Exponential Family Graph Embeddings

Abdulkadir Çelikkanat and Fragkiskos D. Malliaros

Université Paris-Saclay, CentraleSupélec, Inria
Centre de Vision Numérique
France

Introduction

- *Network representation learning* (NRL) aims to encode the structure of a given network into low-dimensional vectors.
- Applications in network analysis: classification, community detection and link prediction.
- In this work:
  - We introduce a novel representation learning model, called EFGE, which generalizes classical *Skip-Gram*-based approaches to exponential family distributions.
  - We show that the objective functions of some existing unsupervised and representation learning models can be re-interpreted under the EFGE model.
  - We perform extensive performance evaluation of the proposed method in two downstream tasks.

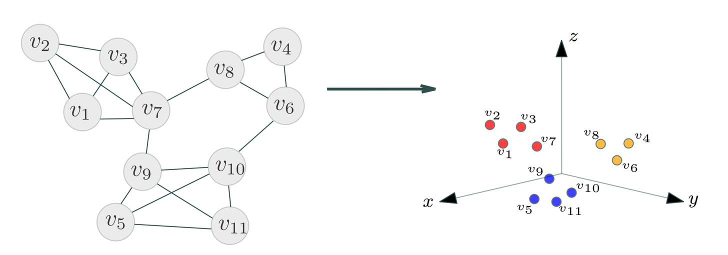


Figure 1: Schematic representation of node embeddings.

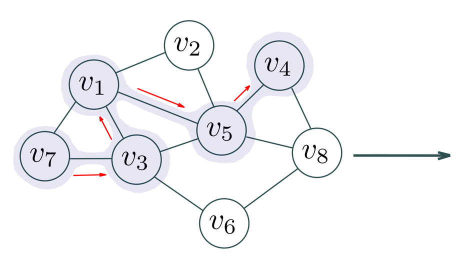


Figure 2: Illustration of generating random walks.

Exponential Families

- A class of probability distribution is called *exponential family* if it can be expressed as

$$p(y) = h(y)\exp\Big(\eta T(y) - A(\eta)\Big) \qquad where \quad \eta_{v,u} := f\big(\beta[v]^\top \cdot \alpha[u]\big)$$

Proposed Approach

- We define a generic objective function to learn node embeddings in the following way:

$$\mathcal{L}(\alpha,\beta) := \arg\max_{\Omega} \sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \sum_{v\in\mathcal{V}} \log p\left(y^l_{w_l,v}; \Omega = (\alpha,\beta)\right) \tag{1}$$

- By assuming that each $y_{w_l,v}$ follows an exponential family distribution, we can rewrite as follows:

$$\arg\max_{\Omega} \sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \sum_{v\in\mathcal{V}} \log h(y_{w_l,v}) + \eta_{w_l,v} T(y_{w_l,v}) - A(\eta_{w_l,v}) \tag{2}$$

- We have examined three particular instances of our approach, EFGE-BERN, EFGE-POIS and EFGE-NORM models corresponding to well known exponential family distributions.

center node

$(a, e, c, b, c, a, \boxed{d, b, d, a, d, c, b, a, b}, a, e, a, b, c, e)$

Figure 3: Illustration of a center node and its context set.

The EFGE-BERN Model

- We assume that each $y_{w_l,v}$ follows a Bernoulli distribution which is equal to 1 if node v appears in the context set of w_l in the walk $\mathbf{w}\in\mathcal{W}$.
- It can be written as $y_{w_l,v} = x^{l-\gamma}_{w_l,v} \vee \cdots \vee x^{l-1}_{w_l,v} \vee x^{l+1}_{w_l,v} \vee \cdots \vee x^{l+\gamma}_{w_l,v}$, where $x^{l+j}_{w_l,v}$ indicates the appearance of v in the context of w_l at the specific position $l+j$ $(-\gamma \leq j \leq \gamma)$.

center node

$(\boxed{a, e, c, b, c, a, d, b, d, a}, d, c, b, a, b, a, e, a, b, c, e)$

Figure 4: Illustration of how EFGE-BERN approach models the context set of a center node.

- The objective function of the EFGE-BERN model can be written by dividing Eq. (2) into two parts with respect to the values of $y_{w_l,v}$ and $x^{l+j}_{w_l,v}$:

$$\sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \Bigg[\sum_{v\in\mathcal{N}_\gamma(w_l)} \log p(y_{w_l,v}) + \sum_{v\notin\mathcal{N}_\gamma(w_l)} \log p(y_{w_l,v}) \Bigg] = \sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \Bigg[\sum_{\substack{|j|\leq\gamma \\ u^+:=w_j}} \log p(x^{l+j}_{w_l,u^+}) + \sum_{\substack{|j|\leq\gamma \\ u^-:\neq w_j}} \log p(x^{l+j}_{w_l,u^-}) \Bigg]$$

Relationship to negative sampling [2]

Lemma 1. *For large values of k, the log-likelihood function defined above converges to*

$$\sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \sum_{|j|\leq\gamma} \Bigg[\log p(x^{l+j}_{w_l,w_{l+j}}) + \sum_{s=1}^{k} \mathop{\mathbb{E}}_{u\sim q^-} \log p(x^{l+j}_{w_l,u}) \Bigg]$$

The EFGE-POIS Model

- We assume $y_{w_l,v}$ indicates the number of occurrences of v in $\mathcal{N}^{\mathbf{w}}_\gamma(w_l)$ and follows a Poisson distribution, with the mean $\tilde{\lambda}_{w_l,v}$ being the number of appearances of node v in the context $\mathcal{N}^{\mathbf{w}}_\gamma(w_l)$.
- It can be expressed as $y_{w_l,v} = x^{l-\gamma}_{w_l,v} + \cdots + x^{l-1}_{w_l,v} + x^{l+1}_{w_l,v} + \cdots + x^{l+\gamma}_{w_l,v}$, where $x^{l+j}_{w_l,v} \sim Pois(\lambda_{w_l,v})$ for $-\gamma \leq j \leq \gamma$ so we have $\tilde{\lambda}_{w_l,v} = \sum_{j=-\gamma}^{\gamma} \lambda^{l+j}_{w_l,v}$.

center node

$(a, e, c, b, c, a, d, b, d, a, d, \boxed{c, b, a, b, a, e, a, b, c}, e)$

Figure 5: Illustration of how EFGE-POIS approach models the context set of a center node.

- By plugging the exponential form of Poisson distribution into Eq. (1), the objective function can be expressed with respect to the values of $y_{w_l,v}$ for $y_{w_l,v} > 0$ and $y_{w_l,v} = 0$.

$$\sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \sum_{\substack{|j|\leq\gamma \\ u:=w_j}} \Big[-\log(x^{l+j}_{w_l,u}!) + \eta_{w_l,u} x^{l+j}_{w_l,u} - \exp(\eta_{w_l,u}) \Big] + \sum_{\substack{|j|\leq\gamma \\ u:\neq w_j}} \Big[-\exp(\eta_{w_l,u}) \Big]$$

Relationship to overlapping community detection (BIGCLAM algorithm [4])

Lemma 2. *Let $Z_{w_l,v}$ be independent random variables following Poisson distribution with natural parameter $\eta_{w_l,v}$ defined by $\log(\beta[w_l]\cdot\alpha[v])$. Then, the objective function of* EFGE-BERN *model becomes equal to*

$$\sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \Bigg[\sum_{v\in\mathcal{N}_\gamma(w_l)} \log\Big(1 - \exp\big(-\beta[w_l]^\top\cdot\alpha[v]\big)\Big) - \sum_{v\notin\mathcal{N}_\gamma(w_l)} \beta[w_l]^\top\cdot\alpha[v] \Bigg]$$

if the model parameter $\pi_{w_l,v}$ defined by $p(Z_{w_l,v} > 0)$.

The EFGE-NORM Model

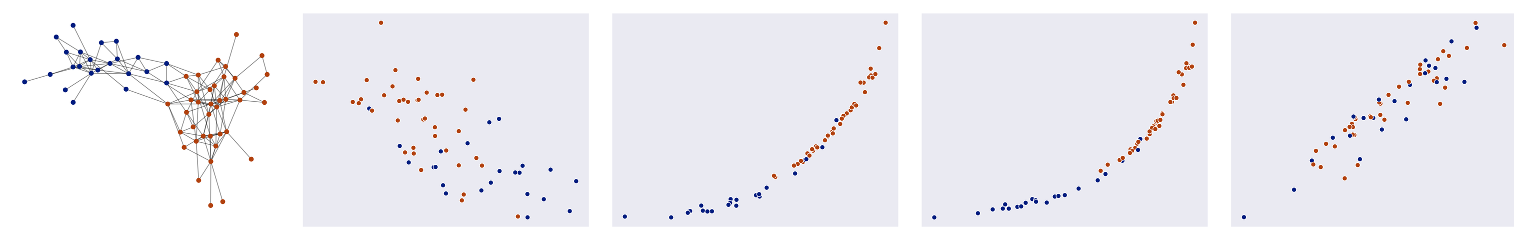
center node

$(a, e, c, b, c, a, \boxed{d, b, d, a, d, c, b, a, b}, a, e, a, b, c, e)$

Figure 6: Illustration of how EFGE-NORM approach models the context set of a center node.

- We consider each $y_{w_l,v}$ as an edge weight indicating the relationship between nodes w_l and v, and we assume that $x^{l+j}_{w_l,v} \sim \mathcal{N}(1,\sigma^2_+)$ if $v\in\mathcal{N}_\gamma(w_l)$, and $x^{l+j}_{w_l,v} \sim \mathcal{N}(0,\sigma^2_-)$ otherwise.
- We have chosen the link function as $\exp(-x)$ so the objective function of the model is defined by

$$\sum_{\mathbf{w}\in\mathcal{W}} \sum_{1\leq l\leq L} \sum_{\substack{|j|\leq\gamma \\ u:=w_j}} \Bigg[\log h(x^{l+j}_{w_l,u}) + \Big(x^{l+j}_{w_l,u} \frac{\eta_{w_l,u}}{\sigma^+} - \frac{\eta^2_{w_l,u}}{2} \Big) \Bigg] + \sum_{\substack{|j|\leq\gamma \\ u:\neq w_j}} \Bigg[\log h(x^{l+j}_{w_l,u}) + \Big(x^{l+j}_{w_l,u} \frac{\eta_{w_l,u}}{\sigma^-} - \frac{\eta^2_{w_l,u}}{2} \Big) \Bigg]$$



(a) Network **(b)** DEEPWALK **(c)** EFGE-BERN **(d)** EFGE-POIS **(e)** EFGE-NORM

Figure 7: The *Dolphins* network composed by 2 communities and the corresponding embeddings for $d=2$.

Experimental Evaluation

	2%	4%	6%	8%	10%	30%	50%	70%	90%
DEEPWALK	0.416	0.460	0.489	0.505	0.517	0.566	0.584	0.595	0.592
NODE2VEC	0.450	0.491	0.517	0.530	0.541	0.585	0.597	0.601	0.599
LINE	0.323	0.387	0.423	0.451	0.466	0.532	0.551	0.560	0.564
HOPE	0.196	0.205	0.210	0.204	0.219	0.256	0.277	0.299	0.320
NETMF	0.451	0.496	0.526	0.540	0.552	0.590	0.603	0.604	0.608
EFGE-BERN	0.461	0.493	0.517	0.536	0.549	0.588	0.603	0.609	0.609
EFGE-POIS	0.484	0.514	0.537	0.551	**0.562**	0.595	**0.606**	0.611	0.613
EFGE-NORM	**0.493**	**0.525**	**0.542**	**0.553**	0.561	**0.596**	**0.606**	**0.612**	**0.616**

Table 1: *CiteSeer*

	2%	4%	6%	8%	10%	30%	50%	70%	90%
DEEPWALK	0.545	0.585	0.600	0.608	0.613	0.626	0.628	0.628	0.633
NODE2VEC	0.575	0.600	0.611	0.619	0.622	0.636	0.638	0.639	0.639
LINE	0.554	0.580	0.590	0.597	0.603	0.618	0.621	0.623	0.623
HOPE	0.379	0.378	0.379	0.379	0.379	0.379	0.379	0.378	0.380
NETMF	0.577	0.589	0.596	0.601	0.605	0.617	0.620	0.623	0.623
EFGE-BERN	0.573	0.598	0.610	0.617	0.622	0.634	0.638	0.638	0.638
EFGE-POIS	0.588	0.605	0.614	0.620	0.624	0.635	0.637	0.636	0.638
EFGE-NORM	**0.603**	**0.614**	**0.622**	**0.624**	**0.628**	**0.637**	**0.640**	**0.642**	**0.641**

Table 2: *DBLP*

Table 3: Micro-F_1 scores for the node classification experiment for varying training sizes of networks.

	DEEPWALK	NODE2VEC	LINE	HOPE	NETMF	EFGE-BERN	EFGE-POIS	EFGE-NORM
Citeseer	0.770	0.780	0.717	0.744	0.742	0.815	**0.834**	0.828
Cora	0.739	0.757	0.686	0.712	0.755	0.769	0.797	**0.807**
DBLP	0.919	0.954	0.933	0.873	0.930	0.950	0.950	**0.955**
AstroPh	0.911	0.969	0.971	0.931	0.897	0.963	0.922	**0.973**
HepTh	0.843	0.896	0.854	0.836	0.882	**0.898**	0.885	0.896
Facebook	0.980	**0.992**	0.986	0.975	0.987	0.991	0.991	**0.992**
GrQc	0.921	**0.940**	0.909	0.902	0.928	0.938	0.937	**0.940**

Table 4: Area Under Curve (AUC) scores for link prediction.

Parameter Sensitivity

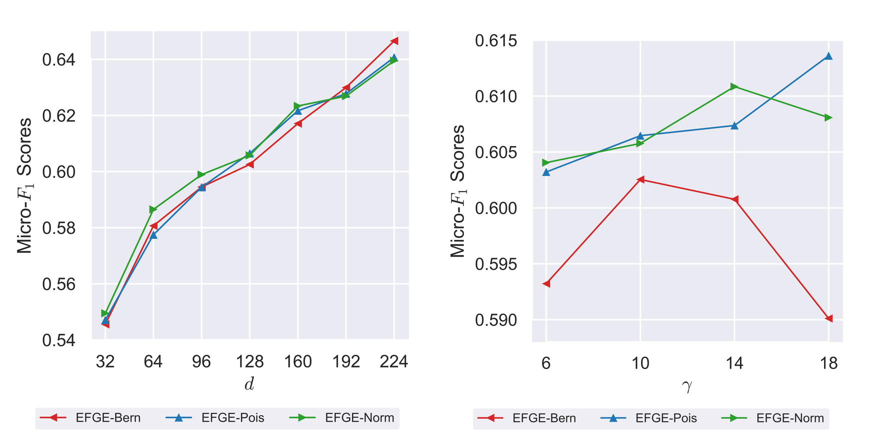


Figure 8: Influence of dimension d and window size γ over *CiteSeer*.

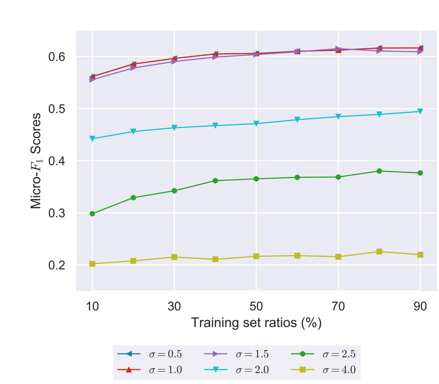


Figure 9: Effect of σ for EFGE-NORM.

References

[1] Li-Ping Liu et al. Context selection for embedding models. In *NIPS*, 2017.

[2] Tomas Mikolov et al. Distributed representations of words and phrases and their compositionality. In *NIPS*, 2013.

[3] Maja Rudolph et al. Exponential family embeddings. In *NIPS*. 2016.

[4] Jaewon Yang et al. Overlapping community detection at scale: A nonnegative matrix factorization approach. In *WSDM*, 2013.

SCAN ME