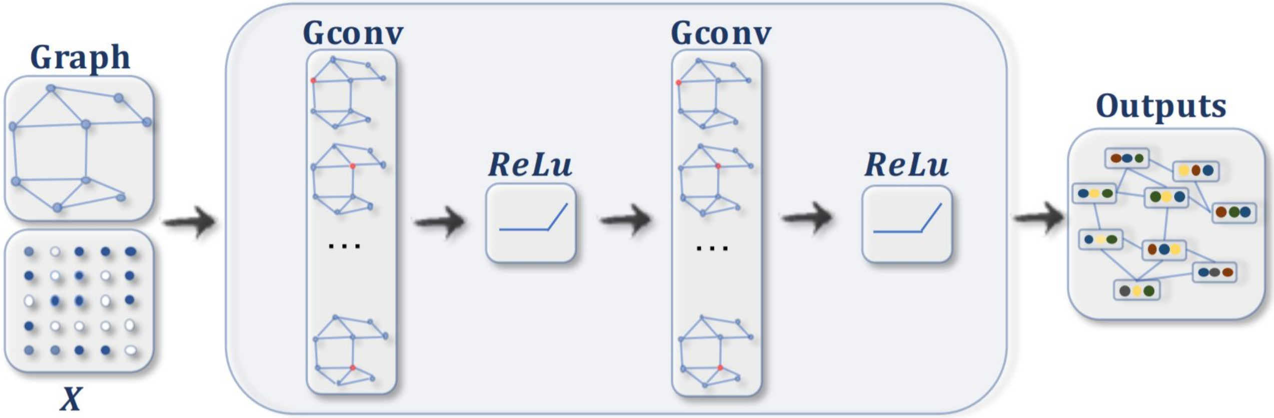
MultiSage: Empowering GCN with Contextualized Multi-Embeddings on Web-Scale Multipartite Networks

Алексей Медведев

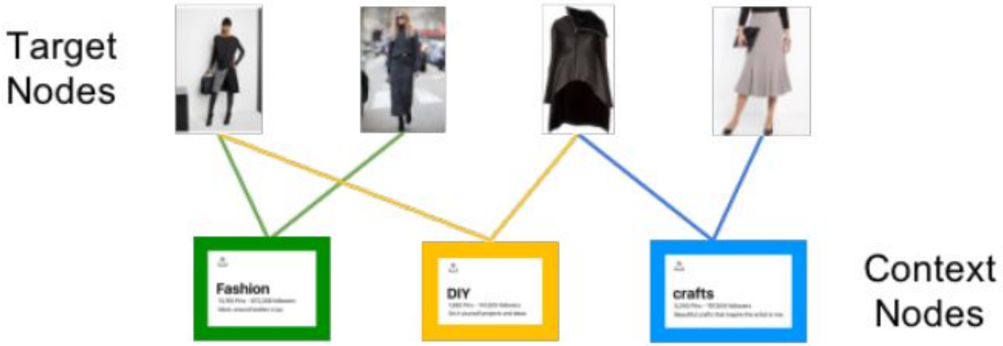
Graph Neural Networks?



$$\mathbf{h}^k_{\mathcal{N}(v)} \leftarrow \text{AGGREGATE}_k(\{\mathbf{h}^{k-1}_u, \forall u \in \mathcal{N}(v)\})$$

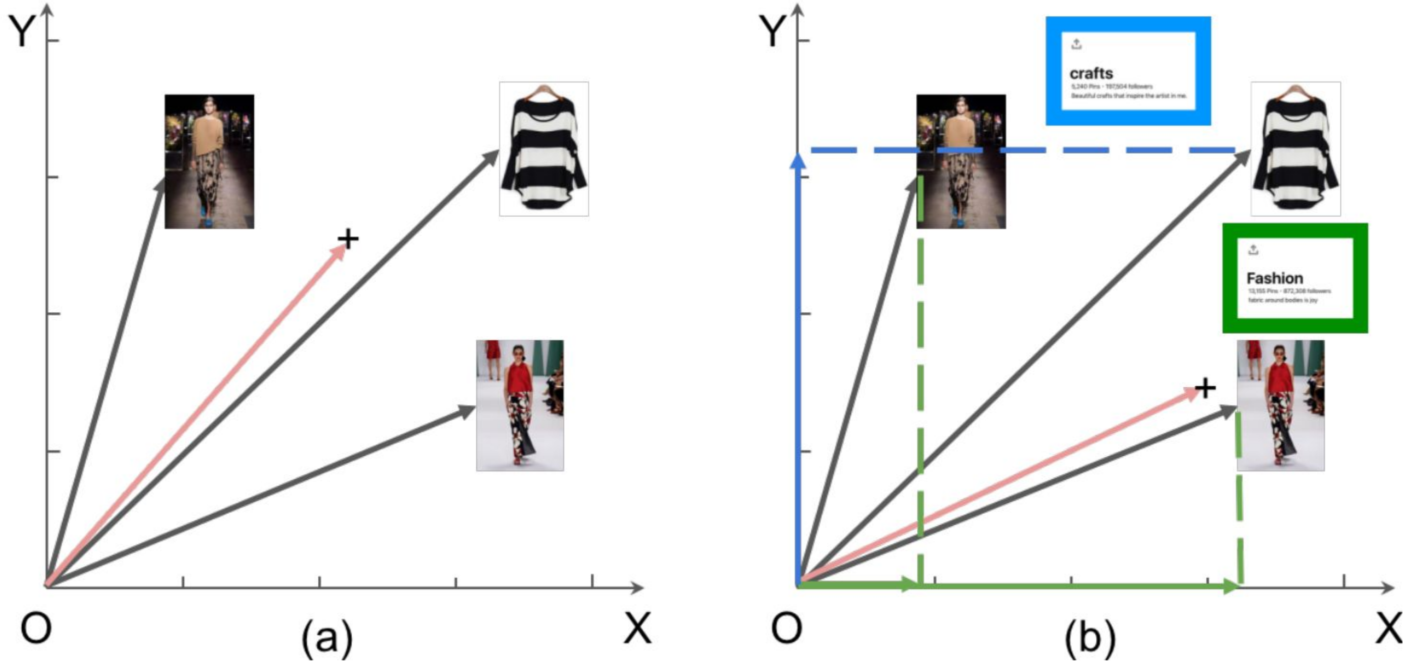
$$\mathbf{h}^k_v \leftarrow \sigma\left(\mathbf{W}^k \cdot \text{CONCAT}(\mathbf{h}^{k-1}_v, \mathbf{h}^k_{\mathcal{N}(v)})\right)$$

Target and Context Nodes



Dataset	Target	Context	Other context
IMDB [40]	movie	genre	director, actor
TCGA [43]	gene	pathway	disease, species
OAG [27]	paper	venue	author, keyword
Pinterest [6]	pin	board	user, session

Contextual Graph Embeddings



(a) (b)

Architecture

Raw feature transformation:

$$\mathbf{z}_t = \mathrm{ReLU}\big(\mathbf{W}_t^{(K)} \ldots \mathrm{ReLU}(\mathbf{W}_t^{(1)}\mathbf{x}_t + \mathbf{b}_t^{(1)}) \ldots + \mathbf{b}_t^{(K)}\big)$$
$$\mathbf{z}_c = \mathrm{ReLU}\big(\mathbf{W}_c^{(K)} \ldots \mathrm{ReLU}(\mathbf{W}_c^{(1)}\mathbf{x}_c + \mathbf{b}_c^{(1)}) \ldots + \mathbf{b}_c^{(K)}\big)$$

Contextual masking:

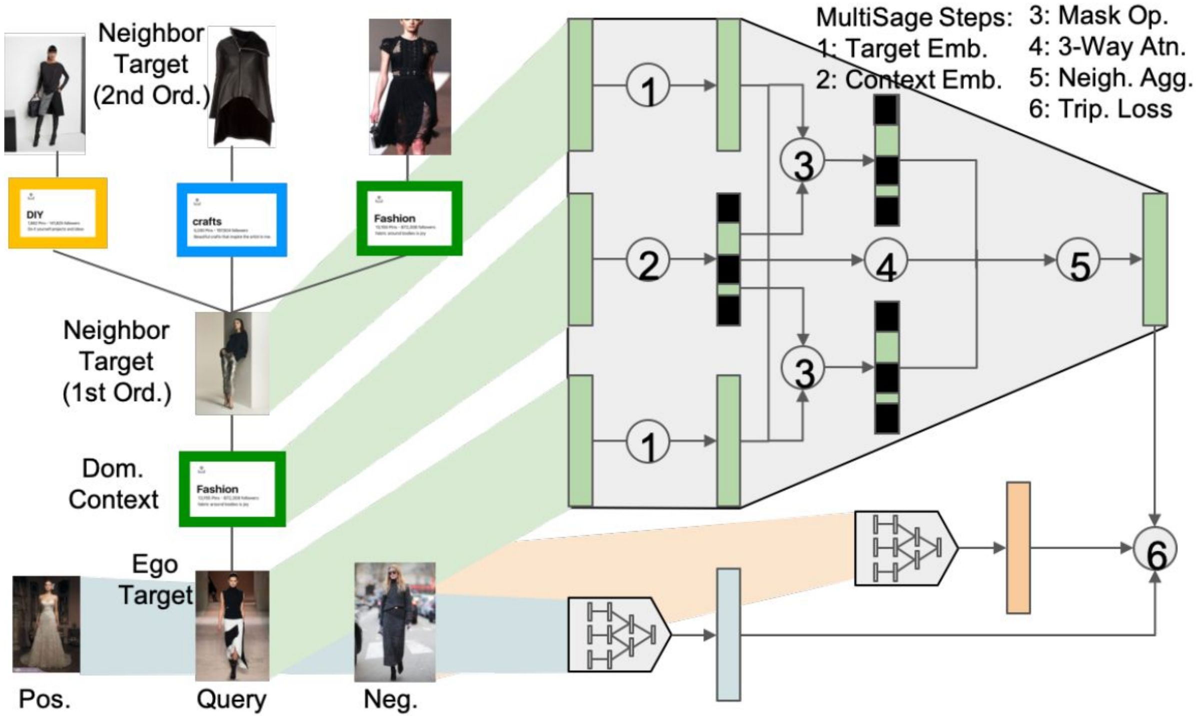
$$\mathbf{z}_{t|c} = \mathbf{z}_t \otimes \mathbf{z}_c$$

Contextual attention:

$$\alpha(v,o,u) = \frac{\exp\Big(\tau\big(\mathbf{a}^T[\mathbf{W}_{at}\mathbf{z}_t(v) \odot \mathbf{W}_{ac}\mathbf{z}_c(o) \odot \mathbf{W}_{at}\mathbf{z}_t(u)]\big)\Big)}{\sum_{u'\in\mathcal{N}_v, o'\sim(v,u')} \exp\Big(\tau\big(\mathbf{a}^T[\mathbf{W}_{at}\mathbf{z}_t(v) \odot \mathbf{W}_{ac}\mathbf{z}_t(o') \odot \mathbf{W}_{at}\mathbf{z}_t(u')]\big)\Big)}$$

$$\mathbf{z}_{\mathcal{N}_v}(x) = \sigma\Big(\frac{1}{D}\sum_{d=1}^{D}\sum_{u\in\mathcal{N}_v, o\sim(v,u)} \alpha^{(d)}(v,o,u)\ \mathbf{z}_{t|c}(x,o)\Big)$$

Architecture



Learning pipeline

Loss: $$\mathcal{J}(v_q, v_p, v_n) = \max\{0, \mathbf{h}_{v_q}^T \mathbf{h}_{v_n} - \mathbf{h}_{v_q}^T \mathbf{h}_{v_p}^L + \delta\}$$

Negative sampling with PersPageRank

Query

Positive Example

Random Negative

Hard Negative

Personalized PageRank

$$\boldsymbol{\pi}_{\mathrm{pr}} = \boldsymbol{A}_{\mathrm{rw}} \boldsymbol{\pi}_{\mathrm{pr}}$$

$$\boldsymbol{A}_{\mathrm{rw}} = \boldsymbol{A}\boldsymbol{D}^{-1}$$

$$\boldsymbol{\pi}_{\mathrm{ppr}}(\boldsymbol{i}_x) = (1-\alpha)\hat{\tilde{\boldsymbol{A}}}\boldsymbol{\pi}_{\mathrm{ppr}}(\boldsymbol{i}_x) + \alpha \boldsymbol{i}_x$$

$$\boldsymbol{\pi}_{\mathrm{ppr}}(\boldsymbol{i}_x) = \alpha \left(\boldsymbol{I}_n - (1-\alpha)\hat{\tilde{\boldsymbol{A}}} \right)^{-1} \boldsymbol{i}_x$$

$$\boldsymbol{\Pi}_{\mathrm{ppr}} = \alpha(\boldsymbol{I}_n - (1-\alpha)\hat{\tilde{\boldsymbol{A}}})^{-1}$$

$$\boldsymbol{A} \in \mathbb{R}^{n \times n}. \ \tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_n$$

$$\hat{\tilde{\boldsymbol{A}}} = \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}} \tilde{\boldsymbol{D}}^{-\frac{1}{2}}$$

$$\boldsymbol{Z}^{(0)} = \boldsymbol{H} = f_\theta(\boldsymbol{X}),$$

$$\boldsymbol{Z}^{(k+1)} = (1-\alpha)\hat{\tilde{\boldsymbol{A}}}\boldsymbol{Z}^{(k)} + \alpha \boldsymbol{H},$$

$$\boldsymbol{Z}^{(K)} = \mathrm{softmax}\left((1-\alpha)\hat{\tilde{\boldsymbol{A}}}\boldsymbol{Z}^{(K-1)} + \alpha \boldsymbol{H} \right)$$

Neighborhood Sampling

- vis[v] - visiting count
- ctx[v] - dominant context
- dom[v] - ‘dominance’ of the context
- path[1] - the first context node
- path[- 1] - the last target node

Algorithm 2 Parallel Contextualized Random Walk

Input: graph $\mathcal{G} = \{\mathcal{T}, \mathcal{C}, \mathcal{E}\}$, walk length ζ, number of walks κ, number of threads ξ, number of neighbors to keep s

Output: sampled contextualized neighbor lists $\mathcal{S} = \{\{v : [(c, u), \forall u \in \mathcal{N}_v, c \sim (v, u)]\}, \forall v \in \mathcal{T}\}$

$\forall v \in \mathcal{T}, \mathcal{S}[v] \leftarrow \emptyset$, vis[$v$], ctx[$v$], dom[$v$]$\leftarrow$ defaultdict(int)
Start a pool Ω of ξ threads in parallel
for v in $|\mathcal{T}|$ **do**
    with ω = rand(Ω)
    for i in κ **do**
        path $\leftarrow$ rand_walk(v, ζ)
        vis[v][path[-1]] ++
        if ctx[v][path[-1]] == path[1] **then**
            dom[v][path[-1]] ++
        else
            if dom[v][path[-1]]>0 **then**
                dom[path[-1]] - -
            else
                dom[v][path[-1]] ++
                ctx[v][path[-1]] = path[1]
            end if
        end if
    end for
    $\mathcal{S}[v] \leftarrow$ top s target-context pairs *w.r.t.* vis[v] & ctx[v]
end for

Results

Pinterest	MRR	REC@1	REC@10	DC+	DC*	DE+	DE*	INT	UNI	JAC
Visual	0.4406	0.1710	0.3606	0.4194	0.6337	0.9101	1.1255	23.32	174.67	0.1506
Textual	0.5741	0.1888	0.4965	0.3414	0.7614	0.7549	1.2050	31.78	166.21	0.1917
Combined	0.4438	0.1731	0.3635	0.4190	0.6340	0.9096	1.1258	23.44	174.53	0.1512
Pixie	0.3093	0.0418	0.2169	N/A	N/A	N/A	N/A	21.32	176.65	0.1351
PinSage	0.8759	0.4928	0.8234	0.2655	0.9279	0.7161	1.3593	47.30	150.69	0.3302
GAT	0.8880	0.5357	0.8665	0.2532	0.9343	0.7060	1.3618	48.70	149.24	0.3572
HAN	0.9013	0.5653	0.8838	0.2501	0.9415	0.6907	1.3558	50.29	148.83	0.3672
MultiSage-2	**0.9569**	**0.6215**	**0.9326**	**0.2316**	**0.9655**	**0.6660**	**1.3871**	**53.95**	**144.04**	**0.3906**
OAG	**MRR**	**REC@1**	**REC@10**	**DC+**	**DC***	**DE+**	**DE***	**INT**	**UNI**	**JAC**
Textual	0.1418	0.0273	0.0399	0.1081	0.4788	0.2814	1.0557	33.10	164.87	0.2193
Pixie	0.3126	0.1054	0.2642	N/A	N/A	N/A	N/A	36.58	160.76	0.2517
PinSage	0.5682	0.1845	0.5193	0.1238	0.6381	0.3179	1.1577	41.13	156.80	0.2935
GAT	0.6059	0.2355	0.5498	0.1104	0.6416	0.2908	1.2022	43.02	155.70	0.3144
HAN	0.6214	0.2641	0.5749	0.1005	0.6543	0.2869	1.2383	44.96	154.49	0.3200
MultiSage-2	0.6874	0.3270	0.6455	0.0836	0.6989	**0.2542**	1.2769	48.63	148.97	0.3602
MultiSage-3	**0.7026**	**0.3614**	**0.6875**	**0.0814**	**0.7127**	0.2583	**1.3058**	**51.63**	**145.40**	**0.3891**

Table 2: Performance of state-of-the-art large-scale embedding methods for general recommendation.

Method	Home Decoration			Women's Fashion		
	MRR	REC@1	REC@10	MRR	REC@1	REC@10
PinSage	0.8021/0.8067	0.4195/0.4257	0.7439/0.7460	0.7537/0.7545	0.3754/0.3759	0.6838/0.6976
MultiSage	**0.8407/0.8488**	**0.4899/0.5160**	**0.7954/0.8146**	**0.8058/0.8294**	**0.4363/0.4711**	**0.7533/0.7806**

Table 3: Off-task utility of embeddings produced by PinSage and MultiSage in shopping recommendation.

Context impact