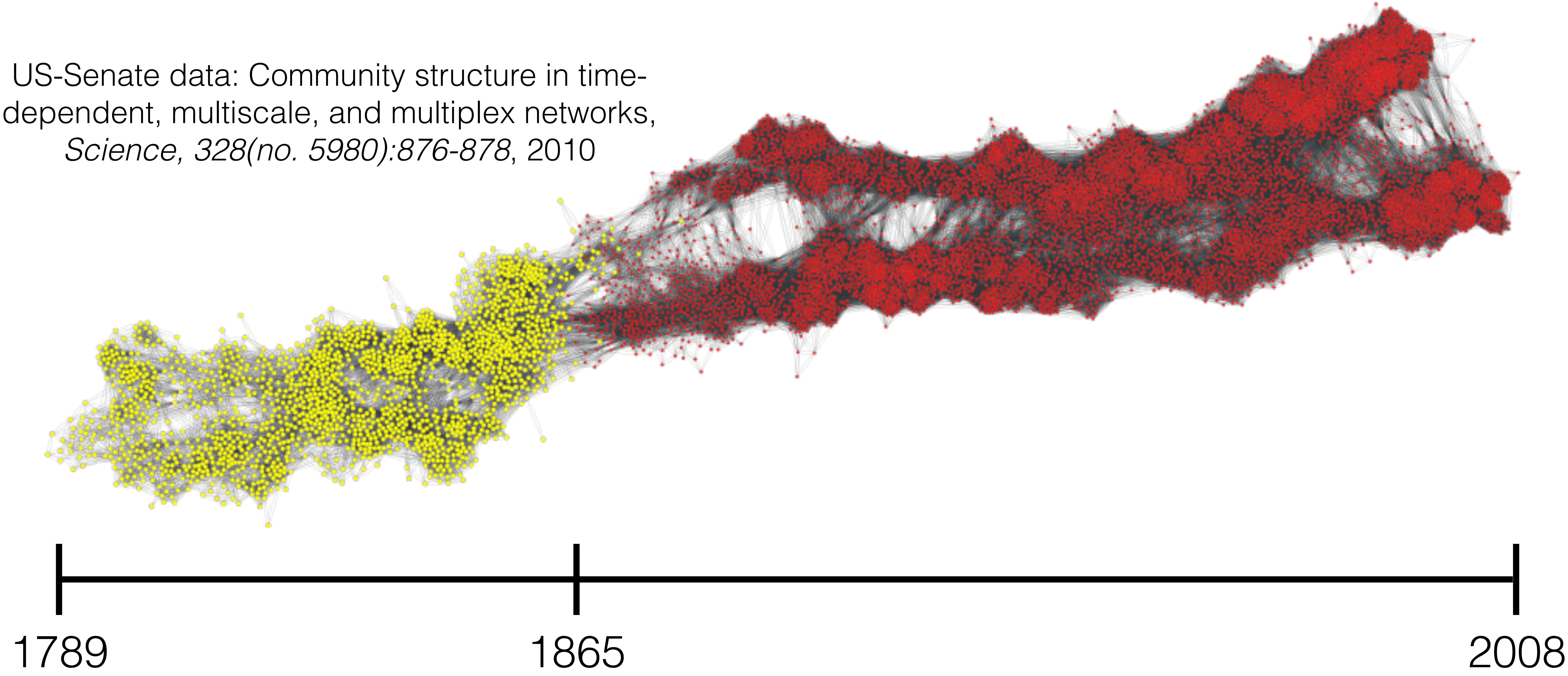
Variational Perspective on Local Graph Clustering

Kimon Fountoulakis, joint work with J. Shun, X. Cheng, F. Roosta-Khorasani, M. Mahoney

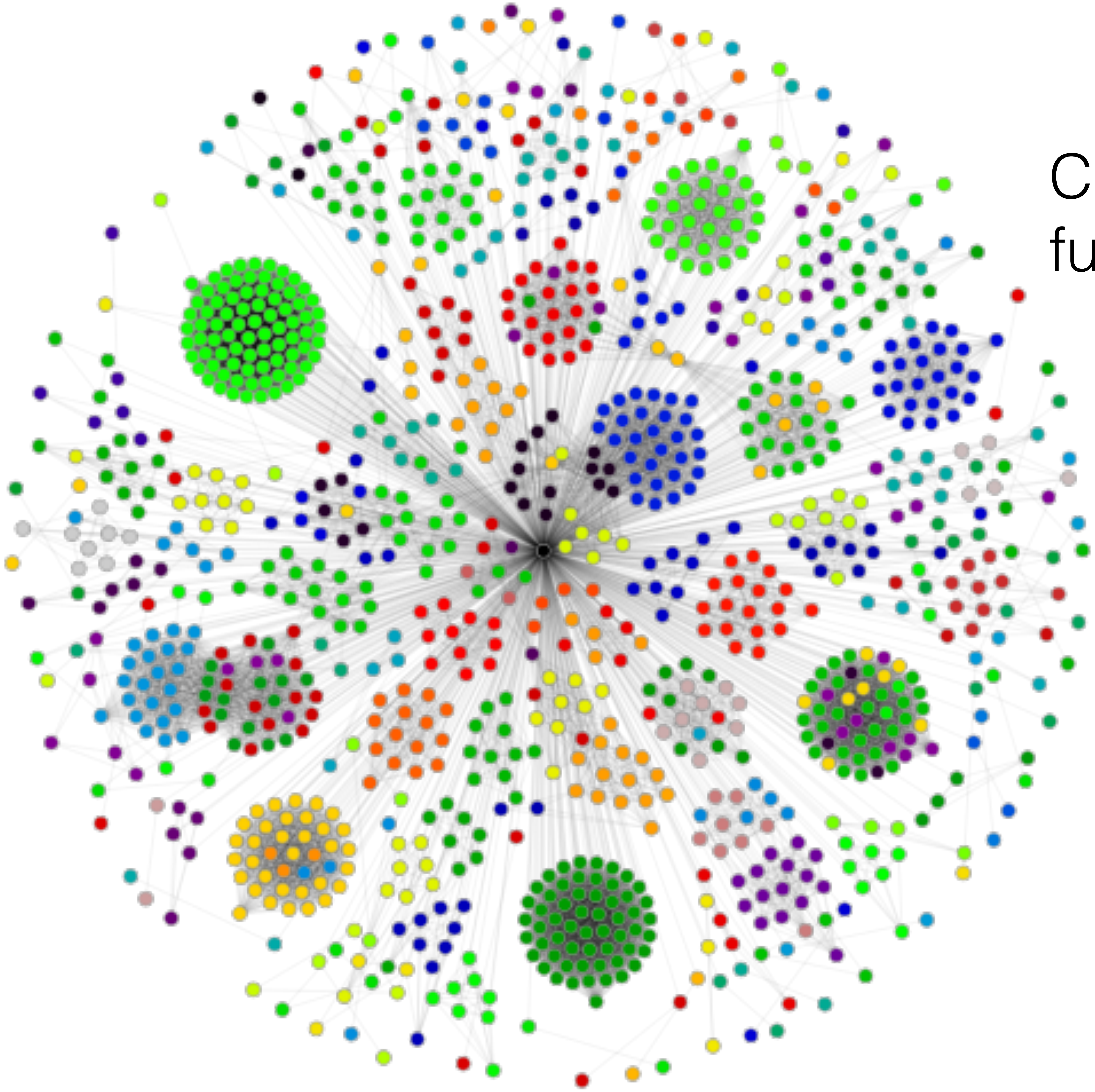
Past and present studies focus on **global** trends of the data

US-Senate data: Community structure in time-dependent, multiscale, and multiplex networks, *Science, 328(no. 5980):876-878*, 2010



The American civil war ended in 1865

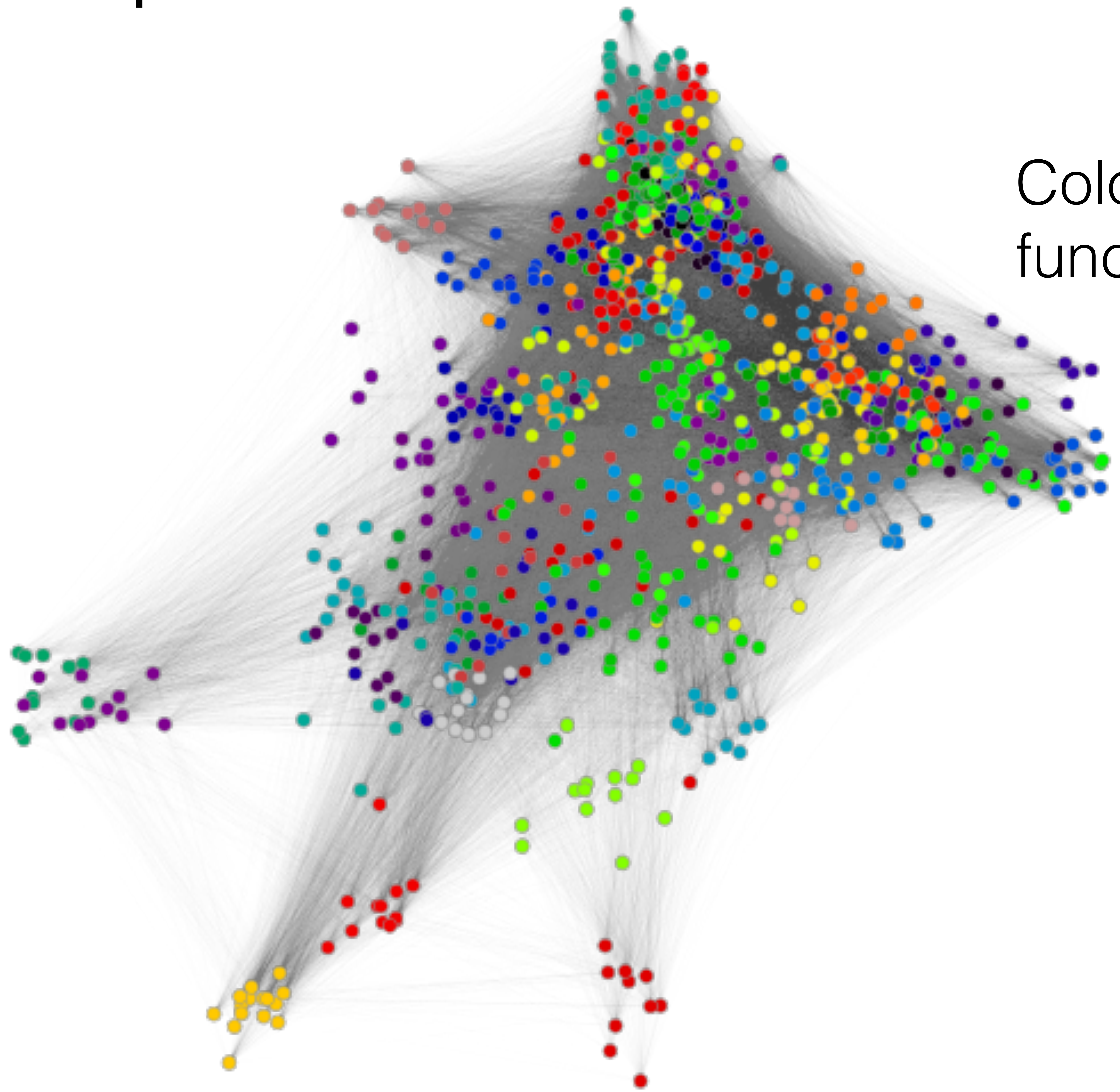
But, most real data have rich **local** structure



Color denotes similar function

Data: The MIPS mammalian protein-protein interaction database. *Bioinformatics, 21(6):832-834*, 2005

And can be very complex



Color denotes similar function

Data: The MIPS mammalian protein-protein interaction database. *Bioinformatics, 21(6):832-834*, 2005

Outline

1. Local graph clustering, definition, examples and software

2. Example of a state-of-the-art method

3. Variational model

4. Proximal gradient descent

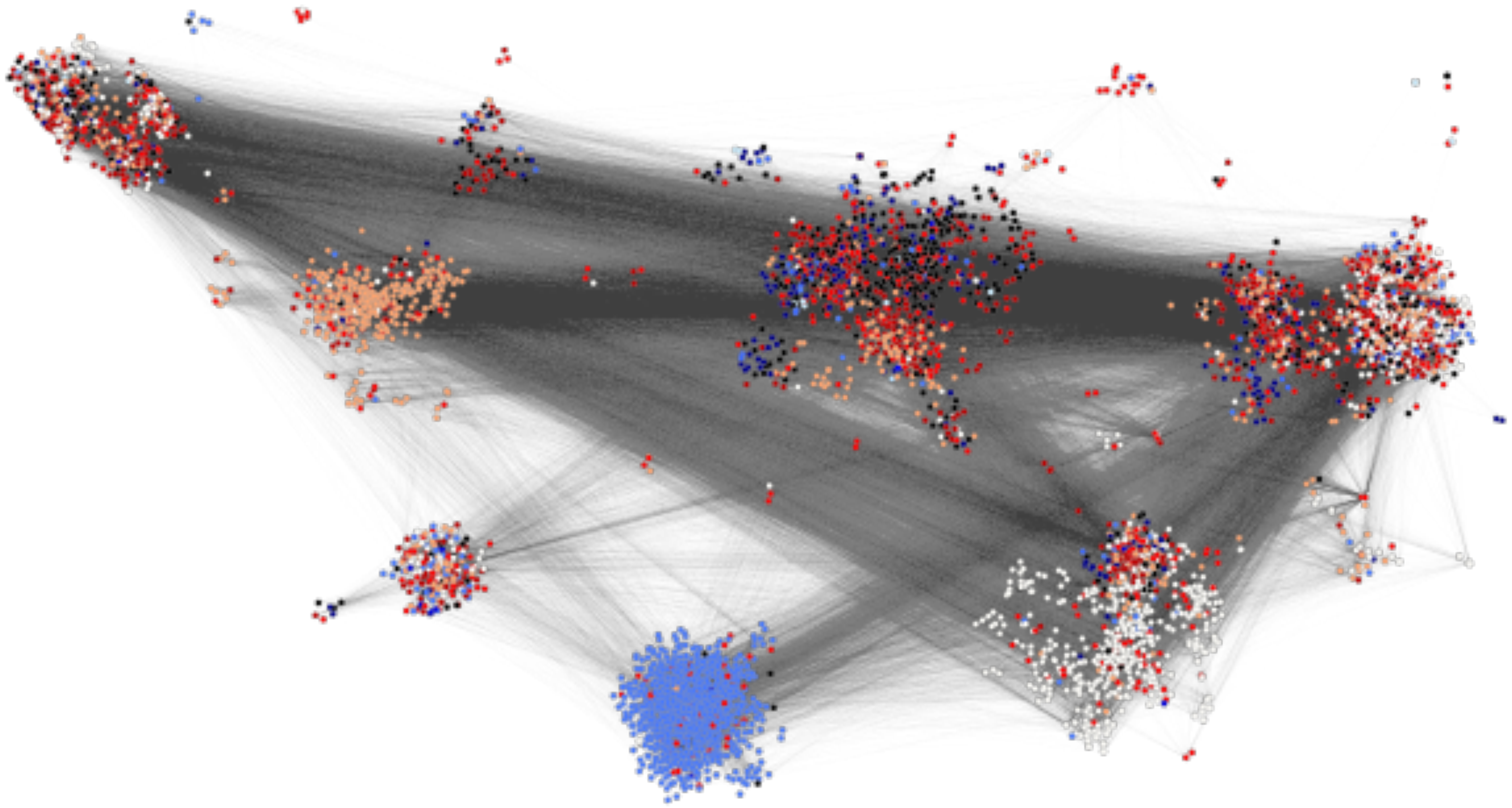
What is local graph clustering and why is it useful?

- Definition: find set of nodes A given a seed node in set B
  - Set A has good precision/recall w.r.t set B
  - The running time depends on A instead of the whole graph

- Scalable to graphs with billions of edges

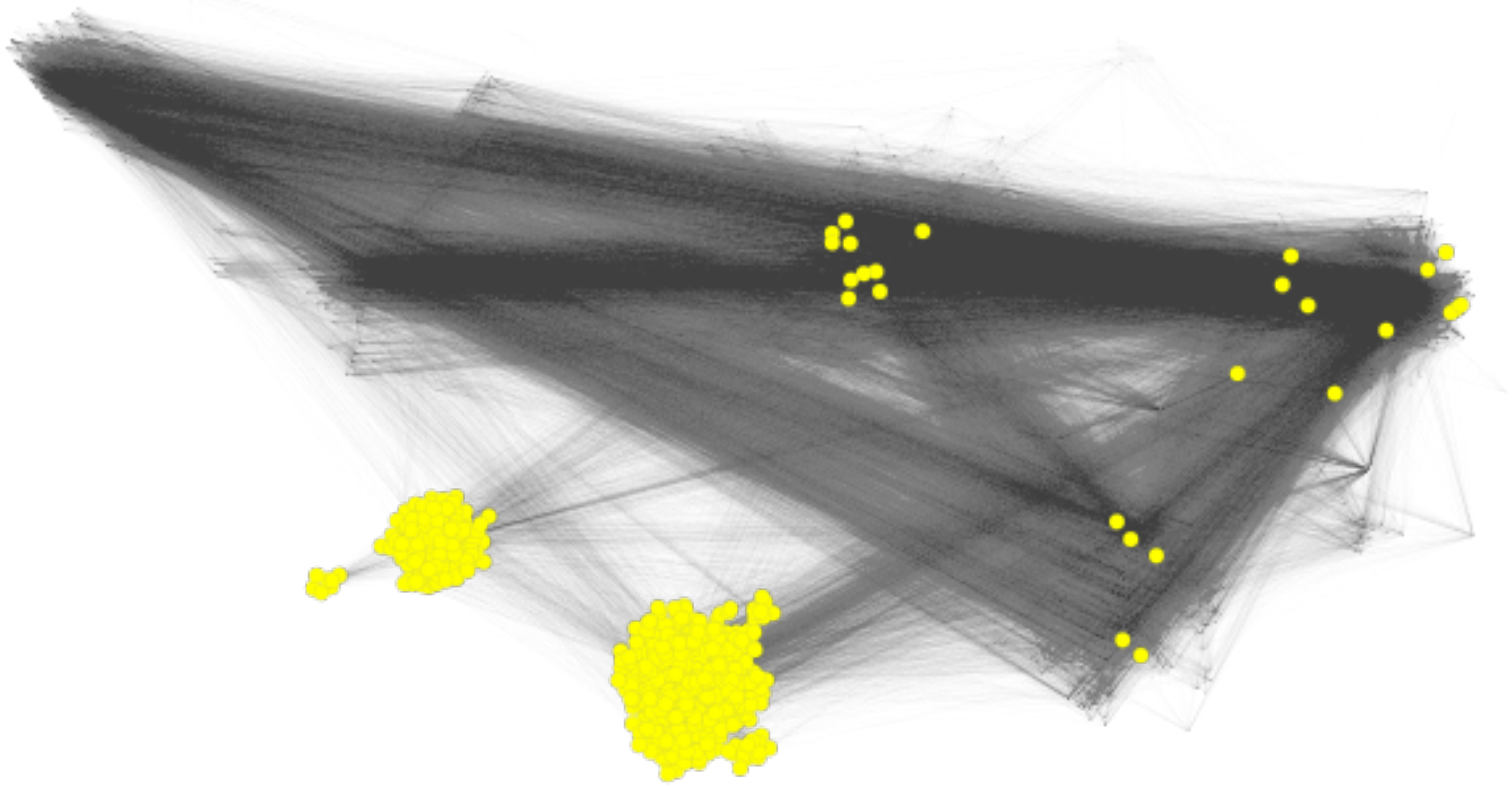
- Ideal for finding small clusters and small neighborhoods

Facebook social network: colour denotes class year



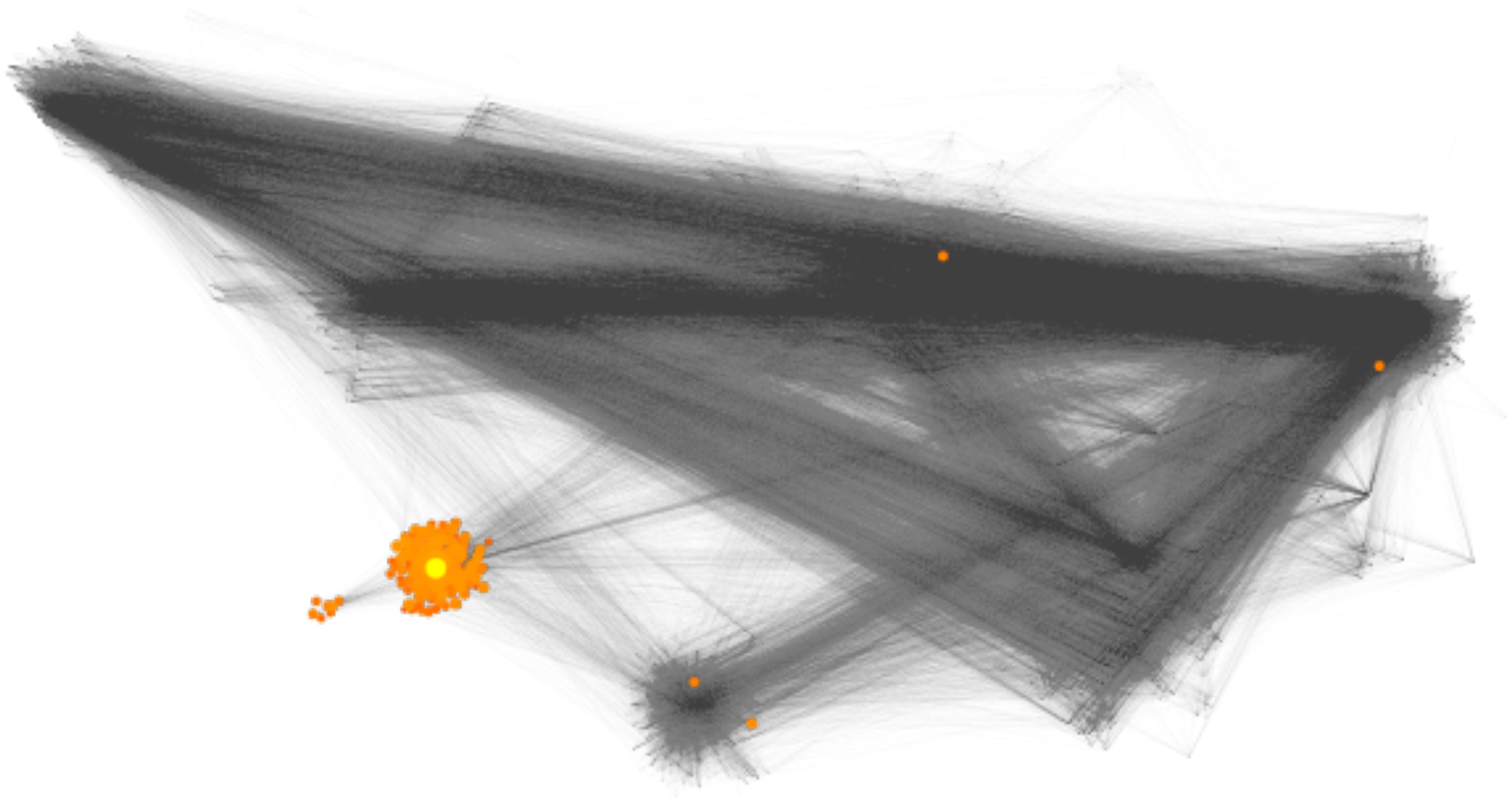
Data: Facebook John Hopkins, A. L. Traud, P. J. Mucha and M. A. Porter, Physica A, 391(16), 2012

Global spectral: finds 20% of the graph



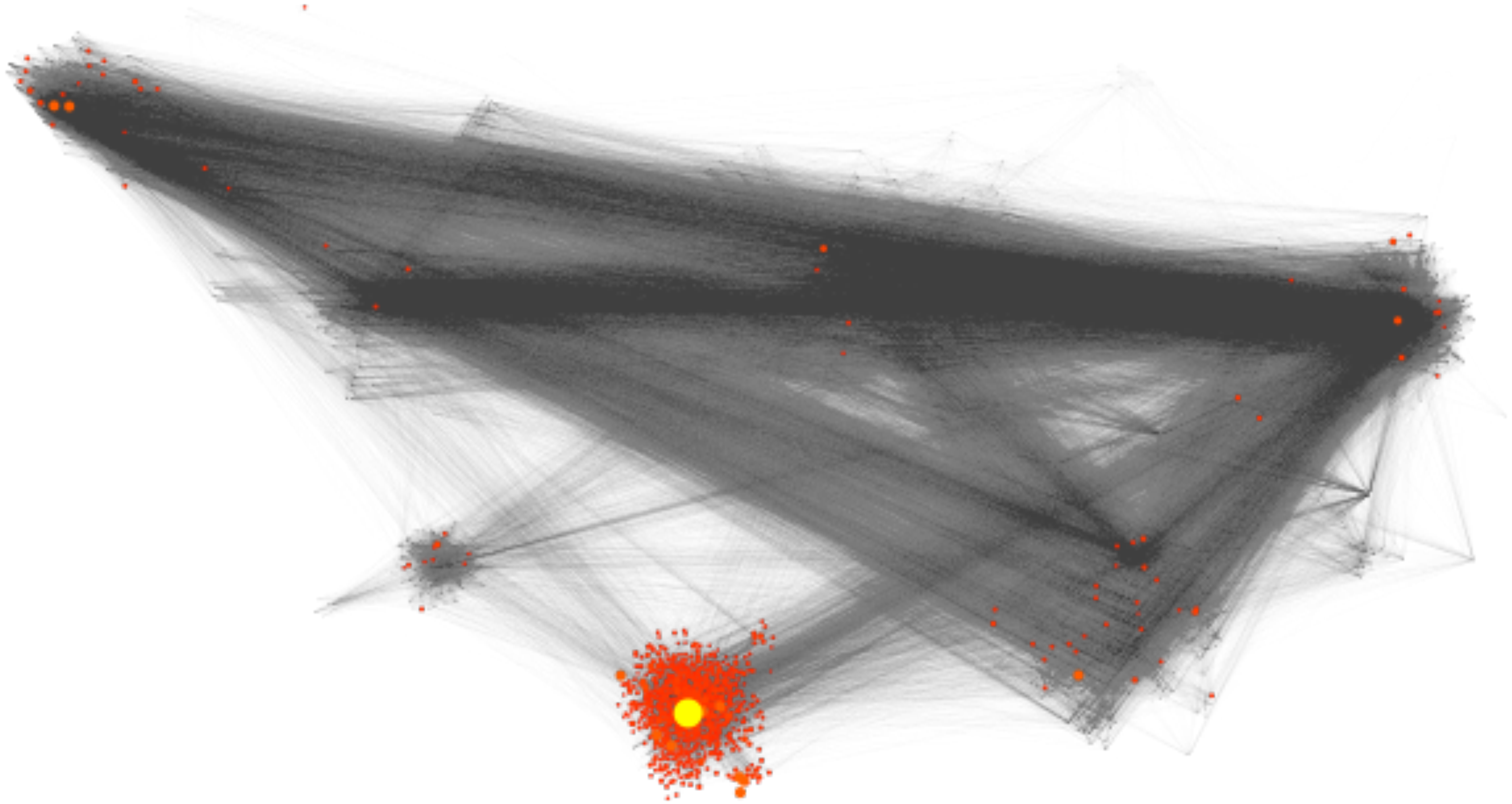
Data: Facebook John Hopkins, A. L. Traud, P. J. Mucha and M. A. Porter, Physica A, 391(16), 2012

Local graph clustering: finds 3% of the graph



Data: Facebook John Hopkins, A. L. Traud, P. J. Mucha and M. A. Porter, Physica A, 391(16), 2012

Local graph clustering: finds 17% of the graph



Data: Facebook John Hopkins, A. L. Traud, P. J. Mucha and M. A. Porter, Physica A, 391(16), 2012

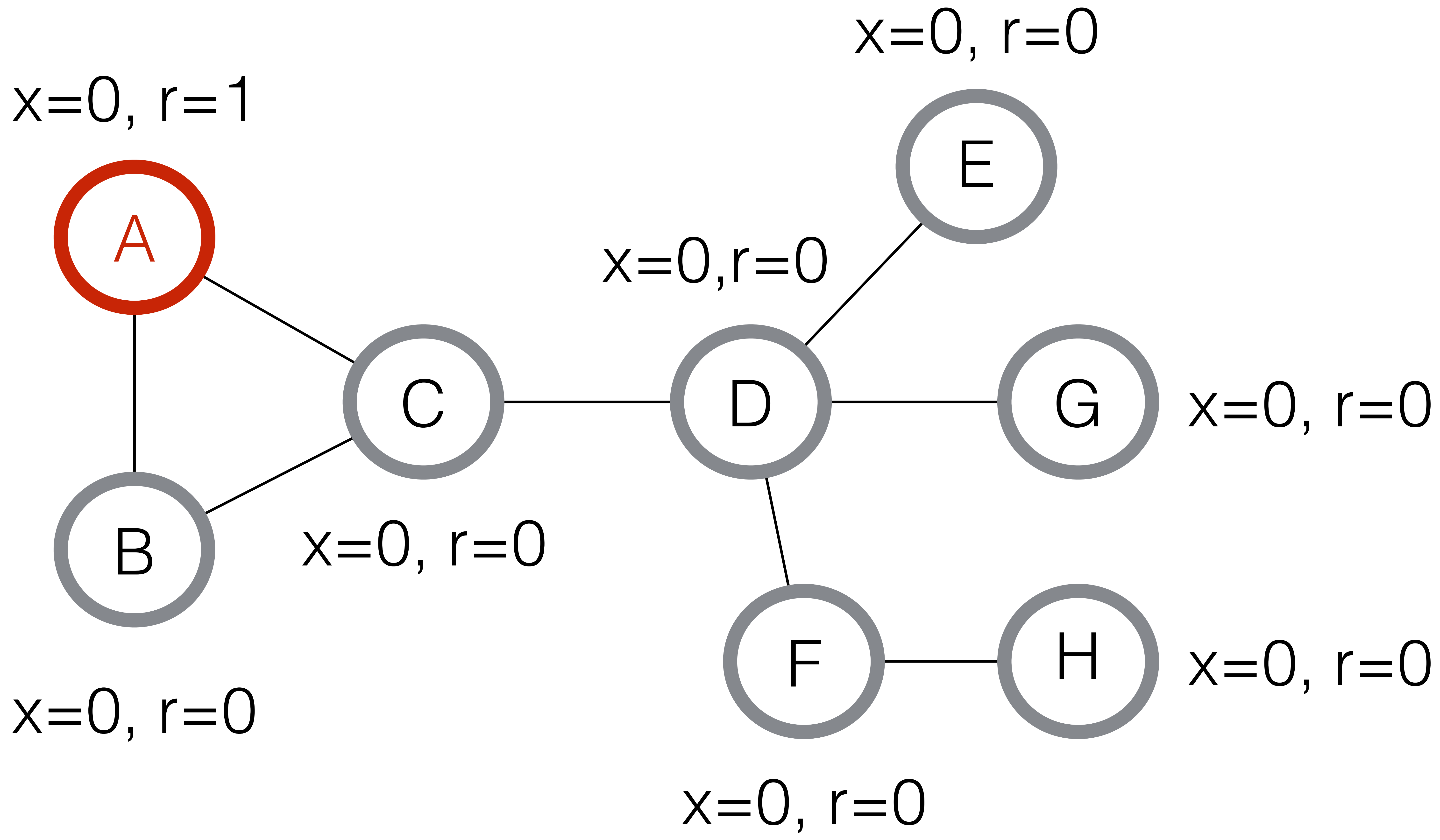
Software

LocalGraphClustering on GitHub

- Written in Python with C++ routines when required
- Graph analytics on 100 million edges graph on a 16GB RAM laptop
- Demonstrations on social and bioinformatics networks
- 8 Python notebooks with numerous examples and graph visualizations
- Video presentations
- 12 methods and pipelines

Approximate Personalized Page-Rank

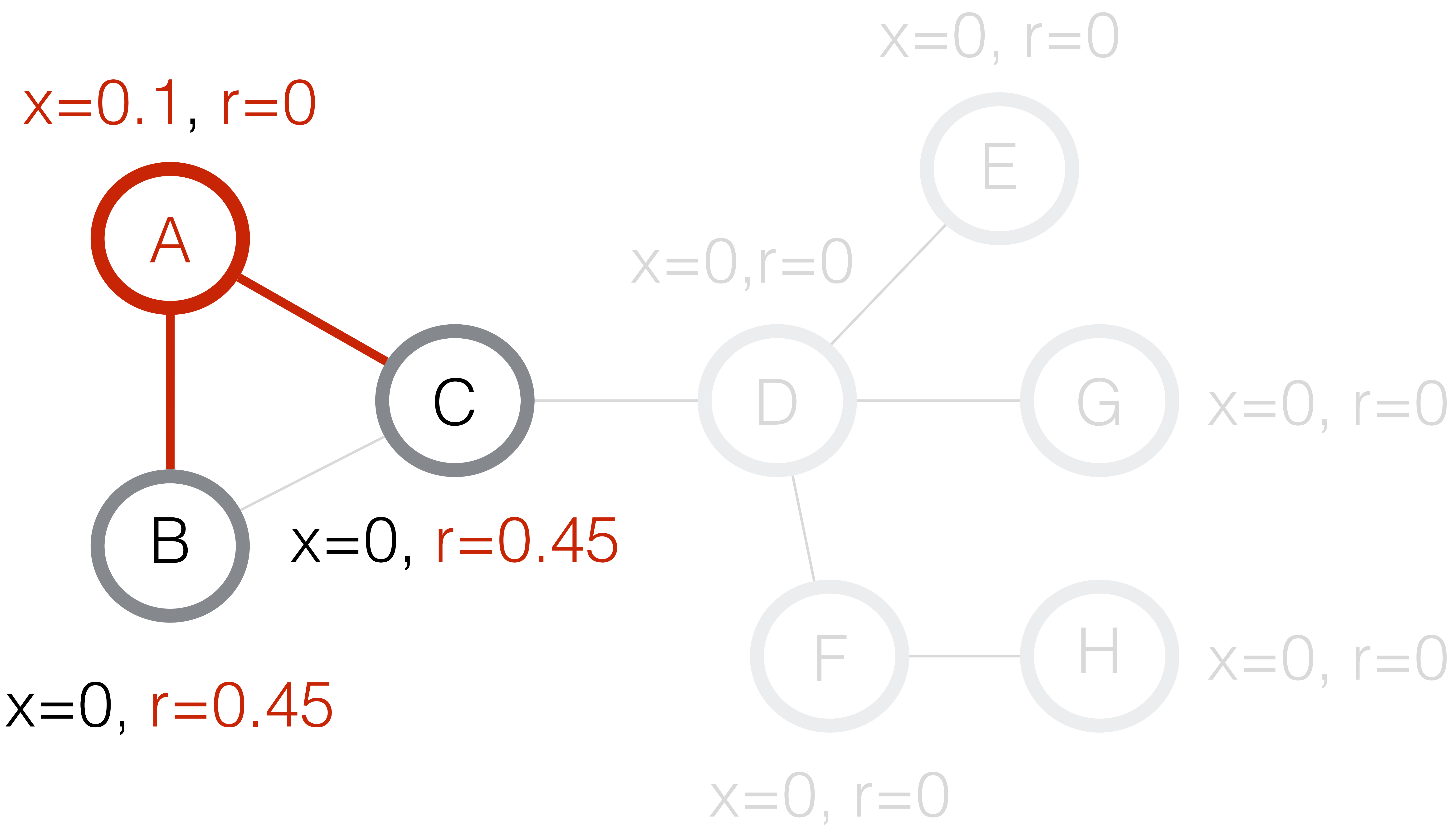
R. Andersen, F. Chung and K. Lang. Local graph partitioning using Page-Rank, FOCS, 2006



Algorithm idea: iteratively spread probability mass around the graph.

Approximate Personalized Page-Rank

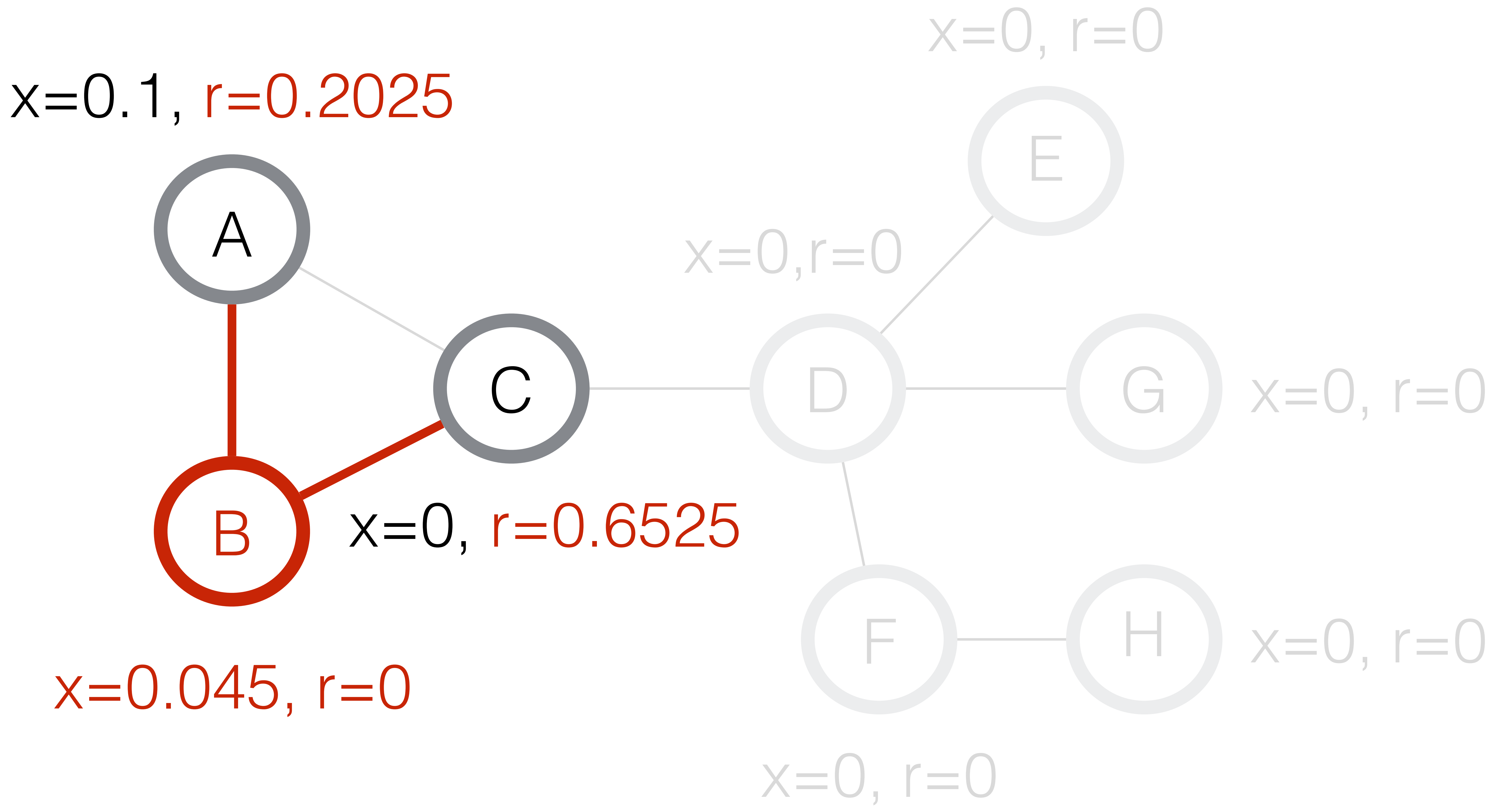
R. Andersen, F. Chung and K. Lang. Local graph partitioning using Page-Rank, FOCS, 2006



Algorithm idea: iteratively spread probability mass around the graph.

Approximate Personalized Page-Rank

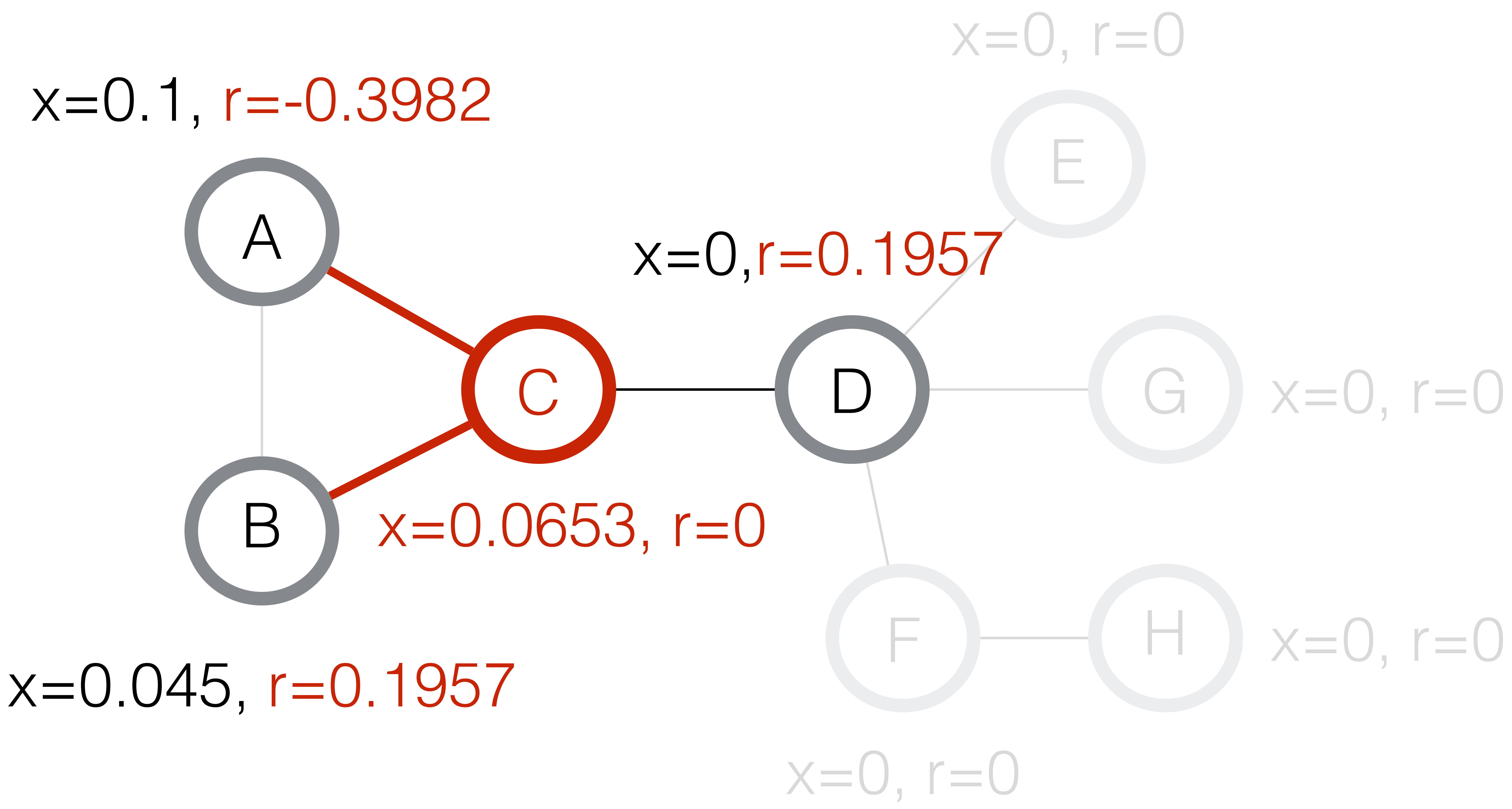
R. Andersen, F. Chung and K. Lang. Local graph partitioning using Page-Rank, FOCS, 2006



Algorithm idea: iteratively spread probability mass around the graph.

Approximate Personalized Page-Rank

R. Andersen, F. Chung and K. Lang. Local graph partitioning using Page-Rank, FOCS, 2006



Algorithm idea: iteratively spread probability mass around the graph.

Approximate Personalized Page-Rank

Algorithm idea: iteratively spread probability mass around the graph until

$$\max_i \frac{r_i}{d_i} \leq \rho\alpha$$

- α: teleportation parameter
- ρ: hyper-parameter
- d: degrees vector

Variational model of APPR

Observation: The optimality conditions of an l1-regularized convex problem imply the termination condition of APPR.

$$\text{minimize} \ \frac{1-\alpha}{2}\|Bx\|_2^2 + \alpha\|H(\mathbf{1}-x)\|_2^2 + \alpha\|Zx\|_2^2 + \rho\alpha\|Dx\|_1$$

where

- B: is the incidence matrix
- D: Degree matrix
- H = diag(initial prob. dist. over nodes)
- Z = D - H
- α: teleportation parameter
- ρ: l1-reg. hyper-parameter

Termination conditions vs optimality conditions

Termination criteria of Approximate Personalized PageRank

$$\max_i \frac{r_i}{d_i} \leq \rho\alpha$$

Optimality conditions of the variational model

$$\frac{r_i}{d_i} = \rho\alpha, \;\; x_i \neq 0$$

$$\frac{r_i}{d_i} \leq \rho\alpha, \;\; x_i = 0$$

Properties of the variational problem

- **Theorem:** The volume of the optimal solution is bounded by $1/\rho$

- **Theorem:** Same combinatorial theoretical guarantees for local graph clustering

- **Crucial:** The model decouples the output from the algorithm.

Proximal gradient descent for local graph clustering

$$f(x) := \frac{1-\alpha}{2}\|Bx\|_2^2 + \alpha\|H(\mathbf{1}-x)\|_2^2 + \alpha\|Zx\|_2^2 \qquad g(x) := \rho\alpha\|Dx\|_1$$

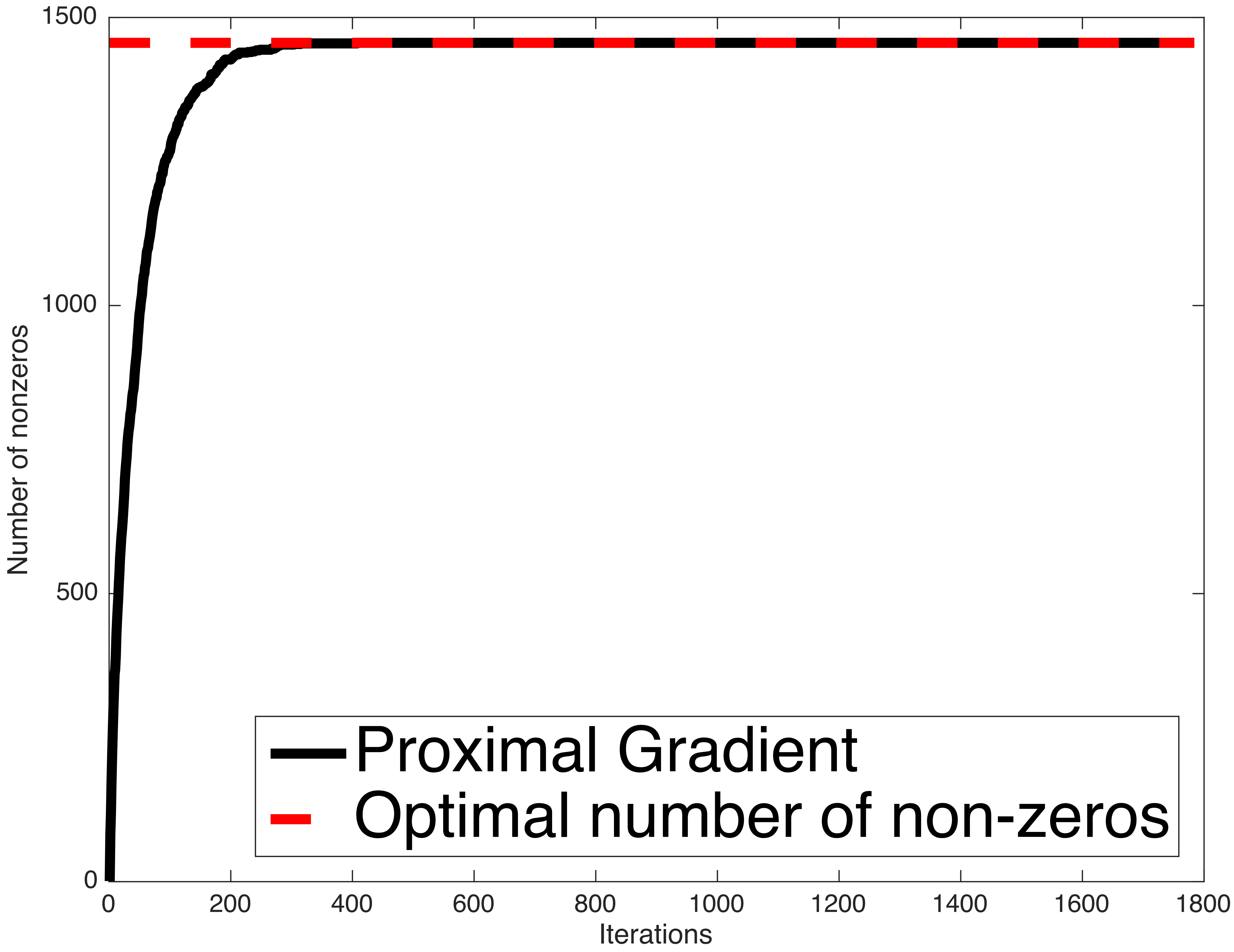
Proximal gradient descent

$$x_{k+1} := \text{argmin } g(x) + \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle}_{\text{first-order Taylor approximation}} + \underbrace{\frac{1}{2}\|x - x_k\|_2^2}_{\substack{\text{upper bound on the} \\ \text{approximation error}}}$$

Requires careful implementation to avoid excessive running time

- Need to maintain a set of non-zero nodes
- Update x and gradient only for non-zero nodes and their neighbors at each iteration

Theorem: non-decreasing non-zero nodes



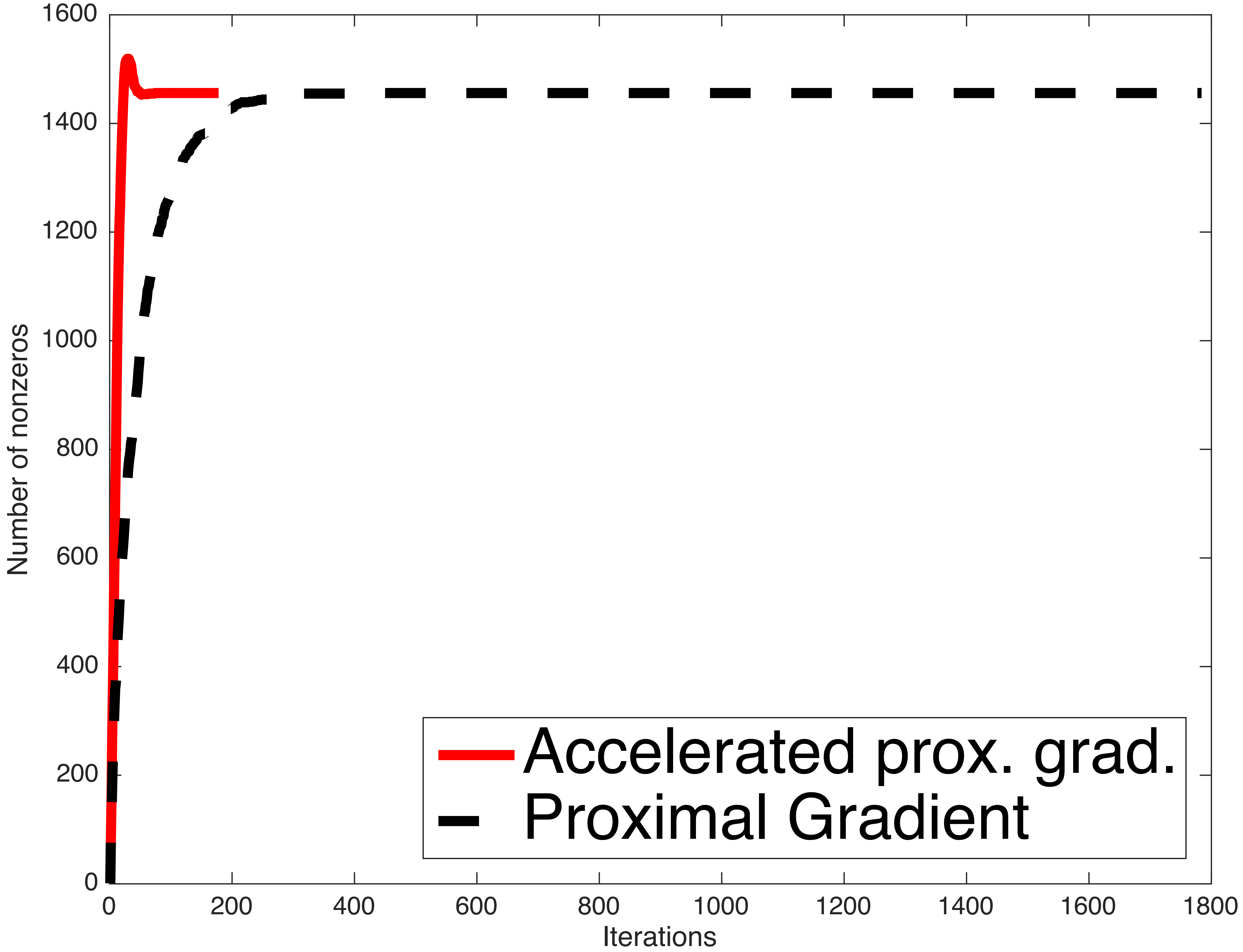
Worst-case running times

	Weighted graphs	Unweighted graphs
Prox. grad.	$\mathcal{O}\left(\frac{(\lvert\mathcal{S}_*\rvert + \widehat{\mathrm{vol}}(\mathcal{S}_*))}{\mu}\log\left(\frac{2}{\epsilon^2\rho^2\alpha^2\min_j d_j}\right)\right)$	$\mathcal{O}\left(\frac{2}{\rho\mu}\log\left(\frac{2}{\epsilon^2\rho^2\alpha^2\min_j d_j}\right)\right).$
APPR	✘	$\frac{1}{\alpha\rho}$

$$\mu := \alpha + \frac{1-\alpha}{4}\lambda_{min}(\mathcal{L}_{\mathcal{S}_*})$$

$\mathcal{L}_{\mathcal{S}_*}$: sub-matrix of normalized Laplacian

Open problem: is accelerated prox. grad. a local algorithm?



Gradient descent running time

- Inversely proportional to the strong-convexity parameter

- Each iteration is provably sparse

Accel. gradient descent

- Inverse proportional to the **square root** of the strong-convexity parameter

- Is each iteration sparse??

Thank you!