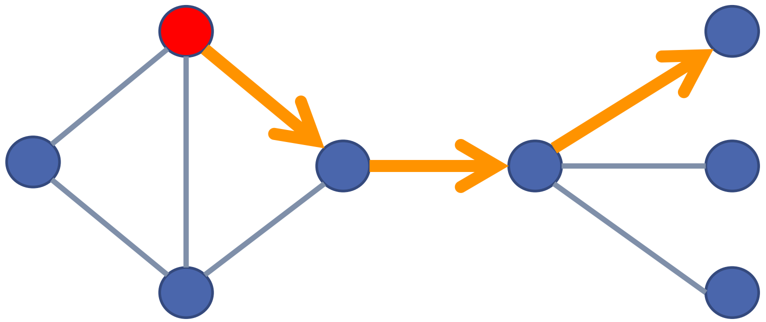
Eccentricity Heuristics via Parallel Set Cover

Tal Wagner

MIT

Graph Eccentricities

$$\boldsymbol{e}(v) = 3$$



- Let $G(V, E)$ be a graph
- Shortest-path metric: $\Delta: V \times V \to \mathbb{R}$
- **Eccentricities:**

$$\boldsymbol{e}(v) = \max_{u \in V} \Delta(v, u)$$

- Many applications in graph mining, network analysis

Computing All Eccentricities

- Exact computation: $O(mn)$ (e.g. BFS from each node)
- Approximate algorithms
  - **Theoretical**:

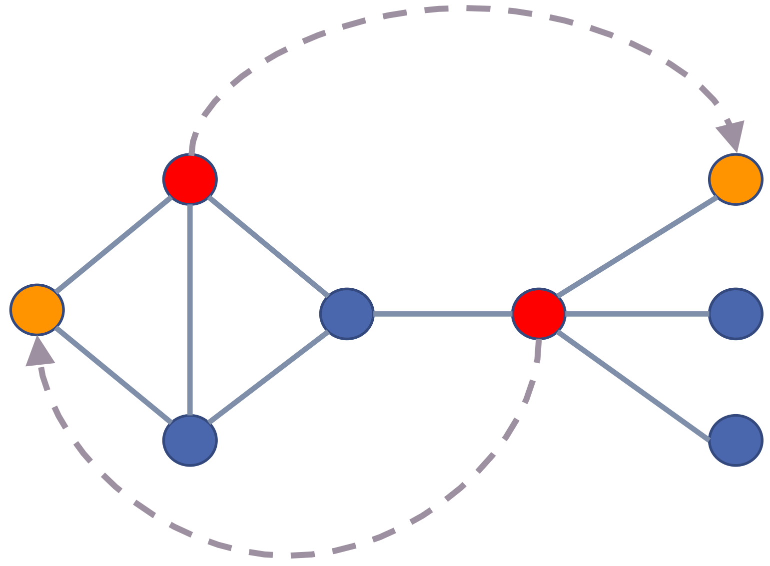
4-approx.	$O(m)$ time	[One BFS]
$(2+\delta)$-approx.	$\tilde{O}(m/\delta)$ time	[Backurs-Roditty-Segal-V.Williams-Wein'18]
$(5/3)$-approx.	$\tilde{O}(m^{1.5})$ time	[Chechik-Larkin-Roditty-Schoenebeck-Tarjan-V.Williams'14]

Tight under SETH

  - **Empirical**: [Kang et al.'11], [Boldi et al.'11], […], **[Shun'15]**

k-BFS$_2$ [Shun'15]

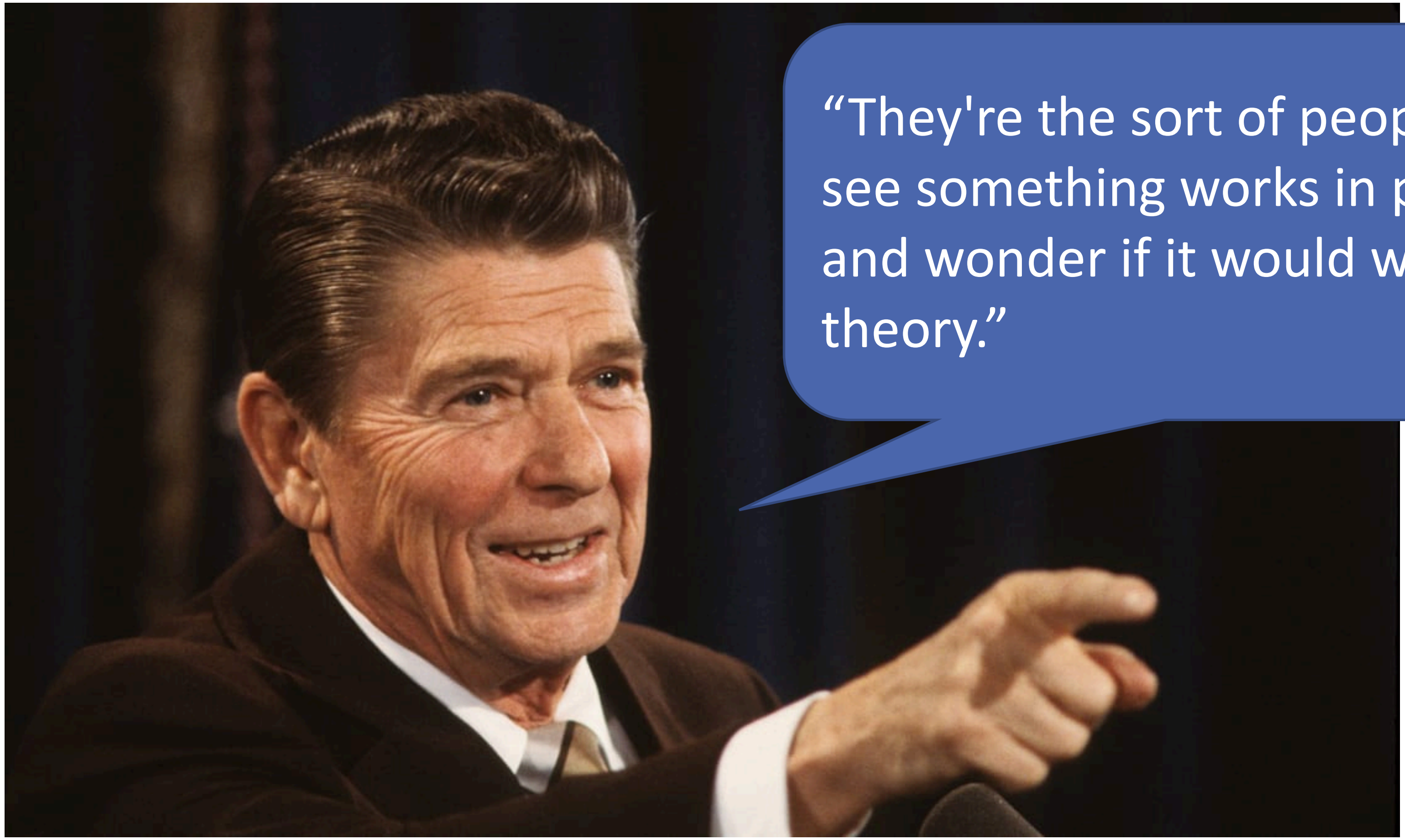
- $S_1 \leftarrow k$ random nodes
- Compute BFS from each $u \in S_1$
- $S_2 \leftarrow k$ furthest nodes from S_1
- Compute BFS from each $u \in S_2$
- Return $\hat{\boldsymbol{e}}(v) \leftarrow$ max distance from $S_1 \cup S_2$



Beats all other methods by orders of magnitude

Why?

Reagan’s Principle



“They're the sort of people who see something works in practice and wonder if it would work in theory.”

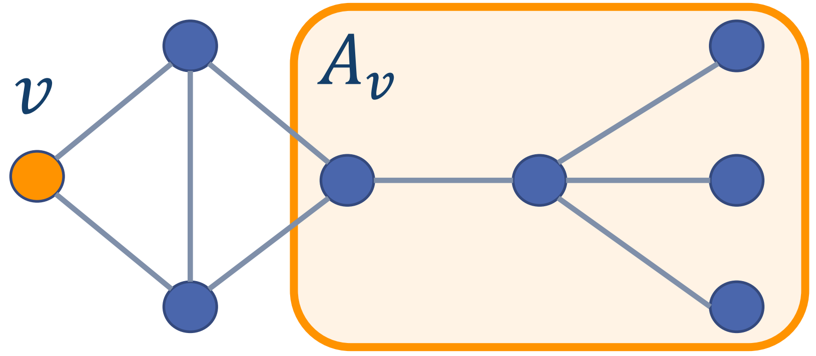
This Talk

- k-**BFS$_2$** is **nearly identical** to the streaming Set Cover algorithm of [Demaine-Indyk-Mahabadi-Vakilian'14]
- Arguably **explains** performance of k-**BFS$_2$**
- Gives **provable** variant with **even better** performance

Set Cover Formulation

- **<u>Set Cover</u>**: Given elements V and subsets $\mathcal{S} \subset 2^V$, find smallest cover $C \subset \mathcal{S}$ of V.

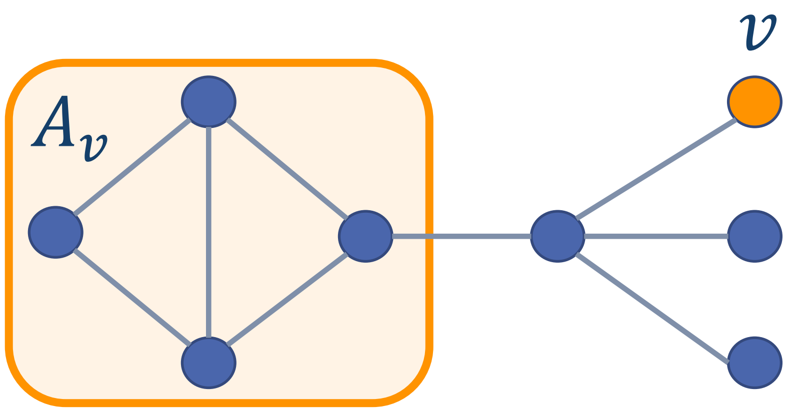
- **<u>Eccentricities as Set Cover:</u>**
  - Nodes are elements
  - Nodes are sets: $\mathcal{S} = \{A_v : v \in V\}$



$$A_v = \{u \in V : \boldsymbol{e}(u) = \Delta(v, u)\}$$

Set Cover Formulation

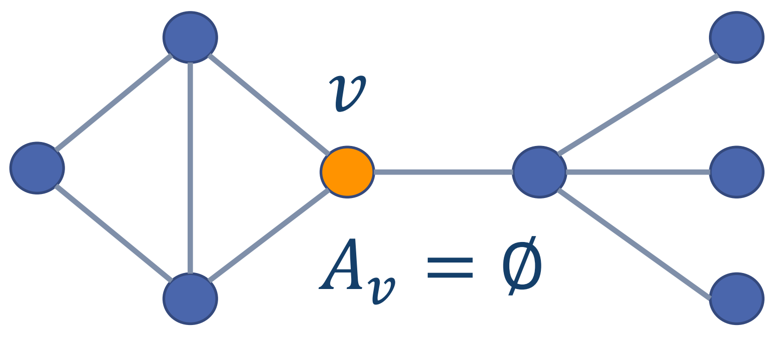
- **Set Cover**: Given elements V and subsets $\mathcal{S} \subset 2^V$, find smallest cover $C \subset \mathcal{S}$ of V.
- **Eccentricities as Set Cover:**
  - Nodes are elements
  - Nodes are sets: $\mathcal{S} = \{A_v : v \in V\}$



$$A_v = \{u \in V : \boldsymbol{e}(u) = \Delta(v, u)\}$$

Set Cover Formulation

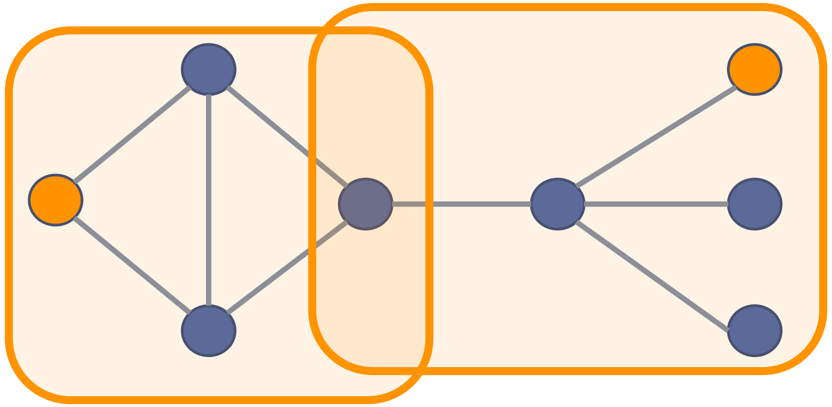
- **Set Cover**: Given elements V and subsets $\mathcal{S} \subset 2^V$, find smallest cover $C \subset \mathcal{S}$ of V.
- **Eccentricities as Set Cover:**
  - Nodes are elements
  - Nodes are sets: $\mathcal{S} = \{A_v : v \in V\}$



$$A_v = \{u \in V : \boldsymbol{e}(u) = \Delta(v, u)\}$$

Set Cover Formulation

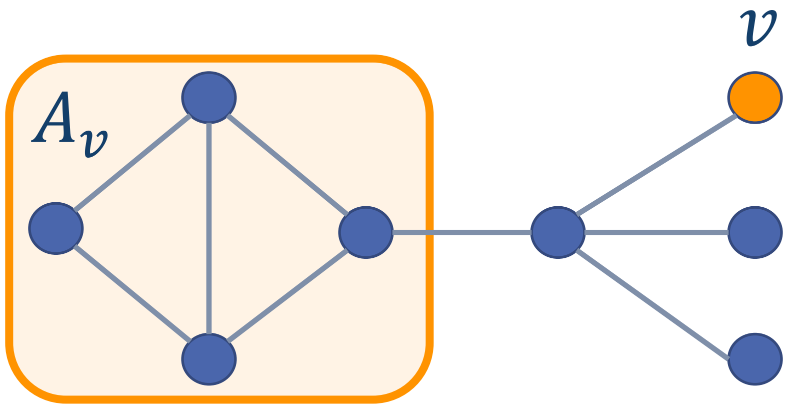
- **<u>Set Cover</u>**: Given elements V and subsets $\mathcal{S} \subset 2^V$, find smallest cover $C \subset \mathcal{S}$ of V.
- **<u>Eccentricities as Set Cover:</u>**
  - Nodes are elements
  - Nodes are sets: $\mathcal{S} = \{A_v : v \in V\}$
- Cover computes all eccentricities
- Optimal cover = "eccentric cover", $\boldsymbol{\kappa}$



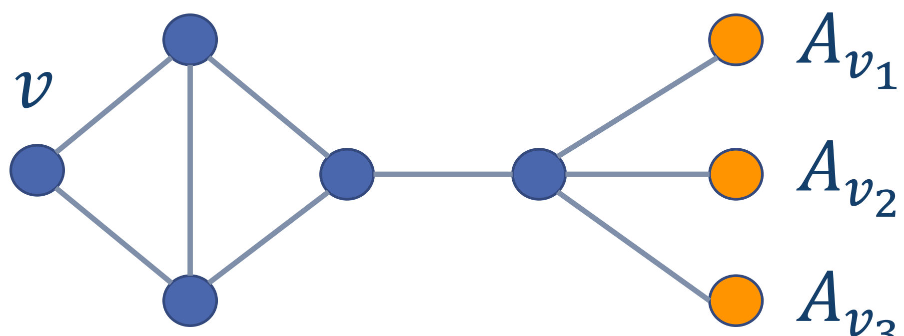
$$A_v = \{u \in V : \boldsymbol{e}(u) = \Delta(v, u)\}$$

Computational Constraints

- Computing a set A_v is **prohibitive**
  - $O(mn)$ work



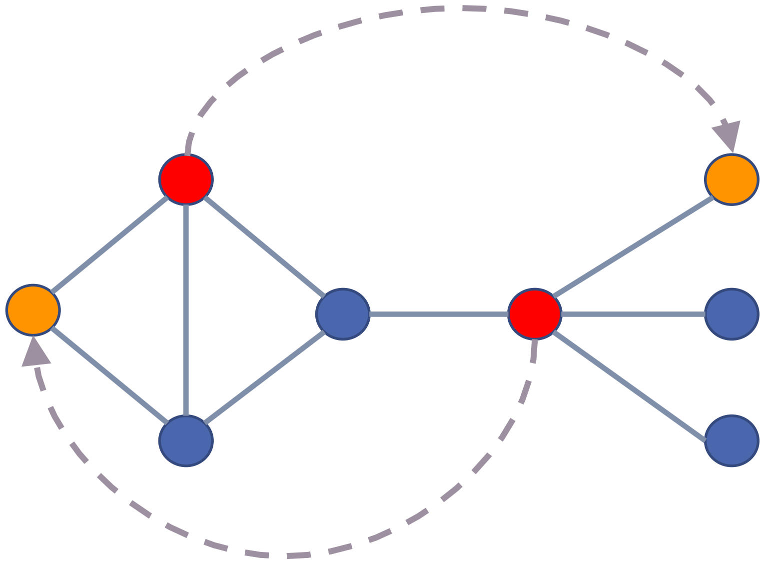
- Computing which sets cover v is **expensive**
  - Single BFS, $O(m)$ work



- Known Set Cover algorithms? **Yes**

Streaming Set Cover [Demaine-Indyk-Mahabadi-Vakilian'14]

- $S_1 \leftarrow k$ random elements
- $C \leftarrow$ Cover for sample (e.g. greedy)



Element Sampling Lemma:

If global optimum is small, C covers almost all elements.

k-BFS$_2$ vs. DIMV

k-BFS$_2$		Streaming Set Cover [DIMV'14]
$S \leftarrow$ Random sample	$\Longleftrightarrow$	$S \leftarrow$ Random sample
Compute BFS from each $v \in S$	$\Longleftrightarrow$	Compute covering sets for each $v \in S$
$C \leftarrow k$ nodes with max $\Delta(v, S)$	$\not\approx$	$C \leftarrow$ Greedy cover for S

k-BFS$_{SC}$

k-BFS$_2$		Streaming Set Cover [DIMV'14]
$S \leftarrow$ Random sample	$\Leftrightarrow$	$S \leftarrow$ Random sample
Compute BFS from each $v \in S$	$\Leftrightarrow$	Compute covering sets for each $v \in S$
~~$C \leftarrow k$ nodes with max $\Delta(v, S)$~~	$\not\approx$	$C \leftarrow$ Greedy cover for S

$C \leftarrow$ Parallel greedy cover for S

[Blelloch-Peng-Tangwongsan'11]

[Blelloch-Simhadri-Tangwongsan'12]

k-BFS$_{\text{SC}}$

Theorem:

Suppose $G(V, E)$ has eccentric cover size $\boldsymbol{\kappa}$.

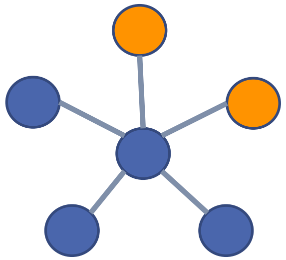
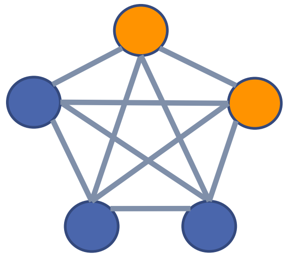
k-**BFS$_{\text{SC}}$** with $k = \tilde{O}(\boldsymbol{\kappa} \cdot \epsilon^{-1} \log n)$ satisfies:

- Expected work: $O(km)$, expected depth: $\tilde{O}(\mathrm{diam}(G))$
- Computes ***exact eccentricities*** of all but an ϵ-fraction of nodes w.h.p.

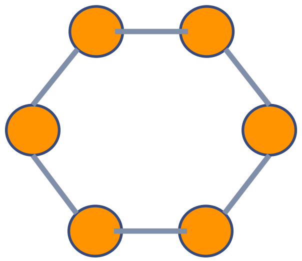
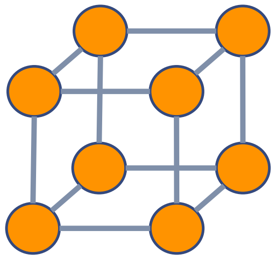
Also approximate version under weaker assumption.

Eccentric Cover: Warm-Up

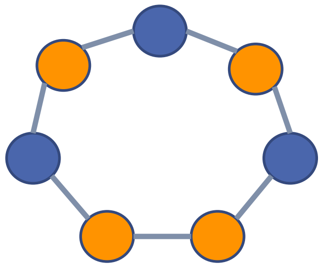
- Path, star, clique: $\boldsymbol{\kappa} = 2$

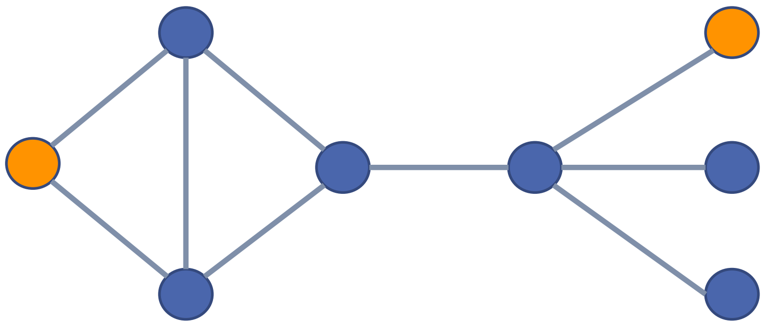
- Even cycle, hypercube: $\boldsymbol{\kappa} = n$

- Odd cycle: $\boldsymbol{\kappa} = \frac{1}{2}(n+1)$



Eccentric Cover in the Wild



- 8 real-world graphs in [Shun'15]
- **1M-4M** nodes each
- Upper bounds on eccentric cover size:
  - 2 graphs: $\boldsymbol{\kappa \leq 128}$
  - 5 graphs: $\boldsymbol{\kappa \lesssim 1,000}$
  - 1 graph: $\boldsymbol{\kappa \lesssim 10,000}$

Real-world graphs have small eccentric covers

Experiments

k-**BFS$_2$** vs. k-**BFS$_{SC}$**

(Real-world graphs from Stanford Network Analysis Project)