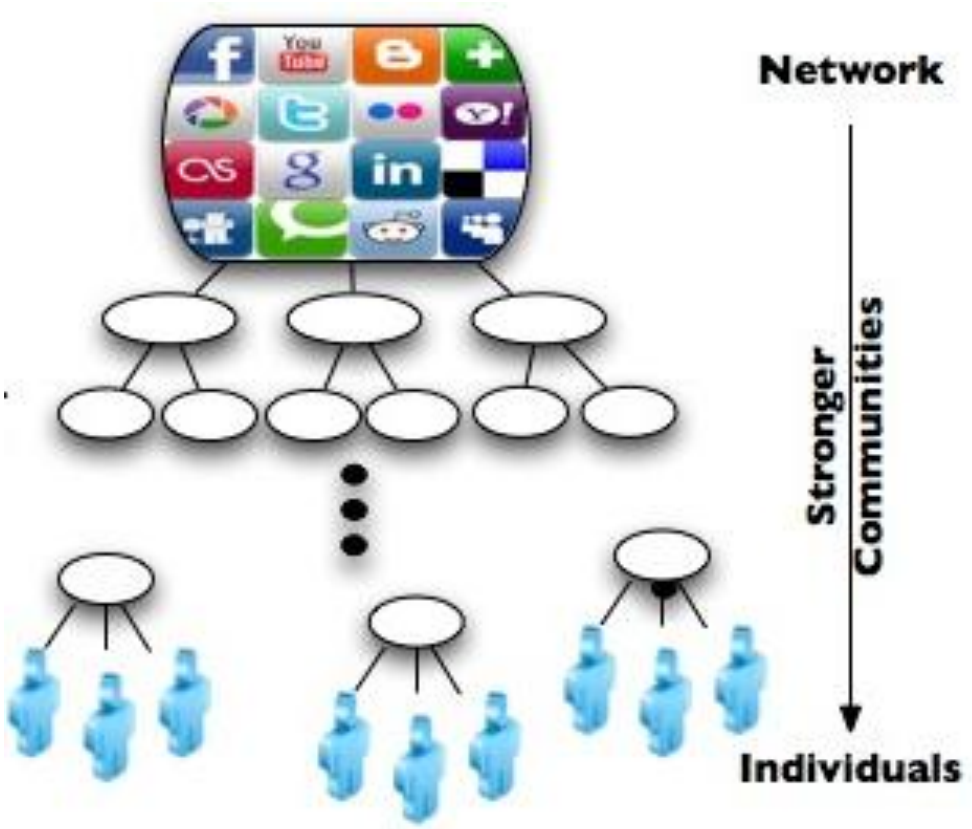
Hierarchical Community Decomposition Via Oblivious Routing Techniques



Sean Kennedy

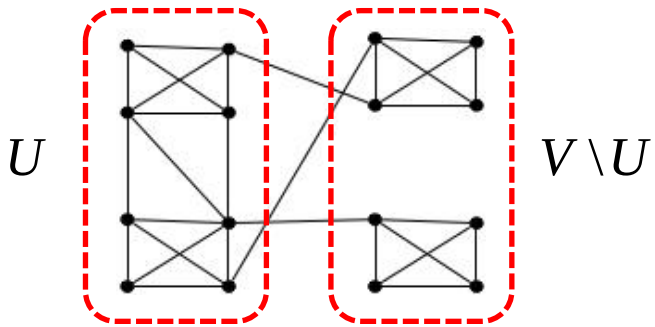
Bell Labs Research (Murray Hill, NJ, USA)
Oct 8th, 2013

Joint work Jamie Morgenstern (CMU), Gordon Wilfong (BLR), Liza Zhang (BLR)

What make a "good" community in a social network?

- Small diameter (k-center, k-median, k-means)
  - Need to choose number of communities before hand
  - Results dependant on starting centers
- High density (high max or average degree)
- Small cut (minimum conductance)
  - NP-hard
- Modularity based methods
  - Louvain method [Blondel, Guillaume, Lambiotte, Lefebvre, 2008]
  - Resolution problems
- ….

Conductance



$$\varphi(U) := \frac{\mathbf{cap}(U, V \setminus U)}{\min\{\mathbf{cap}(U, V), \mathbf{cap}(V \setminus U, V)\}}$$

$$\mathbf{cap}(A, B) = \sum_{x \in A} \sum_{x \in B} w_{xy}$$

Most NP-hard to compute.

Can approximate:

O(log n)-approx [Leighton, Rao, 1999]

O(√log n)-approx [Aurora, Rao Vazirani, 2004]

Spectral cuts heuristically find low conductance cuts

[Leskovec, Lang, Mahoney 2010] show all cut based metrics are similar in practice

Focus is on best cut of minimum conductance

Related

- Bi-criteria result of [Kannan, Vempala, Vetta, 2004]

  (α,ε)-clustering (each community has conductance at most α and at most ε fraction of edges between communities)

  Thm: Exists an (α,ε)-clustering then can find (f(α),g(ε))-clustering.

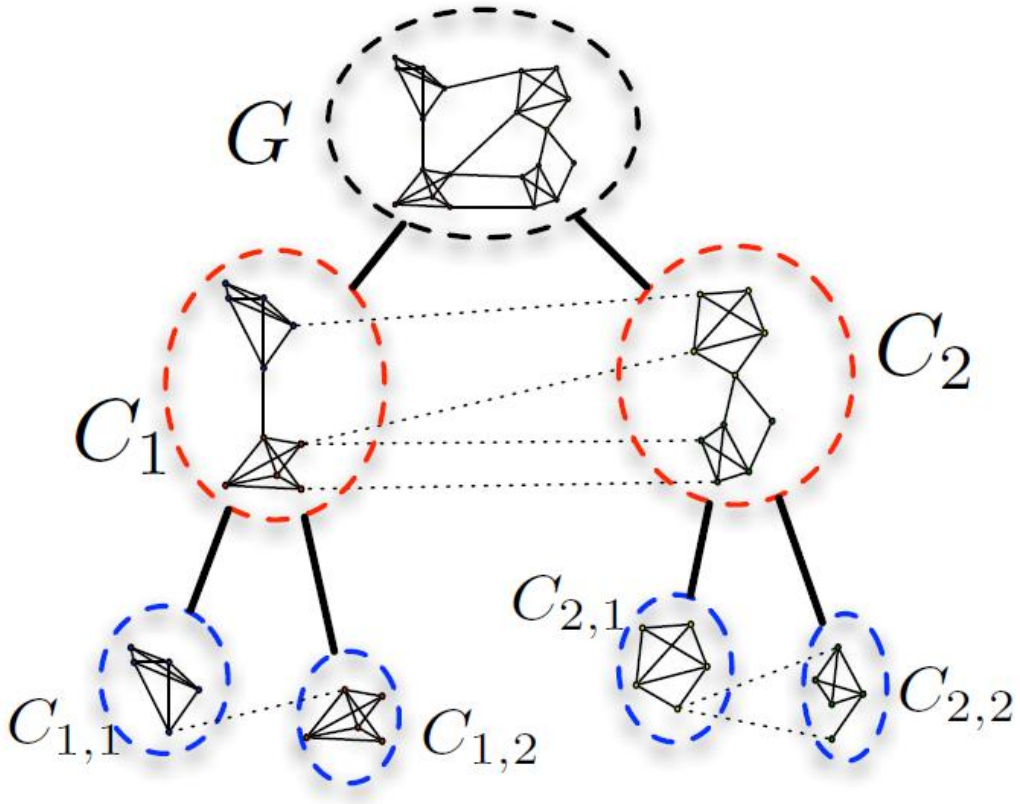
Hierarchical Community Decomposition

Decomposition tree T

- root labeled by V
- children form a partition

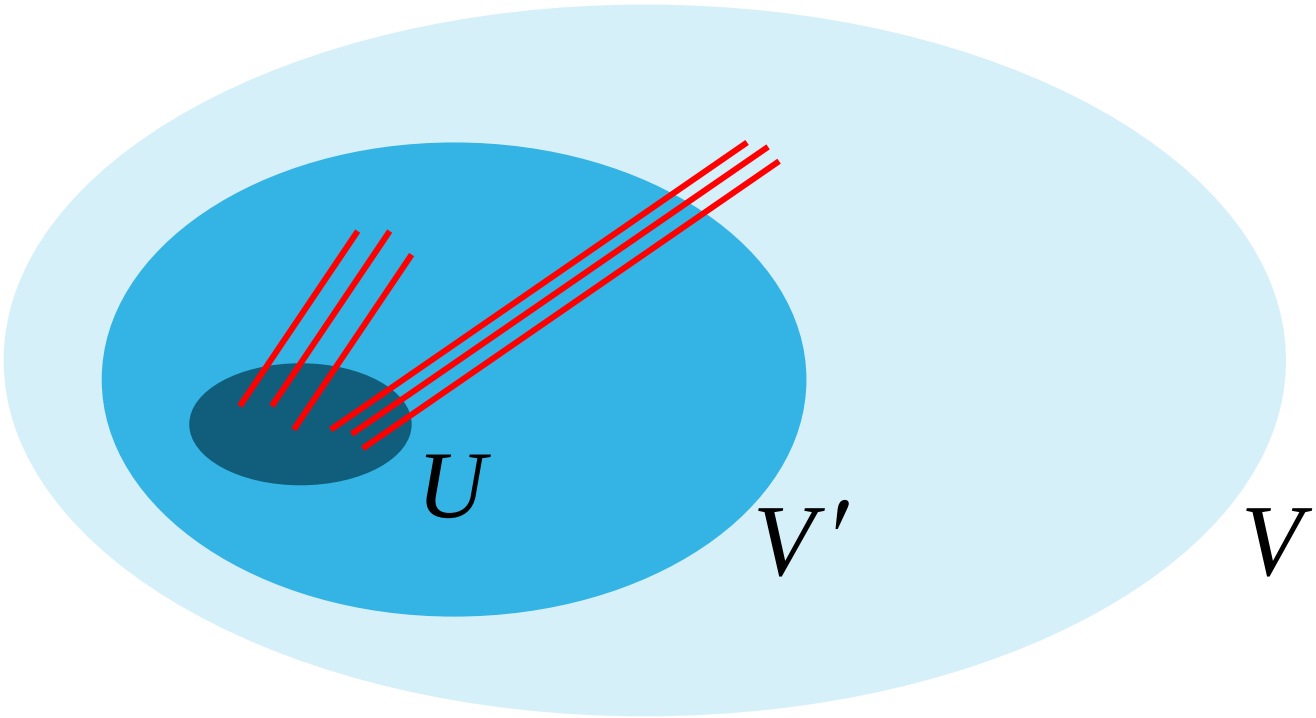
A Good Community:

a) must be internally well-connected



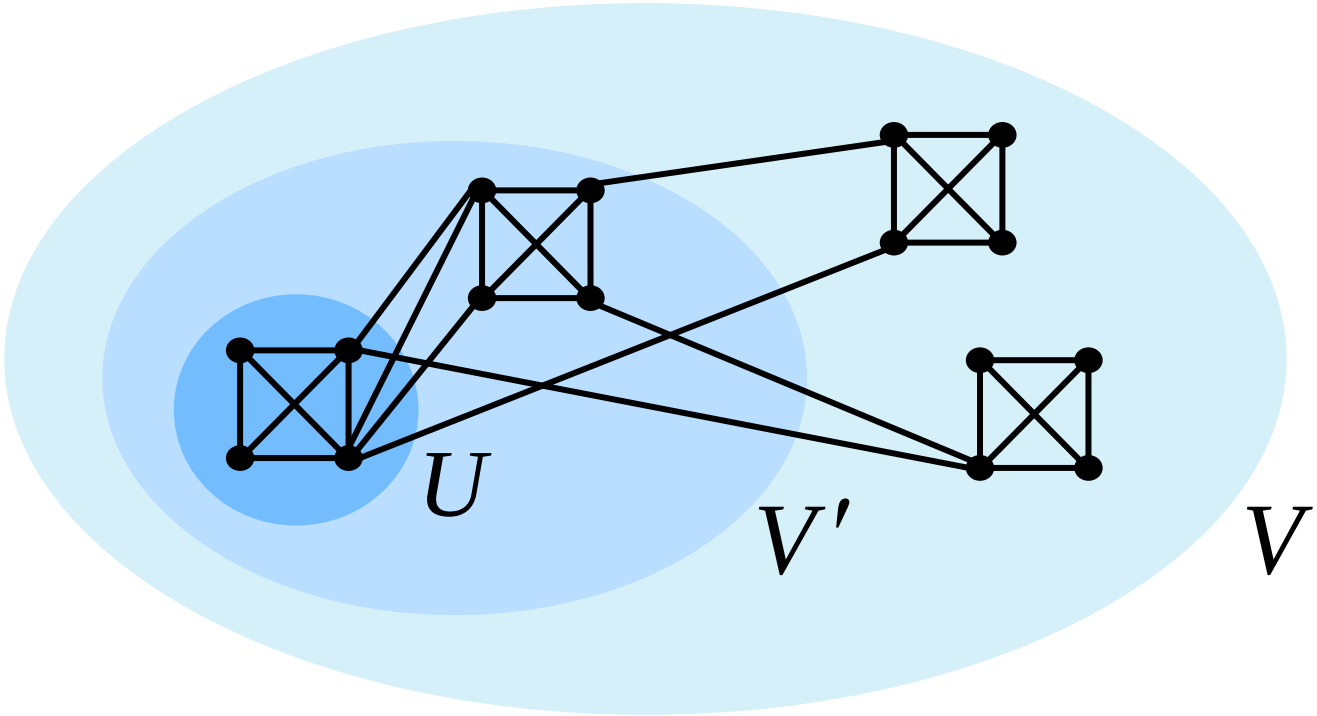
λ-Detached

$$|U| \leq \frac{|V'|}{2} \quad \text{and} \quad \frac{\mathbf{cap}(U, V \setminus U)}{\mathbf{cap}(U, V' \setminus U)} \geq \lambda$$



U is λ-Detached if it is ***small enough*** and has ***many external connections***

λ-Detached (example)

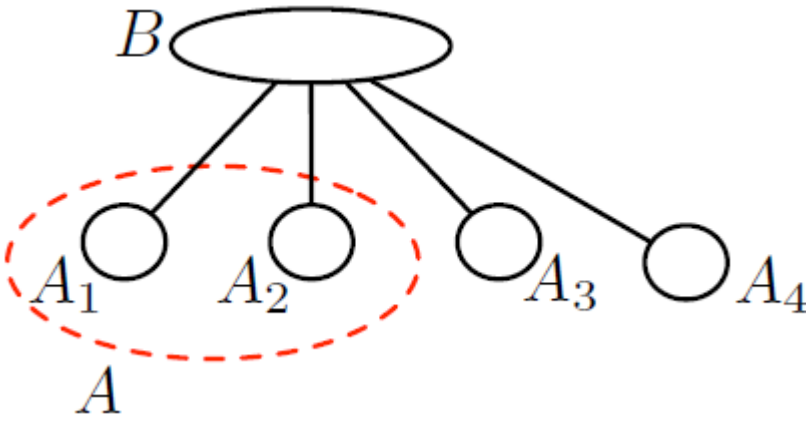


$$|U| \leq \frac{|V'|}{2} \quad \text{and} \quad \frac{\mathbf{cap}(U, V \setminus U)}{\mathbf{cap}(U, V' \setminus U)} = \frac{7}{3}$$

Hierarchical Community Decomposition

A ***decomposition tree T***

- root labeled by V
- children form a partition

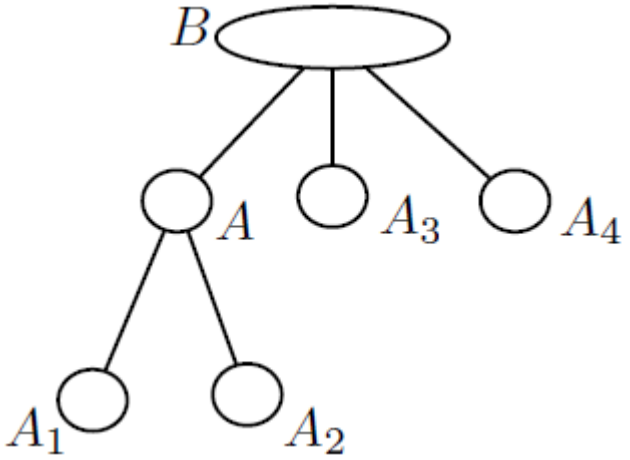


A Good Community:

a) does not contain λ-detached subgraphs.

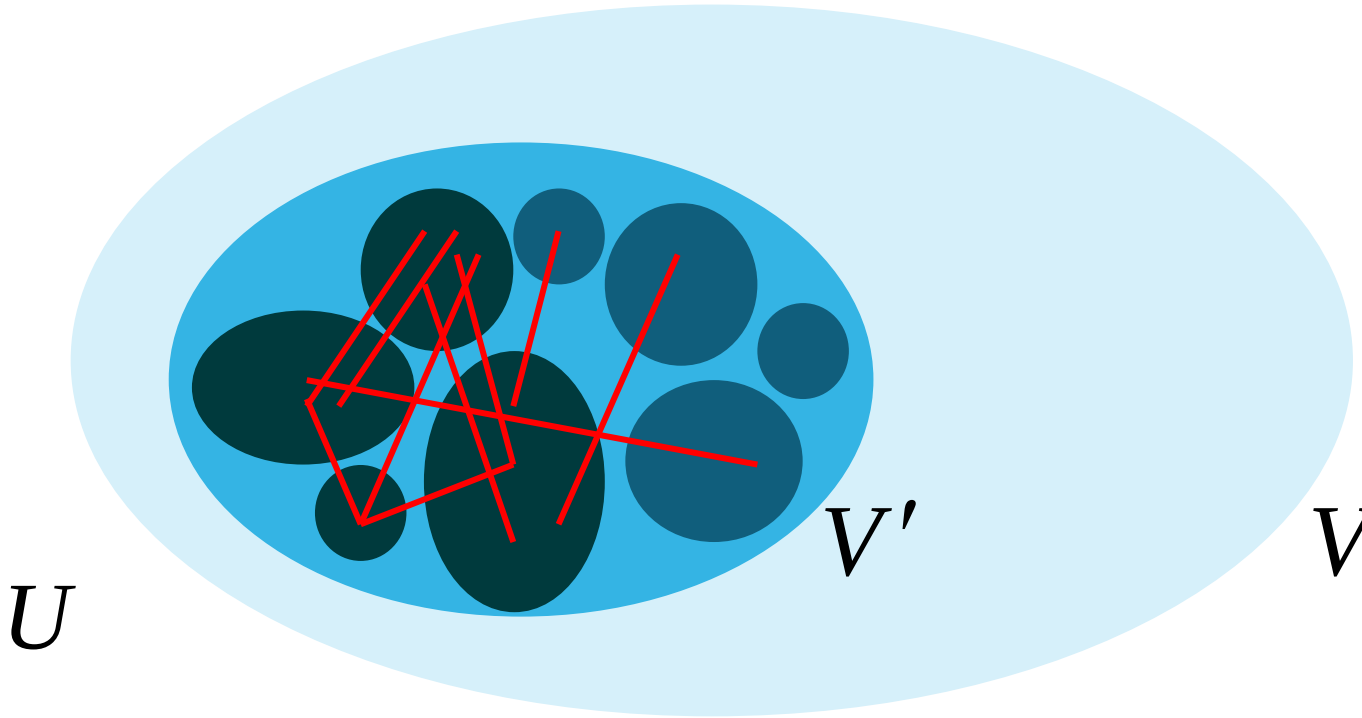
A Good Decomposition:

b) contains all levels of hierarchy.



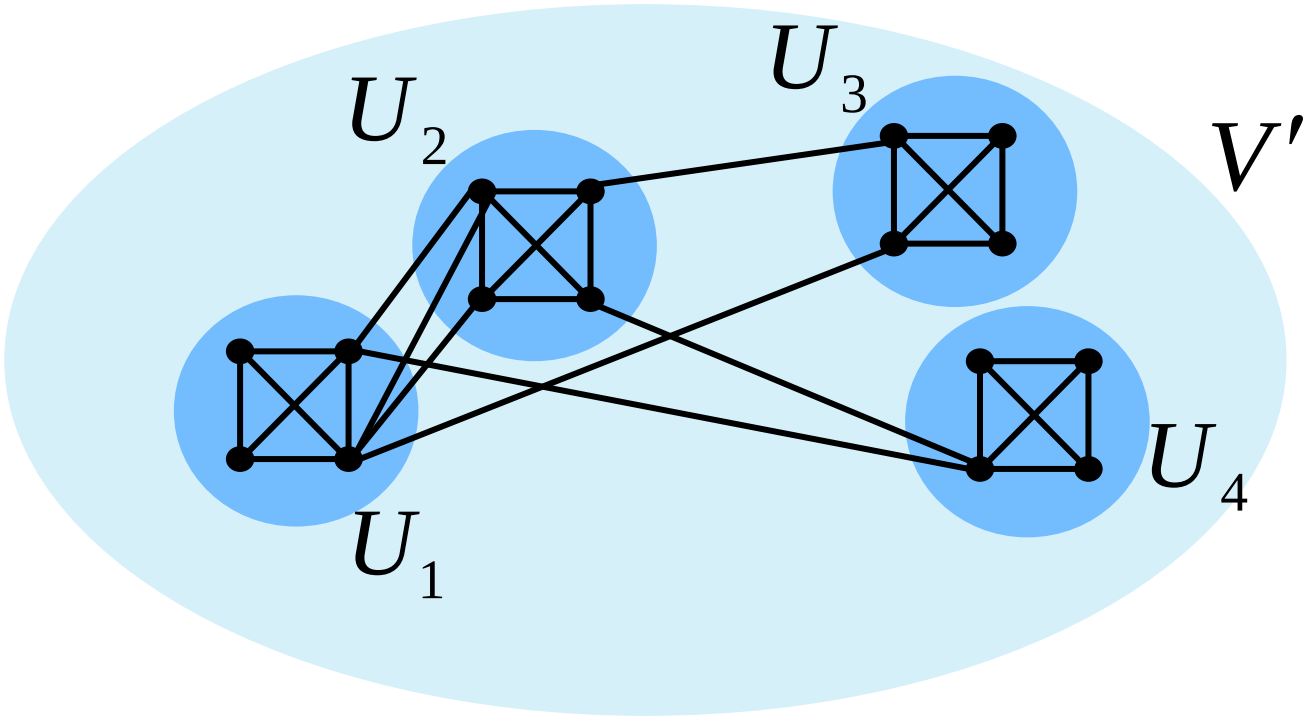
δ-Linked

$$\sum_{i \in I} |U_i| \leq \frac{|V'|}{2} \quad \text{and} \quad \frac{\sum_{i,j \in I} \mathbf{cap}(U_i, U_j)}{\sum_{i \in I, j \notin I} \mathbf{cap}(U_i, U_j)} > \delta$$



$U_1,\ldots,U_k$ is δ-linked if their union is ***small*** &
has ***many internal connections***

δ-Linked (example)

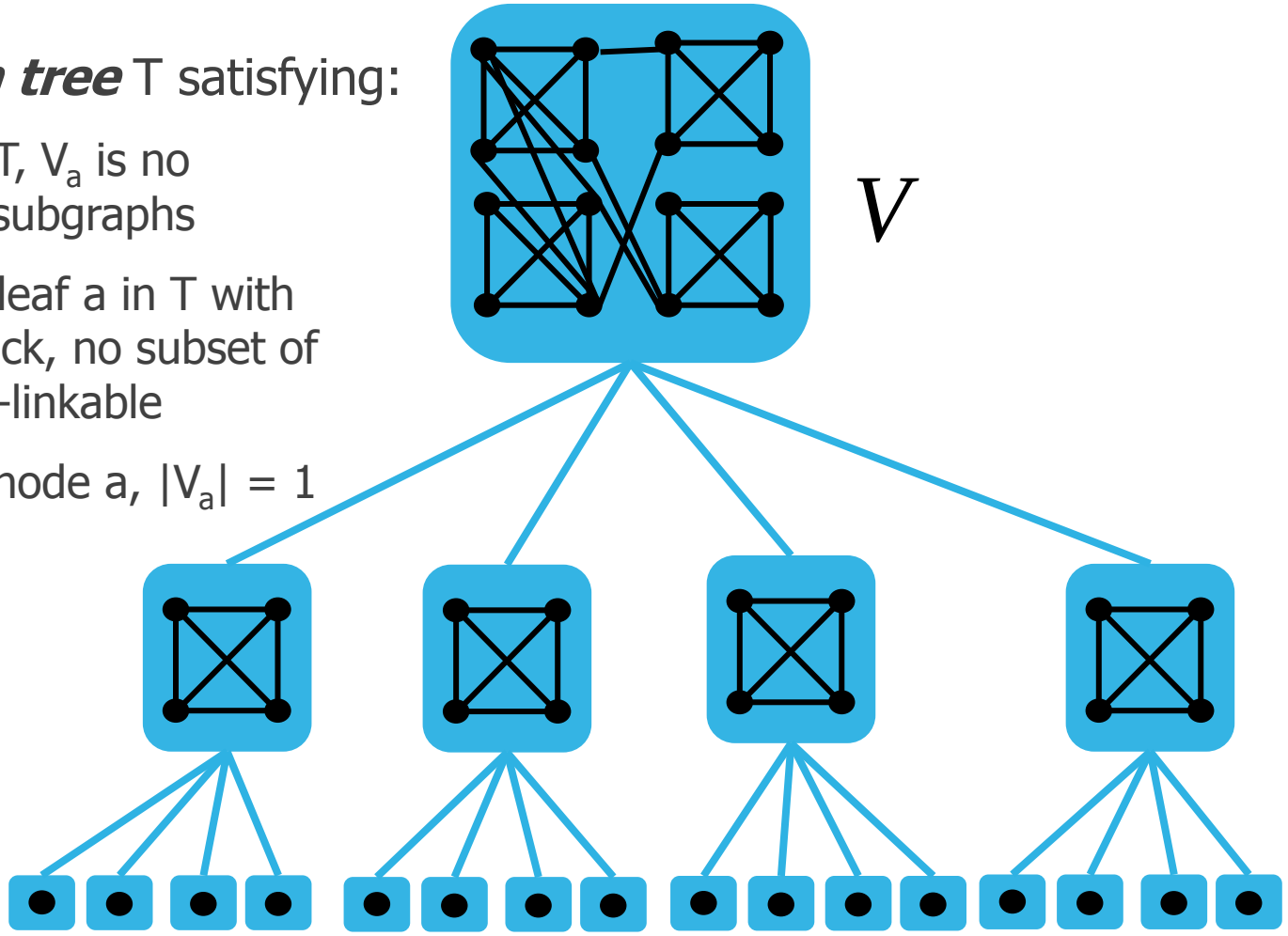


$$|U_1| + |U_2| \leq \frac{|V'|}{2} \quad \text{and} \quad \frac{\sum_{i,j \in I} \mathbf{cap}(U_i, U_j)}{\sum_{i \in I, j \notin I} \mathbf{cap}(U_i, U_j)} = \frac{6}{4}$$

(λ ,δ)-Hierarchical Community Decomposition

A ***decomposition tree*** T satisfying:

1. For each a in T, V_a is no λ-detachable subgraphs
2. For each non-leaf a in T with children c1,…,ck, no subset of V_{c1},…, V_{ck} is δ-linkable
3. For each leaf node a, $|V_a| = 1$



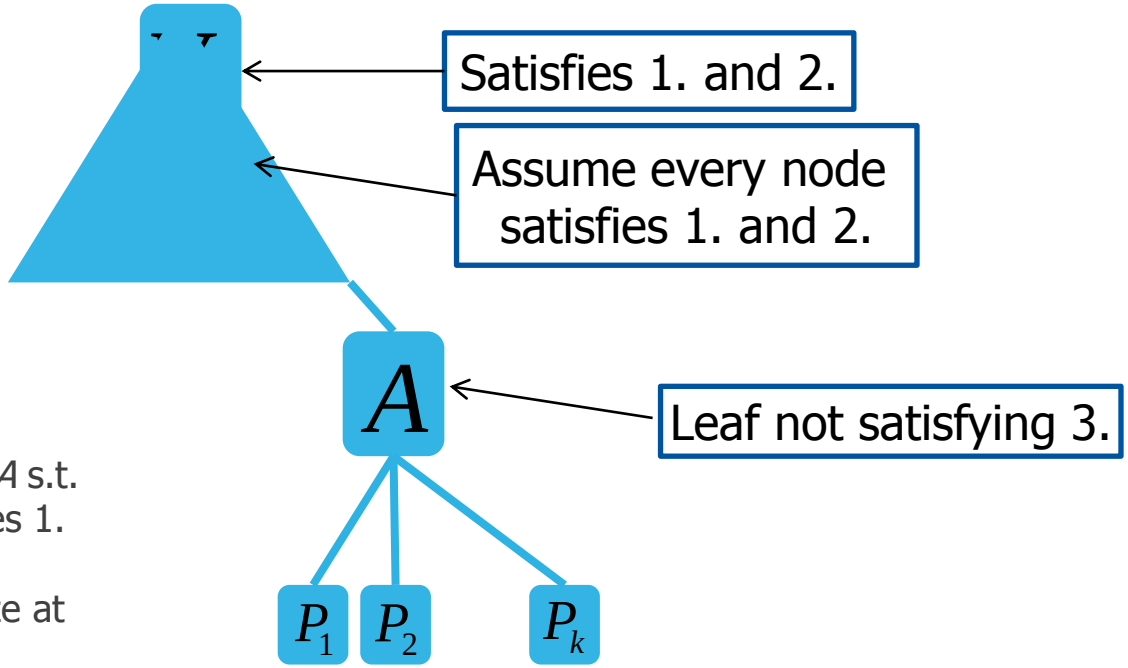
Finding (λ ,δ)-HCDs

(λ ,δ)-HCD is a tree T s.t.:

1. For all a, V_a is no λ-detachable sugraphs
2. Non-leafs: no subset of children is δ-linkable
3. Leaves: $|V_a| = 1$

[Räcke 02] There exists a $(O(\log n), O(\log n))-HCD$ tree T of height at most $\log n$.

Theorem : There exists a $(O(\log n), O(1))-HCD$ tree T.



Find partition $P_1,\ldots, P_k$ of A s.t.

i. Each P_i satisfies 1.
ii. A satisfies 2.
iii. Each P_i has size at most $|A|/2$

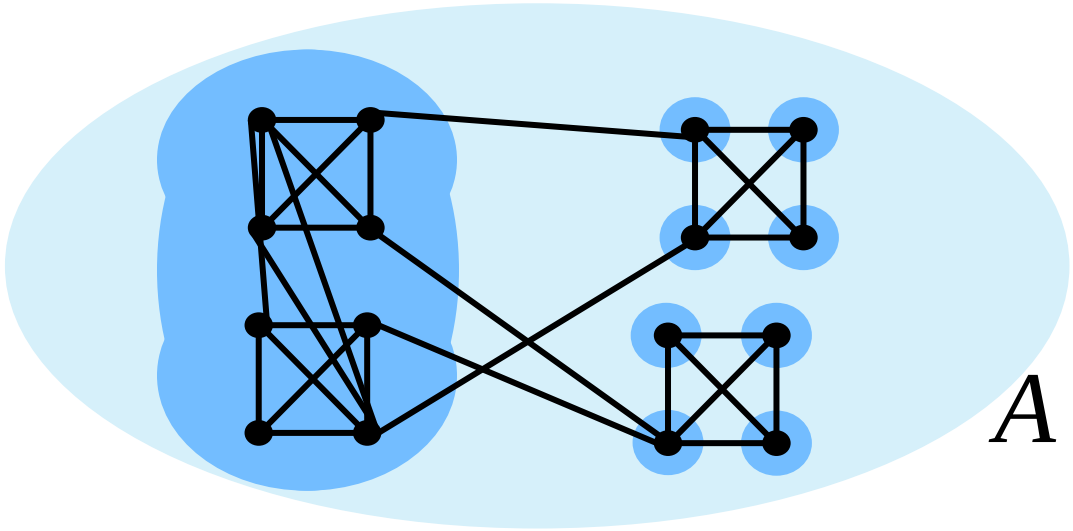
Finding (λ ,δ)-HCDs continued

(λ ,δ)-HCD:

1. For all a, V_a is no λ-detachable subgraphs
2. Non-leafs: no subset of children is δ-linkable
3. Leaves: $|V_a| = 1$

Find partition $P_1,\ldots, P_k$ of A s.t.

i. Each P_i satisfies 1.
ii. A satisfies 2.
iii. Each P_i has size at most $|A|/2$



ALG 1: while subset S of $P_1,\ldots, P_k$ is δ-linkable:

- Group parts of S
- Ensure S satisfies 1. (ALG 2)

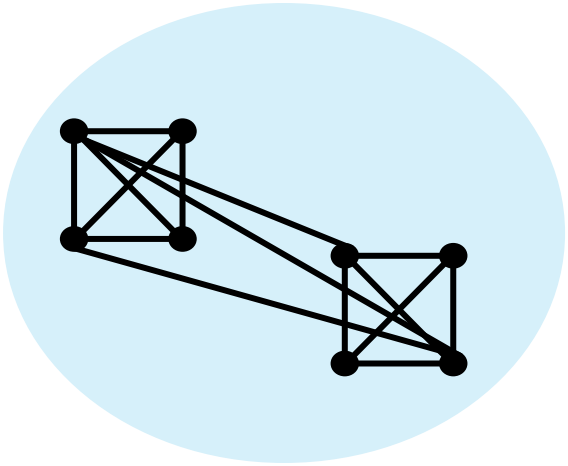
ALG 2: while subset S′ of S is λ-detachable:

- Replace S with S′ and S \ S′
- Ensure S′ and S \ S′ satisfy 1. (ALG 2)

Determining if subset S' of S is λ-detachable

(Determining if subset S of $P_1, \ldots, P_k$ is δ-linkable is similar.)

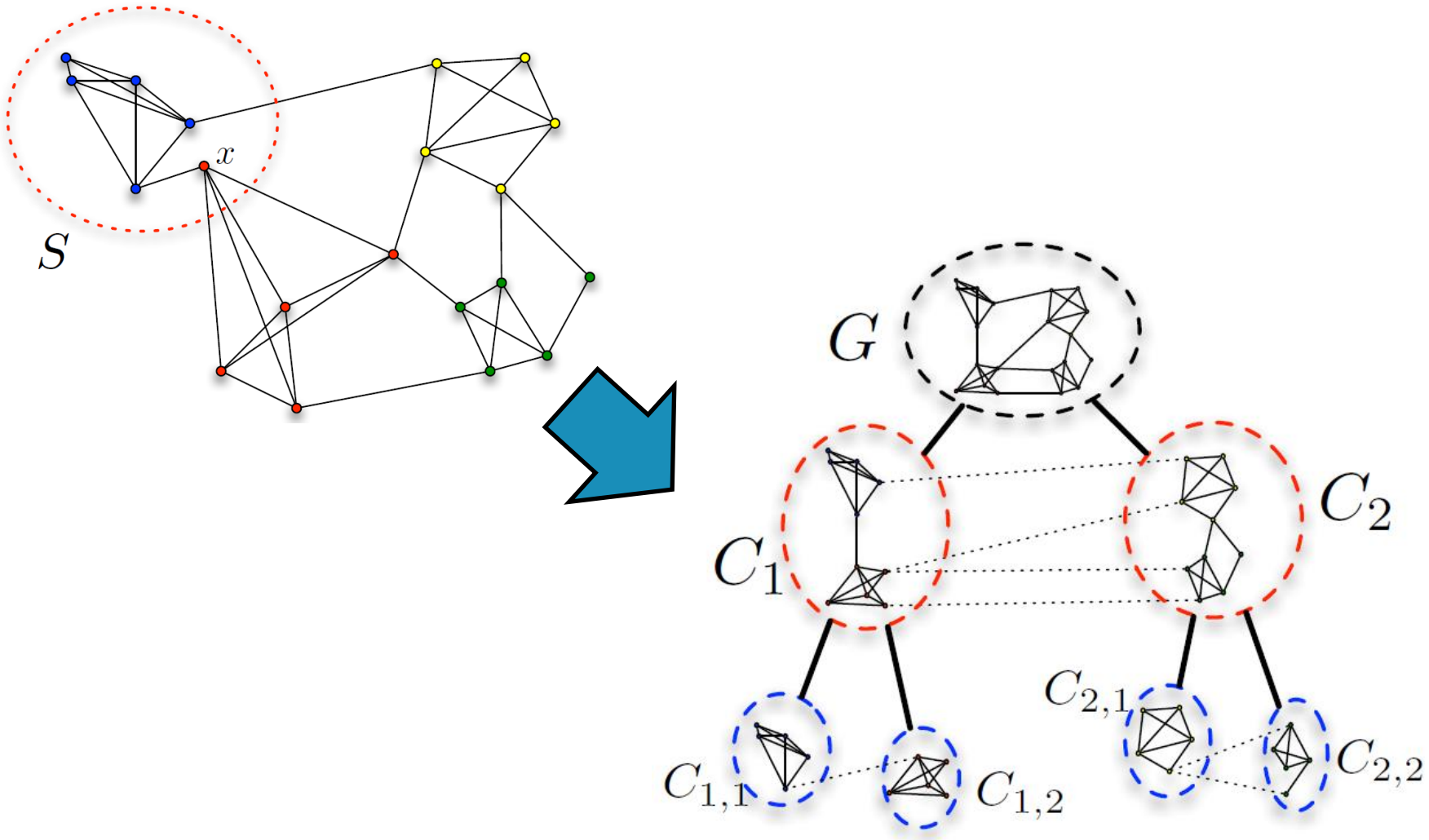
Räcke: Try all subsets
(subsequently improved for routing [Harrelson *et al.* 2003] [Bienkowski *et al.* 2003]



S

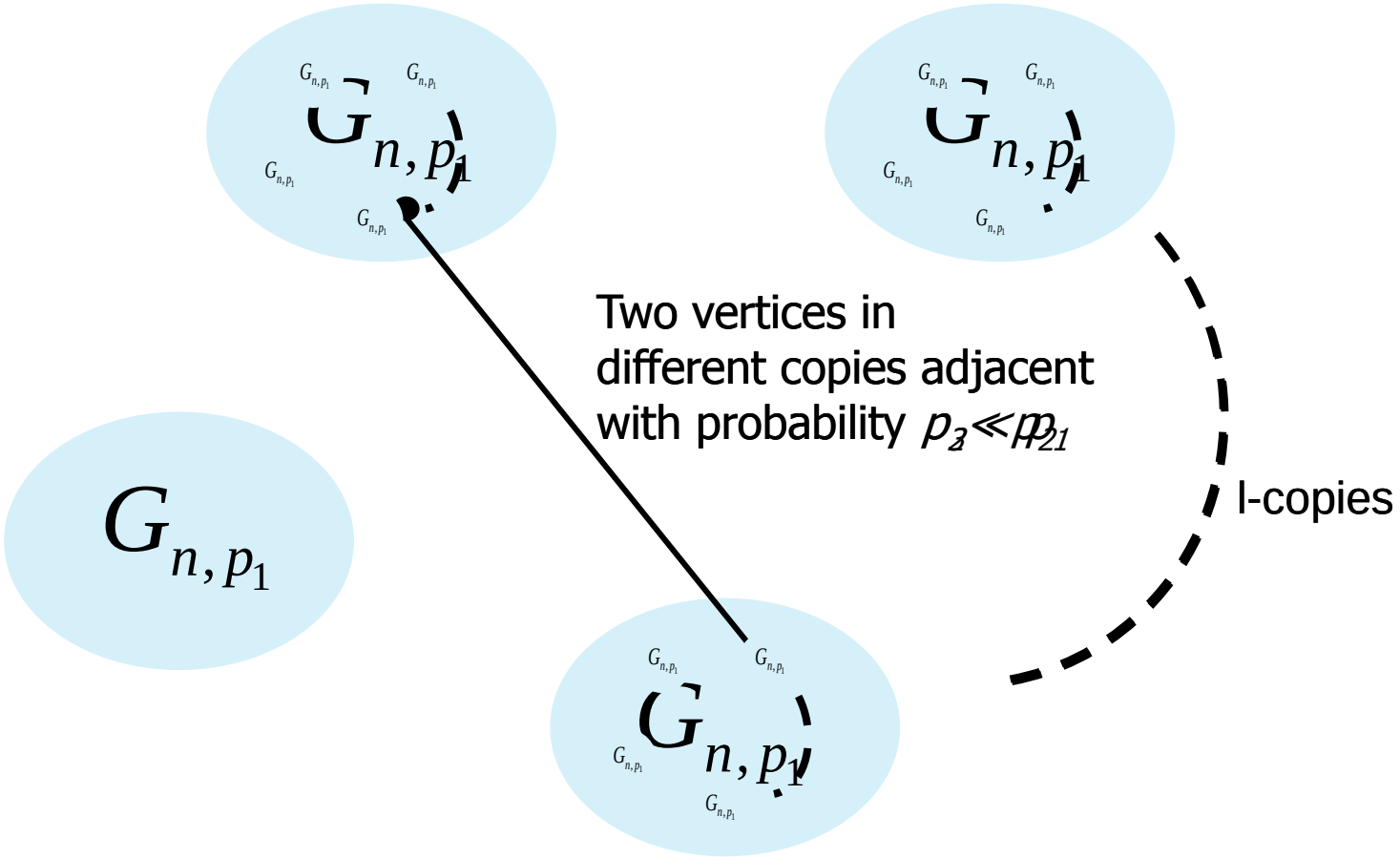
Our Heuristic: Check a fixed number of spectral cuts.

Experiments (fixing a spectral cut)



Generating Synthetic Hierarchical Networks

k-Level Hierarchical Planted l-Partition model → (k,l)-HPPM

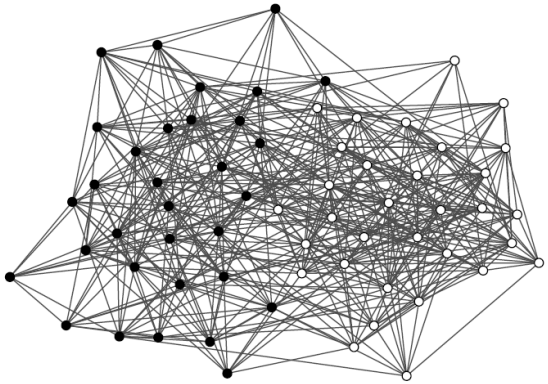


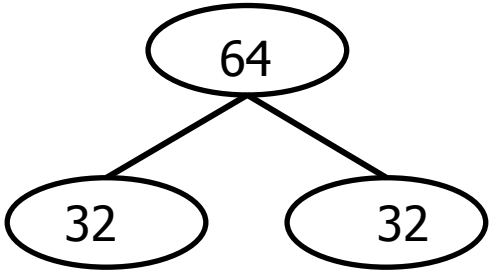
Planted l-Partition model [Codon, Karp 01]

Recursively apply k-times for $p_k < \ldots < p_2 < p_1$

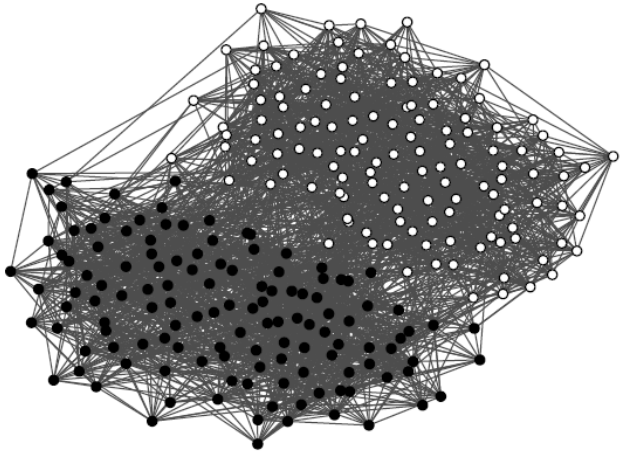
Experiments (Easy cases)

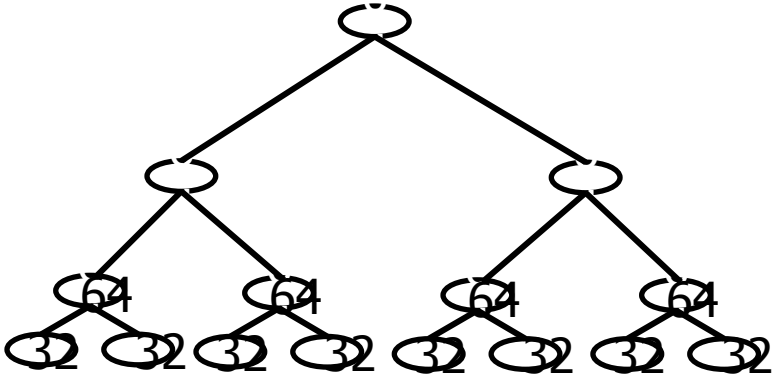
(1,2)-HPPM





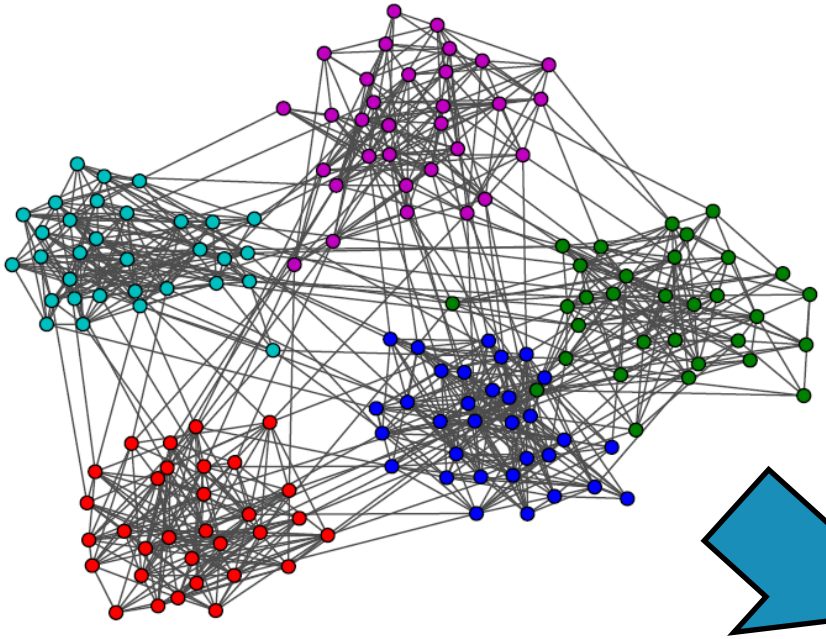
(3,2)-HPPM



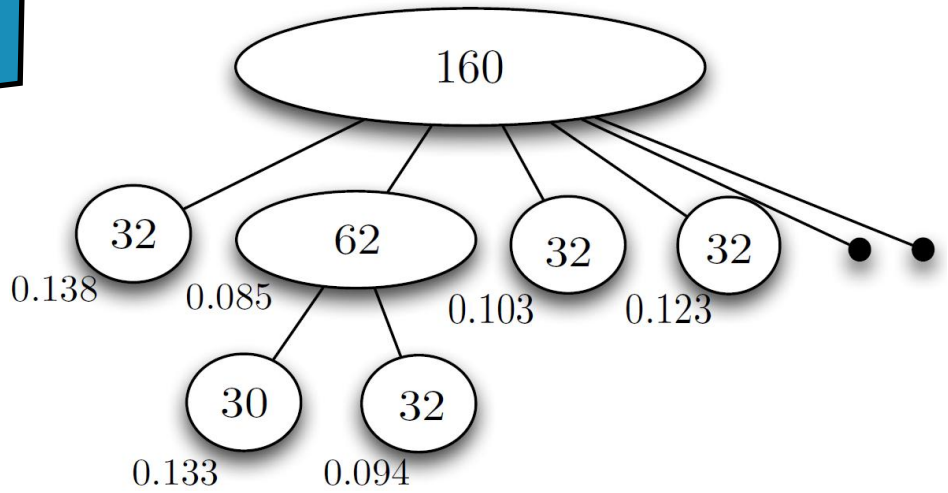


In fact, (*k*,2)-HPPM and (*k*,3)-HPPM are easy for all *k*

Experiments continued



(1,5)-HPPM

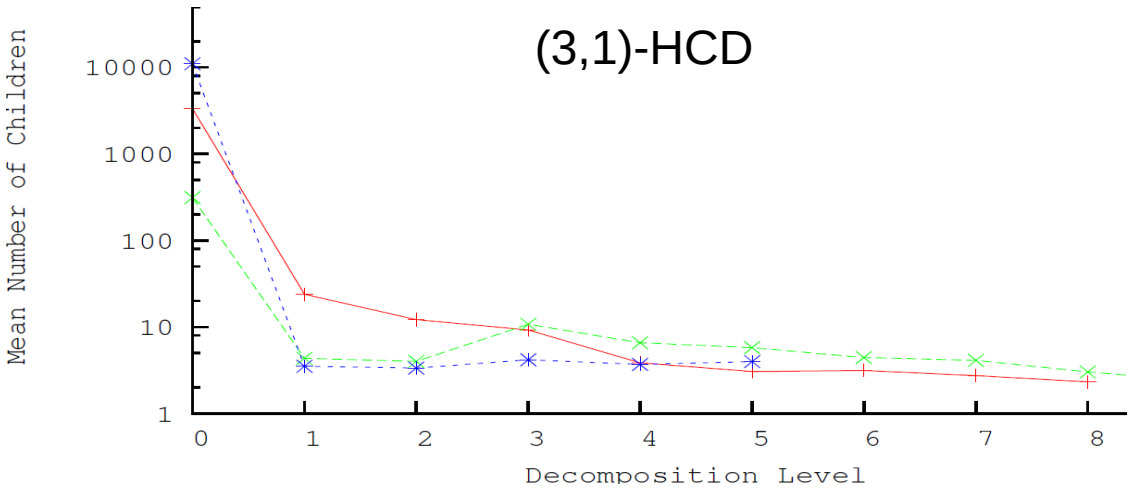


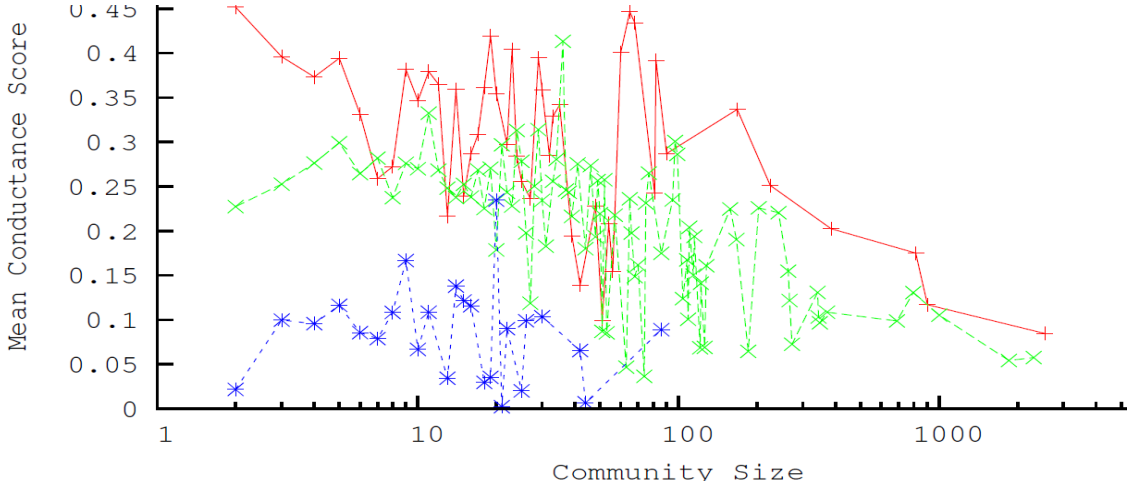
Real-world Network Experiments

Facebook
N= 6K,E=31K

arXiv CA-HepPh
N=12K,E= 118K

arXiv CA-GrQc
N=5K,E=14K





Future Directions

a. How best to choose λ and δ?

b. Trade-offs between λ and δ?

$$\text{Theorem}: \text{There exists a } (\lambda, \delta) - HCD \text{ tree T, where}$$

$$\lambda \geq 4\log \text{n}$$

$$\delta \geq (1 - \frac{2}{\lambda}\log n)^{-1}$$

c. Scalability

- choice of δ
- Local spectral methods, etc.