Global Vectors for Text-Enhanced Networks

Robin Brochier[1,2] Adrien Guille[1]
Benjamin Rothan[2] Julien Velcin[1]

[1] Université de Lyon, Lyon 2, ERIC EA 3083, France
{robin.brochier,julien.velcin, adrien.guille}@univ-lyon2.fr
`https://eric.ish-lyon.cnrs.fr/`

[2] Digital Scientific Research Technology, Lyon, France *{robin, benjamin}@peer.us*
`https://peer.us/`

09/07/2018

Table of Contents

Introduction

GloVeNet

Adding Textual Information

Future Works

References

1 Introduction

2 GloVeNet

3 Adding Textual Information

4 Future Works

About Me

peerus

INSPIRING RESEARCH

- PhD in collaboration with Digital Scientific Research Technology.
- **Peerus**: a tool to be up to date with the scientific papers of your field [1].
- **Peerus Review**: a tool for publishers to find reviewers [2].

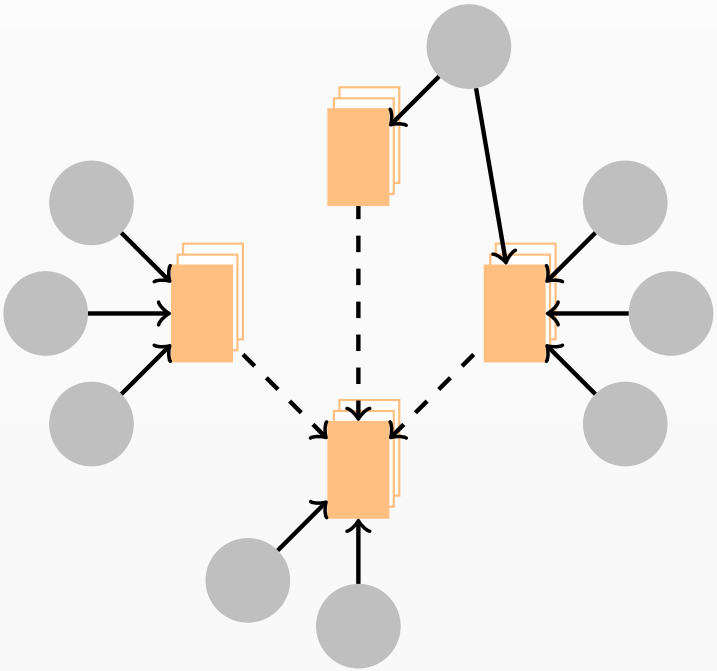
[1] https://peer.us/

[2] https://review.peer.us/

Scientific and Industrial Context

The Heterogeneous Scientific Network

$G = (V, E)$ où $V = R \cup A \cup W$ et $E = E_{RA} \cup E_{AA} \cup E_{AW}$
R: researchers
A: articles
W: words
V_{RA}: authorships
V_{AA}: citations
V_{AW}: articles textual contents



Researchers

Articles

Goals

1. Adapt GloVe [Pennington, Socher, and Manning 2014] for network embedding instead of Skip-Gram [Mikolov et al. 2013].
2. Build embeddings for a network that take into account the textual content linked with the vertices.
3. Embed different types of vertices in the same vector space.

GloVe

Builds a co-occurrence matrix X of words by sliding a "harmonic" window function over a corpus and learns 2 sets of embeddings W (target) and $\widetilde{W}$ (context) and bias B and $\widetilde{B}$:

$$J = \sum_{i,j} f\left(X_{ij}\right)\left(w_i^T \widetilde{w}_j + b_i + \widetilde{b}_j - log(X_{ij})\right)^2 \quad (1)$$

$$f(x) = \left(\frac{x}{x_{\max}}\right)^{\frac{3}{4}} \text{ if } x < x_{\max}$$

$$f(x) = 1 \text{ if } x > x_{\max}$$

DeepWalk Matrix

Algorithm 1 Generating a corpus of vertices sequences given an adjacency matrix of a graph.

Require: A, η, l
  $T \leftarrow L_1$_normalize_rows(A)
  $V \leftarrow$ number_of_rows(A)
  corpus = [ ]
  for i=1,...,V **do**
    for j=1,...,η **do**
      current_node = i
      walk = [i]
      for k=1,...,l **do**
        current_node = random_sample(T[*current_node*])
        Append *current_node* to *walk*
      end for
      Append *walk* to *corpus*

Shifted and Filtered Log Matrix

- Low values of X can be considered as noisy since they are the results of a limited number of walks.
- The difference between values close to 0 (where $log(X_{ij}) \to -\infty$) and values close to one (where $log(X_{ij}) \to 0$) is unclear.
- $=>$ We remove values lower or equal to X_{min} and shift the occurrence counts: $log(1 + X_{ij})$.

Negative Sampling

- We need a way to push away dissimilar nodes that never co-occurred.
- We hypothesize that low values of X play this role in GloVe, but their number and distribution are not optimal.
- As in Skip-Gram, we add a negative sampling term in the objective function.

GloVeNet

$$J = \sum_{i,j} f\left(X_{ij}\right)\Big[\left(w_i^T \widetilde{w}_j + b_i + \widetilde{b}_j - log(1 + X_{ij})\right)^2 + \sum_k \left(w_i^T \widetilde{w_k} + b_i + \widetilde{b_k}\right)^2\Big] \quad (2)$$

with: $f(x) = 0$ if $x \leq x_{\text{min}}$

$f(x) = 1$ otherwise

Results on Vertex Classification

Table: Experiments results on Aminer for vertex classification using only graph features.

Score	F1 (Micro)					F1 (Macro)				
% of labeled nodes	10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
GloVe	56.2	61.4	64.5	65.0	65.6	51.4	59.0	62.1	63.0	63.9
GloVeNet ($x_{\min} = 0$)	61.9	67.4	69.6	70.6	70.9	58.0	65.3	67.4	68.9	69.0
GloVeNet ($x_{\min} = 1$)	**70.6**	**74.0**	**75.0**	**75.7**	**76.2**	**67.9**	**72.7**	**73.4**	**74.3**	**74.8**

Table: Experiments results on Cora for vertex classification using only graph features

Score	F1 (Micro)					F1 (Macro)				
% of labeled nodes	10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
GloVe	46.3	53.0	55.4	58.6	58.4	35.2	47.5	50.6	55.1	55.2
GloVeNet ($x_{\min} = 0$)	49.5	55.6	60.7	62.0	62.1	38.0	49.3	55.3	57.3	57.4
GloVeNet ($x_{\min} = 1$)	**61.9**	**69.7**	**72.9**	**75.2**	**76.4**	**50.6**	**64.4**	**67.9**	**70.6**	**72.7**

Robustness to Filtering

Table: Density of X depending on the filter value $x_{\min}$.

$x_{\min}$	0	1	2	5	10	20
AMiner	4.51%	1.74%	1.12%	0.65%	0.46%	0.30%
Cora	2.84%	1.03%	0.71%	0.41%	0.25%	0.16%
Large-Scale	0.066‰	0.0127‰	0.0061‰	0.0041‰	0.0036‰	0.0016‰

Robustness to Filtering

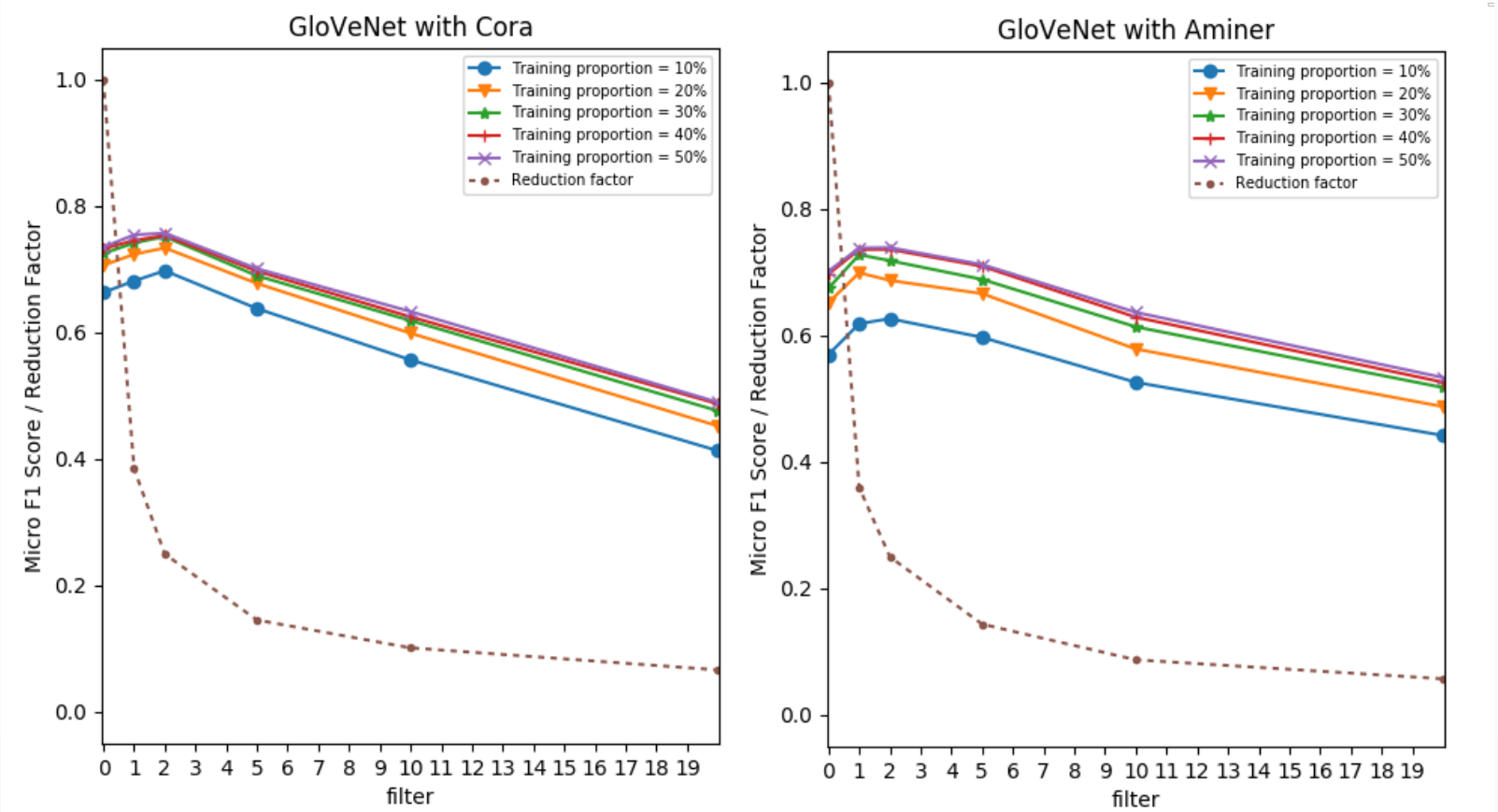


Figure: Robustness to the parameter $x_{\min}$.

From Words to Documents Embeddings

- $\widetilde{W_j^w}$ a word embedding matrix of dimension (M, d).
- d_j a sparse bag of word representation of the textual content linked with the node.
- generate a document embedding: $\widetilde{w_j} = \frac{d_j \widetilde{W_j^w}}{\sum_l d_{jl}}$.

Objective function

$$J = \sum_{i,j} f\left(X_{ij}\right)\Big[\left(w_i^T \frac{d_j \widetilde{W_j^w}}{\sum_l d_{jl}} + b_i - log(1 + X_{ij})\right)^2 + \sum_k \left(w_i^T \frac{d_k \widetilde{W_j^w}}{\sum_l d_{kl}} + b_i\right)^2 + \sum_k \left(w_k^T \frac{d_j \widetilde{W_j^w}}{\sum_l d_{jl}} + b_k\right)^2\Big] \quad (3)$$

with: $f(x) = 0$ if $x \leq x_{\text{min}}$

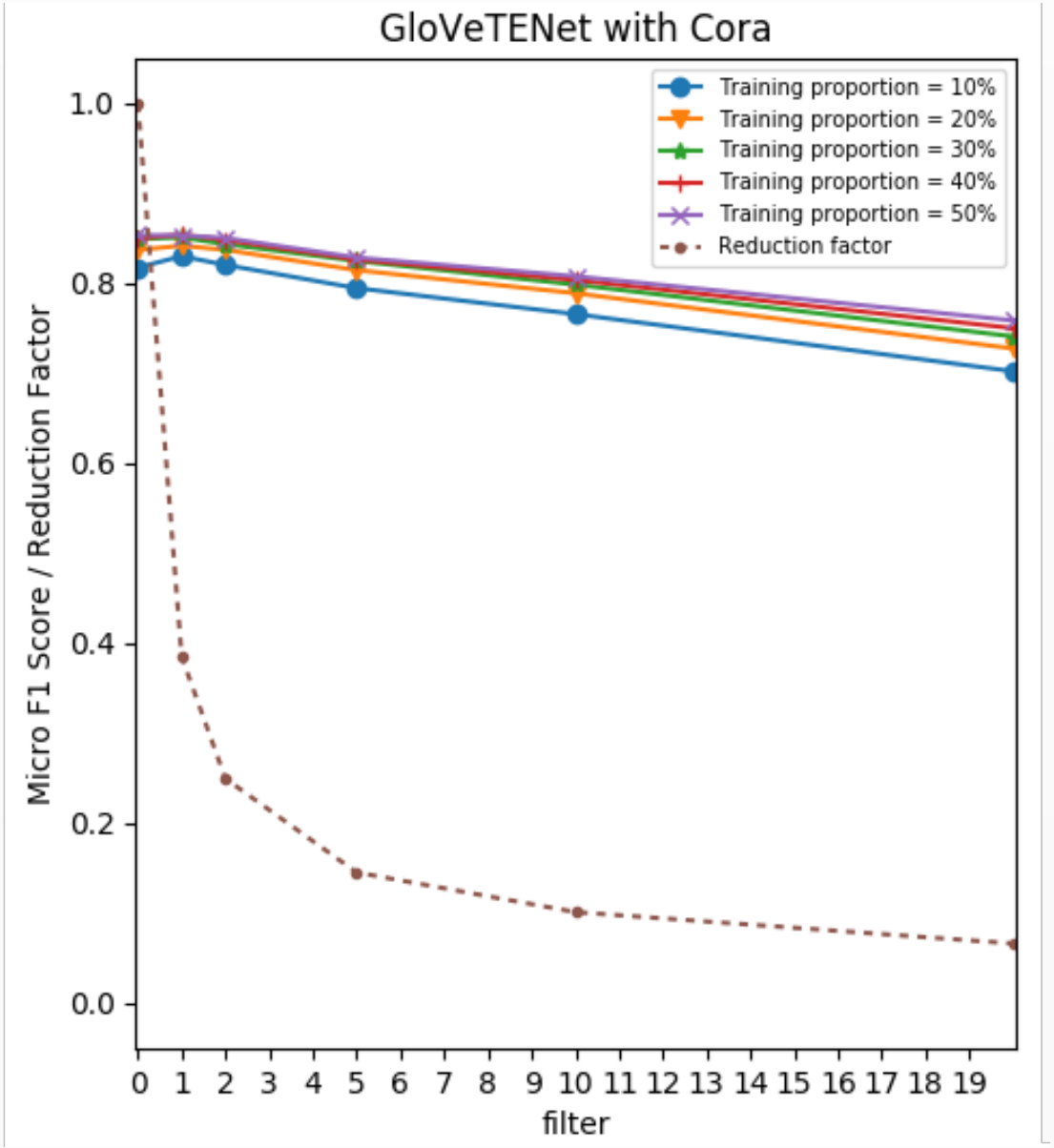
$f(x) = 1$ otherwise

Results on Vertex Classification

Table: Experiments results on Cora for vertex classification using graph and text features

Score	F1 (Micro)					F1 (Macro)				
% of labeled nodes	10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
Binary	60.8	65.9	67.0	68.0	69.4	57.1	62.5	64.0	65.1	66.7
TFIDF	64.3	69.5	71.3	71.6	73.1	60.5	66.7	68.8	69.1	71.0
SVD	58.0	68.1	71.9	73.5	74.7	53.6	65.2	69.3	70.9	72.3
TADW (text)	60.5	69.3	72.7	73.6	74.5	57.4	66.4	70.1	71.1	72.1
TADW (graph)	78.4	82.1	83.9	84.5	85.3	77.4	81.0	82.7	83.2	84.3
GloVeTENet (text)	74.5	76.5	78.5	78.6	79.8	72.8	75.1	77.1	77.2	78.4
GloVeTENet (graph)	75.9	78.5	80.6	81.1	82.6	74.7	77.1	79.3	79.6	81.5
NetMF + SVD	74.3	80.3	83.0	84.1	85.3	72.8	79.0	79.0	82.8	84.2
TADW	81.1	84.1	85.2	85.5	85.9	79.8	82.7	83.9	84.1	84.7
GloVeTENet	**83.6**	**85.0**	**85.8**	**86.1**	**86.4**	**82.3**	**83.9**	**84.7**	**84.9**	**85.4**

Robustness to Filtering



Parameters sensitivity

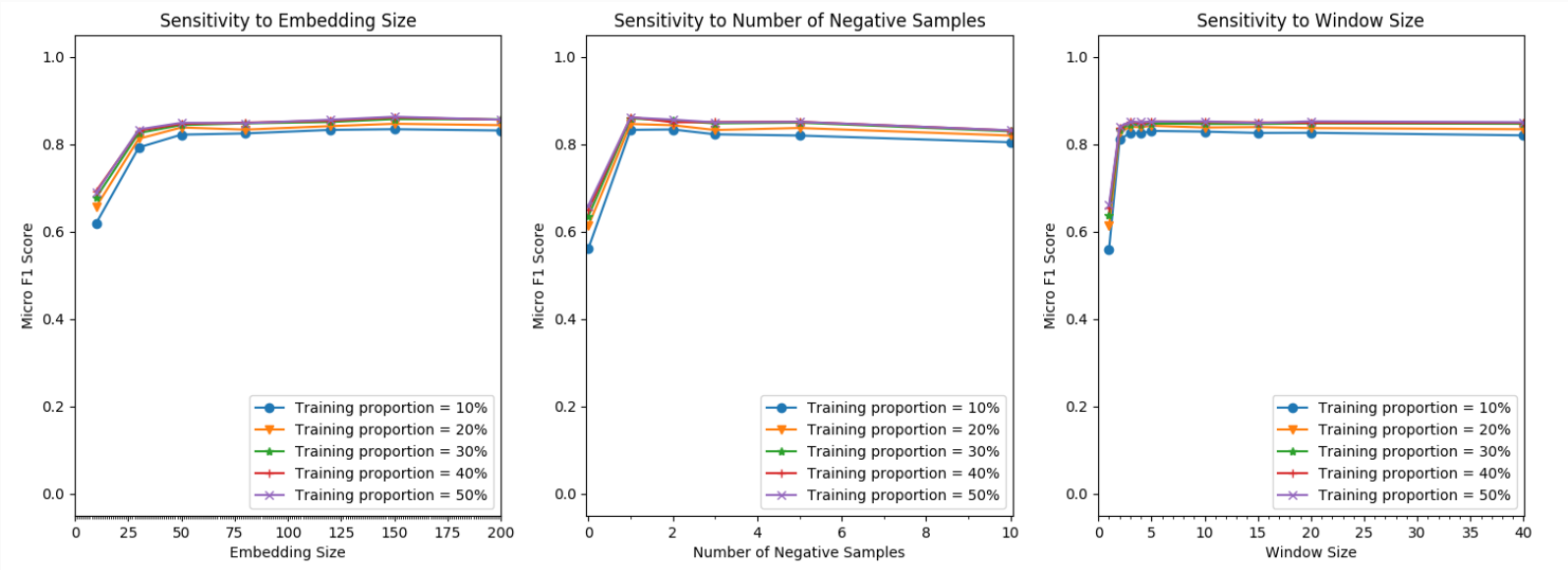


Figure: Sensitivity to the embedding dimension, to the number of negative samples and to the sliding window size.

Working with Heterogeneous Networks and Textual Information

- Generate heterogeneous paths: A1 => D1 => A2 => D2 => A3 => D3 => A4...
- Sliding window: (D2, A2, 1), (D2, A3, 1), (D2, A1, $\frac{1}{2}$), (D2, A4, $\frac{1}{2}$)...
- W are the authors nodes-embeddings and $\widetilde{W}$ are the papers text-embeddings.
- work in progress: application to expert finding and recommendation.

Thank You

Questions ?

Bibliography I

Mikolov, Tomas et al. (2013). "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*, pp. 3111–3119.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.