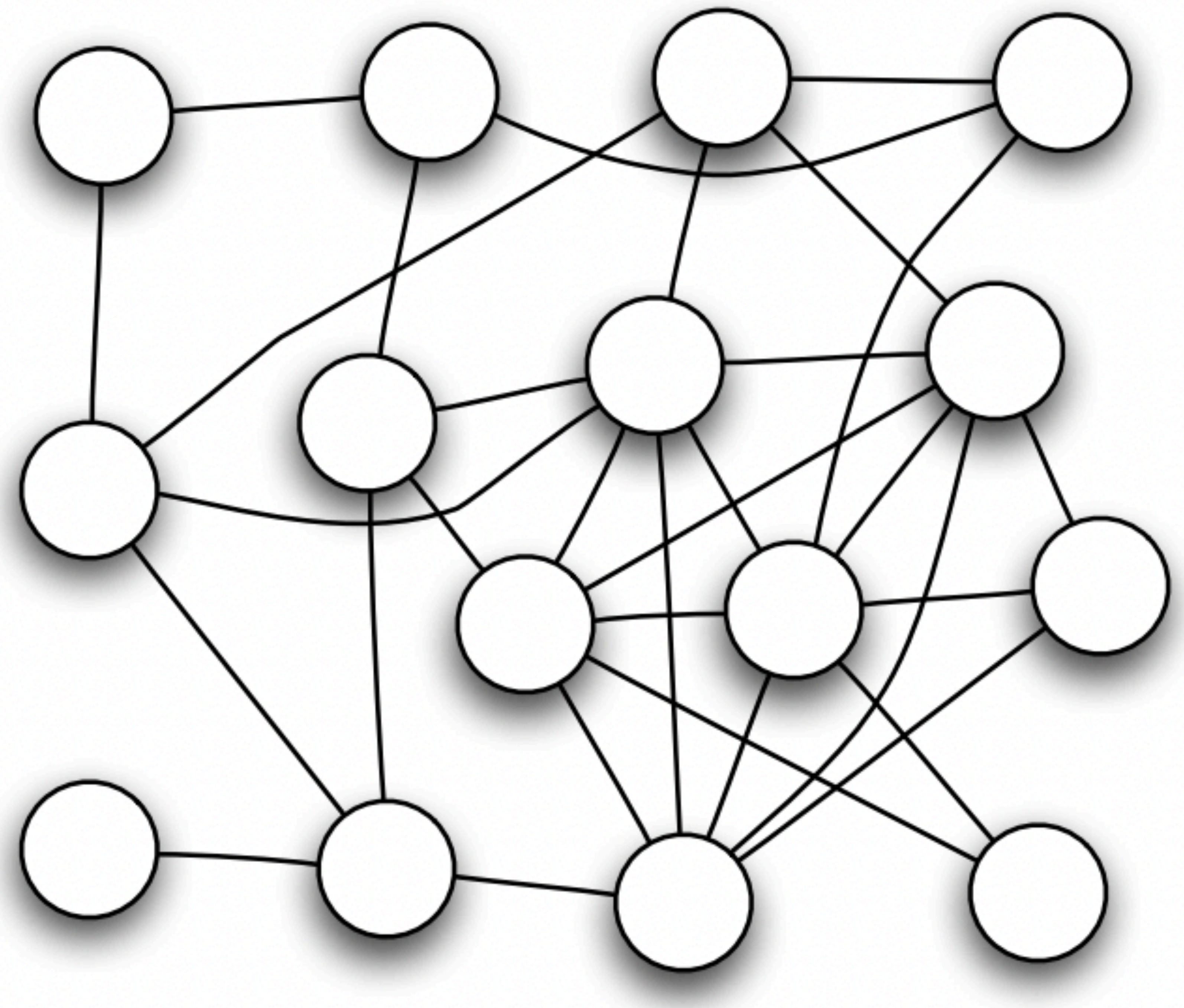
Decentralized optimization over slowly changing time-varying graphs

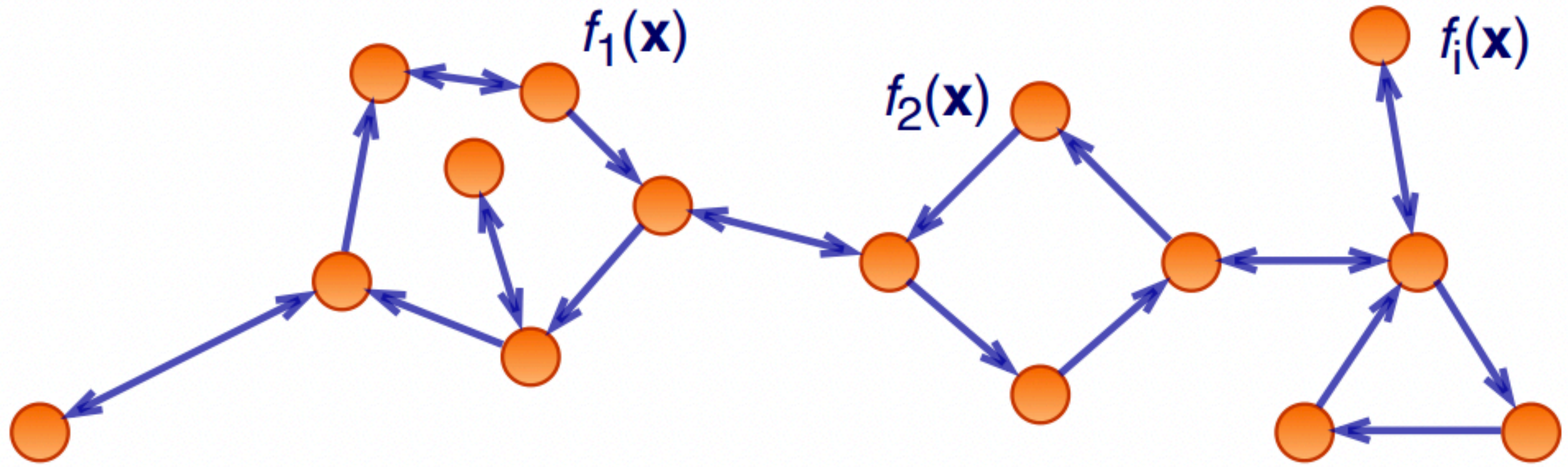
Metelev Dmitry, Sirius, 2022

The problem of decentralized optimization

$$\min_{x \in \mathbb{R}^d} f(x) = \frac{1}{m} \sum_{i=1}^{m} f_i(x).$$



Undirected network



Directed network

Consensus

$$x_i^{k+1} = w_{ii}^k + \sum_{(i,j)\in E^k} w_{ij}^k x_j^k.$$

Consensus from the view of a single agent

$$\mathbf{X}^{k+1} = \mathbf{W}^k \mathbf{X}^k$$

Matrix form of consensus

$$[W^k]_{ij} = \begin{cases} \frac{1}{\max(\deg(i),\deg(j))+1} & \text{if } (i,j) \in \mathcal{E}^k, \\ 0 & \text{if } (i,j) \notin \mathcal{E}^k \text{ and } i \neq j, \\ 1 - \sum_{j\neq i}[W^k]_{ij}, & i = j \end{cases}$$

Metropolis weights matrix

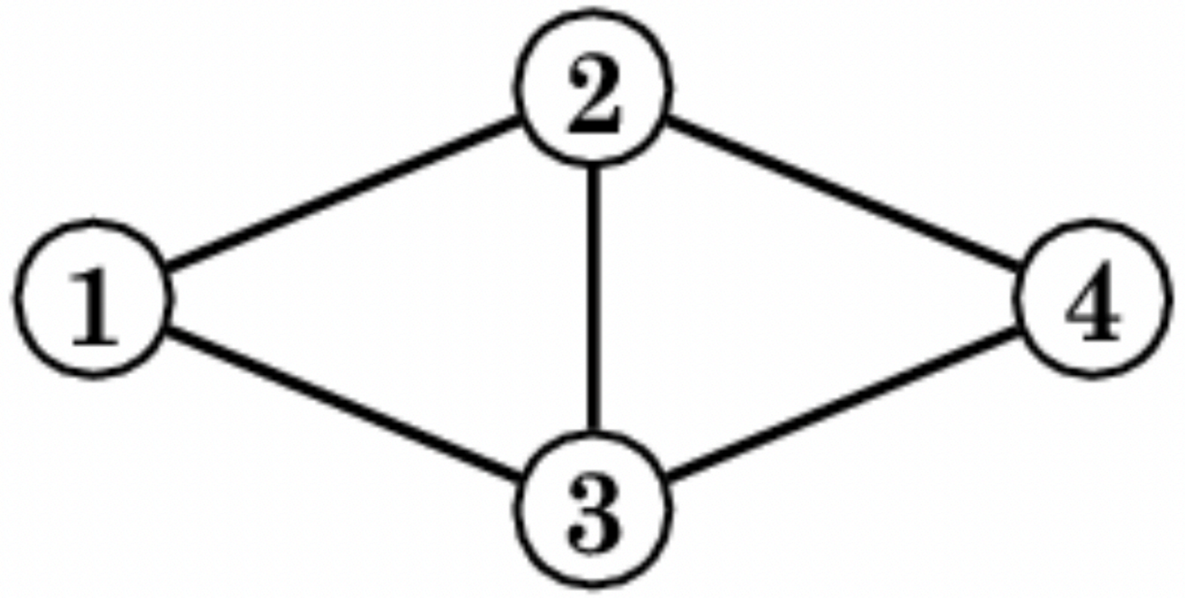
Properties of mixing matrixes and laplacian

- *(Decentralized property) If* $(i,j) \notin E_k$*, then* $[\mathbf{W}^k]_{ij} = 0$.
- *(Double stochasticity)* $\mathbf{W}^k \mathbf{1}_n = \mathbf{1}_n,\ \mathbf{1}_n^\top \mathbf{W}^k = \mathbf{1}_n^\top$.
- *(Contraction property) There exist* $\tau \in \mathbb{Z}_{++}$ *and* $\lambda \in (0,1)$ *such that for every* $k \geqslant \tau - 1$ *it holds*

$$\left\| \mathbf{W}_\tau^k \mathbf{X} - \overline{\mathbf{X}} \right\| \leqslant (1-\lambda) \left\| \mathbf{X} - \overline{\mathbf{X}} \right\|,$$

where $\mathbf{W}_\tau^k = \mathbf{W}^k \ldots \mathbf{W}^{k-\tau+1}$.

$$\mathbf{L}(\mathcal{G}^k) = \begin{cases} \deg(i), & i = j, \\ -1, & (i,j) \in \mathcal{E}^k, \\ 0, & (i,j) \notin \mathcal{E}^k \text{ and } i \neq j. \end{cases}$$



Laplacian Matrix

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}$$

$$\chi = \sup_{k \geq 0} \frac{\lambda_{\max}(\mathbf{L}(\mathcal{G}^k))}{\lambda_{\min}^+(\mathbf{L}(\mathcal{G}^k))}.$$

Static and tame-varying settings

Communication complexities

$$O\left(\sqrt{\chi\kappa}\log\left(\frac{1}{\varepsilon}\right)\right)$$

Upper bound for static graph

$$\Omega\left(\sqrt{\chi\kappa}\log\left(\frac{1}{\varepsilon}\right)\right)$$

Lower bound for static graph

$$O\left(\chi\sqrt{\kappa}\log\left(\frac{1}{\varepsilon}\right)\right)$$

Upper bound for time-varying graph

$$\Omega\left(\chi\sqrt{\kappa}\log\left(\frac{1}{\varepsilon}\right)\right)$$

Lower bound for time-varying graph

Remind on classic lower bound technique

$$f_{\mu,Q_f}(x) = \frac{\mu(Q_f - 1)}{8}\left\{(x^{(1)})^2 + \sum_{i=1}^{\infty}(x^{(i)} - x^{(i+1)})^2 - 2x^{(1)}\right\} + \frac{\mu}{2}\|x\|^2.$$

«Bad» function

$$q = \frac{\sqrt{Q_f} - 1}{\sqrt{Q_f} + 1}. \quad (x^*)^{(k)} = q^k$$

Solution for this problem

$$\|x_k - x^*\|^2 \geqslant \sum_{i=k+1}^{\infty}[(x^*)^{(i)}]^2 = \sum_{i=k+1}^{\infty} q^{2i} = \frac{q^{2(k+1)}}{1-q^2} = q^{2k}\|x_0 - x^*\|^2$$

$$\|x_k - x^*\|^2 \geqslant \left(\frac{\sqrt{Q_f} - 1}{\sqrt{Q_f} + 1}\right)^{2k} \|x_0 - x^*\|^2,$$

$$f(x_k) - f^* \geqslant \frac{\mu}{2}\left(\frac{\sqrt{Q_f} - 1}{\sqrt{Q_f} + 1}\right)^{2k} \|x_0 - x^*\|^2,$$

Lower bound for static graphs

$$\chi \sim n^2$$

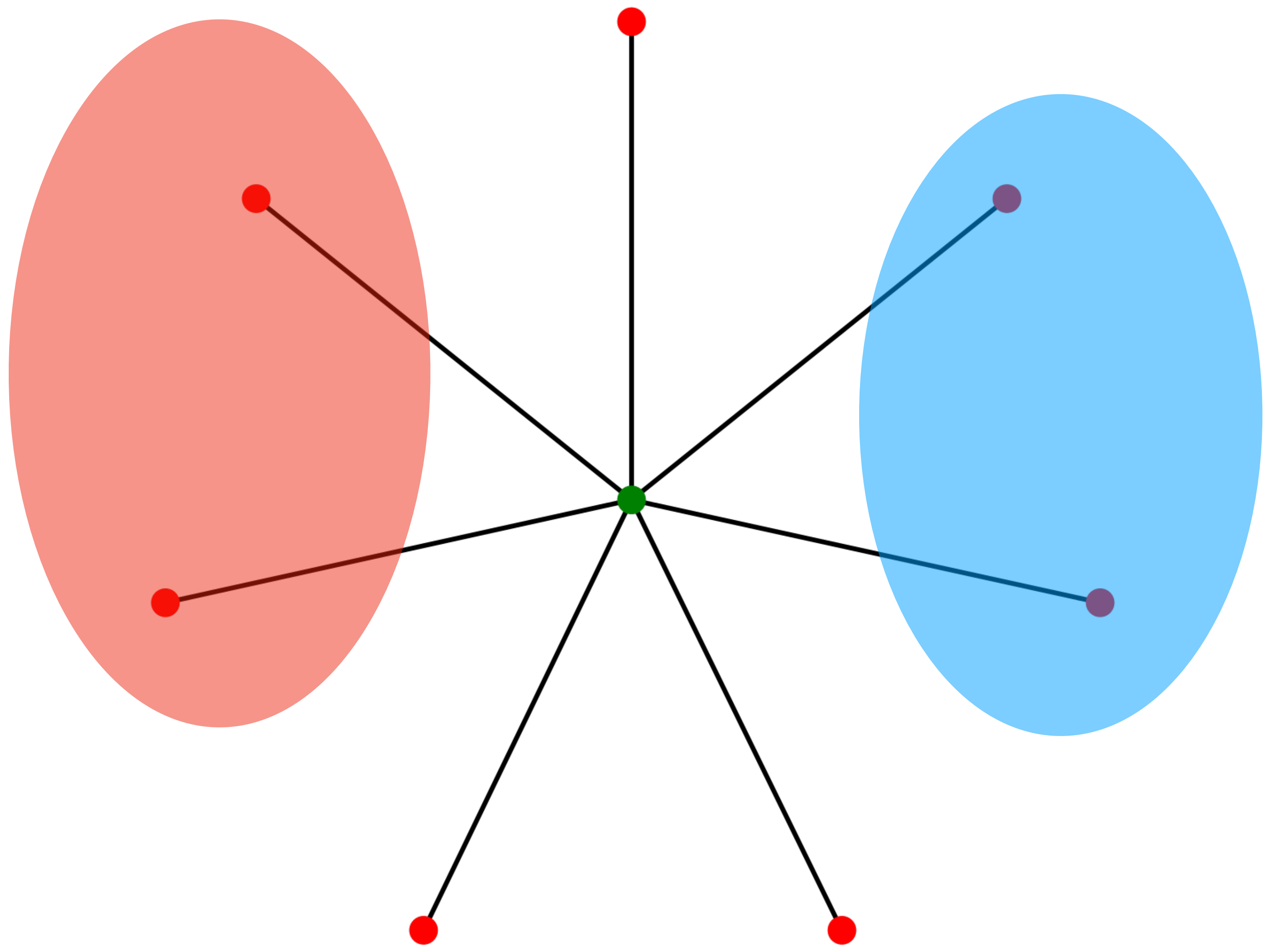
Characteristic number of the path

$$f_i^A(\theta) = \begin{cases} \frac{\alpha}{2n}\|\theta\|_2^2 + \frac{\beta-\alpha}{8|A|}(\theta^\top M_1 \theta - 2\theta_1) & \text{if } i \in A \\ \frac{\alpha}{2n}\|\theta\|_2^2 + \frac{\beta-\alpha}{8|A_d^c|}\theta^\top M_2 \theta & \text{if } i \in A_d^c \\ \frac{\alpha}{2n}\|\theta\|_2^2 & \text{otherwise} \end{cases}$$

Functions at nodes

Time for passing the information from one group to another: $\sqrt{\chi}$

Lower bound for time-varying graphs



$$\chi = n$$

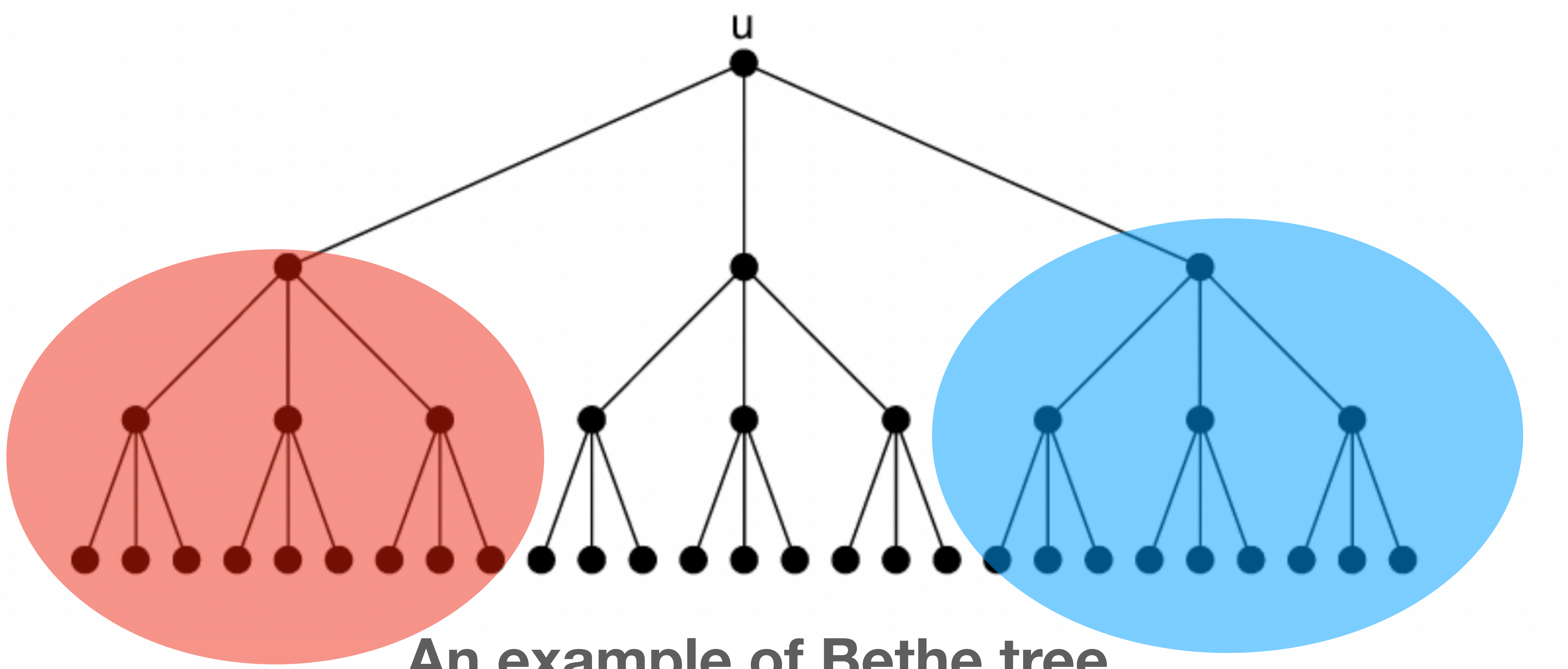
Characteristic number of the star

$$f_i(x) = \begin{cases} \frac{\mu}{2}\|x\|^2 + \frac{L-\mu}{4}\left[(x_{[1]}-1)^2 + \sum_{l=1}^{\infty}(x_{[2l]} - x_{[2l+1]})^2\right], & i \in \mathcal{V}_1 \\ \frac{\mu}{2}\|x\|^2, & i \in \mathcal{V}_2 \\ \frac{\mu}{2}\|x\|^2 + \frac{L-\mu}{4}\sum_{l=1}^{\infty}(x_{[2l-1]} - x_{[2l]})^2, & i \in \mathcal{V}_3 \end{cases}$$

Time for passing the information from one group to another: χ

Slowly changing graphs

A rooted Bethe tree $\mathscr{B}_{d,k}$ is an unweighted rooted tree of k levels in which the vertex root has degree d, the vertices in level 2 to level $(k-1)$ have degree $(d+1)$ and the vertices in level k have degree 1



An example of Bethe tree

$$\chi \sim n$$

Characteristic number of Bethe trees

Time for passing the information from one group to another: χ

Results

For any parameter alpha there exists an increasing sequence of Bethe graphs such that

$$change \leq 2\left(\frac{1}{\alpha}+1\right)n^{\alpha}$$

$$\Omega\left(\alpha\chi\sqrt{\kappa}\log\left(\frac{1}{\varepsilon}\right)\right)$$

For any $L > \mu > 0$, $\chi > 3$ exists a Graph such that

$$change \leq 6\log_2(n)$$

$$\Omega\left(\frac{\chi}{log(\chi)}\sqrt{\kappa}\log\left(\frac{1}{\varepsilon}\right)\right)$$