L$^\gamma$-PageRank for Semi-Supervised Learning on Graphs

Esteban Bautista

ENS de Lyon, LIP

In collaboration with P. Gonçalves, P. Abry, S. de Nigris, R. Fontugne

Jan 25, 2021

Motivation

The big data era

- **4.3 billion** people connected to Internet
- **36 million** web-sites created per minute
- **55 thousand** photos posted per minute
- **5.5 million** videos watched per minute
- **188 million** emails sent per minute

Goal: categorize the data

- Users by Service Provider
- Web-sites by topic
- Photographies by genre
- Videos by content
- Email is spam or not

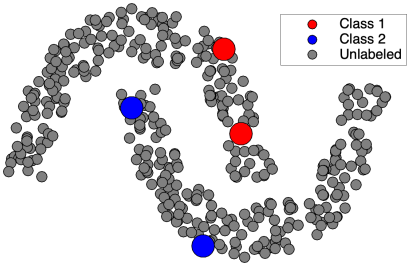
Sources: (1) DOMO's 'Data Never Sleeps 7.0' report, 2019. (2) Z. Zhan et al, 'Fast incremental PageRank on Dynamic networks', 2019.

Annotated data is limited

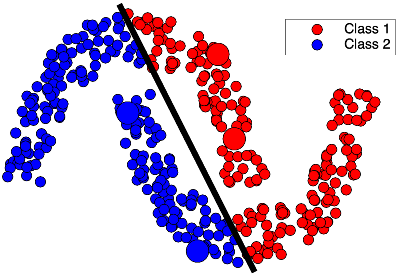
✓ **Classifiers**: learn from annotated data

✗ **Annotated data:** expensive to collect

Raw data



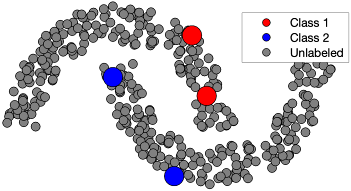
Supervised classifier



Graph-Based Semi-Supervised Learning (G-SSL)

Learning from labelled and unlabelled data

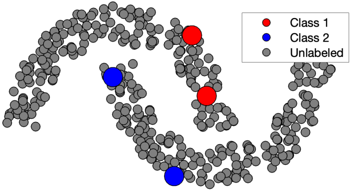
1 **Raw data**



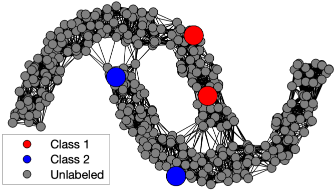
Graph-Based Semi-Supervised Learning (G-SSL)

Learning from labelled and unlabelled data

1 **Raw data**



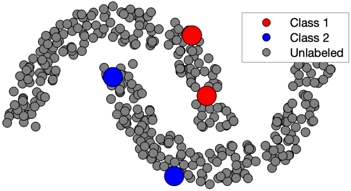
2 **Build similarity graph**



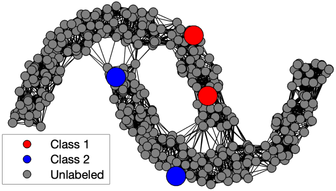
Graph-Based Semi-Supervised Learning (G-SSL)

Learning from labelled and unlabelled data

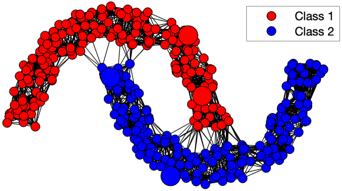
1 Raw data



2 Build similarity graph



3 Learned classes



Proposed works

Proposed methods

- [X. Zhu et al. 2003]
- [M. Belkin et al. 2003]
- [D. Zhou et al. 2004]
- [D. Zhou et al. 2007]
- [A. Subramanya et al. 2011]
- [X. Zhou et al. 2011]
- [M. Sokol et al. 2014]

etc.

Proposed works

Proposed methods

- [X. Zhu et al. 2003]
- [M. Belkin et al. 2003]
- [D. Zhou et al. 2004]
- [D. Zhou et al. 2007]
- [A. Subramanya et al. 2011]
- [X. Zhou et al. 2011]
- [M. Sokol et al. 2014]

etc.

Successful applications

- [A. Subramanya et al. 2008]
- [K. Avrachenkov et al. 2012]
- [M. Zhao et al. 2014]
- [W. Hu et al. 2016]
- [H. Cecotti, 2016]
- [F. De Morsier et al. 2016]
- [X. Li et al. 2018]

etc.

Proposed works

Proposed methods

- [X. Zhu et al. 2003]
- [M. Belkin et al. 2003]
- [D. Zhou et al. 2004]
- [D. Zhou et al. 2007]
- [A. Subramanya et al. 2011]
- [X. Zhou et al. 2011]
- [M. Sokol et al. 2014]

etc.

Successful applications

- [A. Subramanya et al. 2008]
- [K. Avrachenkov et al. 2012]
- [M. Zhao et al. 2014]
- [W. Hu et al. 2016]
- [H. Cecotti, 2016]
- [F. De Morsier et al. 2016]
- [X. Li et al. 2018]

etc.

State-of-the-art

PageRank-based G-SSL

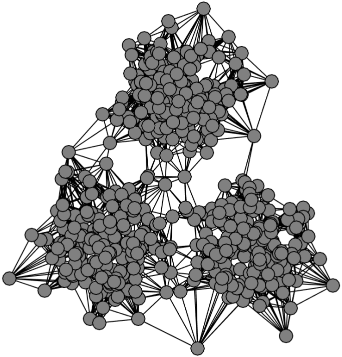
✓ Best overall performance
✓ Deep theoretical understanding
✓ Optimization problem
✓ Connection to the graph topology
✓ Swep-cut and multi-class decisions
✓ Robustness to parameters
✓ Efficient computation
✓ Efficient updating

PageRank-based G-SSL

Problem definition

Problem setup:

- $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$: weighted undirected graph with positive edges

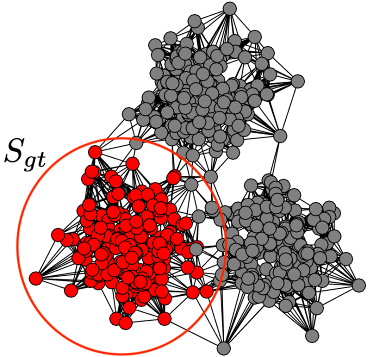


PageRank-based G-SSL

Problem definition

Problem setup:

- $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$: weighted undirected graph with positive edges
- $S_{gt} \subset \mathcal{V}$: ground truth class

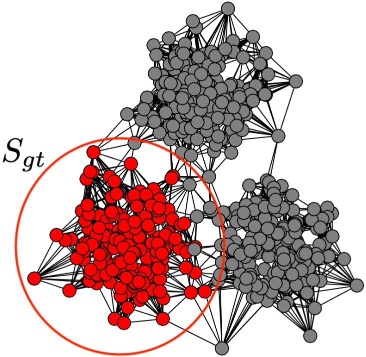


PageRank-based G-SSL

Problem definition

Problem setup:

- $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$: weighted undirected graph with positive edges
- $S_{gt} \subset \mathcal{V}$: ground truth class



Classification challenge:

- $\mathcal{V} = S_{gt} \cup S_{gt}^c$

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}}\{\mathbf{f}^\top \mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1}\mathbf{f} + \mu\left(\mathbf{f}-\mathbf{y}\right)^\top \mathbf{D}^{-1}\left(\mathbf{f}-\mathbf{y}\right)\}$$

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}}\{\mathbf{f}^\top \mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1}\mathbf{f} + \mu\,(\mathbf{f}-\mathbf{y})^\top\,\mathbf{D}^{-1}\,(\mathbf{f}-\mathbf{y})\}$$

- $\mathbf{f}$: personalized PageRank score vector

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}}\{\mathbf{f}^\top \mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1}\mathbf{f} + \mu\left(\mathbf{f}-\mathbf{y}\right)^\top \mathbf{D}^{-1}\left(\mathbf{f}-\mathbf{y}\right)\}$$

- $\mathbf{f}$: personalized PageRank score vector
- $\mathbf{y}$: one-hot encoding of labels of S_{gt} (personalization vector)

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}}\{\mathbf{f}^\top \mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1}\mathbf{f} + \mu\left(\mathbf{f}-\mathbf{y}\right)^\top \mathbf{D}^{-1}\left(\mathbf{f}-\mathbf{y}\right)\}$$

- $\mathbf{f}$: personalized PageRank score vector
- $\mathbf{y}$: one-hot encoding of labels of S_{gt} (personalization vector)
- $\mathbf{D} = diag(d_u)$, with $d_u = \sum_v \mathbf{W}_{uv}$: degree matrix

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}}\{\mathbf{f}^\top \mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1}\mathbf{f} + \mu\left(\mathbf{f}-\mathbf{y}\right)^\top \mathbf{D}^{-1}\left(\mathbf{f}-\mathbf{y}\right)\}$$

- $\mathbf{f}$: personalized PageRank score vector
- $\mathbf{y}$: one-hot encoding of labels of S_{gt} (personalization vector)
- $\mathbf{D} = diag(d_u)$, with $d_u = \sum_v \mathbf{W}_{uv}$: degree matrix
- $\mathbf{L} = \mathbf{D} - \mathbf{W}$: combinatorial Laplacian matrix

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}} \{ \underbrace{\mathbf{f}^\top \mathbf{D}^{-1} \mathbf{L} \mathbf{D}^{-1} \mathbf{f}}_{\text{smoothness}} + \mu \underbrace{(\mathbf{f} - \mathbf{y})^\top \mathbf{D}^{-1} (\mathbf{f} - \mathbf{y})}_{\text{fitting}} \}$$

- $\mathbf{f}$: personalized PageRank score vector
- $\mathbf{y}$: one-hot encoding of labels of S_{gt} (personalization vector)
- $\mathbf{D} = diag(d_u)$, with $d_u = \sum_v \mathbf{W}_{uv}$: degree matrix
- $\mathbf{L} = \mathbf{D} - \mathbf{W}$: combinatorial Laplacian matrix
- μ: regularization parameter

PageRank-based G-SSL

Definition

$$\arg\min_{\mathbf{f}}\{\underbrace{\mathbf{f}^\top\mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1}\mathbf{f}}_{\text{smoothness}}+\mu\underbrace{(\mathbf{f}-\mathbf{y})^\top\mathbf{D}^{-1}(\mathbf{f}-\mathbf{y})}_{\text{fitting}}\}$$

- $\mathbf{f}$: personalized PageRank score vector
- $\mathbf{y}$: one-hot encoding of labels of S_{gt} (personalization vector)
- $\mathbf{D} = diag(d_u)$, with $d_u = \sum_v \mathbf{W}_{uv}$: degree matrix
- $\mathbf{L} = \mathbf{D} - \mathbf{W}$: combinatorial Laplacian matrix
- μ: regularization parameter

Analytic closed form solution

$$\mathbf{f} = \mu\left(\mathbf{L}\mathbf{D}^{-1} + \mu\mathbb{I}\right)^{-1}\mathbf{y}$$

PageRank-based G-SSL

Interpretation as a random walk process

$$\mathbf{f} = \mu \left(\mathbf{L}\mathbf{D}^{-1} + \mu \mathbb{I}\right)^{-1} \mathbf{y} = \sum_{k=0}^{\infty} (1-\alpha)\alpha^k [\mathbf{P}^\top]^k \mathbf{y}$$

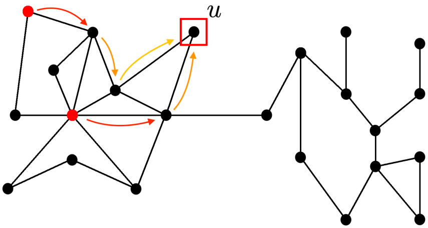
- $\alpha = 1/(1+\mu)$
- $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$

PageRank-based G-SSL

Interpretation as a random walk process

$$\mathbf{f} = \mu \left(\mathbf{L}\mathbf{D}^{-1} + \mu\mathbb{I}\right)^{-1}\mathbf{y} = \sum_{k=0}^{\infty}(1-\alpha)\alpha^k[\mathbf{P}^\top]^k\mathbf{y}$$

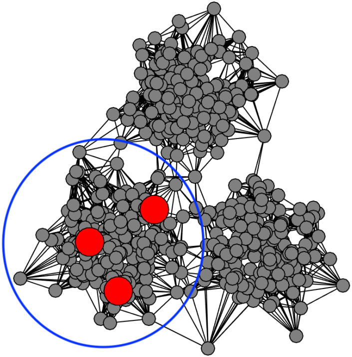
- $\alpha = 1/(1+\mu)$
- $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$

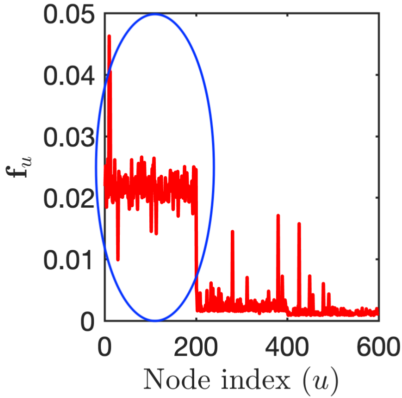


- $k = 0$: walker at label
- $k+1$: continue or restart
- $\mathbf{f}_u \propto \#$ visits to u

PageRank-based G-SSL

Illustration

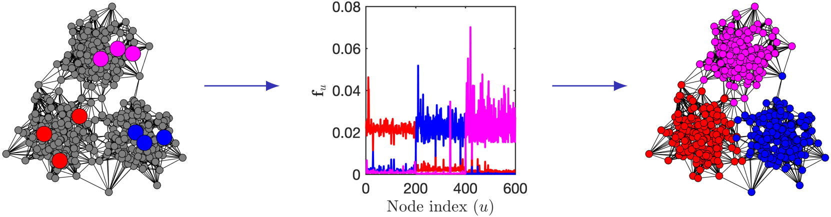




PageRank-based G-SSL

How to classify nodes from f?

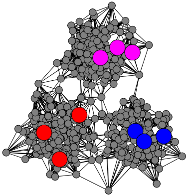
① **Multi-class**

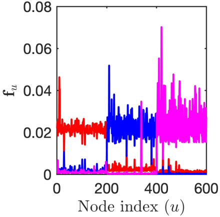


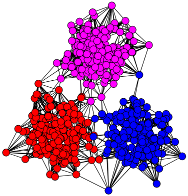
PageRank-based G-SSL

How to classify nodes from f?

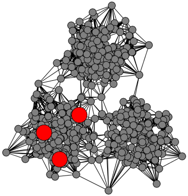
① **Multi-class**

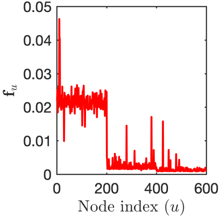


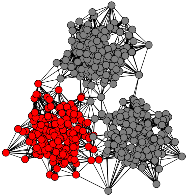




② **Sweep-cut**







PageRank-based G-SSL

Theoretical guarantees

Conductance

$$h_S := \frac{\sum_{u \in S} \sum_{v \in S^c} \mathbf{W}_{uv}}{\min\left(\sum_{u \in S} \mathbf{D}_{uu}, \sum_{v \in S^c} \mathbf{D}_{vv}\right)}$$

- $S \subseteq \mathcal{V}$
- $S^* = \arg\min_S h_S$ maximizes internal connections and minimizes external ones

PageRank-based G-SSL

Theoretical guarantees

Conductance

$$h_S := \frac{\sum_{u\in S}\sum_{v\in S^c} \mathbf{W}_{uv}}{\min\left(\sum_{u\in S}\mathbf{D}_{uu}, \sum_{v\in S^c}\mathbf{D}_{vv}\right)}$$

- $S \subseteq \mathcal{V}$
- $S^* = \arg\min_S h_S$ maximizes internal connections and minimizes external ones

Lemma [F. Chung, Internet Mathematics 2007]

For randomly placed labelled points, PageRank satisfies:

$$\mathbb{E}\left[\sum_{v\in S_{gt}^c} \mathbf{f}_v\right] \leq \frac{h_{S_{gt}}}{\mu}$$

PageRank-based G-SSL

Limitations

- ✗ Only reliable on highly clusterable data
- ✗ Biased in unbalanced labelled situations
- ✗ Bad learning with hubs/skewed graphs

Proposition

$\mathbf{L}^{\gamma}$-PageRank for Semi-supervised Learning

The $\mathbf{L}^{\gamma}$-graphs

Definition

$$\mathbf{L}^{\gamma} = \mathbf{Q}\mathbf{\Lambda}^{\gamma}\mathbf{Q}^{\top} = \mathbf{D}_{\gamma} - \mathbf{W}_{\gamma}$$

- $\mathbf{D}_{\gamma}$ is a new degree matrix: $[\mathbf{D}_{\gamma}]_{uu} = [\mathbf{L}^{\gamma}]_{uu}$
- $\mathbf{W}_{\gamma}$ is a new adjacency matrix: $[\mathbf{W}_{\gamma}]_{uv} = -[\mathbf{L}^{\gamma}]_{uv}$

The $\mathbf{L}^{\gamma}$-graphs

Definition

$$\mathbf{L}^{\gamma} = \mathbf{Q}\mathbf{\Lambda}^{\gamma}\mathbf{Q}^{\top} = \mathbf{D}_{\gamma} - \mathbf{W}_{\gamma}$$

- $\mathbf{D}_{\gamma}$ is a new degree matrix: $[\mathbf{D}_{\gamma}]_{uu} = [\mathbf{L}^{\gamma}]_{uu}$
- $\mathbf{W}_{\gamma}$ is a new adjacency matrix: $[\mathbf{W}_{\gamma}]_{uv} = -[\mathbf{L}^{\gamma}]_{uv}$

Lemma 1

For all $\gamma > 0$, the $\mathbf{L}^{\gamma}$-graphs satisfy the Laplacian property

$$[\mathbf{D}_{\gamma}]_{uu} = \sum_{v} [\mathbf{W}_{\gamma}]_{uv} \geq 0$$

The $\mathbf{L}^\gamma$-graphs

Definition

$$\mathbf{L}^\gamma = \mathbf{Q}\boldsymbol{\Lambda}^\gamma\mathbf{Q}^\top = \mathbf{D}_\gamma - \mathbf{W}_\gamma$$

- $\mathbf{D}_\gamma$ is a new degree matrix: $[\mathbf{D}_\gamma]_{uu} = [\mathbf{L}^\gamma]_{uu}$
- $\mathbf{W}_\gamma$ is a new adjacency matrix: $[\ \mathbf{W}_\gamma]_{uv} = -[\mathbf{L}^\gamma]_{uv}$

Lemma 1

For all $\gamma > 0$, the $\mathbf{L}^\gamma$-graphs satisfy the Laplacian property

$$[\mathbf{D}_\gamma]_{uu} = \sum_v [\mathbf{W}_\gamma]_{uv} \geq 0$$

For every fixed $\gamma > 0$, $\mathbf{L}^\gamma$ codes for a new graph

The $\mathbf{L}^{\gamma}$-PageRank G-SSL

Extending PageRank to $\mathbf{L}^{\gamma}$-graphs

Definition

$$\mathit{arg}\min_{\mathbf{f}} \left\{ \mathbf{f}^{\top}\mathbf{D}_{\gamma}^{-1}\mathbf{L}^{\gamma}\mathbf{D}_{\gamma}^{-1}\mathbf{f} + \mu\left(\mathbf{f}-\mathbf{y}\right)^{\top}\mathbf{D}_{\gamma}^{-1}\left(\mathbf{f}-\mathbf{y}\right) \right\}$$

The $\mathbf{L}^{\gamma}$-PageRank G-SSL

Extending PageRank to $\mathbf{L}^{\gamma}$-graphs

Definition

$$\mathit{arg}\min_{\mathbf{f}} \left\{ \mathbf{f}^{\top}\mathbf{D}_{\gamma}^{-1}\mathbf{L}^{\gamma}\mathbf{D}_{\gamma}^{-1}\mathbf{f} + \mu\left(\mathbf{f}-\mathbf{y}\right)^{\top}\mathbf{D}_{\gamma}^{-1}\left(\mathbf{f}-\mathbf{y}\right) \right\}$$

- $\mathbf{f}$: $\mathbf{L}^{\gamma}$-PageRank score vector

The $\mathbf{L}^{\gamma}$-PageRank G-SSL

Extending PageRank to $\mathbf{L}^{\gamma}$-graphs

Definition

$$\textit{arg}\min_{\mathbf{f}} \left\{ \mathbf{f}^{\top} \mathbf{D}_{\gamma}^{-1} \mathbf{L}^{\gamma} \mathbf{D}_{\gamma}^{-1} \mathbf{f} + \mu \left(\mathbf{f} - \mathbf{y} \right)^{\top} \mathbf{D}_{\gamma}^{-1} \left(\mathbf{f} - \mathbf{y} \right) \right\}$$

- $\mathbf{f}$: $\mathbf{L}^{\gamma}$-PageRank score vector

Analytic closed form solution

$$\mathbf{f} = \mu \left(\mathbf{L}^{\gamma} \mathbf{D}_{\gamma}^{-1} + \mu \mathbb{I} \right)^{-1} \mathbf{y}$$

The $\mathbf{L}^\gamma$-PageRank G-SSL

Extending PageRank to $\mathbf{L}^\gamma$-graphs

Definition

$$\mathit{arg}\min_{\mathbf{f}} \left\{ \mathbf{f}^\top \mathbf{D}_\gamma^{-1} \mathbf{L}^\gamma \mathbf{D}_\gamma^{-1} \mathbf{f} + \mu \left(\mathbf{f} - \mathbf{y} \right)^\top \mathbf{D}_\gamma^{-1} \left(\mathbf{f} - \mathbf{y} \right) \right\}$$

- $\mathbf{f}$: $\mathbf{L}^\gamma$-PageRank score vector

Analytic closed form solution

$$\begin{aligned} \mathbf{f} &= \mu \left(\mathbf{L}^\gamma \mathbf{D}_\gamma^{-1} + \mu \mathbb{I} \right)^{-1} \mathbf{y} \\ &= \sum_{k=0}^{\infty} (1-\alpha) \alpha^k [\mathbf{P}_\gamma{}^\top]^k \mathbf{y} \end{aligned}$$

- $\alpha = 1/(1+\mu)$
- $\mathbf{P}_\gamma = \mathbf{D}_\gamma^{-1} \mathbf{W}_\gamma$: Generalized random walk transition matrix?

Two new regimes arise

1. $\gamma = 1$: **Standard PageRank**
   - $\mathbf{P}_\gamma$: Standard random walk

2. $\gamma < 1$: **Lévy Flights for Classification**
   - $\mathbf{P}_\gamma$: Lévy flight random walk

3. $\gamma > 1$: **Signed Graphs for Classification**
   - $\mathbf{P}_\gamma$: Not a stochastic matrix ($\mathbf{W}_\gamma$ signed)

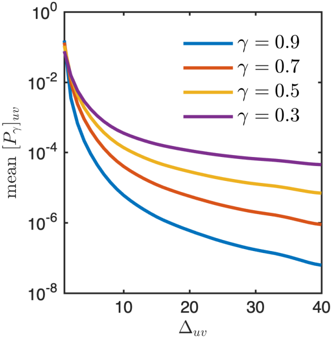
Regime $0 < \gamma < 1$

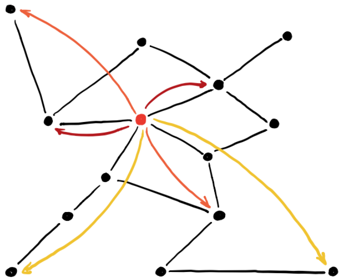
Lévy flights for classification

The Lévy flight random walk

The long range transitions

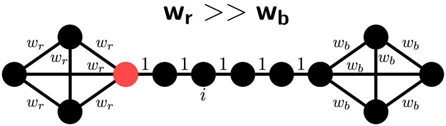
$$[\mathbf{P}_\gamma]_{uv} \sim \Delta_{uv}^{-(2\gamma+1)}$$

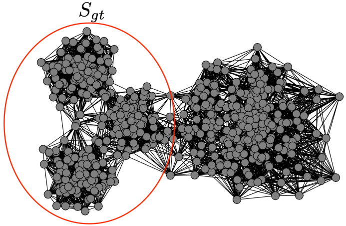


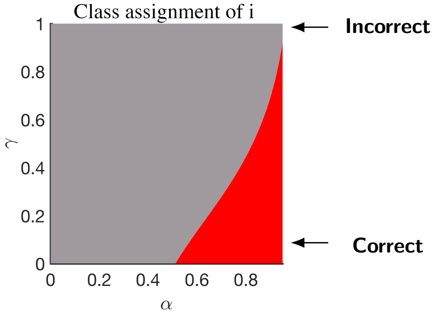


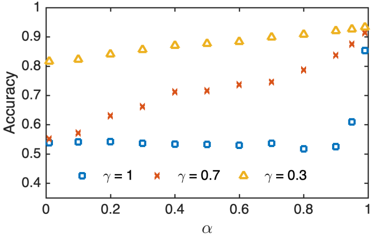
The L^γ-PageRank G-SSL

The Lévy flight random walk for classification









Regime $\gamma > 1$

Signed graphs for classification

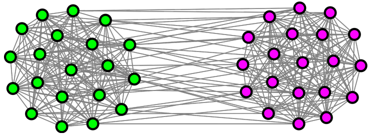
The graph arising from $\mathbf{L}^2 = \mathbf{D}_2 - \mathbf{W}_2$

$$[\mathbf{W}_2]_{uv} = \underbrace{(\mathbf{D}_{uu} + \mathbf{D}_{vv})\mathbf{W}_{uv}}_{\text{1 hop (positive)}} - \sum_{l \neq u,v} \underbrace{\mathbf{W}_{ul}\mathbf{W}_{lv}}_{\text{2 hop (negative)}}$$

The graph arising from $\mathbf{L}^2 = \mathbf{D}_2 - \mathbf{W}_2$

$$[\mathbf{W}_2]_{uv} = \underbrace{(\mathbf{D}_{uu} + \mathbf{D}_{vv})\mathbf{W}_{uv}}_{\text{1 hop (positive)}} - \sum_{l \neq u,v} \underbrace{\mathbf{W}_{ul}\mathbf{W}_{lv}}_{\text{2 hop (negative)}}$$

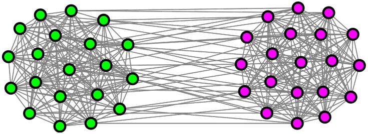
1. **Initial graph**



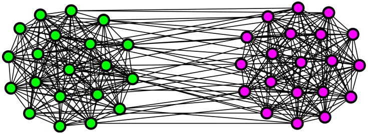
The graph arising from $\mathbf{L}^2 = \mathbf{D}_2 - \mathbf{W}_2$

$$[\mathbf{W}_2]_{uv} = \underbrace{(\mathbf{D}_{uu} + \mathbf{D}_{vv})\mathbf{W}_{uv}}_{\text{1 hop (positive)}} - \sum_{l \neq u,v} \underbrace{\mathbf{W}_{ul}\mathbf{W}_{lv}}_{\text{2 hop (negative)}}$$

① **Initial graph**



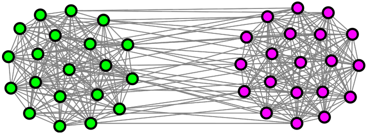
② **Positive edges in $\mathbf{W}_2$**
(Agreements)



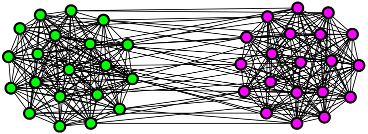
The graph arising from $\mathbf{L}^2 = \mathbf{D}_2 - \mathbf{W}_2$

$$[\mathbf{W}_2]_{uv} = \underbrace{(\mathbf{D}_{uu} + \mathbf{D}_{vv})\mathbf{W}_{uv}}_{\text{1 hop (positive)}} - \sum_{l \neq u,v} \underbrace{\mathbf{W}_{ul}\mathbf{W}_{lv}}_{\text{2 hop (negative)}}$$

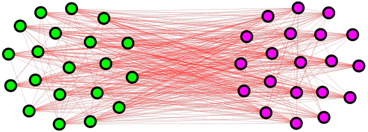
1. **Initial graph**



2. **Positive edges in $\mathbf{W}_2$**
   (Agreements)



3. **Negative edges in $\mathbf{W}_2$**
   (Disagreements)



Clusters in L$^\gamma$-graphs

Definition

Group of nodes $S \subset \mathcal{V}$ with:

- Large

$$\mathcal{A}_{in}(S) = \sum_{u \in S} \sum_{w \in S} |[\mathbf{W}_\gamma^+]_{uw}|$$

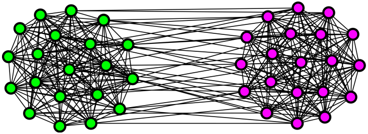
$$\mathcal{D}_{out}(S) = \sum_{u \in S} \sum_{v \in S^c} |[\mathbf{W}_\gamma^-]_{uv}|$$

- Small

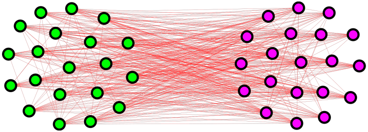
$$\mathcal{A}_{out}(S) = \sum_{u \in S} \sum_{v \in S^c} |[\mathbf{W}_\gamma^+]_{uv}|$$

$$\mathcal{D}_{in}(S) = \sum_{u \in S} \sum_{w \in S} |[\mathbf{W}_\gamma^-]_{uw}|$$

Agreements



Disagreements



Generalizing the conductance

Definition

$$h_S^{(\gamma)} = \frac{\sum_{u \in S} \sum_{v \in S^c} [\mathbf{W}_\gamma]_{uv}}{\min\left(\sum_{u \in S} [\mathbf{D}_\gamma]_{uu}, \sum_{v \in S^c} [\mathbf{D}_\gamma]_{vv}\right)} \geq 0$$

Generalizing the conductance

Definition

$$h_S^{(\gamma)} = \frac{\sum_{u\in S}\sum_{v\in S^c}[\mathbf{W}_\gamma]_{uv}}{\min\left(\sum_{u\in S}[\mathbf{D}_\gamma]_{uu}, \sum_{v\in S^c}[\mathbf{D}_\gamma]_{vv}\right)} \geq 0$$

Lemma 2

For a fixed γ, let $S^* = \arg min_S h_S^{(\gamma)}$. Then, S^* also

- Maximizes $\mathcal{A}_{in}(S^*)$ and $\mathcal{D}_{out}(S^*)$
- Minimizes $\mathcal{A}_{out}(S^*)$ and $\mathcal{D}_{in}(S^*)$

L$^\gamma$-PageRank as a dynamic process

L$^\gamma$-PageRank satisfies the following properties

1. **Mass preservation**: $\sum_{u \in \mathcal{V}} \mathbf{f}_u = \sum_{u \in \mathcal{V}} \mathbf{y}_u$

2. **Stationarity**: $\mathbf{f} = \pi_\gamma$ if $\mathbf{y} = \pi_\gamma$
   where $\pi_\gamma = [\mathbf{D}_\gamma]_{uu} / \sum_{u \in S} [\mathbf{D}_\gamma]_{uu}$

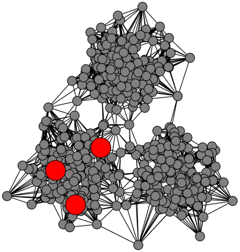
3. **Limit behavior**: $\mathbf{f} \to \pi_\gamma$ as $\mu \to 0$ and $\mathbf{f} \to \mathbf{y}$ as $\mu \to \infty$

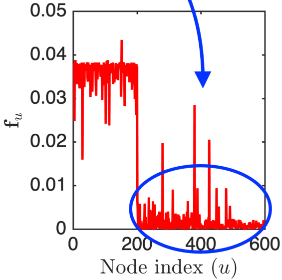
Confinement of scores

Lemma 3

For randomly placed labelled points, the $\mathbf{L}^\gamma$-PageRank vector satisfies:

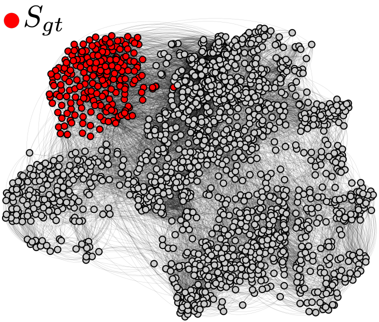
$$\mathbb{E}\left[\sum_{u \in S_{gt}^c} \mathbf{f}_u\right] \leq \frac{h_{S_{gt}}^{(\gamma)}}{\mu}$$

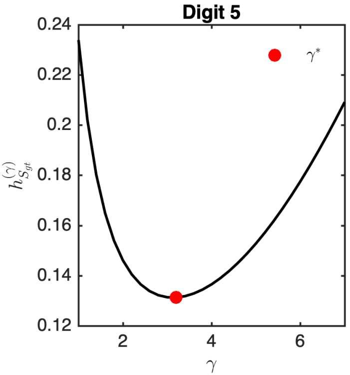




The influence of γ

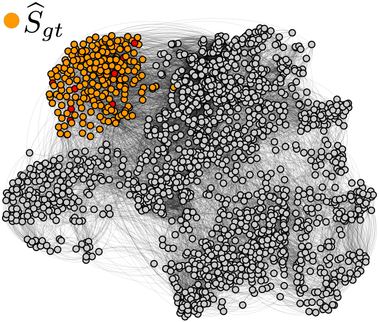
Optimal value appears: $\gamma^* = arg\min_\gamma h_{S_{gt}}^{(\gamma)}$

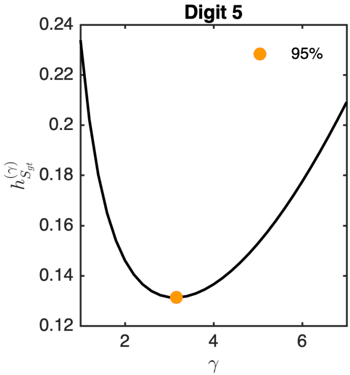




How to find γ^*?

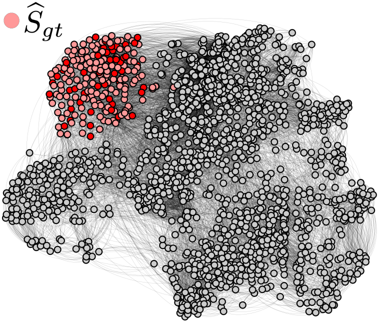
Optimal value on subsets of S_{gt}

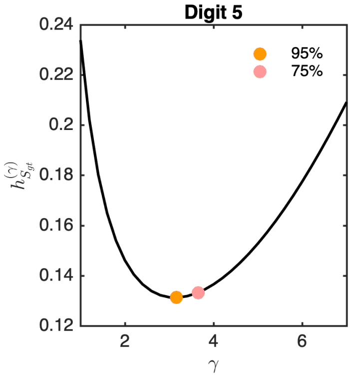




How to find γ^*?

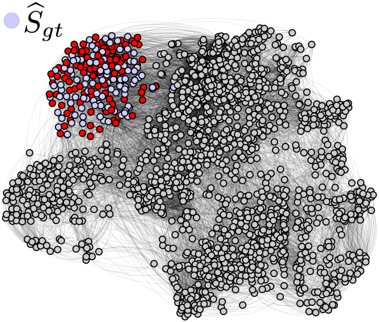
Optimal value on subsets of S_{gt}

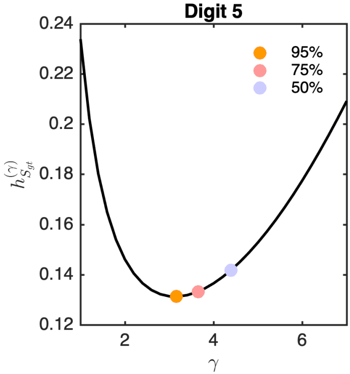




How to find γ^*?

Optimal value on subsets of S_{gt}





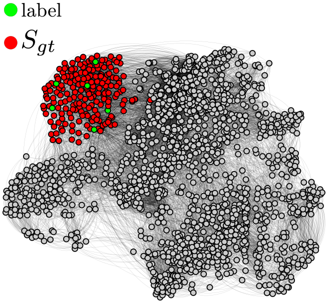
The estimation of γ^*

1. **Compute k: maximum distance between labeled points**

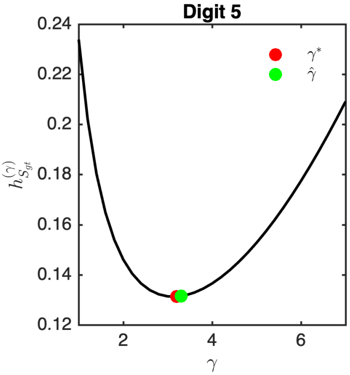
2. **Run walkers starting on the labels for k steps**

3. **Use nodes where it is 0.7 more likely to find the walkers as proxy $\widehat{S}_{gt}$**

4. **Compute $\widehat{\gamma}$ on $\widehat{S}_{gt}$**



Algorithm assessment



$\hat{\gamma}$ by algorithm

Algorithm assessment

Digit	1	2	3	4	5	6	7	8	9
γ^*	7.0	3.0	7.0	3.2	3.2	7.0	7.0	3.2	4.2
$\hat{\gamma}$	5.45 (0.15)	3.10 (0.14)	6.41 (0.11)	4.92 (0.16)	3.21 (0.14)	6.04 (0.15)	4.98 (0.17)	4.40 (0.18)	5.08 (0.15)
$h_{S_{gt}}^{(\gamma^*)}$	0.065	0.166	0.035	0.141	0.131	0.011	0.052	0.116	0.135
$h_{S_{gt}}^{(\hat{\gamma})}$	0.073 (9e-4)	0.174 (8e-4)	0.041 (1e-3)	0.185 (4e-3)	0.148 (2e-3)	0.017 (1e-3)	0.074 (2e-3)	0.142 (2e-3)	0.149 (9e-4)
$h_{S_{gt}}^{(1)}$	0.175	0.248	0.216	0.258	0.233	0.107	0.203	0.215	0.285

Table: Evaluation of Algorithm on the MNIST Dataset. Mean values (95% confidence interval) are shown.

Retrieving S_{gt} via a sweep-cut

Reducing complexity when searching S_{gt}

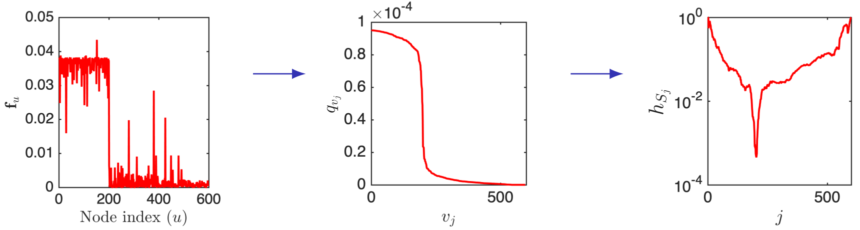
- Let $v_1, \ldots, v_N$ be the permutation: $\mathbf{q}_{v_i} = \mathbf{f}_{v_i}/[\mathbf{D}_\gamma]_{v_i,v_i} \geq \mathbf{q}_{v_{i+1}} = \mathbf{f}_{v_{i+1}}/[\mathbf{D}_\gamma]_{v_{i+1},v_{i+1}}$
- Let $S_j = \{v_1, \ldots, v_j\}$
- Retrieve $\hat{S}_{gt} = S_j$ for the set S_j achieving $\min_j h_{S_j}^{(\gamma)}$

Retrieving S_{gt} via a sweep-cut

Reducing complexity when searching S_{gt}

- Let $v_1, \ldots, v_N$ be the permutation: $\mathbf{q}_{v_i} = \mathbf{f}_{v_i}/[\mathbf{D}_\gamma]_{v_i,v_i} \geq \mathbf{q}_{v_{i+1}} = \mathbf{f}_{v_{i+1}}/[\mathbf{D}_\gamma]_{v_{i+1},v_{i+1}}$
- Let $S_j = \{v_1, \ldots, v_j\}$
- Retrieve $\hat{S}_{gt} = S_j$ for the set S_j achieving $\min_j h_{S_j}^{(\gamma)}$

$\mathbf{L}^\gamma$-PR vector **Permutation** **Sweep**

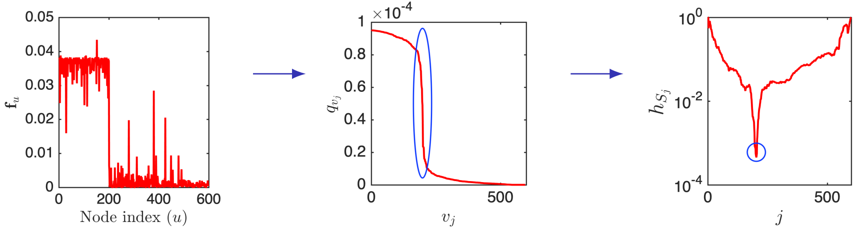


Retrieving S_{gt} via a sweep-cut

Reducing complexity when searching S_{gt}

- Let $v_1, \dots, v_N$ be the permutation: $\mathbf{q}_{v_i} = \mathbf{f}_{v_i}/[\mathbf{D}_\gamma]_{v_i,v_i} \geq \mathbf{q}_{v_{i+1}} = \mathbf{f}_{v_{i+1}}/[\mathbf{D}_\gamma]_{v_{i+1},v_{i+1}}$
- Let $S_j = \{v_1, \dots, v_j\}$
- Retrieve $\hat{S}_{gt} = S_j$ for the set S_j achieving $\min_j h_{S_j}^{(\gamma)}$

$\mathbf{L}^\gamma$-PR vector **Permutation** **Sweep**



A sharp drop implies a good cut

If there is sharp drop between $\mathbf{q}_j$ and $\mathbf{q}_{j+1}$, then S_j has small $h_{S_j}^{(\gamma)}$

Performance evaluation

Performance evaluation

Real world datasets (Sweep)

	S_{gt}	$\gamma = 1$	$\gamma = 2$	$\gamma = \hat{\gamma}$	$\gamma = \gamma^*$
MNIST	Digit 1	0.67 (0.075)	0.78 (0.032)	0.78 (0.034) [5.4]	0.80 (0.027) [7.0]
	Digit 2	0.38 (0.042)	0.60 (0.064)	0.64 (0.059) [3.3]	0.64 (0.059) [3.0]
	Digit 3	0.47 (0.040)	0.61 (0.032)	0.61 (0.028) [6.0]	0.61 (0.028) [7.0]
	Digit 4	0.39 (0.022)	0.48 (0.036)	0.53 (0.044) [4.7]	0.53 (0.037) [3.2]
	Digit 5	0.44 (0.036)	0.56 (0.046)	0.61 (0.036) [3.3]	0.64 (0.035) [3.2]
	Digit 6	0.90 (0.039)	0.94 (0.003)	0.94 (0.002) [6.0]	0.94 (0.002) [7.0]
	Digit 7	0.43 (0.027)	0.66 (0.043)	0.71 (0.042) [4.8]	0.75 (0.032) [7.0]
	Digit 8	0.47 (0.062)	0.65 (0.057)	0.74 (0.038) [4.8]	0.72 (0.050) [3.2]
	Digit 9	0.43 (0.020)	0.52 (0.026)	0.53 (0.023) [4.9]	0.56 (0.026) [4.2]
Gender images	Female	0.51 (0.039)	0.57 (0.028)	0.57 (0.020) [3.0]	0.57 (0.028) [2.0]
	Male	0.55 (0.028)	0.61 (0.021)	0.60 (0.022) [3.3]	0.61 (0.021) [2.4]
BBC articles	Business	0.80 (0.020)	0.53 (0.038)	0.72 (0.040) [1.3]	0.81 (0.021) [1.1]
	Entmt.	0.84 (0.027)	0.57 (0.040)	0.76 (0.047) [1.5]	0.86 (0.025) [1.3]
Phoneme	Nasal	0.37 (0.030)	0.41 (0.028)	0.43 (0.025) [2.9]	0.43 (0.025) [3.0]
	Oral	0.41 (0.025)	0.44 (0.022)	0.46 (0.019) [2.8]	0.46 (0.019) [3.0]

Table: γ enhances performance. Cells: MCC, 95% confidence interval (parenthesis) and the value of γ [squared brackets].

Performance evaluation

Ratio of labelled points: 3 to 1 (Multi-class)

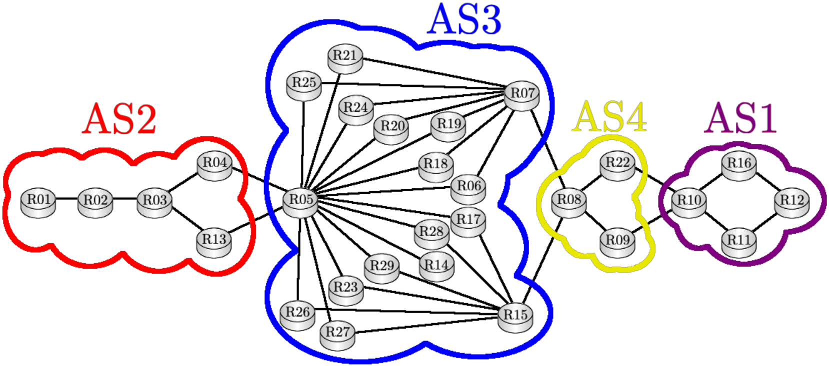
	Planted Partition	MNIST 4vs9	MNIST 3vs8	BBC articles	Gender images	Phoneme
$\gamma = 1$	0.81 (1.1e-2)	0.51 (1.5e-2)	0.70 (1.4e-2)	0.66 (1.8e-2)	0.63 (2.1e-2)	0.44 (2.3e-2)
$\gamma = 2$	0.87 (8.7e-3)	0.56 (1.5e-2)	0.76 (1.2e-2)	0.92 (5.0e-3)	0.73 (1.6e-2)	0.48 (1.4e-2)
$\gamma =$ Best	0.90 (7.0e-3) [6]	0.57 (1.5e-2) [3]	0.78 (1.2e-2) [4]	0.93 (1.5e-3) [3]	0.75 (1.7e-2) [3]	0.48 (1.4e-2) [1.9]

Table: γ enhances performance. Cells: MCC, 95% confidence interval (parenthesis) and the value of γ [squared brackets].

Application to Internet routing

L^γ-PageRank for Internet routing

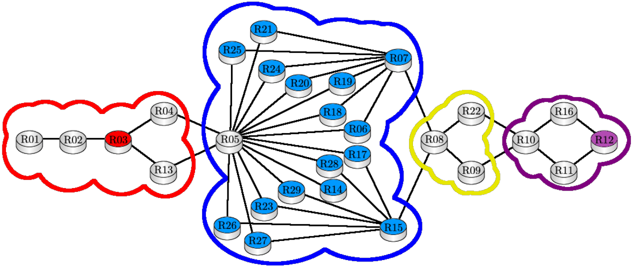
The Internet graph



L$^\gamma$-PageRank for Internet routing

Semi-supervision

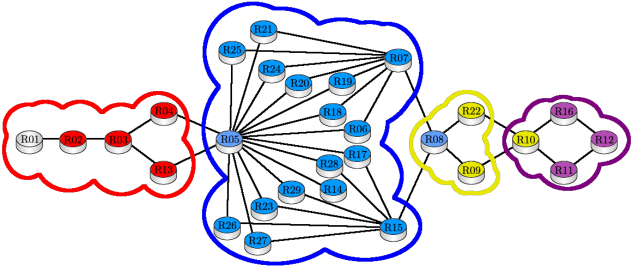
- Strict expert
  - ▷ High trust
  - ▷ Few labels



L^γ-PageRank for Internet routing

Semi-supervision

- Strict expert
  - ▷ High trust
  - ▷ Few labels
- Loose expert
  - ▷ Low trust
  - ▷ Lots of labels



L$^\gamma$-PageRank for Internet routing

Why interesting?

Problem arises when

- Congestion due to mismatch between capacity and demand
- Vulnerability to DDoS attacks
- Traffic changes due to facilities outages

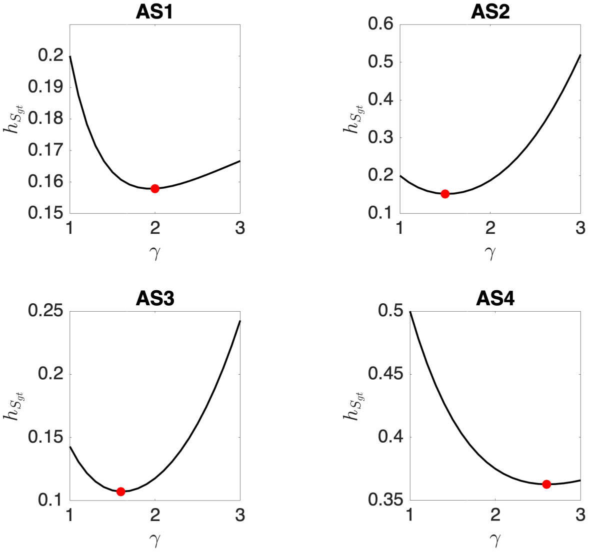
Current approaches:

✗ Extremely hard to compute and error prone

✗ Retrieve results every six months

L^γ-PageRank for Internet routing

The influence of γ



L^{γ}-PageRank for Internet routing

Classification using strict expert

ID	IP	True AS	Labels				Sweep-cut		Multi-class	
			AS1	AS2	AS3	AS4	$\gamma = 1$	$\gamma = 2$	$\gamma = 1$	$\gamma = 2$
(1)	10.6.66.1	AS2	0	0	0	0	AS2	AS2	AS2	AS2
(2)	62.214.63.145	AS2	0	0	0	0	AS2	AS2	AS2	AS2
(3)	62.214.36.177	AS2	0	1	0	0	AS2	AS2	AS2	AS2
(4)	62.214.37.130	AS2	0	0	0	0	AS2	AS2	AS2	AS2
(5)	213.155.129.188	AS3	0	0	0	0	AS2, AS3	AS3	AS3	AS3
(6)	62.115.141.236	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(7)	62.115.120.0	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(8)	213.248.68.71	AS4	0	0	0	0	AS1	n.a	AS3	AS3
(9)	63.223.34.74	AS4	0	0	0	0	AS1	n.a	AS3	AS1
(10)	63.217.25.146	AS1	0	0	0	0	AS1	AS1	AS1	AS1
(11)	139.162.0.10	AS1	0	0	0	0	AS1	AS1	AS1	AS1
(12)	139.162.27.28	AS1	1	0	0	0	AS1	AS1	AS1	AS1
(13)	62.214.37.134	AS2	0	0	0	0	AS2	AS2	AS2	AS2
(14)	62.115.137.168	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(15)	62.115.120.6	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(16)	139.162.0.2	AS1	0	0	0	0	AS1	AS1	AS1	AS1
(17)	62.115.137.166	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(18)	62.115.121.2	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(19)	62.115.137.164	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(20)	62.115.141.238	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(21)	62.115.141.240	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(22)	63.223.34.138	AS4	0	0	0	0	AS1	n.a	AS3	AS1
(23)	62.115.121.8	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(24)	62.115.121.4	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(25)	62.115.141.234	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(26)	62.115.121.10	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(27)	62.115.116.159	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(28)	62.115.116.163	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(29)	62.115.121.6	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3

Green: Correct inference. Red: Wrong inference

L^{γ}-PageRank for Internet routing

Classification using loose expert

ID	IP	True AS	Labels				Sweep-cut		Multi-class	
			AS1	AS2	AS3	AS4	$\gamma = 1$	$\gamma = 2$	$\gamma = 1$	$\gamma = 2$
(1)	10.6.66.1	AS2	0	0	0	0	AS2	AS2	AS2	AS2
(2)	62.214.63.145	AS2	0	1	0	0	AS2	AS2	AS2	AS2
(3)	62.214.36.177	AS2	0	1	0	0	AS2	AS2	AS2	AS2
(4)	62.214.37.130	AS2	0	1	0	0	AS2	AS2	AS2	AS2
(5)	213.155.129.188	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(6)	62.115.141.236	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(7)	62.115.120.0	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(8)	213.248.68.71	AS4	0	0	1	0	AS1, AS4	AS4	AS3	AS4
(9)	63.223.34.74	AS4	0	0	0	1	AS1, AS4	AS4	AS4	AS4
(10)	63.217.25.146	AS1	0	0	0	1	AS1, AS4	AS1, AS4	AS4	AS1
(11)	139.162.0.10	AS1	1	0	0	0	AS1, AS4	AS1	AS1	AS1
(12)	139.162.27.28	AS1	1	0	0	0	AS1, AS4	AS1	AS1	AS1
(13)	62.214.37.134	AS2	0	1	0	0	AS2	AS2	AS2	AS2
(14)	62.115.137.168	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(15)	62.115.120.6	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(16)	139.162.0.2	AS1	1	0	0	0	AS1, AS4	AS1	AS1	AS1
(17)	62.115.137.166	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(18)	62.115.121.2	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(19)	62.115.137.164	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(20)	62.115.141.238	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(21)	62.115.141.240	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(22)	63.223.34.138	AS4	0	0	0	1	AS1, AS4	AS4	AS4	AS4
(23)	62.115.121.8	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(24)	62.115.121.4	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(25)	62.115.141.234	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(26)	62.115.121.10	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(27)	62.115.116.159	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(28)	62.115.116.163	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3
(29)	62.115.121.6	AS3	0	0	1	0	AS2, AS3	AS3	AS3	AS3

Green: Correct inference. Red: Wrong inference

Conclusions

- New degree of freedom γ into PageRank
- Rewires graph and induces two regimes
- For $\gamma < 1$ embeds Lévy Flights into PageRank
- For $\gamma > 1$ makes opposite clusters repel themselves
- Optimal topology to perform classification
- Significant improvements in accuracy
- Promising AS inferences using the proposed approach