p-Norm Flow Diffusion for Local Graph Clustering

Kimon Fountoulakis[1], Di Wang[2], **Shenghao Yang**[1]

[1]University of Waterloo [2]Google Research

IOS 2022

UNIVERSITY OF WATERLOO

Google Research

Sublinear-time Coordinate Descent Algorithm
~~p-Norm Flow Diffusion~~
for Local Graph Clustering

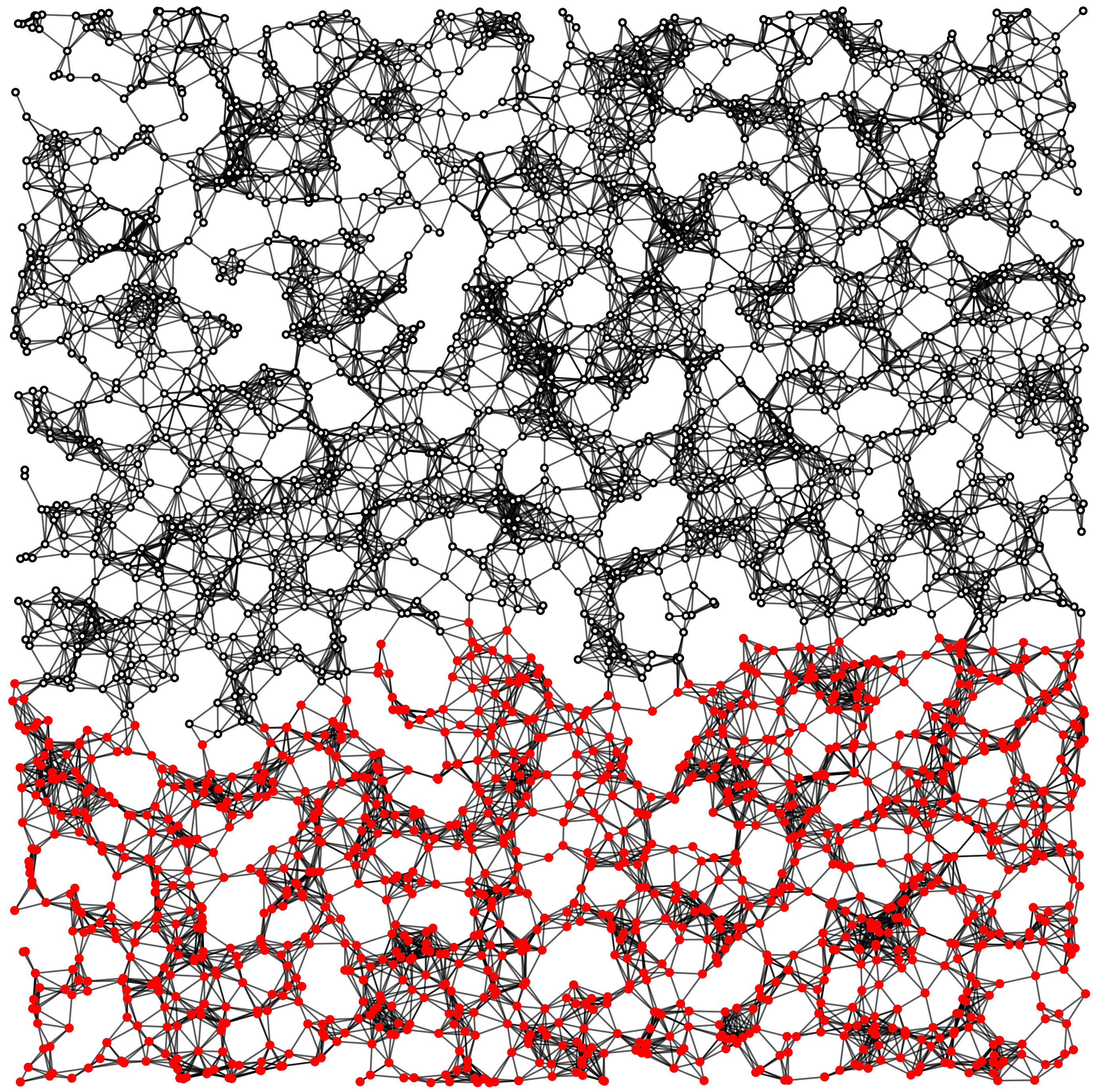
Kimon Fountoulakis[1], Di Wang[2], **Shenghao Yang**[1]

[1]University of Waterloo [2]Google Research

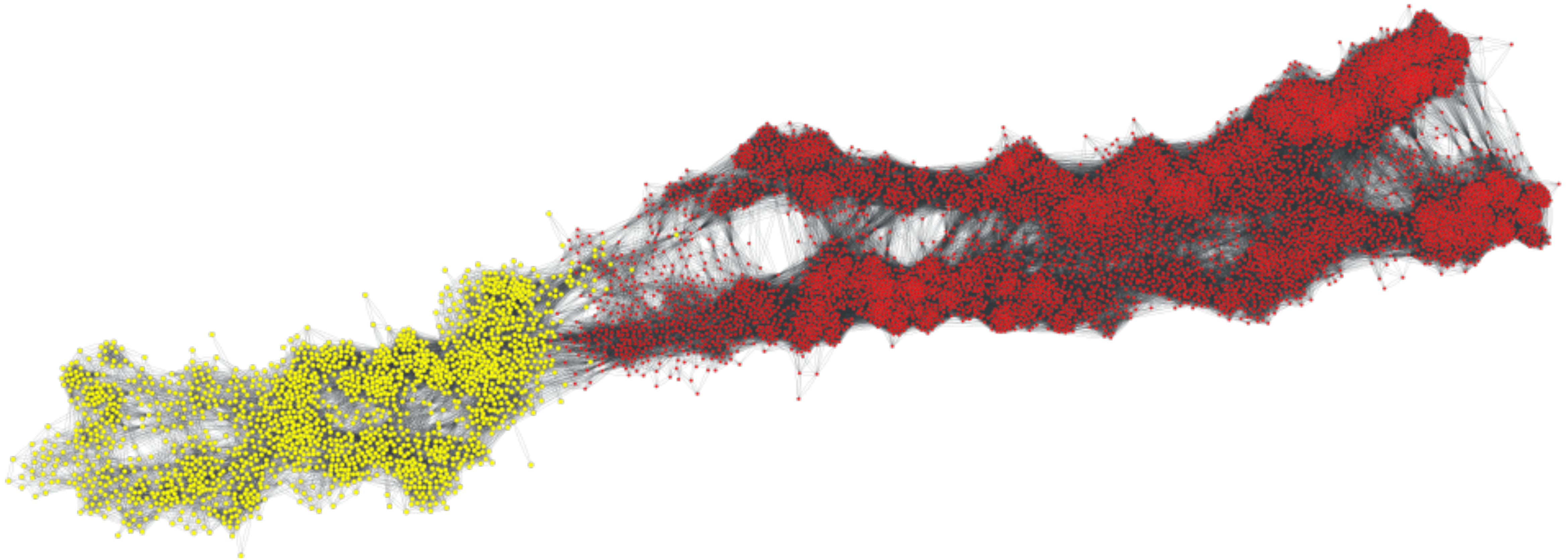
IOS 2022

Motivation: detection of small clusters in large and noisy graphs

- Graph clustering partitions nodes into closely connected clusters



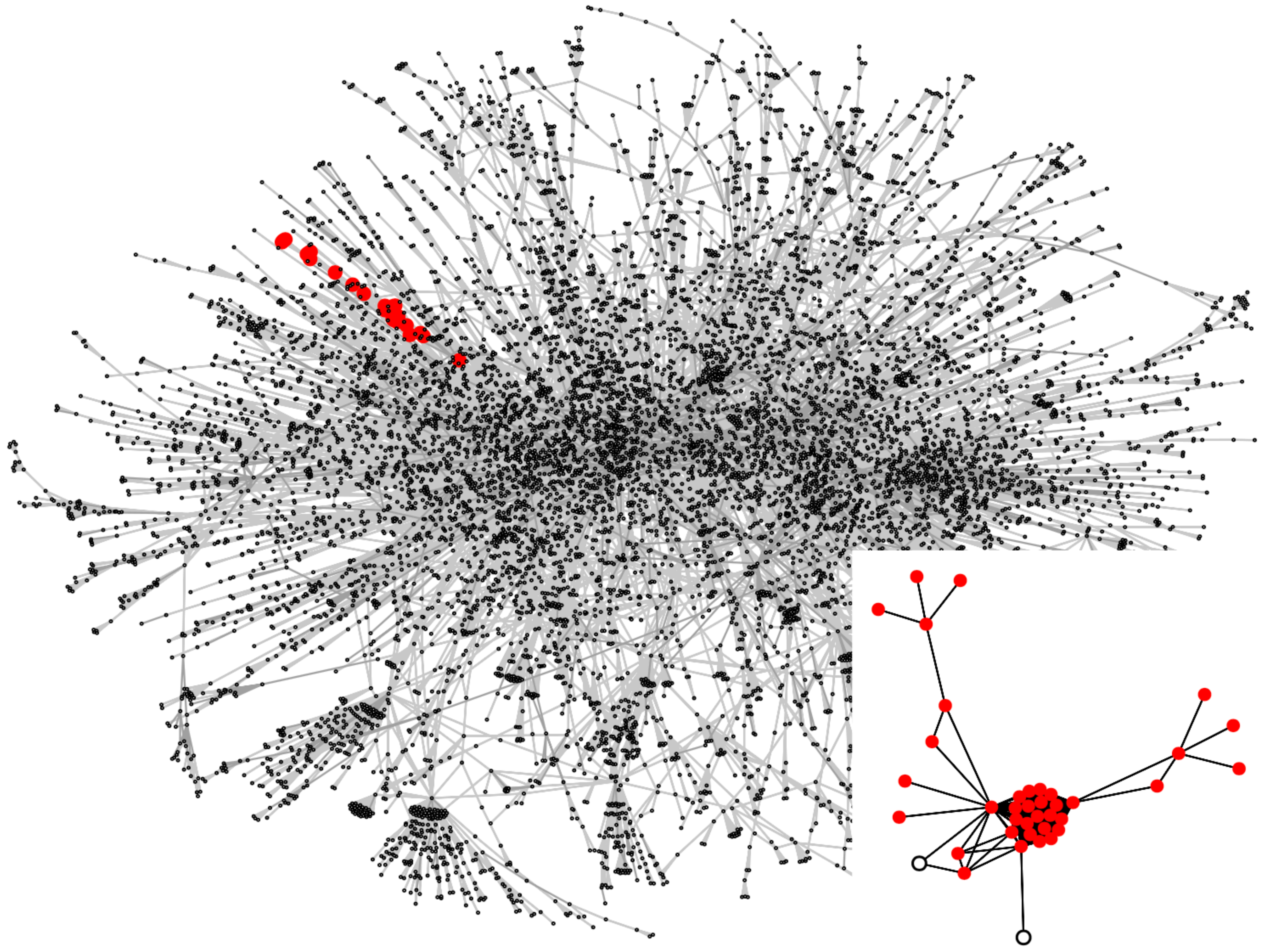
A random geometric graph
partitions into two well-balanced pieces



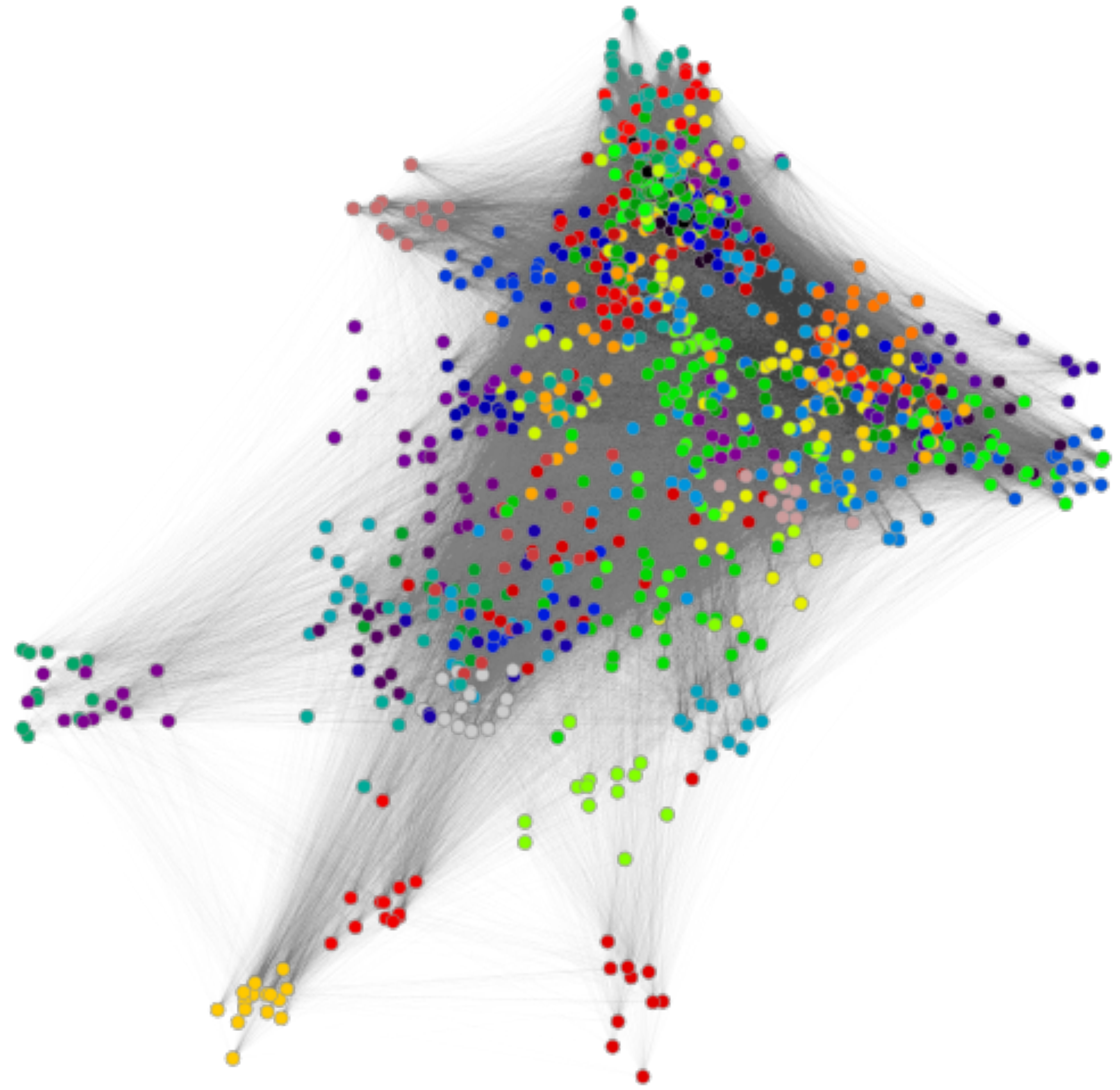
US-Senate graph,
nice bi-partition in year 1865 around the end of
the American civil war

Motivation: detection of small clusters in large and noisy graphs

- Graph clustering partitions nodes into closely connected clusters
- Most real-world graphs lack nice global structures - they have rich local structures



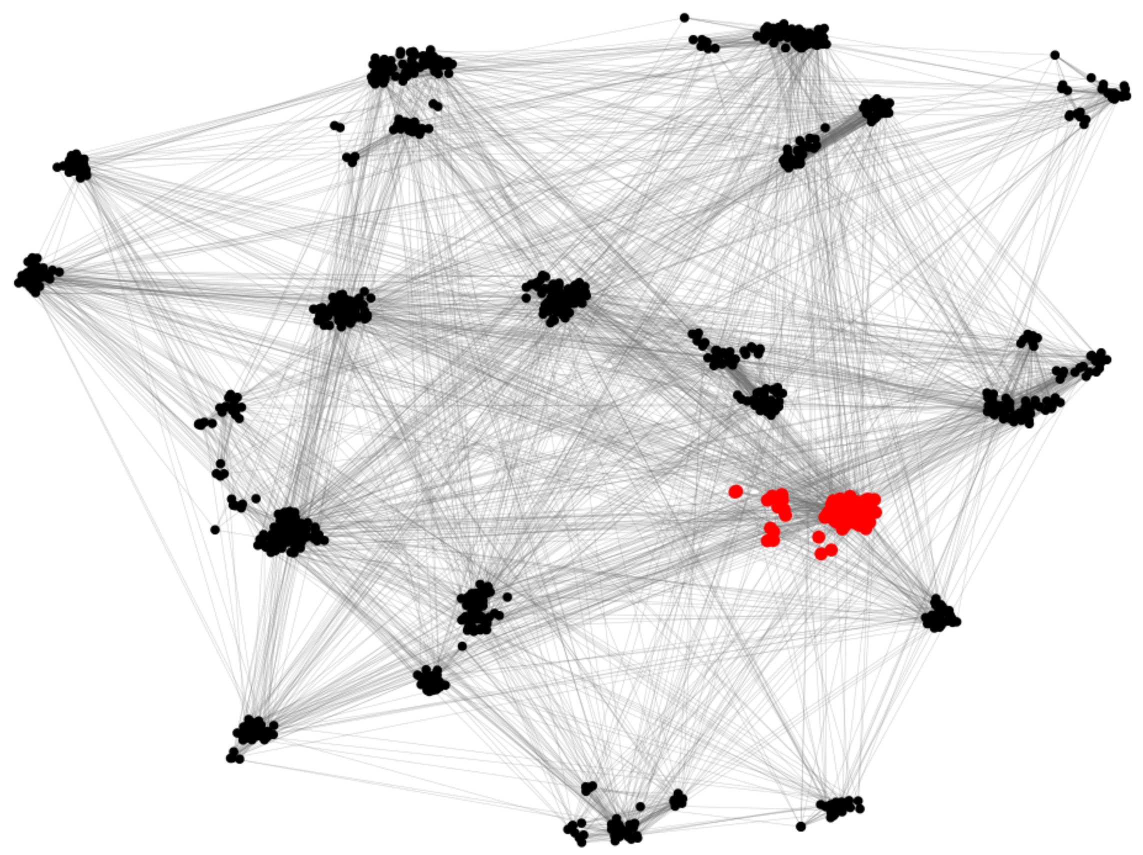
A typical real-world graph
has a classic hairball layout



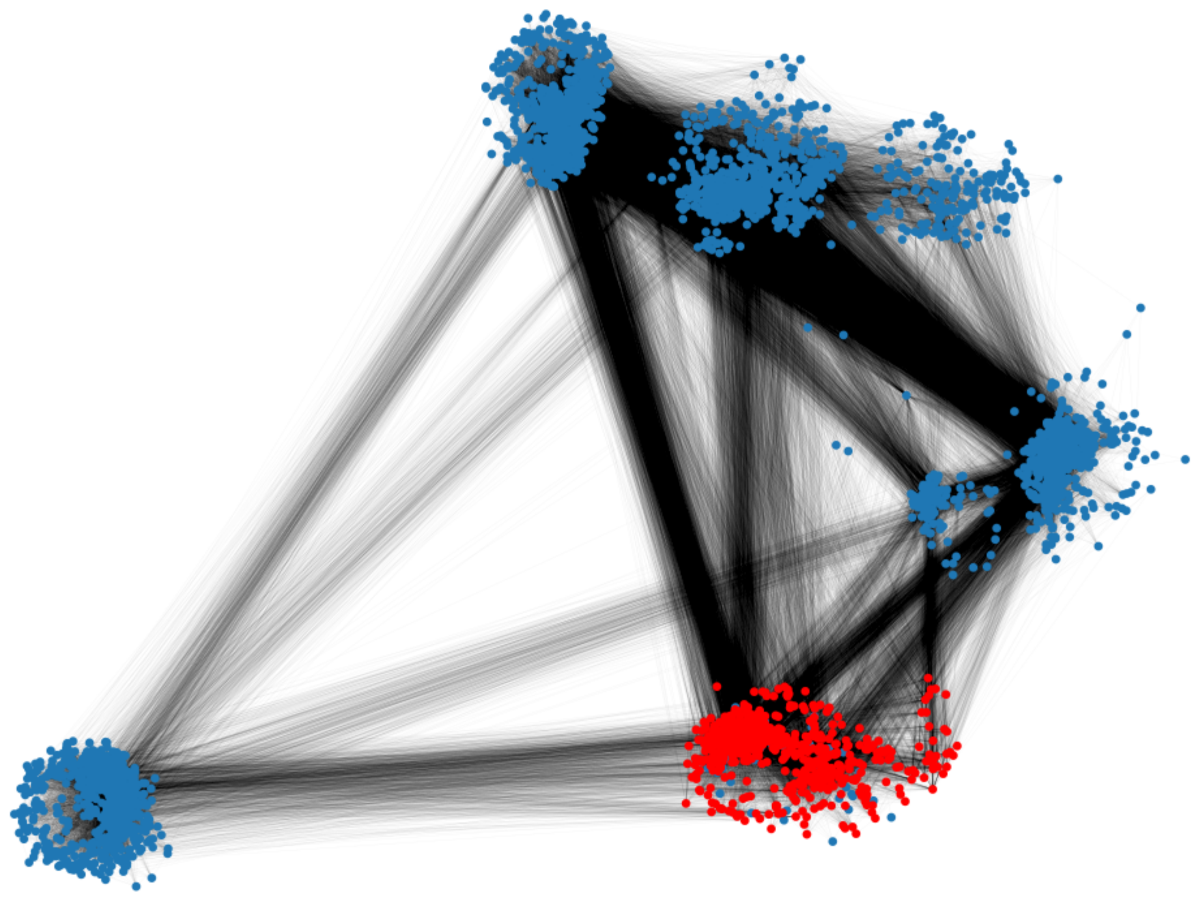
Protein-protein interaction graph,
color denotes similar functionality

Our goals: local algorithm with good theoretical guarantees

- Local graph clustering
  - we often have to detect small-scale clusters in large and noisy graphs
  - given a small set of seed nodes, identify a good cluster around it



LFR benchmark, red nodes
form a target cluster

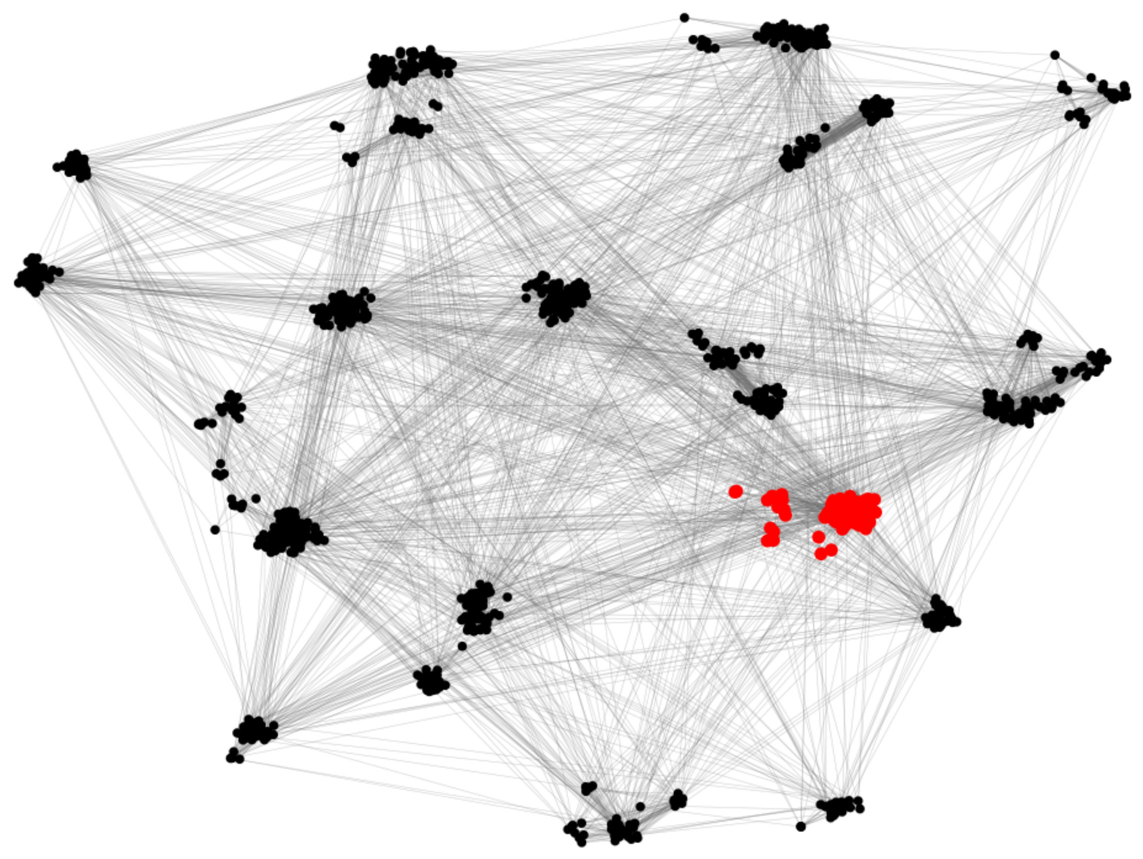


Facebook friendship graph at a liberal arts collage,
red nodes are students of year 2008

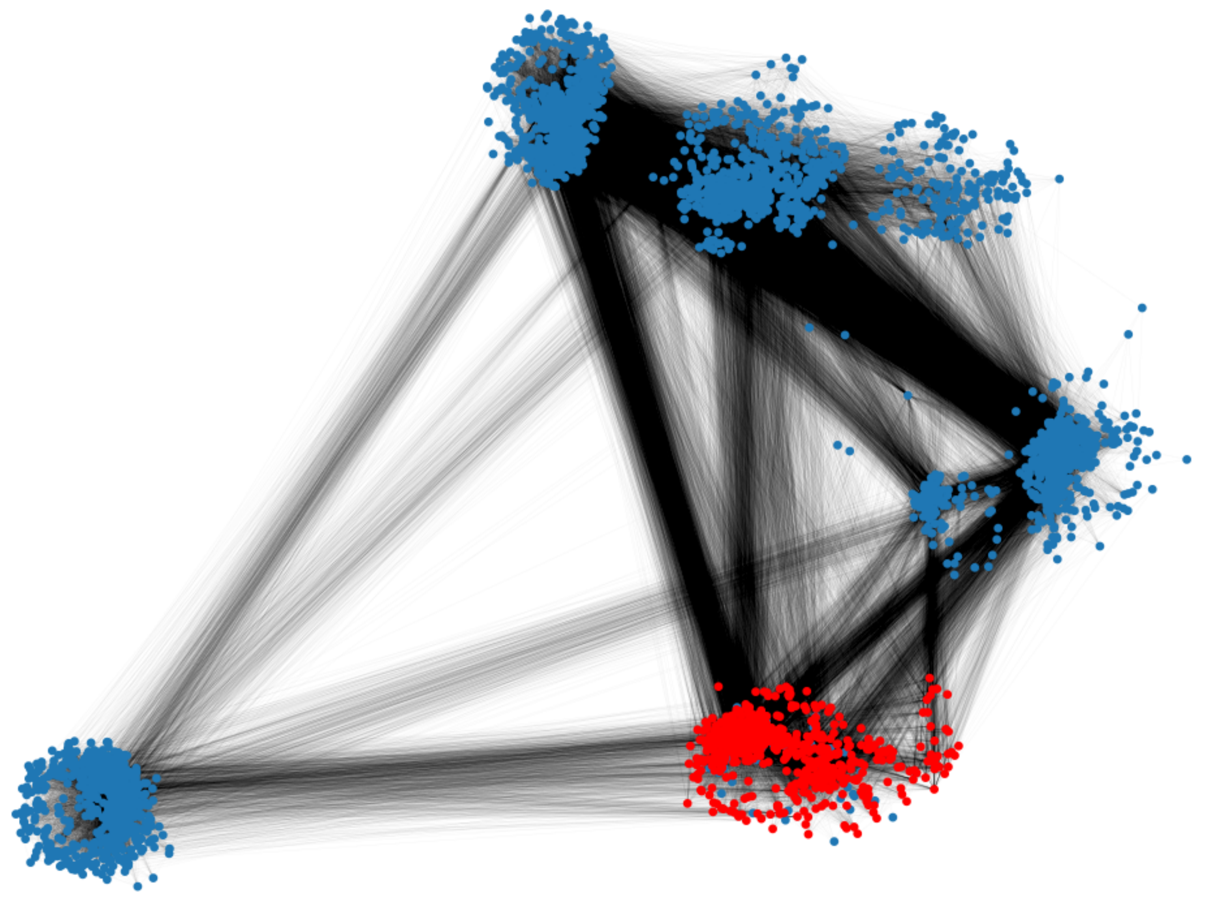
Our goals: local algorithm with good theoretical guarantees

- Local graph clustering
  - we often have to detect small-scale clusters in large and noisy graphs
  - given a small set of seed nodes, identify a good cluster around it

- This requires new algorithms that
  - (local) has a running time that only depends on the size of the output
  - (simple) has fewer tuning parameters, is easy to implement
  - (tight) is supported by good approximation guarantees



LFR benchmark, red nodes form a target cluster



Facebook friendship graph at a liberal arts collage, red nodes are students of year 2008

Existing local graph clustering methods

Spectral diffusions

based on the dynamics of
random walks

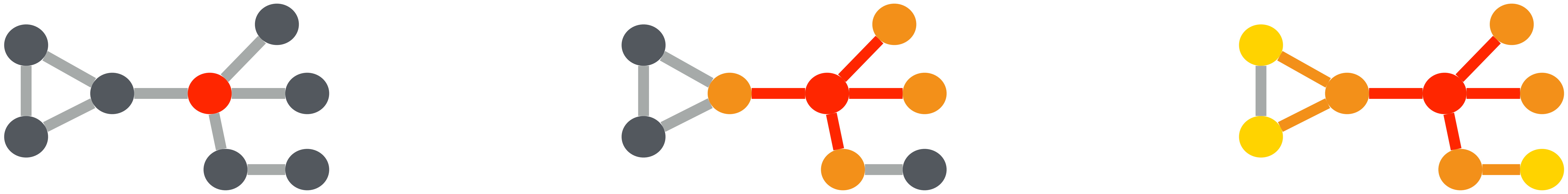
e.g., Approx. PageRank [ACL 2006]

Combinatorial diffusions

based on the dynamics of
maximum flows

e.g., Capacity Releasing Diffusion [WFH+ 2017]

- Diffusion as a physical phenomenon: paint spills, spreads and settles



1 ***2*** ***3***

Existing local graph clustering methods

Spectral diffusions

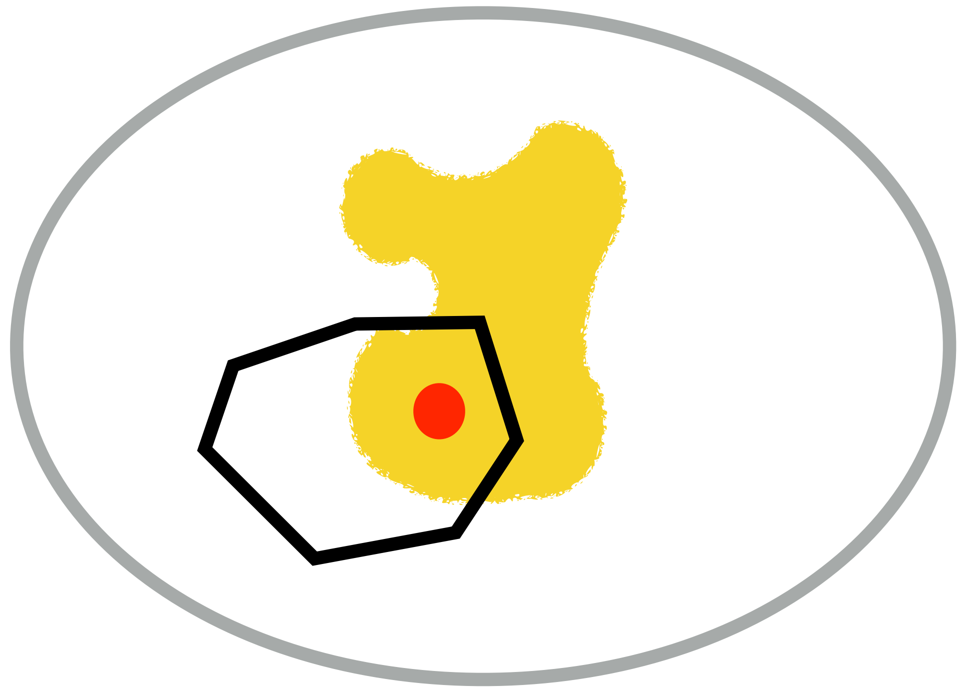
based on the dynamics of
random walks

e.g., Approx. PageRank [ACL 2006]

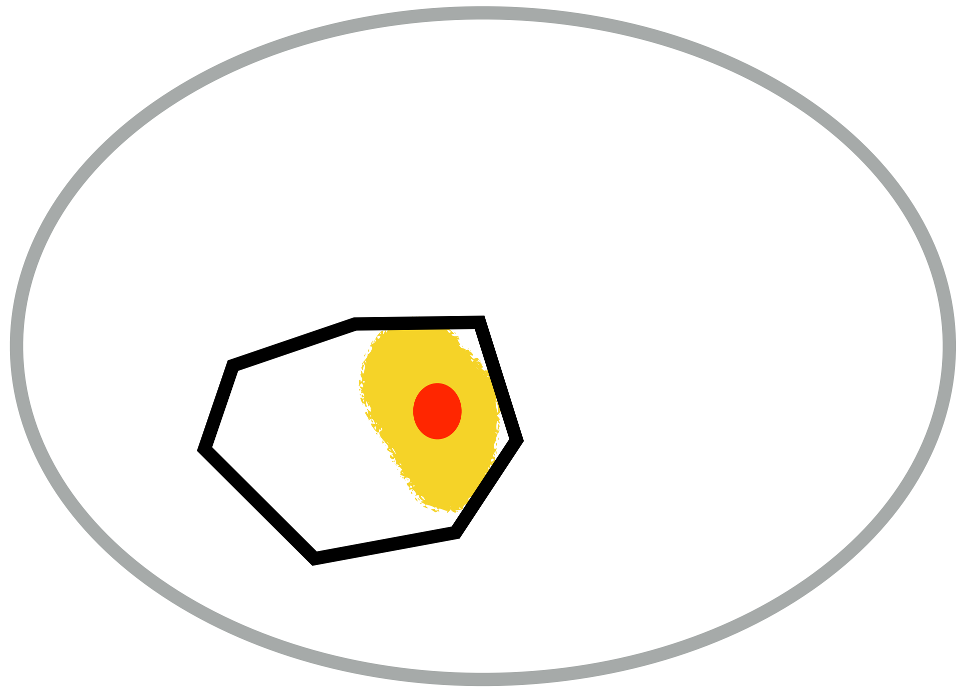
Combinatorial diffusions

based on the dynamics of
maximum flows

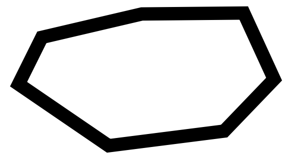
e.g., Capacity Releasing Diffusion [WFH+ 2017]



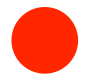
Spectral diffusions leak mass



Combinatorial diffusions are
hard to tune



target cluster

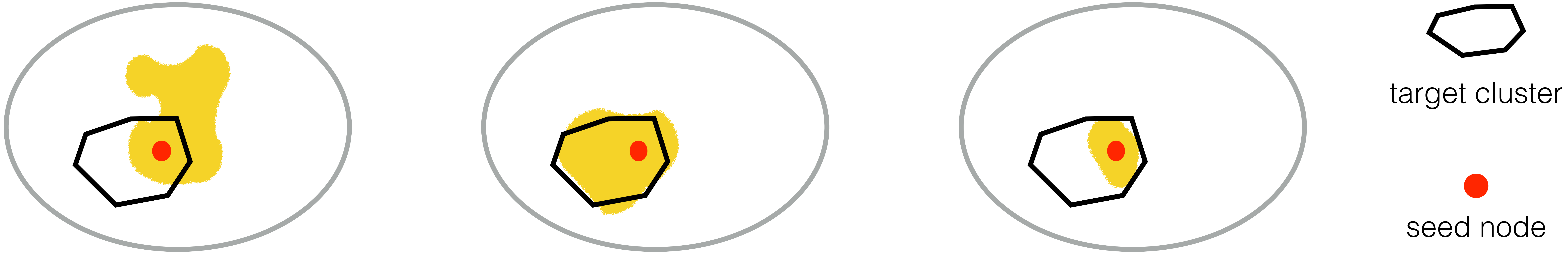


seed node

Existing local graph clustering methods

Spectral diffusions **Combinatorial diffusions**

p-Norm flow diffusions
based on the idea of
minimizing p-norm network flows



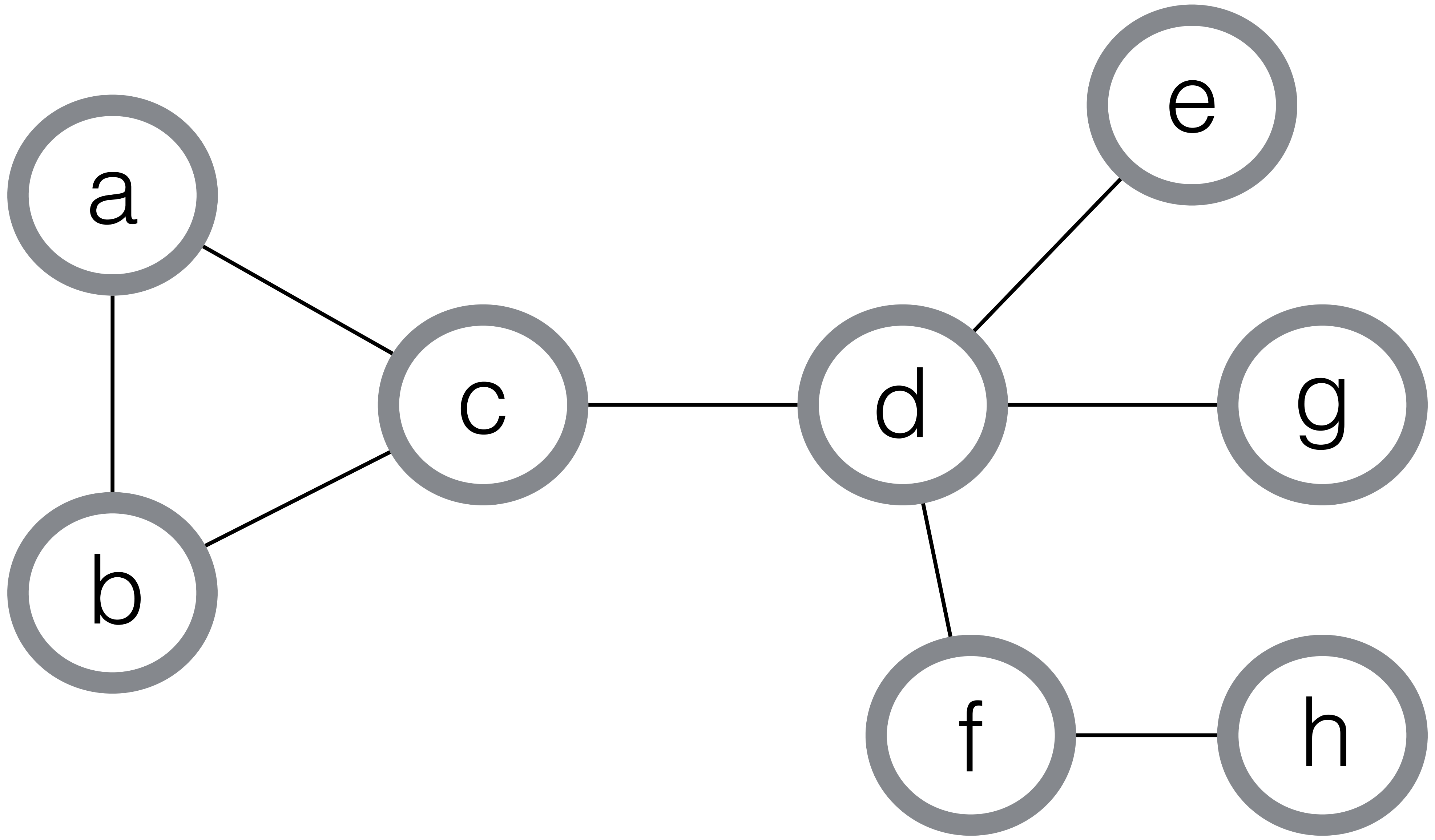
Spectral diffusions leak mass

p-Norm flow diffusion combines the best of both worlds

Combinatorial diffusions are hard to tune

Notations and definitions

- Undirected graph $G = (V, E)$



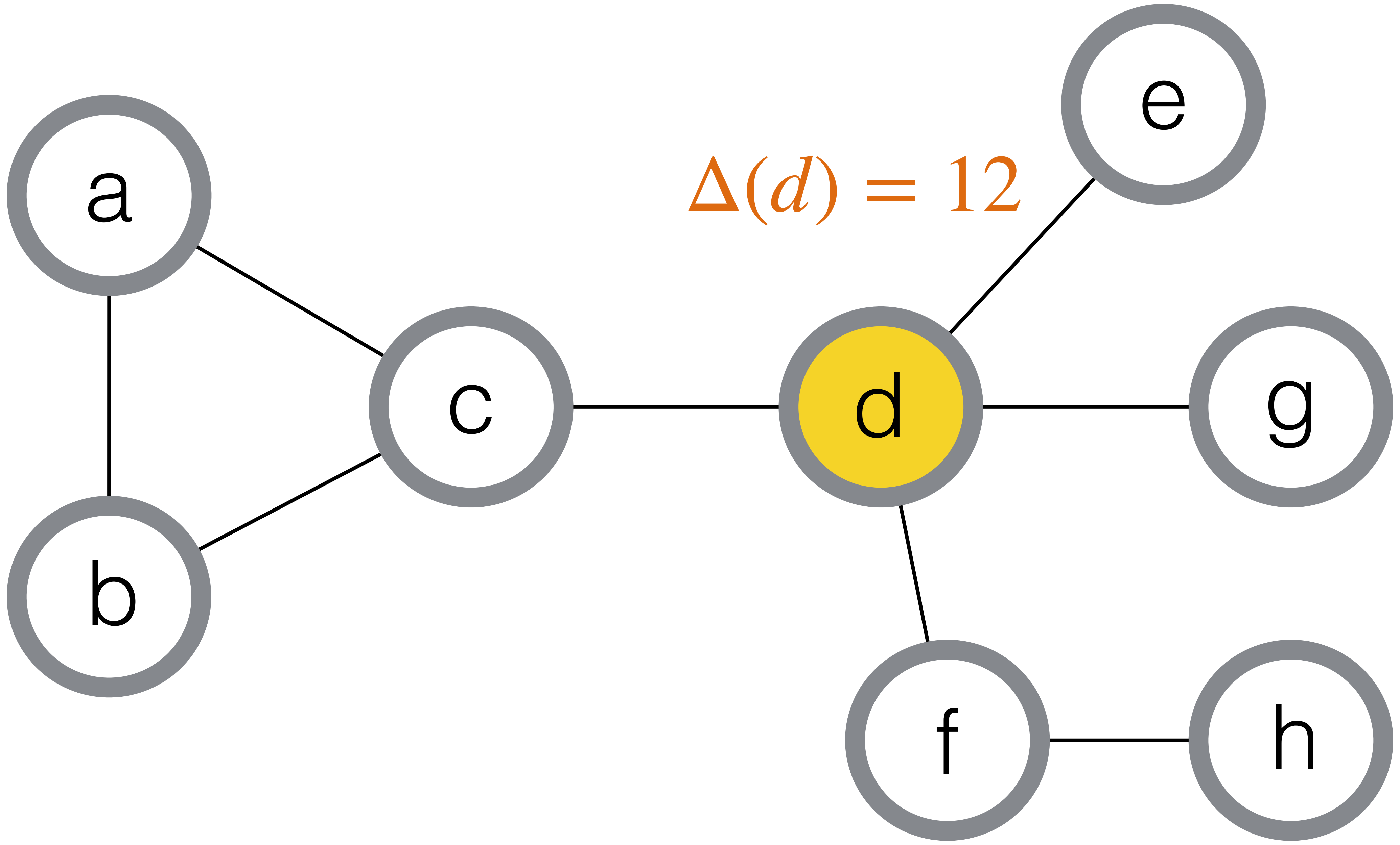
Incidence matrix B

	a	b	c	d	e	f	g	h
(a,b)	1	-1						
(a,c)	1		-1					
(b,c)		1	-1					
(c,d)			1	-1				
(d,e)				1	-1			
(d,f)				1		-1		
(d,g)				1			-1	
(f,h)						1		-1

- B is $|E| \times |V|$ *signed incidence matrix* where the row of edge (u, v) has two non-zero entries, -1 at column u and 1 at column v
- Ordering of edges and direction is arbitrary

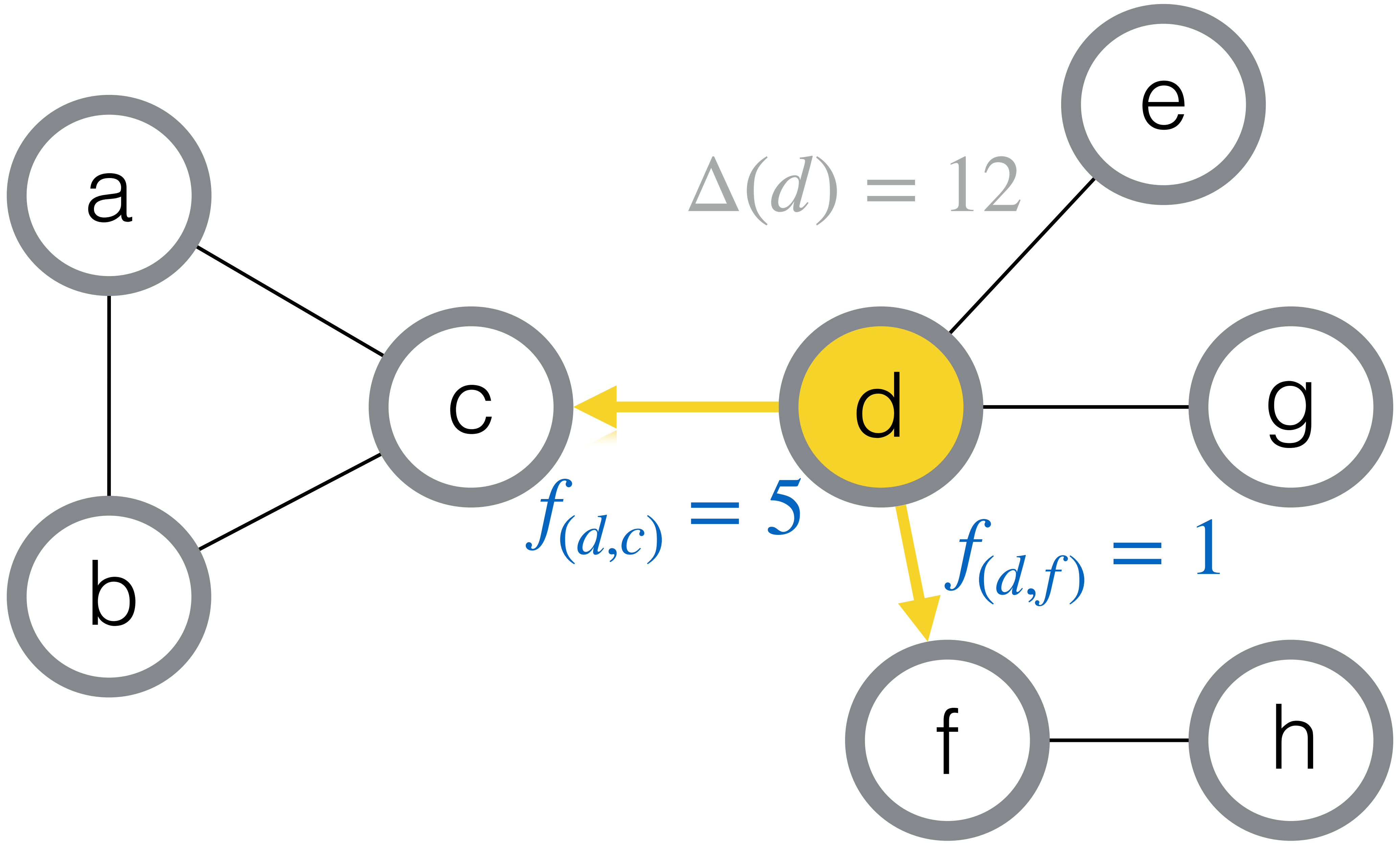
Notations and definitions

- $\Delta \in \mathbb{R}_{+}^{|V|}$ specifies **initial mass** on nodes.



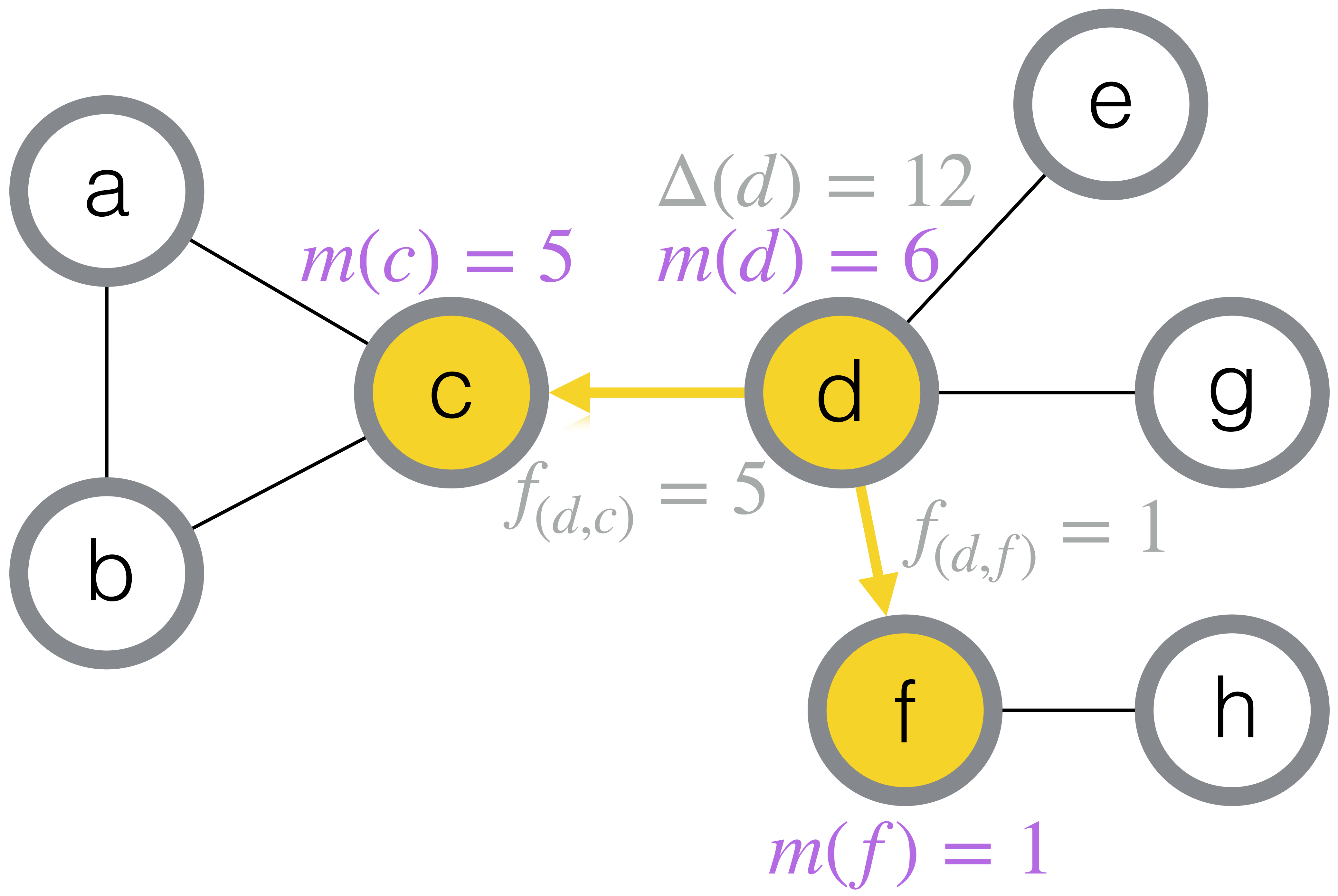
Notations and definitions

- $\Delta \in \mathbb{R}_+^{|V|}$ specifies **initial mass** on nodes.
- $f \in \mathbb{R}^{|E|}$ specifies the **amount of flow**.



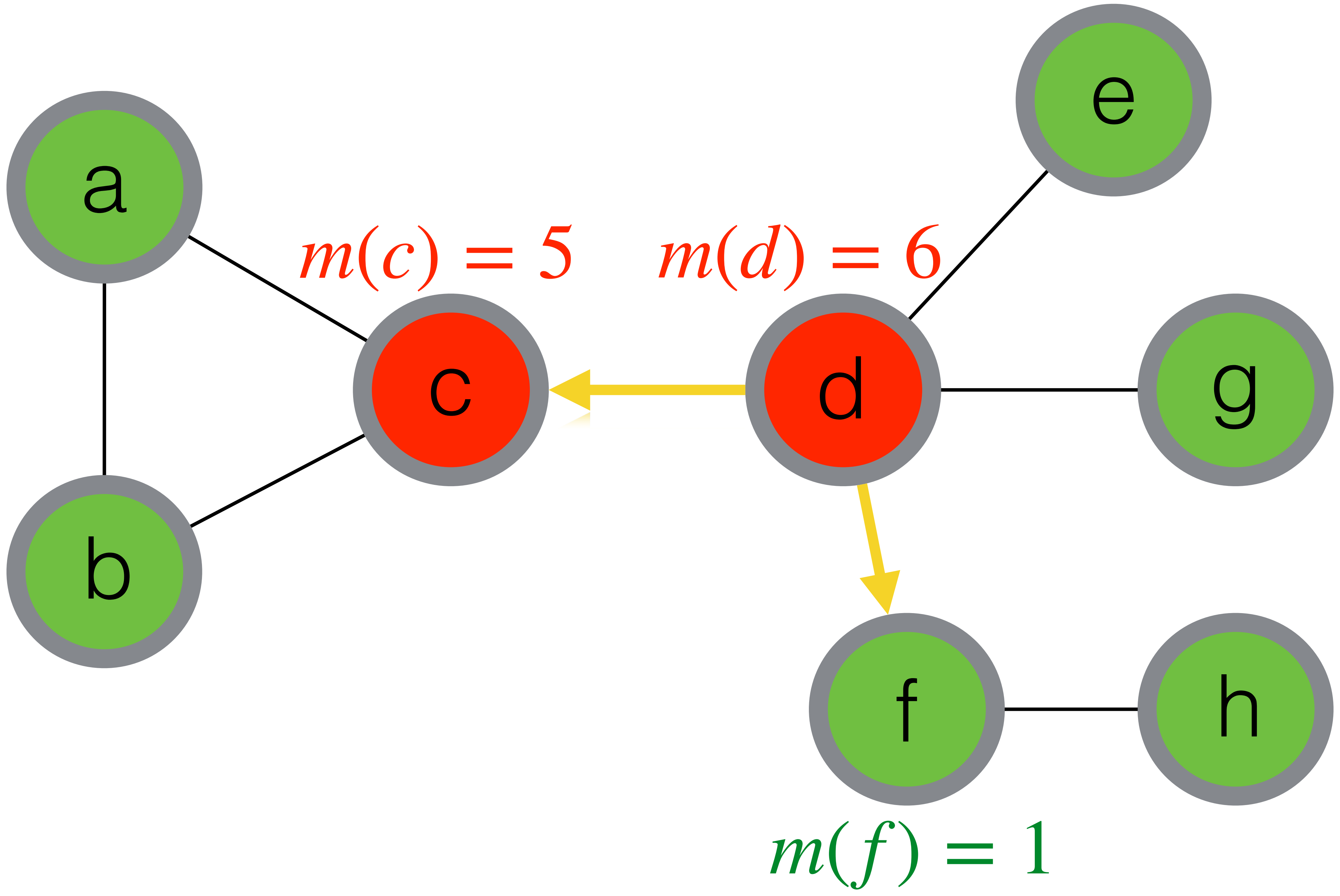
Notations and definitions

- $\Delta \in \mathbb{R}_{+}^{|V|}$ specifies **initial mass** on nodes.
- $f \in \mathbb{R}^{|E|}$ specifies the **amount of flow**.
- $m := B^T f + \Delta$ specifies **net mass** on nodes.



Notations and definitions

- $\Delta \in \mathbb{R}_+^{|V|}$ specifies **initial mass** on nodes.
- $f \in \mathbb{R}^{|E|}$ specifies the **amount of flow**.
- $m := B^T f + \Delta$ specifies **net mass** on nodes.
- Each node v has **capacity** equal to its degree $d(v)$.
- A flow f is **feasible** if $[B^T f + \Delta](v) \leq d(v), \forall v$.



Diffusion as optimization

- Diffusion process on graph as optimization

$$\text{minimize } \|f\|_p^p$$
$$\text{subject to: } B^T f + \Delta \leq d$$

- Out of all feasible flows f, we are interested in the one having minimum p-norm, $p \in [2,\infty)$.
  - Different p-norms lead to diffusions that explore different structures in a graph
- In practice if only one seed node $s \in V$ is given, we set $\Delta = t \cdot \mathbf{1}_s$ where $\mathbf{1}_s$ is the indicator vector of node s, and $t > 0$ is a tuning parameter.

Diffusion as optimization

- Diffusion process on graph as optimization

$$\text{minimize } \|f\|_p^p$$
$$\text{subject to: } B^T f + \Delta \le d$$

- Out of all feasible flows f, we are interested in the one having minimum p-norm, $p \in [2,\infty)$.
  - Different p-norms lead to diffusions that explore different structures in a graph
- In practice if only one seed node $s \in V$ is given, we set $\Delta = t \cdot \mathbf{1}_s$ where $\mathbf{1}_s$ is the indicator vector of node s, and $t > 0$ is a tuning parameter.
- The optimal solution is sparse: $\text{nnz}(f^*) \le \|\Delta\|_1$
- The total running time of an optimization algorithm should depend on $\text{nnz}(f^*)$, not $\text{len}(f^*)$.
  - In local clustering applications, usually $\text{nnz}(f^*) \ll \text{len}(f^*)$

Diffusion as optimization

- Diffusion process on graph as optimization

$$\text{minimize } \|f\|_p^p$$
$$\text{subject to: } B^T f + \Delta \leq d$$

- The dual problem is (for $1/p + 1/q = 1$)

$$\text{minimize}_{x \geq 0} \; \frac{1}{q}\|Bx\|_q^q - x^T(\Delta - d)$$

- When $q \in (1,2)$, smooth it by replacing $|x_i - x_j|^q$ with $\left((x_i - x_j)^2 + \mu^2\right)^{q/2}$

Diffusion as optimization

- Diffusion process on graph as optimization

$$\text{minimize } \|f\|_p^p$$
$$\text{subject to: } B^T f + \Delta \leq d$$

- The dual problem is (for $1/p + 1/q = 1$)

$$\text{minimize}_{x \geq 0} \frac{1}{q}\|Bx\|_q^q - x^T(\Delta - d)$$

- When $q \in (1,2)$, smooth it by replacing $|x_i - x_j|^q$ with $\left((x_i - x_j)^2 + \mu^2\right)^{q/2}$
  - Lipschitz coordinate gradient with parameter $d(i)\mu^{q-2}$
  - Strongly convex objective with parameter $\gamma > \left(p-1)\|\Delta\|_1^p\right)^{-1}$ (over a restricted domain)
  - Setting $\mu = O((\epsilon/\|\Delta\|_1)^{1/q})$ coordinate descent finds ϵ-suboptimal solution in time

$$O\left(\frac{\|\Delta\|_1 d_{\max}}{\gamma}\left(\frac{\|\Delta\|_1}{\epsilon}\right)^{2/q-1} \log\frac{1}{\epsilon}\right)$$

Diffusion as optimization

Algorithm 1 Coordinate solver for smoothed dual problem

Initialize: $x_0 = 0$

For $k = 0, 1, 2, \ldots,$ **do**

Set $S_k = \{i \in V \mid \nabla_i F_\mu(x_k) < 0\}$.

Pick $i_k \in S_k$ uniformly at random.

Update $x_{k+1} = x_k - \frac{\mu^{2-q}}{\deg(i_k)} \nabla_{i_k} F_\mu(x_k) e_{i_k}$.

If $S_k = \emptyset$ **then return** x_k.

The only difference from the standard randomized CD

- $0 = x_0 \leq x_1 \leq \cdots \leq x_k \leq x^*$ for all $k \geq 0$
  - A curial property that guarantees convergence and rate of convergence

Diffusion as optimization

Algorithm 1 Coordinate solver for smoothed dual problem

Initialize: $x_0 = 0$

For $k = 0, 1, 2, \ldots,$ **do**

Set $S_k = \{i \in V \mid \nabla_i F_\mu(x_k) < 0\}$.

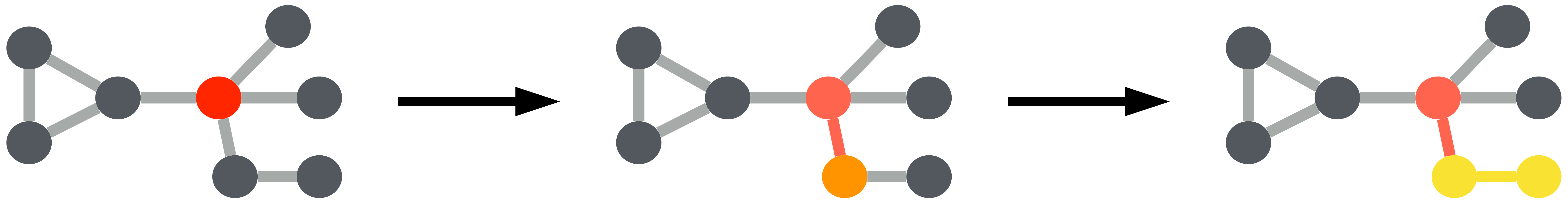
Pick $i_k \in S_k$ uniformly at random.

Update $x_{k+1} = x_k - \frac{\mu^{2-q}}{\deg(i_k)} \nabla_{i_k} F_\mu(x_k) e_{i_k}$.

If $S_k = \emptyset$ **then return** x_k.

The only difference from the standard randomized CD

- $0 = x_0 \leq x_1 \leq \cdots \leq x_k \leq x^*$ for all $k \geq 0$
  - A curial property that guarantees convergence and rate of convergence
  - Intuitively, diffusion is a local phenomenon



Local clustering guarantee

- Conductance of target cluster $C \subset V$

$$\phi(C) = \frac{|\{(u,v) \in E : u \in C, v \notin C\}|}{\min\{\mathbf{vol}(C), \mathbf{vol}(V \backslash C)\}} \quad \text{where } \mathbf{vol}(C) := \sum_{v \in C} d(v)$$

- Seed set $S :=$ supp(Δ)
- Assumption (sufficient overlap): $\mathbf{vol}(S \cap C) \geq \beta \mathbf{vol}(S)$, $\mathbf{vol}(S \cap C) \geq \alpha \mathbf{vol}(C)$, $\alpha, \beta \geq 1/\log^t \mathbf{vol}(C)$ for some t

Local clustering guarantee

- Conductance of target cluster $C \subset V$

$$\phi(C) = \frac{|\{(u,v) \in E : u \in C, v \notin C\}|}{\min\{\mathbf{vol}(C), \mathbf{vol}(V \backslash C)\}} \quad \text{where } \mathbf{vol}(C) := \sum\nolimits_{v \in C} d(v)$$

- Seed set $S := \mathrm{supp}(\Delta)$
- Assumption (sufficient overlap): $\mathbf{vol}(S \cap C) \geq \beta \mathbf{vol}(S)$, $\mathbf{vol}(S \cap C) \geq \alpha \mathbf{vol}(C)$, $\alpha, \beta \geq 1/\log^t \mathbf{vol}(C)$ for some t
- Apply the sweepcut rounding procedure to x^* returns a cluster $\tilde{C}$ satisfying

$$\phi(\tilde{C}) \leq \tilde{O}(\phi(C)^{1-1/p})$$

  - Cheeger-type bound $\phi(\tilde{C}) \leq \tilde{O}(\sqrt{\phi(C)})$ for $p = 2$
  - Constant approximation $\phi(\tilde{C}) \leq \tilde{O}(\phi(C))$ for $p \to \infty$

Local clustering guarantee

- Conductance of target cluster $C \subset V$

$$\phi(C) = \frac{|\{(u,v) \in E : u \in C, v \notin C\}|}{\min\{\mathbf{vol}(C), \mathbf{vol}(V \backslash C)\}} \quad \text{where } \mathbf{vol}(C) := \sum_{v \in C} d(v)$$

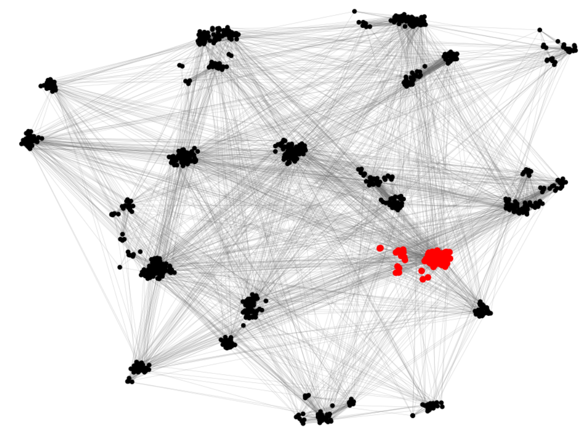
- Seed set $S := \text{supp}(\Delta)$
- Assumption (sufficient overlap): $\mathbf{vol}(S \cap C) \geq \beta \mathbf{vol}(S)$, $\mathbf{vol}(S \cap C) \geq \alpha \mathbf{vol}(C)$, $\alpha, \beta \geq 1/\log^t \mathbf{vol}(C)$ for some t
- Apply the sweepcut rounding procedure to x^* returns a cluster $\tilde{C}$ satisfying

$$\phi(\tilde{C}) \leq \tilde{O}(\phi(C)^{1-1/p})$$

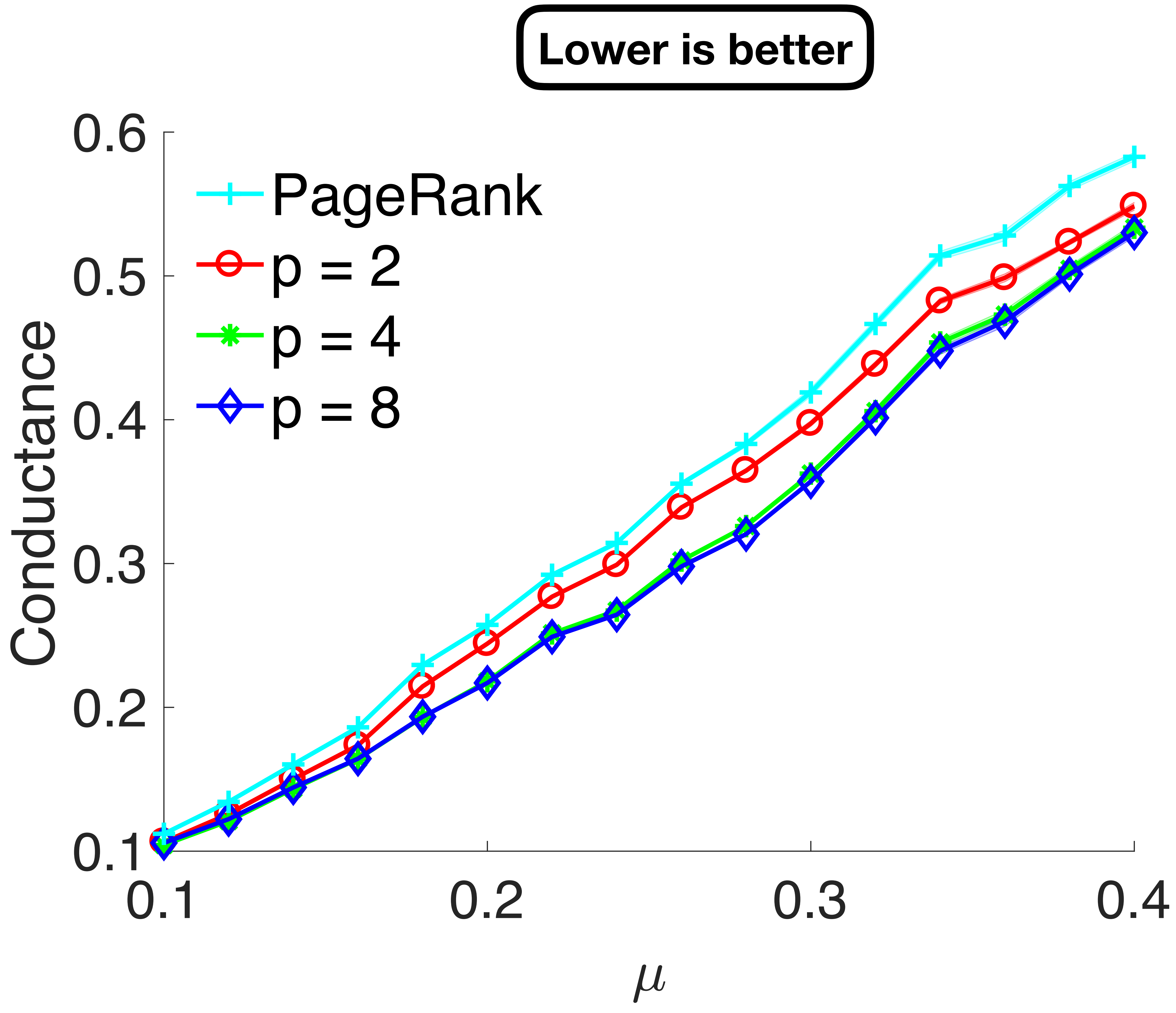
  - Cheeger-type bound $\phi(\tilde{C}) \leq \tilde{O}(\sqrt{\phi(C)})$ for $p = 2$
  - Constant approximation $\phi(\tilde{C}) \leq \tilde{O}(\phi(C))$ for $p \to \infty$

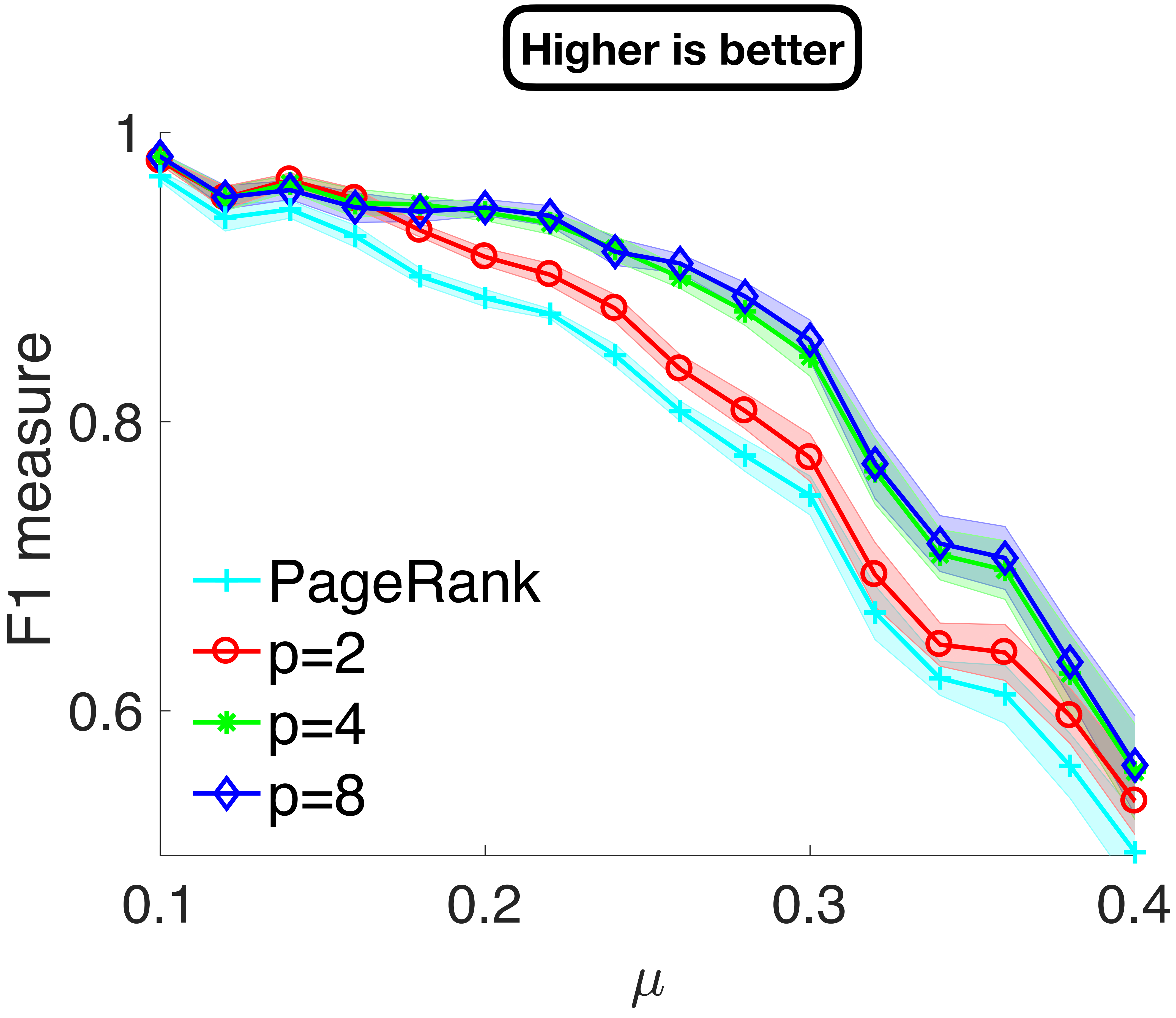
- Tradeoff between running time $O\left(\frac{\|\Delta\|_1 d_{\max}}{\gamma}\left(\frac{\|\Delta\|_1}{\epsilon}\right)^{2/q-1} \log\frac{1}{\epsilon}\right)$ and approximation $\tilde{O}(\phi(C)^{1/q})$

Empirical performance



- LFR synthetic graphs
  - μ is a parameter that controls noise, higher μ means more noisy clustering structure





Empirical performance

- Facebook social network for Colgate University, students in Class of 2009

	PageRank	p = 2	p = 4
Conductance	0.13	0.13	**0.12**
F1 measure	0.96	0.96	**0.97**

very clean ground truth

- Facebook social network for Johns Hopkins University, students of the same major

	PageRank	p = 2	p = 4
Conductance	0.25	0.23	**0.22**
F1 measure	0.83	0.85	**0.87**

average ground truth

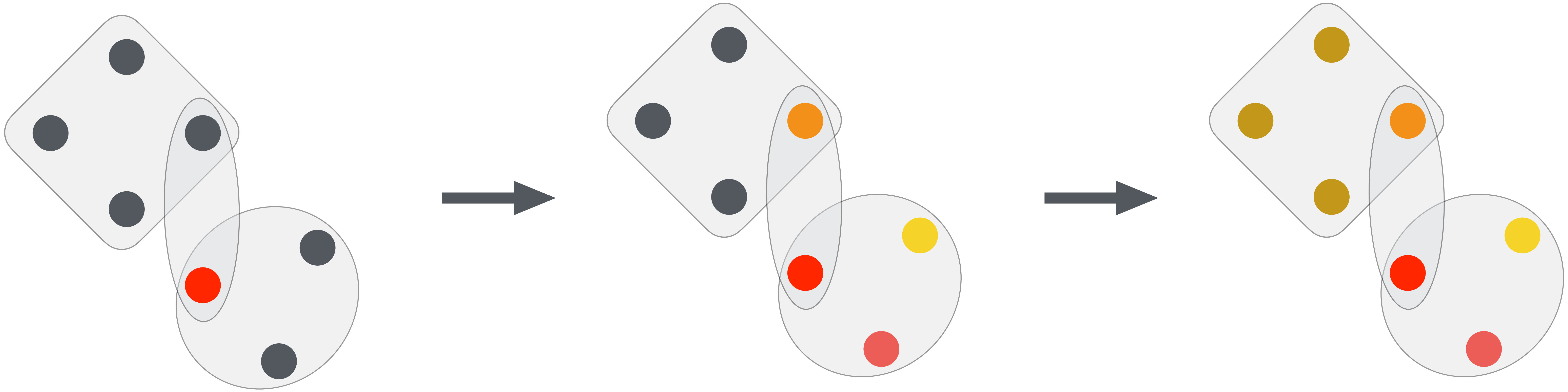
- Orkut, large-scale on-line social network, user-defined group

	PageRank	p = 2	p = 4
Conductance	0.37	0.35	**0.33**
F1 measure	0.66	0.71	**0.73**

very noisy ground truth

Extensions

- New network centrality with applications in network epidemic intervention [YSW+ 2021]
  - Orders of magnitude faster than traditional measures
- (Submodular) Hypergraph diffusion [FLY 2021]
  - Submodular function minimization, alternating minimization, duality …



References

p-Norm Flow Diffusion for Local Graph Clustering. K. Fountoulakis, D. Wang, S. Yang. ICML 2020

Local Hyper-Flow Diffusion. K. Fountoulakis, P. Li, S. Yang. NeurIPS 2021

Thank you!