)

Example: Cost associated with the indicated flow (not necessarily optimal)

How do we compute the optimal transport (OT) cost?

Previously known for directed graphs only. We extend to the undirected setting.

$$\min_{\substack{J^+, J^- \in \mathbb{R}^{|E|} \\ \mathbf{0} \preceq J^+, J^-}} \sum_{e \in E} c(e)(J^+(e) + J^-(e))$$

$$\text{s.t.} \quad F^\top (J^- - J^+) = \rho_1 - \rho_0$$

$c(e)$: cost of transporting unit flow on edge e

$J^+(e), J^-(e)$: flow on e in two directions

F : unoriented (unsigned) incidence matrix

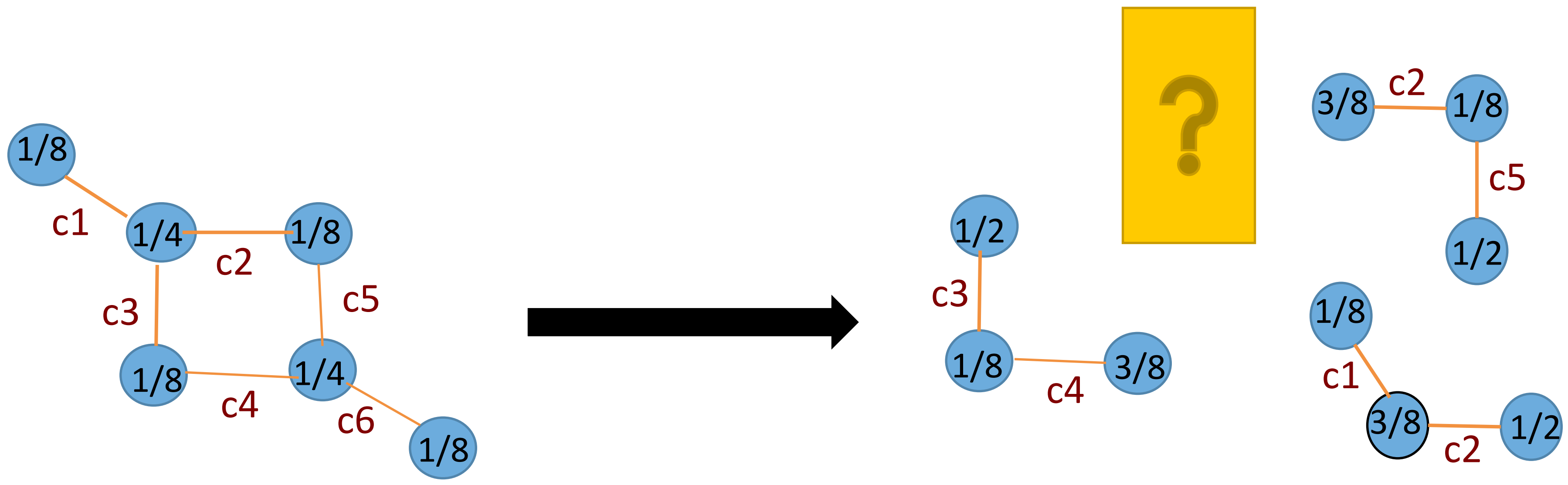
ρ_0, ρ_1 : initial and target distributions

Outline of our approach

- Define OT on a (directed/annotated) graph
  - cost depends on specified prior information
    - e.g. importance of nodes and their labels or attributes
    - thus can be informed by the downstream task
- Optimize the target distribution (its support)
  - using a regularized OT cost as the criterion
  - key step: we show how subgraph selection is found (yet to be illustrated)

Challenge: Target distribution is not known

- Combinatorial problem! Need to compute optimal cost relative to every target distribution over the specified size of support



Optimization formulation

$$\min_{\substack{\rho_1 \in \Delta(V) \\ ||\rho_1||_0 \leq k}} \max_{\substack{t \in \mathbb{R}^{|V|} \\ -c \preceq Ft \preceq c}} t^\top(\rho_1 - \rho_0) + \frac{\lambda}{2}||\rho_1||^2$$

Regularization

Control the size of target support

Optimal transport cost in dual form

Solving the optimization

- Non-convex due to sparsity constraint $||\rho_1||_0 \leq k$

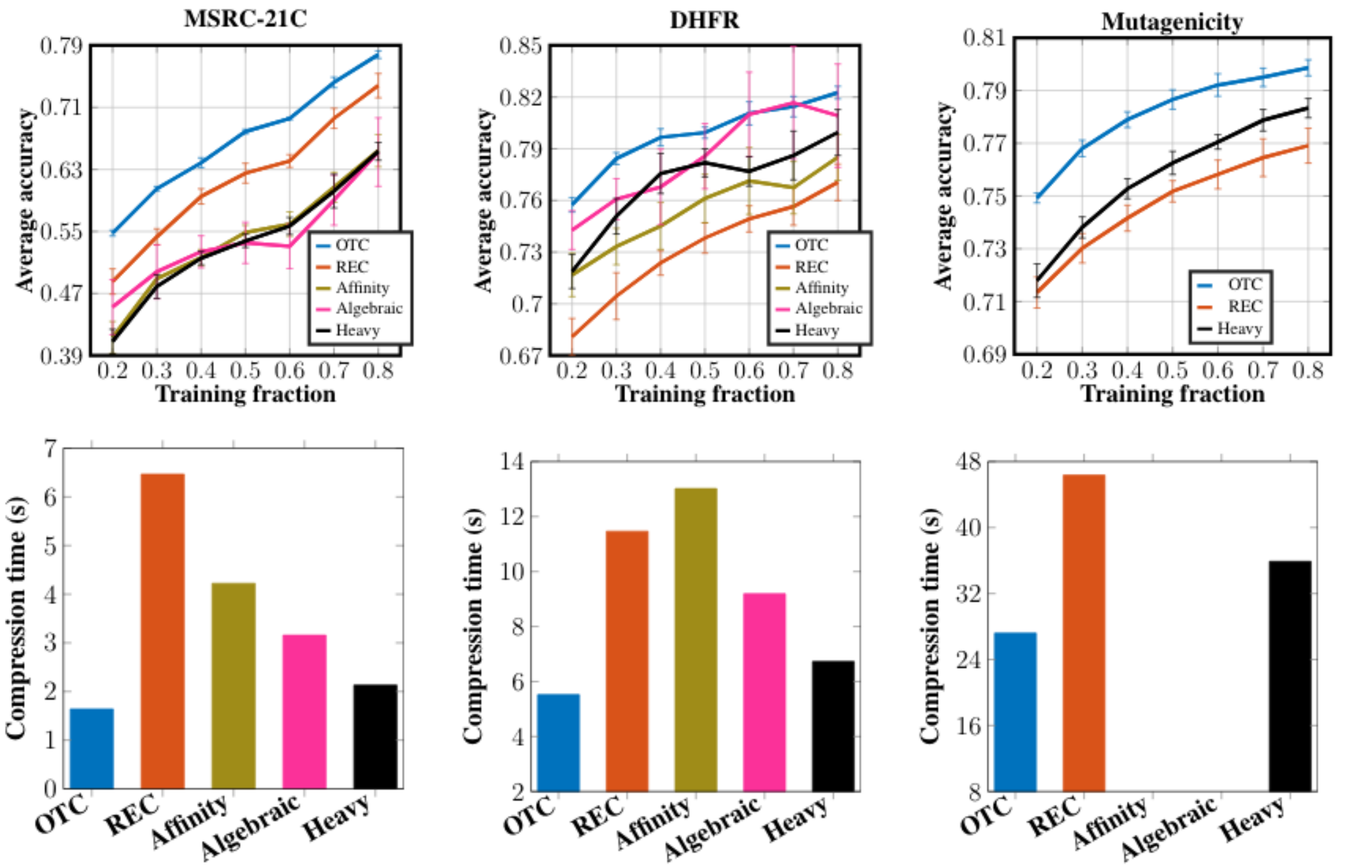
$$\min_{\substack{\rho_1 \in \Delta(V) \\ ||\rho_1||_0 \leq k}} \max_{\substack{t \in \mathbb{R}^{|V|} \\ -c \preceq Ft \preceq c}} t^\top(\rho_1 - \rho_0) + \frac{\lambda}{2}||\rho_1||^2$$

- Introduce Boolean variables ϵ and deduce

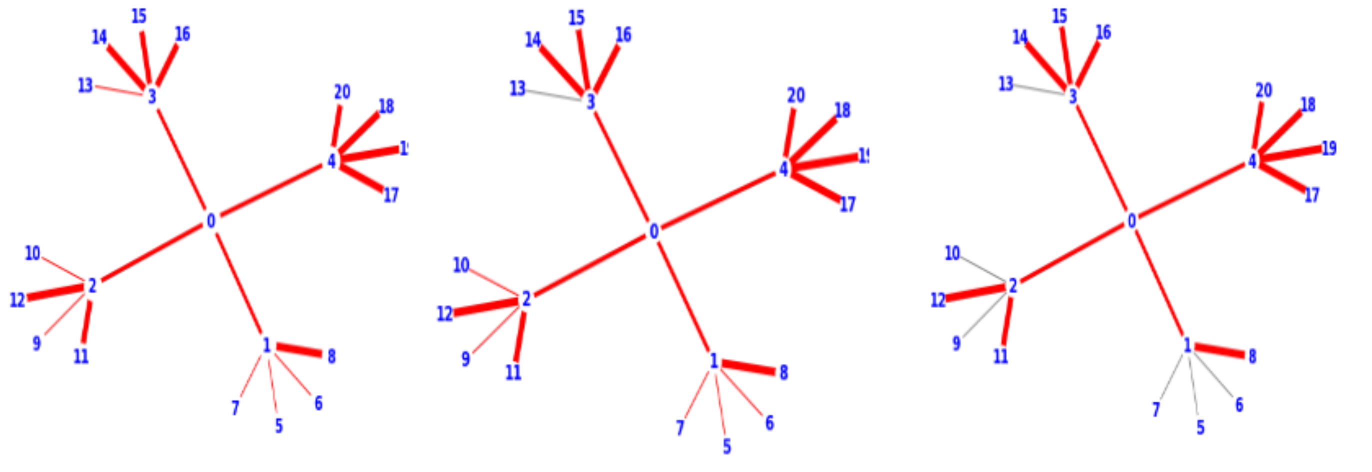
$$\min_{\epsilon \in \{0,1\}^{|V|}} \min_{\substack{\overline{\rho}_1 \in \mathbb{R}^{|V|} \\ \overline{\rho}_1 \odot \epsilon \in \Delta(V)}} \max_{\substack{t \in \mathbb{R}^{|V|} \\ -c \preceq Ft \preceq c}} t^\top(\overline{\rho}_1 \odot \epsilon - \rho_0) + \frac{\lambda}{2}||\overline{\rho}_1||^2$$

- Relax each coordinate of ϵ to [0, 1] and solve. Perform rounding to have at most k vertices. The spanned subgraph is our compressed graph.

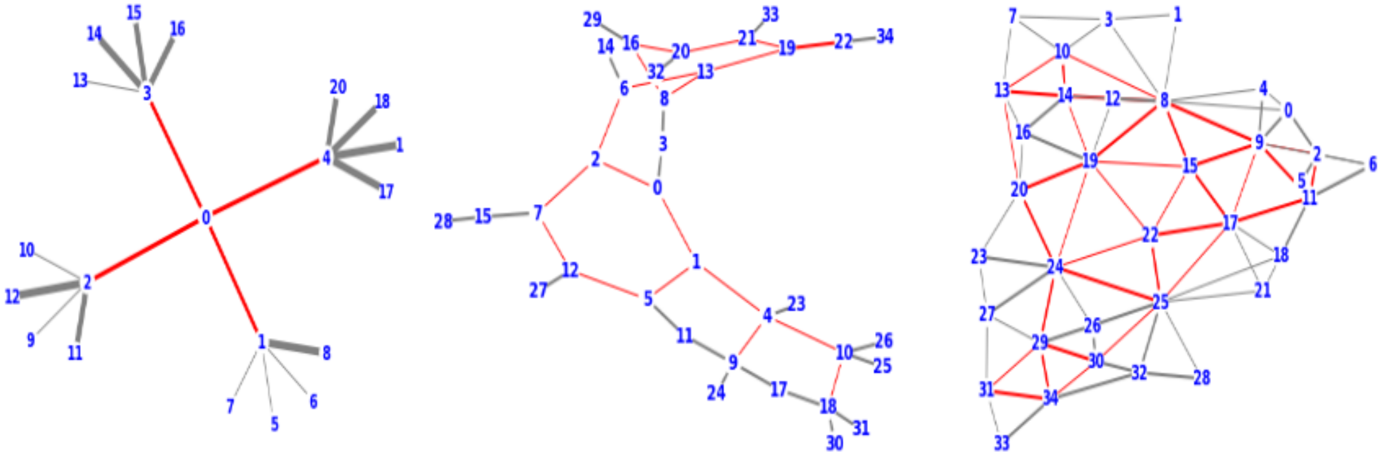
Performance on standard graph datasets



Qualitative results on synthetic and real data



(a) Synthetic graph (b) Compressed graph ($k = 20$) (c) Compressed graph ($k = 15$)



(d) Compressed graph ($k = 5$) (e) Mutagenicity (f) MSRC-21C

Conclusion

- A new principled approach to compressing graphs
- Prior information can be seeded easily
- Suitable for downstream tasks such as classification
- Interesting directions
  - complement encoding (compression) with decoding (decompression)
  - expand the framework to allow additional constraints (e.g. requiring the compressed graph to be connected)
  - higher order, structured graph compression