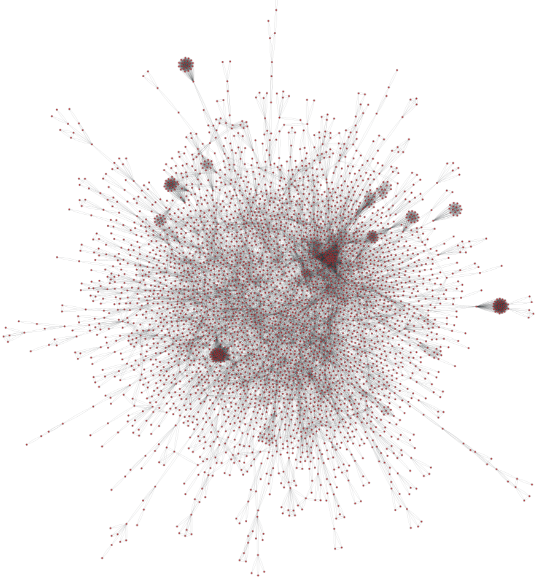
Local graph clustering and optimization

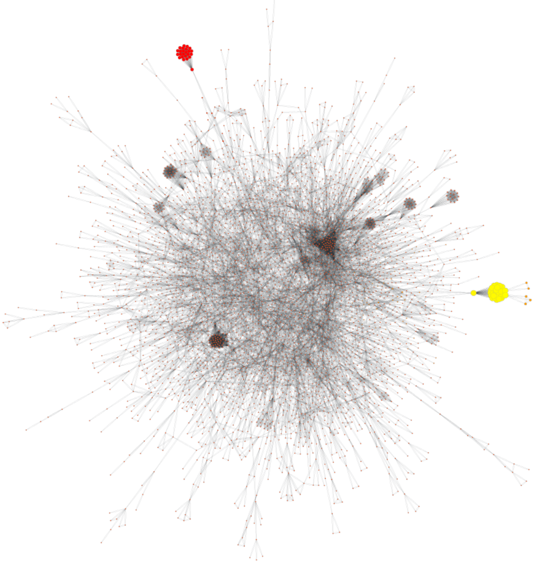
Kimon Fountoulakis

joint work with: X. Cheng, J. Demmel, A. Devarakonda, D. Gleich, M. Mahoney, F. Roosta-Khorasani and J. Shun

Local graph clustering: motivation

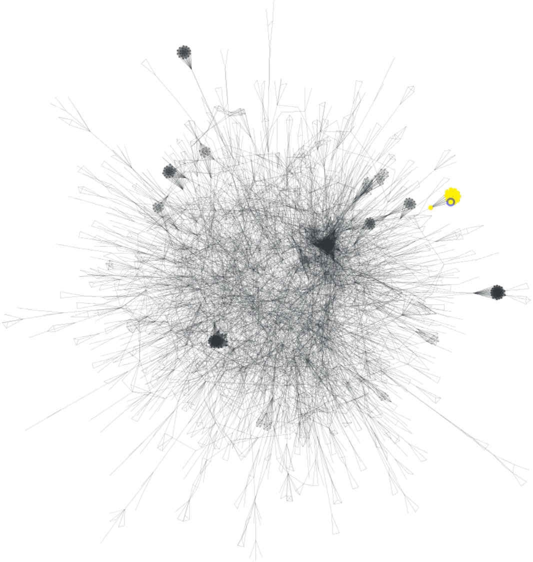


Data: general relativity and quantum cosmology collaboration network, J. Leskovec, J. Kleinberg and C. Faloutsos, ACM TKDD, 1(1), 2007



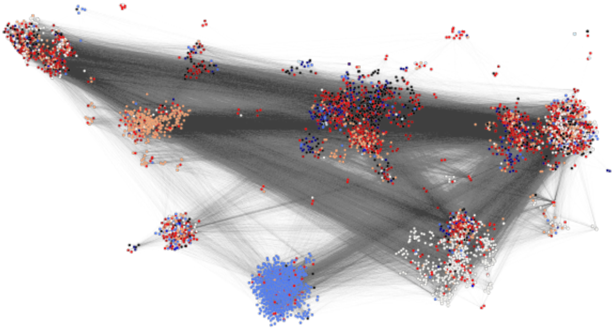
Global graph clustering: normalized cuts

Data: general relativity and quantum cosmology collaboration network, J. Leskovec, J. Kleinberg and C. Faloutsos, ACM TKDD, 1(1), 2007



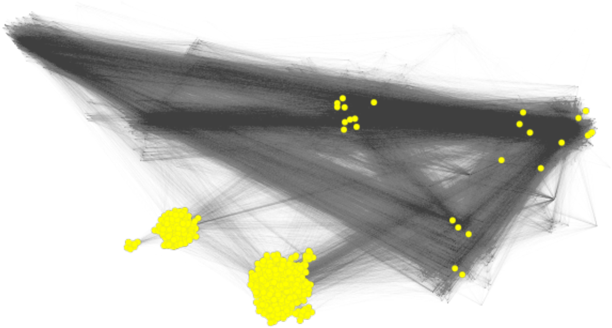
Local graph clustering: small clusters!

Data: general relativity and quantum cosmology collaboration network, J. Leskovec, J. Kleinberg and C. Faloutsos, ACM TKDD, 1(1), 2007



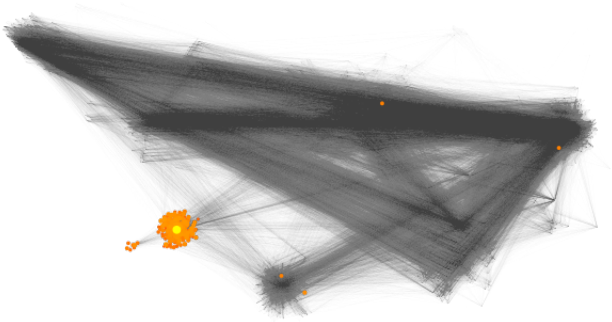
Colour denotes class year

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



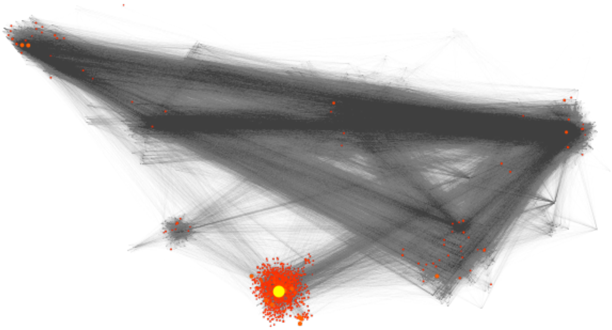
Global graph clustering (normalized cuts) finds 20% of the graph

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



Local graph clustering can cut 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 can cut 17% of the graph

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

Current algorithms and running time

Local algorithms

- **MQI**: K. Lang and S. Rao, IPCO (2004)
- **ACL**: R. Andersen, F. Chung and K. Lang, FOCS (2006)
- **spectral MQI**: F. Chung, Linear Algebra and its Applications 423(1), 22–32 (2007)
- **Flow-Improve**: R. Andersen and K. Lang, SODA (2008)
- **MOV**: Mahoney, Orecchia, Vishnoi, 13, 2339–2365 (2012)
- **Nibble**: D. A. Spielman and S-H. Teng, SIAM J. Comput., 42(1), 1–26 (2013)
- **Local Flow-Improve**: L. Orecchia and Z. A. Zhu, SODA (2014)

Global, weakly and strongly local methods

Global algorithms

- The workload depends on the size of the graph, i.e., $\mathcal{O}(vol(G))$.

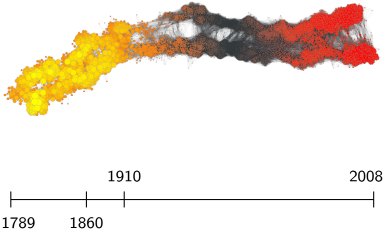
Weakly local algorithms

- A seed set of nodes is given.
- The solution is locally biased to the input seed set.
- The workload depends on the size of the graph, i.e., $\mathcal{O}(vol(G))$.

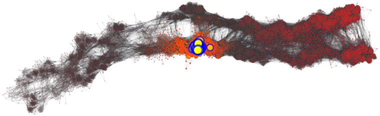
Strongly local algorithms

- A seed set of nodes is given.
- The solution is locally biased to the input seed set.
- The workload of the algorithm is $\mathcal{O}(\text{volume of output cluster})$.

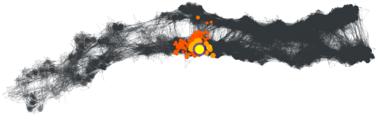
Global, weakly and strongly local methods



Global spectral



Weakly local



Strongly local

Data: US Senate, P. Mucha, T. Richardson, K. Macon, M. Porter, and J. Onnela, Science, vol. 328, no. 5980, pp. 876–878, 2010

Local algorithms

MOV (weakly local)

- second eigenvector
- subject to a constraint that represents locality

ACL (strongly local)

- random walk
- uses the personalized PageRank transition matrix

Flow-Improve (weakly local)

- input solution from ACL
- solves a sequence of max-flow/min-cut computations

Local Flow-Improve (strongly local)

- similar to Flow-Improve but with strongly local max-flow/min-cut computation

Computer science and optimization perspectives

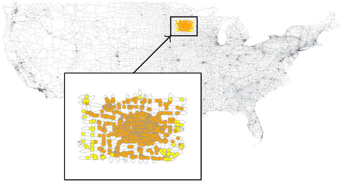
Computer science perspective

Cluster quality is captured by

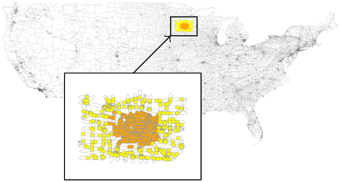
$$\text{Conductance} := \frac{\text{external connectivity}}{\text{internal connectivity}}$$

- Minimum conductance is a NP-hard problem.
- Question: how can we find an approximate solution? Use local algorithms!

Knowing that we have an approximation methodology is good, but... What is the actual solution that we find?



Flow, seed: Minneapolis and suburban areas. Conductance $= 0.010$.



Spectral, seed: Minneapolis and suburban areas. Conductance $= 0.011$.

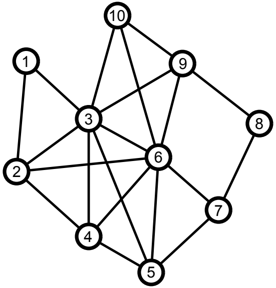
This talk

- Optimization perspective - a unified framework and regularization properties
  - *"An optimization approach to locally-biased graph algorithms"*, K. Fountoulakis, D. Gleich and M. Mahoney
- Three common factors in the optimization problems
  - Localizing bias
  - Distance metrics, i.e., ℓ_1 or ℓ_2
  - Strongly or weakly local algorithms
- Leverage state-of-the-art optimization algorithms
  - *"Exploiting optimization for local graph clustering"*, K. Fountoulakis, X. Cheng, J. Shun, F. Roosta and M. Mahoney
- Parallel and distributed implementations
  - *"Parallel local graph clustering"*, J. Shun, F. Roosta-Khorasani, K. Fountoulakis and M. Mahoney
  - *"Communication-avoiding coordinate descent methods for linear systems"*, A. Devarakonda, K. Fountoulakis, J. Demmel and M. Mahoney

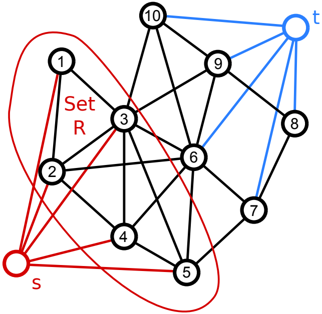
First factor: localizing bias

Localizing bias

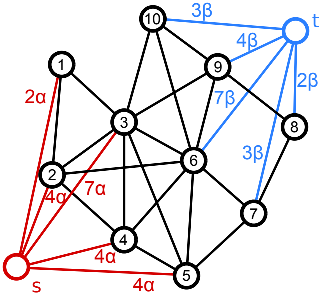
We describe locally-biased graph methods in terms of an augmented graph.



Simple unweighted graph



Adding the source s and sink t

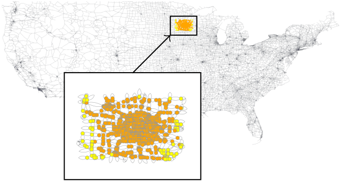


Augmented graph, with weights indicated

Second factor: distance metric

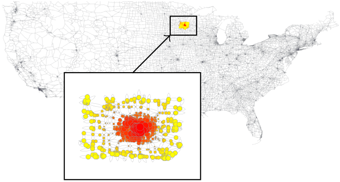
Distance metric: US roads

- $\mathcal{E}$: the set of edges of the augmented graph
- B, C: incidence and edge weights matrices of the augmented graph



Flow

$$\|Bx\|_{1,C} = \sum_{(i,j)\in\mathcal{E}} C_{ij}|x_i - x_j|$$

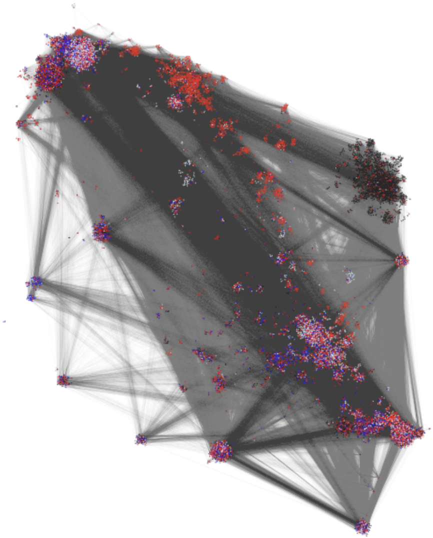


Spectral

$$\|Bx\|_{2,C} := \sum_{(i,j)\in\mathcal{E}} C_{ij}(x_i - x_j)^2$$

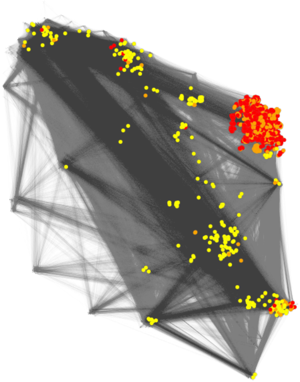
- Spectral is good in finding well-connected internally clusters

Distance metric: Berkeley Facebook



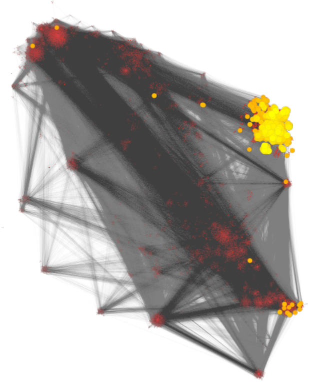
Colour denotes class year

Distance metric: Berkeley Facebook



Flow

$\|Bx\|_{1,C} = \sum_{(i,j)\in\mathcal{E}} C_{ij}|x_i - x_j|$

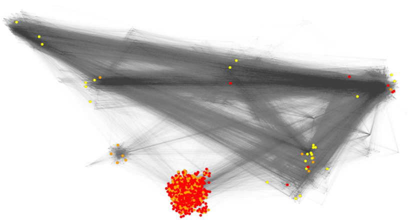


Spectral

$\|Bx\|_{2,C} := \sum_{(i,j)\in\mathcal{E}} C_{ij}(x_i - x_j)^2$

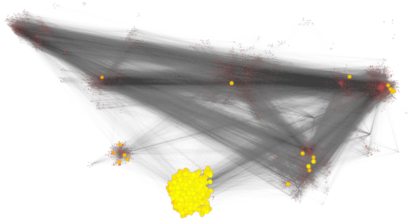
- Spectral leaks mass outside the boundaries
- Flow finds good boundaries

Distance metric: John Hopkins Facebook



Flow

$\|Bx\|_{1,C} = \sum_{(i,j)\in\mathcal{E}} C_{ij}|x_i - x_j|$

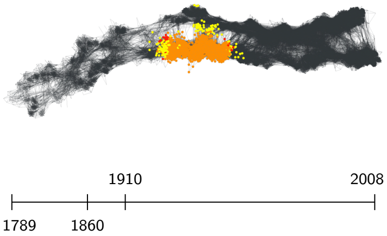


Spectral

$\|Bx\|_{2,C} := \sum_{(i,j)\in\mathcal{E}} C_{ij}(x_i - x_j)^2$

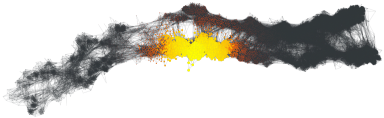
- Spectral leaks mass outside the boundaries
- Flow finds good boundaries

Distance metric: US Senate



Flow

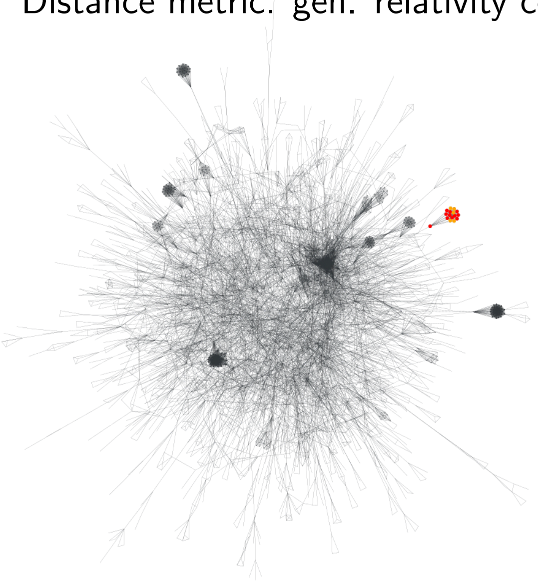
$$\|Bx\|_{1,C} = \sum_{(i,j)\in\mathcal{E}} C_{ij}|x_i - x_j|$$



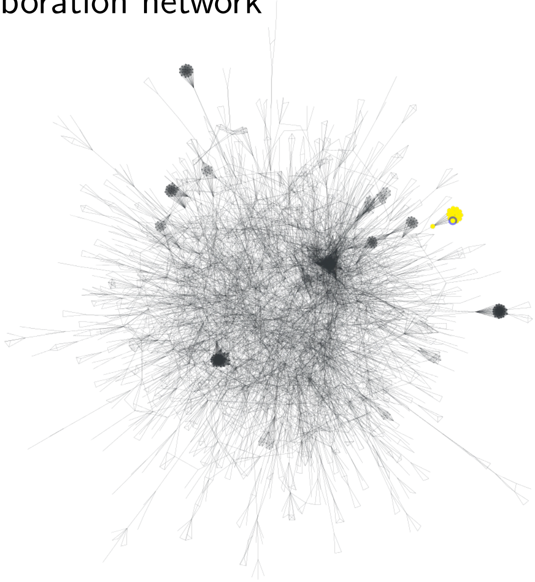
Spectral

$$\|Bx\|_{2,C} := \sum_{(i,j)\in\mathcal{E}} C_{ij}(x_i - x_j)^2$$

Distance metric: gen. relativity collaboration network



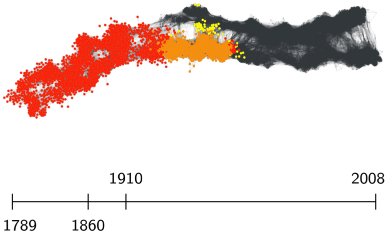
Flow



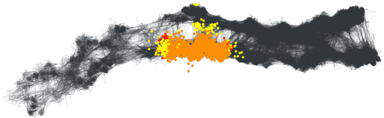
Spectral

Third factor: weak and strong locality

Weakly vs. strongly local flow methods



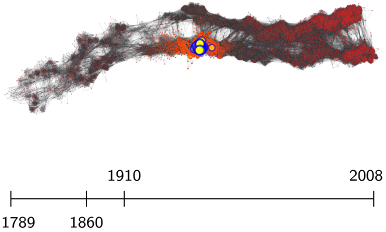
Weakly local flow: touches the whole graph



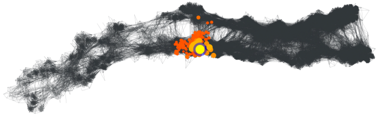
Strongly local flow: touches part of the graph

- Weakly local scales up to thousands of nodes, i.e., $\mathcal{O}(vol(G))$.
- Strongly local scales up to billions of nodes, i.e., $\mathcal{O}(\text{volume of output cluster})$.

Weakly vs. strongly local spectral methods



Weakly local spectral: touches the whole graph



Strongly local spectral: touches part of the graph

- Weakly local scales up to thousands of nodes, i.e., $\mathcal{O}(vol(G))$.
- Strongly local scales up to billions of nodes, i.e., $\mathcal{O}(\text{volume of output cluster})$.

Optimization problems: putting all three factors together

Optimization problems

Flow (locally biased solution)

$$\begin{aligned} &\text{minimize } \|Bx\|_{1,C} \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Spectral (locally biased solution)

$$\begin{aligned} &\text{minimize } \|Bx\|_{2,C} \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Optimization problems

Flow (locally biased solution)

$$\begin{aligned} &\text{minimize } \|Bx\|_{1,C} \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Spectral (locally biased solution)

$$\begin{aligned} &\text{minimize } \|Bx\|_{2,C} \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Flow (sparse locally biased sol.)

$$\begin{aligned} &\text{minimize } \|Bx\|_{1,C} + \epsilon\|x\|_1 \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Spectral (sparse locally biased sol.)

$$\begin{aligned} &\text{minimize } \|Bx\|_{2,C} + \epsilon\|Dx\|_1 \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Optimization problems and CS algorithms

Flow-Improve (weakly local)

$$\begin{aligned} &\text{minimize} \|Bx\|_{1,C} \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

MOV (weakly local spectral)

$$\begin{aligned} &\text{minimize } \|Bx\|_{2,C} \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Local Flow-Improve (strongly local)

$$\begin{aligned} &\text{minimize } \|Bx\|_{1,C} + \epsilon\|x\|_1 \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

ACL (strongly local spectral)

$$\begin{aligned} &\text{minimize } \|Bx\|_{2,C} + \epsilon\|Dx\|_1 \\ &\text{subject to } x_s = 1, x_t = 0 \end{aligned}$$

Optimization algorithms

ℓ_1-regularized spectral clustering and proximal gradient descent

"Exploiting optimization for local graph clustering", K. Fountoulakis, X. Cheng, J. Shun, F. Roosta and M. Mahoney (submitted)

The termination criterion of ACL is implied by the optimality conditions of

$$\begin{aligned} &\text{minimize } \|Bx\|_{2,C} + \rho\alpha\|Dx\|_1 \\ &\text{subject to: } x_s = 1, x_t = 0. \end{aligned}$$

In terms of algorithms we

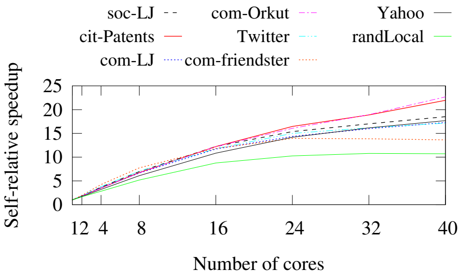
- prove that **proximal gradient descent is a strongly local algorithm**, which is not true in general.
- Running time: $\mathcal{O}$(volume at optimality) instead of $\mathcal{O}(vol(G))$
- Opens the door to accelerated prox. gradient descent

Parallel gradient descent

"Parallel local graph clustering", J. Shun, F. Roosta-Khorasani, K. Fountoulakis and M. Mahoney (submitted)

- State-of-the-art parallel implementations
- Up to 10x-20x speed up for 40 cores.

Input Graph	Num. Vertices	Num. Edges†
soc-LJ	4,847,571	42,851,237
cit-Patents	6,009,555	16,518,947
com-LJ	4,036,538	34,681,189
com-Orkut	3,072,627	117,185,083
Twitter	41,652,231	1,202,513,046
com-friendster	124,836,180	1,806,607,135
Yahoo	1,413,511,391	6,434,561,035
randLocal (synthetic)	10,000,000	49,100,524



Distributed coordinate descent and local spectral clustering

- Many spectral algorithms reduce to solving a linear system
  - *"Communication-avoiding coordinate descent methods for linear systems", A. Devarakonda, K. Fountoulakis, J. Demmel and M. Mahoney (will be made available soon)*

- Works for distributed memory parallel machines
- Decreases the number of messages and words moved by a factor of s
- s is a hyper-parameter

Let P be the number of machines available

Algorithm	Flops	Latency	Bandwidth
CD	$O\left(\frac{vol}{\alpha P}\right)$	$O(\frac{1}{\alpha}\log P)$	$O(\frac{vol}{\alpha}\log P)$
CA-CD	$O\left(\frac{vol \cdot s}{\alpha P} + \frac{s^2}{\alpha}\right)$	$O(\frac{1}{\alpha}\frac{\log P}{s})$	$O(\frac{s}{\alpha}\log P + \frac{vol}{\alpha s}\log P)$

Large-scale experiments will be in the paper!

Summary and future work

Summary

- We put all methods under a single optimization framework
  - which allows us to understand regularization properties of the solutions.
- We studied new optimization algorithms, parallel and distributed versions
  - which are scalable up billions of nodes.

Short term objectives

- Combine flow and spectral, i.e., good boundaries and well-connected internally clusters?

Long term objectives

- Better empirical understanding of each optimization problem
- Characterize the solutions space of each optimization problem
- Stochastic models for local graph clustering?
- Better visualization tools? Avoid the hairball effect.
- New state-of-the-art optimization algorithms
- Beyond social graphs?

Thank you!