Randomized Gossip Algorithms: New Insights

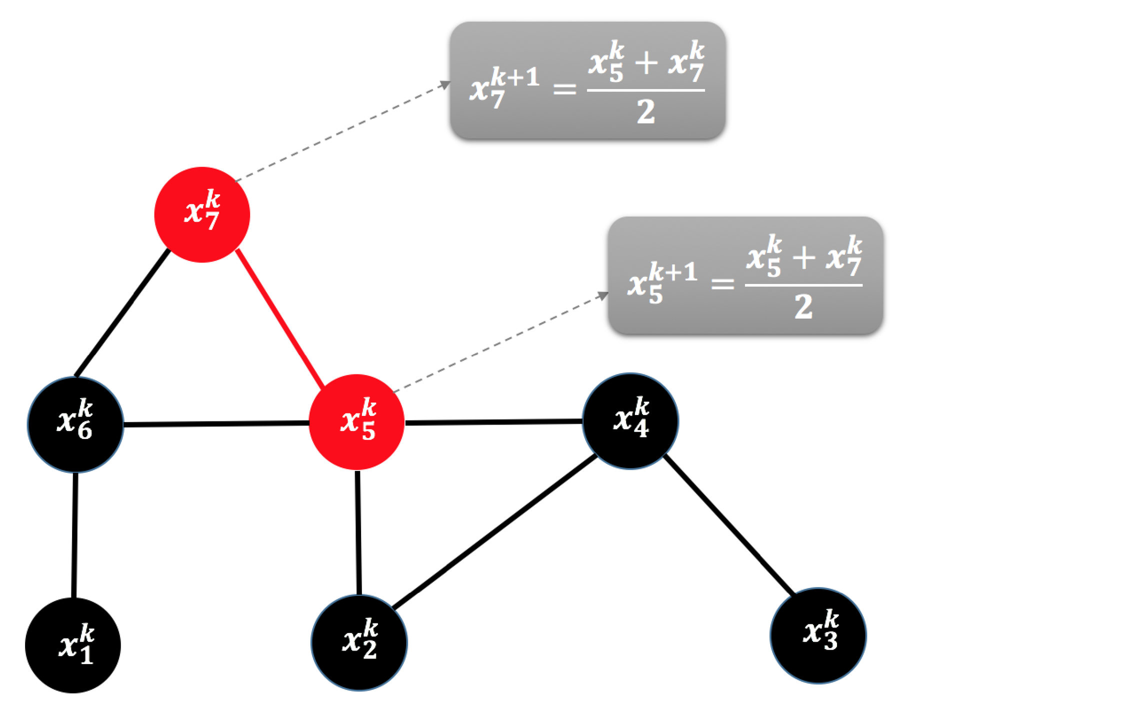
Nicolas Loizou & Peter Richtárik
School of Mathematics, University of Edinburgh

1. Average Consensus Problem (ACP)

SETUP: Let $G = (V, E)$ be a connected network with $|V| = n$ nodes (e.g., people) and $|E| = m$ edges (e.g., friendships). All nodes $i \in V$ store a private value $c_i \in \mathbb{R}$ (e.g., salary).

GOAL: Compute the average of the private values (i.e., the quantity $\bar{c} := \frac{1}{n}\sum_i c_i$) in a **distributed** fashion. That is, exchange of information can only occur along the edges.

2. Randomized Gossip (RG) Algorithm [1]



1. Set $x_i^0 = c_i$ for all nodes $i \in V$
2. Iterate for $k \geq 0$:
   (a) Pick a random edge $e = (i, j)$
   (b) Replace the values x_i^k and x_j^k by their average:

$$x_i^{k+1} \leftarrow \frac{x_i^k + x_j^k}{2}, \quad x_j^{k+1} \leftarrow \frac{x_i^k + x_j^k}{2}$$

Theorem [1]. $x_i^k \to \bar{c}$ for all $i \in V$ (in a probabilistic sense)

3. Optimization Formulation of ACP

The optimal solution of the optimization problem

$$\text{minimize} \quad \tfrac{1}{2}\|x - c\|^2 \quad \text{subject to} \quad x_i = x_j \quad \text{for all} \quad e = (i, j) \in E \qquad (1)$$

is $x_i^* = \bar{c}$ for all i. So, RG solves the above optimization problem. The constraints can be written compactly as $Ax = 0$, with each row of the system enforcing $x_i = x_j$ for one edge $(i, j) \in E$.

QUESTION: Does RG generalize beyond this? Can we get new variants of RG?

4. Duality for Linear Systems

Problem (1) a special case of this more general problem:

PRIMAL PROBLEM:

$$\text{(P)} \ \min_{x \in \mathbb{R}^n} \quad \tfrac{1}{2}\|x - c\|^2 \quad \text{s.t.} \quad Ax = b,$$

where A can be any matrix such that $Ax = b$ has a solution.

DUAL PROBLEM:

$$\text{(D)} \ \max_{y \in \mathbb{R}^m} \quad D(y) := (b - Ac)^\top y - \tfrac{1}{2}\|A^\top y\|^2$$

5. Stochastic Dual Ascent [2]

DUAL METHOD (SDA):

$$\boxed{y^{k+1} \leftarrow y^k + S\lambda^k}$$

where S is a random matrix with m rows, and λ^k is chosen so that $D(y^k + S\lambda^k)$ is maximized.

PRIMAL METHOD: With the dual iterates $\{y^k\}$ we can associate primal iterates $\{x^k\}$:

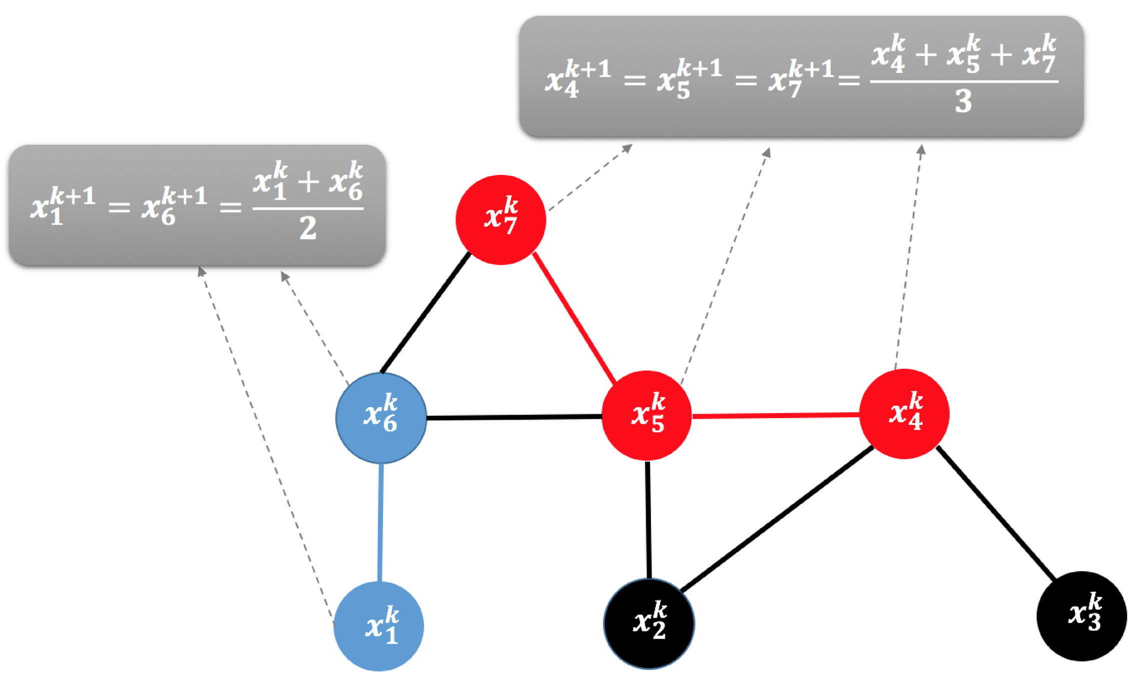
$$\boxed{x^k \leftarrow c + A^\top y^k}$$

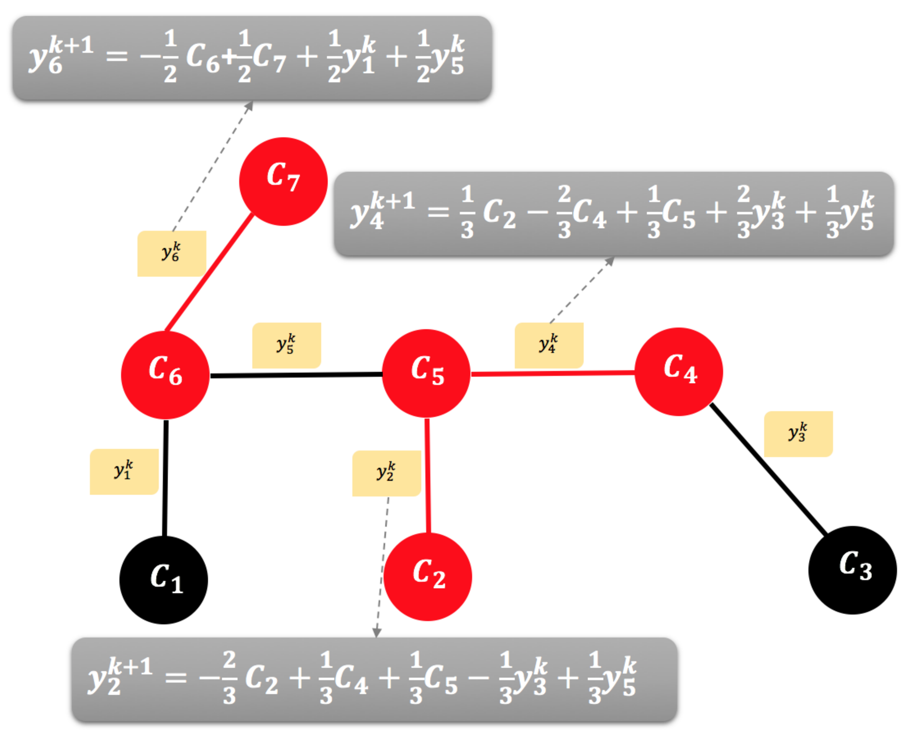
Theorem. If we choose $y^0 = 0$, $S = e^i$ (standard basis vector) with probability $p_i = 1/n$, then the primal iterates $\{x^k\}$ of SDA applied to (1) are identical to the RG method.

6. Randomized Block Kaczmarz (RBK) [4] and Randomized Newton (RN) [5] as Gossip

NEW GOSSIP METHODS: We can now formulate many new variants of RG, by applying SDA to (1) with various other choices of random matrices S. We also naturally obtain dual interpretation of such new gossip methods.

SETUP: Choose $S = I_{S_k}$, where I_{S_k} is a column submatrix of the $m \times m$ identity matrix corresponding to columns e belonging to a random set $S_k \subseteq E$.





Primal Iterates of SDA = Randomized Block Kaczmarz Algorithm

1. Form a subgraph G_k of G by selecting a random set of edges $S_k \subseteq E$
2. For each connected component of G_k, replace node values with their average

Dual Iterates of SDA = Randomized Newton Algorithm

1. Form a subgraph G_k of G by selecting a random set of edges $S_k \subseteq E$
2. Modify the dual variables y_e for $e \in S_k$ (see the image)

7. Convergence Rate

Theorem [2]. RN and RBK converge as:

$$\mathbf{E}[D(y^*) - D(y^k)] \leq \rho^k (D(y^*) - D(y^0)),$$

$$\mathbf{E}[\tfrac{1}{2}\|x^k - x^*\|^2] \leq \rho^k \tfrac{1}{2}\|x^0 - x^*\|^2,$$

where the rate is given by

$$\rho := 1 - \lambda_{\min}^+ \left(A^\top \mathbf{E}[I_{S_k}(I_{S_k}^\top A A^\top I_{S_k})^\dagger I_{S_k}^\top] A\right)$$

8. References

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[2] R. M. Gower and P. Richtárik. Stochastic dual ascent for solving linear systems. *arXiv:1512.06890*, 2015.

[3] N. Loizou and P. Richtárik. Randomized gossip algorithms: Complexity, duality and new variants. *(In Progress)*, 2016.

[4] D. Needell and J. A. Tropp. Paved with good intentions: analysis of a randomized block Kaczmarz method. *Linear Algebra and its Applications*, 441:199–221, 2014.

[5] Z. Qu, P. Richtárik, M. Takáč, and O. Fercoq. SDNA: stochastic dual Newton ascent for empirical risk minimization. *ICML*, 2016.