Supplementary Material

A Closed Form Solutions for Maximum Violation Loss for Simple Implications

Let $\phi_r(\mathbf{h}_1, \mathbf{h}_2)$ be a scoring function for a relation r defined over pairs of entity vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, such as the scoring function in DISTMULT or COMPLEX. In the following, we assume that all entity embeddings live on a subspace $\mathcal{U} \subseteq \mathbb{R}^k$. The subspace $\mathcal{U}$ can either correspond to the unit sphere – *i.e.* $\mathcal{U} \triangleq \{\mathbf{h} \mid \|\mathbf{h}\|_2 = 1\}$ – or to the unit cube – *i.e.* $\mathcal{U} \triangleq \{\mathbf{h} \mid \mathbf{h} \in [0,1]^k\}$.

Let us consider a mapping $\mathcal{S} : \mathcal{V} \mapsto \mathbb{R}^k$ from variables to k-dimensional embeddings, where $\mathbf{h}_i = \mathcal{S}(X_i), \forall i$.

Given a clause expressing a simple implication in the form $b(X_1, X_2) \Rightarrow r(X_1, X_2)$, we would like to maximise the inconsistency loss $\mathcal{J}_\mathcal{I}$ associated to the clause:

$$\begin{aligned}&\mathcal{J}_\mathcal{I}^{\max} = \max\left(0, \mathcal{J}^{\max}\right)\\ &\text{with} \quad \mathcal{J}^{\max} = \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \left(\phi_b(\mathbf{h}_1, \mathbf{h}_2) - \phi_r(\mathbf{h}_1, \mathbf{h}_2)\right).\end{aligned}$$

The following sections show how to directly derive $\mathcal{J}_\mathcal{I}^{\max}$ for various models and entity embedding space restrictions. One way to solve the maximisation problem is via the Karush–Kuhn–Tucker (KKT) conditions – we refer to Boyd and Vandenberghe [2004] for more details.

A.1 DISTMULT

In the following, we focus on the Bilinear-Diagonal model (DISTMULT), proposed by Yang et al. [2015], and provide the corresponding derivations for different choices of the entity embedding subspace $\mathcal{U}$.

We want to solve the following optimisation problem:

$$\begin{aligned}\mathcal{J}^{\max} &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \left(\phi_b(\mathbf{h}_1, \mathbf{h}_2) - \phi_r(\mathbf{h}_1, \mathbf{h}_2)\right)\\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\theta}_b, \mathbf{h}_1, \mathbf{h}_2 \rangle - \langle \boldsymbol{\theta}_r, \mathbf{h}_1, \mathbf{h}_2 \rangle\\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\delta}, \mathbf{h}_1, \mathbf{h}_2 \rangle,\end{aligned}$$

where $\boldsymbol{\delta} \triangleq \boldsymbol{\theta}_b - \boldsymbol{\theta}_r$.

A.1.1 Unit Sphere

Assume that the subspace $\mathcal{U}$ corresponds to a the unit sphere, *i.e.* $\forall x \in \mathcal{E} : \|\mathbf{h}_x\|_2^2 = 1$. The Lagrangian is:

$$\mathcal{L} = -\langle \boldsymbol{\delta}, \mathbf{h}_1, \mathbf{h}_2 \rangle + \lambda_1(\|\mathbf{h}_1\|_2^2 - 1) + \lambda_2(\|\mathbf{h}_2\|_2^2 - 1).$$

Imposing stationarity: $\nabla_{\mathbf{h}_1} \mathcal{L} = 0$ and $\nabla_{\mathbf{h}_2} \mathcal{L} = 0$ gives:

$$\begin{aligned}-\boldsymbol{\delta} \odot \mathbf{h}_2 + 2\lambda_1 \mathbf{h}_1 = 0\\ -\boldsymbol{\delta} \odot \mathbf{h}_1 + 2\lambda_2 \mathbf{h}_2 = 0\end{aligned}$$

in which $\odot$ denotes the component-wise multiplication. For $\lambda_1 \neq 0$, a simple substitution leads to:

$$-\boldsymbol{\delta}^2 \odot \mathbf{h}_2 + 4\lambda_1\lambda_2 \mathbf{h}_2 = 0,$$

with the notation $\boldsymbol{\delta} \odot \boldsymbol{\delta} = \boldsymbol{\delta}^2$. As a result, $4\lambda_1\lambda_2 = \delta_i^2$ for components i with $h_{1,i} \neq 0$. Given the symmetry of the equations, the same requirements $4\lambda_1\lambda_2 = \delta_i^2$ hold for components $h_{2,i} \neq 0$.

We search for $\mathbf{h}_1$ and $\mathbf{h}_2$, such that δ_i^2 is constant for their non-zero components. Construct $\mathbf{h}_1$ and $\mathbf{h}_2$ such that only their component j is non-zero: $\forall i \neq j : h_{1,i} = h_{2,i} = 0$, whereas $h_{1,j} = \pm 1$, $h_{2,j} = \pm 1$ (given the unit sphere constraint). The contribution of component j to $\langle \boldsymbol{\delta}, \mathbf{h}_1, \mathbf{h}_2 \rangle$ depends on $h_{1,i} h_{2,i}$ which can take values ± 1. As a result:

$$\mathcal{J}^{\max} = \max_j |\delta_j|.$$

In the case where several components δ_j have the same value, then all $h_{1,i}$ and $h_{2,i}$ need to be zero for $i \neq j$, but due to the normalisation constraint, the highest value of $\sum_j h_{1_j} h_{2_j}$ is found for a single index j if both components take the value ± 1.

Finally, since $\mathcal{J}^{\max}$ is always non-negative, we find:

$$\boxed{\mathcal{J}_{\mathcal{I}}^{\max} = \max_j |\boldsymbol{\theta}_{b,j} - \boldsymbol{\theta}_{r,j}|}$$

A.1.2 Unit cube

For the entity embeddings to be constrained in the unit cube, their subspace is set to $\mathcal{U} = [0,1]^k \subset \mathbb{R}^k$. This corresponds to reducing the entity embeddings to approximately Boolean embeddings(see Demeester et al. [2016]).

The Lagrangian becomes:

$$\mathcal{L} = -\langle \boldsymbol{\delta}, \mathbf{h}_1, \mathbf{h}_2 \rangle + \sum_i \left[\boldsymbol{\mu}_1 \odot (\mathbf{h}_1 - \mathbf{1}) + \boldsymbol{\mu}_2 \odot (\mathbf{h}_2 - \mathbf{1}) \right]_i$$

with $\forall i : h_{1,i} - 1 \leq 0, h_{2,i} - 1 \leq 0$ (primal feasibility), $\forall i : \mu_{1,i} \geq 0, \mu_{2,i} \geq 0$ (dual feasibility), $\forall i : \mu_{1,i}(h_{1,i} - 1) = 0,\ \mu_{2,i}(h_{2,i} - 1) = 0$ (complementary slackness). In fact, we should add KKT multipliers for the conditions $-h_{1,i} \leq 0, -h_{2,i} \leq 0$ as well. These don't change the results, if we ensure the unit cube restrictions are satisfied.

Imposing stationarity, or $\nabla_{\mathbf{h}_1} \mathcal{L} = 0$ and $\nabla_{\mathbf{h}_2} \mathcal{L} = 0$, we get:

$$-\boldsymbol{\delta} \odot \mathbf{h}_2 + \boldsymbol{\mu}_1 = 0$$
$$-\boldsymbol{\delta} \odot \mathbf{h}_1 + \boldsymbol{\mu}_2 = 0$$

This can be solved component-wise, or for any component i:

$$\mu_{1i} = \delta_i h_{2,i}$$
$$\mu_{2i} = \delta_i h_{1,i}$$

Dual feasibility dictates that $\mu_{1i} \geq 0$ and $\mu_{2i} \geq 0$, such that $h_{2,i} = 0$ and $h_{1,i} = 0$ for any components where $\delta_i < 0$.

For the other components, we have $\delta_i \geq 0$, such that while satisfying the unit cube constraints the highest value of the objective becomes $\mathcal{J}_i^{\max} = \delta_i$ for $h_{1,i} = h_{2,i} = 1$.

Finally, we find:

$$\boxed{\mathcal{J}_{\mathcal{I}}^{\max} = \sum_j \max(0, \boldsymbol{\theta}_{b,j} - \boldsymbol{\theta}_{r,j}),}$$

which is exactly the same expression as the lifted loss for simple implications with unit cube entity embeddings introduced by Demeester et al. [2016] for MODEL F (which can be seen as a special case of DISTMULT where the subject embeddings are replaced by entity pair embeddings, and all object embedding components are set to 1).

A.2 COMPLEX

In COMPLEX, proposed by Trouillon et al. [2016], the scoring function ϕ is defined as follows:

$$\phi_r^{\boldsymbol{\theta}}(\mathbf{h}_1, \mathbf{h}_2) \triangleq \langle \boldsymbol{\theta}_r, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}}$$

where $\mathbf{x}^{\mathrm{R}} \triangleq \mathrm{Re}\,(\mathbf{x})$ and $\mathbf{x}^{\mathrm{I}} \triangleq \mathrm{Im}\,(\mathbf{x})$ denote the real and imaginary part of $\mathbf{x}$, respectively. In the following, we analyse the cases where entity embeddings live on the unit sphere (*i.e.* $\forall x \in \mathcal{E} : \|\mathbf{h}_x\|_2^2 = 1$) and in the unit cube (*i.e.*$\forall x \in \mathcal{E} : \mathbf{h}_x^{\mathrm{R}} \in [0,1]^k,\ \mathbf{h}_x^{\mathrm{I}} \in [0,1]^k$).

We want to solve the following maximisation problem:

$$\begin{aligned}\mathcal{J} &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \big(\phi_b(\mathbf{h}_1, \mathbf{h}_2) - \phi_r(\mathbf{h}_1, \mathbf{h}_2)\big) \\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}},\end{aligned}$$

with the complex vector $\boldsymbol{\delta} = \boldsymbol{\theta}_b - \boldsymbol{\theta}_r$.

To keep notations simple, we will continue to work with complex vectors as much as possible. For the Lagrangian $\mathcal{L}$, the pair of stationarity equations with respect to the real and imaginary part of a variable (say $\mathbf{h}_1$), can be taken together as follows:

$$\nabla_{\mathbf{h}_1^{\mathrm{R}}} \mathcal{L} + j \nabla_{\mathbf{h}_1^{\mathrm{I}}} \mathcal{L} = 0$$

and we introduce the notation $\nabla_{\mathbf{h}_1} \mathcal{L} \triangleq \nabla_{\mathbf{h}_1^{\mathrm{R}}} \mathcal{L} + j \nabla_{\mathbf{h}_1^{\mathrm{I}}} \mathcal{L}$, with j the imaginary unit. For the remainder, we need the following:

$$\begin{aligned}
\nabla_{\mathbf{h}_1} \langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}} &= \frac{1}{2} \nabla_{\mathbf{h}_1} \left(\langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle + \langle \overline{\boldsymbol{\delta}}, \overline{\mathbf{h}_1}, \mathbf{h}_2 \rangle \right) \\
&= \frac{1}{2} \left(\boldsymbol{\delta} \odot \overline{\mathbf{h}_2} + j \left(j \boldsymbol{\delta} \odot \overline{\mathbf{h}_2} \right) + \overline{\boldsymbol{\delta}} \odot \mathbf{h}_2 + j \left(-j \overline{\boldsymbol{\delta}} \odot \mathbf{h}_2 \right) \right) \\
&= \overline{\boldsymbol{\delta}} \odot \mathbf{h}_2 \\
\nabla_{\mathbf{h}_2} \langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}} &= \boldsymbol{\delta} \odot \mathbf{h}_1
\end{aligned}$$

A.2.1 Unit sphere

We now restrict the subject and object embeddings to live on the unit sphere, *i.e.* $\forall x \in \mathcal{E} : \|\mathbf{h}_x\|_2^2 = 1$. Given that the adversarial entity embeddings need to live on the unit sphere, the Lagrangian can be defined as follows:

$$\begin{aligned}
\mathcal{L}(\mathbf{h}_1, \mathbf{h}_2, \lambda_1, \lambda_2) = &- \langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}} \\
&+ \lambda_1 \left(\|\mathbf{h}_1\|_2^2 - 1 \right) \\
&+ \lambda_2 \left(\|\mathbf{h}_2\|_2^2 - 1 \right)
\end{aligned}$$

with real-valued Lagrange multipliers λ_1 and λ_2, and in which $\|\mathbf{h}_1\|_2^2 \triangleq \left\|\mathbf{h}_1^{\mathrm{R}}\right\|_2^2 + \left\|\mathbf{h}_1^{\mathrm{I}}\right\|_2^2$ is the L2 norm of the complex vector $\mathbf{h}_1$. With the expressions for $\nabla_{\mathbf{h}_1} \langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}}$ and $\nabla_{\mathbf{h}_2} \langle \boldsymbol{\delta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\mathrm{R}}$ above, the stationarity conditions can be written out as follows:

$$\begin{aligned}
-\overline{\boldsymbol{\delta}} \odot \mathbf{h}_2 + 2\lambda_1 \mathbf{h}_1 = 0 \\
-\boldsymbol{\delta} \odot \mathbf{h}_1 + 2\lambda_2 \mathbf{h}_2 = 0
\end{aligned}$$

Substituting into each other, we find:

$$\begin{aligned}
4\lambda_1\lambda_2 \mathbf{h}_1 &= \overline{\boldsymbol{\delta}} \odot \boldsymbol{\delta} \odot \mathbf{h}_1, \qquad (\lambda_2 \neq 0) \\
4\lambda_1\lambda_2 \mathbf{h}_2 &= \boldsymbol{\delta} \odot \overline{\boldsymbol{\delta}} \odot \mathbf{h}_2, \qquad (\lambda_1 \neq 0)
\end{aligned}$$

As a result, we require $4\lambda_1\lambda_2 = |\delta_i|^2$ for components i with $h_{1,i} \neq 0$ or $h_{2,i} \neq 0$. As in the case with DistMult, take $h_{1,i} = h_{2,i} = 0$ for each component $i \neq j$, such that for component j, we need $|h_{1,j}| = |h_{2,j}| = 1$, such that $|h_{1,j} h_{2,j}| = 1$. In order to maximise the contribution of that component to the loss, we choose the argument of the complex number $h_{1,j} \overline{h_{2,j}}$ such that $\delta_j h_{1,j} \overline{h_{2,j}}$ falls on the positive real axis. As a result:

$$\boxed{\mathcal{J}_{\mathcal{I}}^{\max} = \max_j |b_j - r_j| = \max_j \sqrt{(\boldsymbol{\theta}_{b,j}^{\mathrm{R}} - \boldsymbol{\theta}_{r,j}^{\mathrm{R}})^2 + (\boldsymbol{\theta}_{b,j}^{\mathrm{I}} - \boldsymbol{\theta}_{r,j}^{\mathrm{I}})^2}}$$

A.2.2 Unit cube

This case can be solved with the KKT conditions again, but instead we provide a shorter, less formal, derivation. It is clear that we can maximise the objective by maximising each component independently. For component i we need to optimise the following:

$$\delta_i^{\mathrm{R}} {h_{1,i}}^{\mathrm{R}} {h_{2,i}}^{\mathrm{R}} + \delta_i^{\mathrm{R}} {h_{1,i}}^{\mathrm{I}} {h_{2,i}}^{\mathrm{I}} + \delta_i^{\mathrm{I}} {h_{1,i}}^{\mathrm{R}} {h_{2,i}}^{\mathrm{I}} - \delta_i^{\mathrm{I}} {h_{1,i}}^{\mathrm{I}} {h_{2,i}}^{\mathrm{R}}$$

Regrouping gives:

$$\alpha \delta_i^{\mathrm{R}} + \beta \delta_i^{\mathrm{I}},$$

with $\alpha = {h_{1,i}}^{\text{R}}{h_{2,i}}^{\text{R}} + {h_{1,i}}^{\text{I}}{h_{2,i}}^{\text{I}}$ and $\beta = {h_{1,i}}^{\text{R}}{h_{2,i}}^{\text{I}} - {h_{1,i}}^{\text{I}}{h_{2,i}}^{\text{R}}$. We know $0 \leq \alpha \leq 2$, $-1 \leq \beta \leq 1$ and $\alpha + |\beta| \leq 2$. This allows maximising the objective as follows:

$$\boxed{\mathcal{J}_{\mathcal{I}}^{\max} = \sum_i \max(\delta_i^{\text{R}}, 0) + \max(\delta_i^{\text{R}}, |\delta_i^{\text{I}}|)}$$

B Simple Implications with Swapped Arguments

Given a clause expressing a simple implication with swapped arguments, in the form $b(X_1, X_2) \Rightarrow r(X_2, X_1)$, we would like to maximise the inconsistency loss $\mathcal{J}_{\mathcal{I}}$ associated to the clause, *i.e.*:

$$\begin{aligned} &\mathcal{J}_{\mathcal{I}}^{\max} = \max\left(0, \mathcal{J}^{\max}\right) \\ &\text{with} \quad \mathcal{J}^{\max} = \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \left(\phi_b(\mathbf{h}_1, \mathbf{h}_2) - \phi_r(\mathbf{h}_2, \mathbf{h}_1)\right). \end{aligned}$$

B.1 DISTMULT

Due to symmetry, the same close form expressions as for the simple implications hold.

B.2 COMPLEX

We want to solve the following maximisation problem:

$$\begin{aligned} \mathcal{J}^{\max} &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \left(\phi_b(\mathbf{h}_1, \mathbf{h}_2) - \phi_r(\mathbf{h}_2, \mathbf{h}_1)\right) \\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\theta}_b, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\text{R}} - \langle \boldsymbol{\theta}_r, \mathbf{h}_2, \overline{\mathbf{h}_1} \rangle^{\text{R}} \\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\theta}_b, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\text{R}} - \langle \overline{\boldsymbol{\theta}_r}, \overline{\mathbf{h}_2}, \mathbf{h}_1 \rangle^{\text{R}} \\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\theta}_b - \overline{\boldsymbol{\theta}_r}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\text{R}} \\ &= \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \langle \boldsymbol{\zeta}, \mathbf{h}_1, \overline{\mathbf{h}_2} \rangle^{\text{R}} \end{aligned}$$

This has the same form as the simple implications case, but with $\boldsymbol{\theta}_r$ replaced by $\overline{\boldsymbol{\theta}_r}$, or, more specifically, $\boldsymbol{\theta}_r^{\text{I}}$ by $-\boldsymbol{\theta}_r^{\text{I}}$.

B.2.1 Unit sphere

Under unit sphere constraints, $\mathcal{J}_{\mathcal{I}}^{\max}$ has the following value:

$$\boxed{\mathcal{J}_{\mathcal{I}}^{\max} = \max_i \sqrt{(\boldsymbol{\theta}_{b,i}^{\text{R}} - \boldsymbol{\theta}_{r,i}^{\text{R}})^2 + (\boldsymbol{\theta}_{b,i}^{\text{I}} + \boldsymbol{\theta}_{r,i}^{\text{I}})^2}} \tag{4}$$

B.2.2 Unit cube

With $\zeta_i^{\text{R}} = \boldsymbol{\theta}_{b,i}^{\text{R}} - \boldsymbol{\theta}_{r,i}^{\text{R}}$ and $\zeta_i^{\text{I}} = \boldsymbol{\theta}_{b,i}^{\text{I}} + \boldsymbol{\theta}_{r,i}^{\text{I}}$:

$$\boxed{\mathcal{J}_{\mathcal{I}}^{\max} = \sum_i \max(\zeta_i^{\text{R}}, 0) + \max(\zeta_i^{\text{R}}, |\zeta_i^{\text{I}}|)}$$

C Symmetry

Given a clause expressing a simple implication with swapped arguments, in the form $r(X_1, X_2) \Rightarrow r(X_2, X_1)$, we would like to maximise the inconsistency loss $\mathcal{J}_{\mathcal{I}}$ associated to the clause, *i.e.*:

$$\begin{aligned} &\mathcal{J}_{\mathcal{I}}^{\max} = \max\left(0, \mathcal{J}^{\max}\right) \\ &\text{with} \quad \mathcal{J}^{\max} = \max_{\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{U}} \left(\phi_r(\mathbf{h}_1, \mathbf{h}_2) - \phi_r(\mathbf{h}_2, \mathbf{h}_1)\right). \end{aligned}$$

Note that this is a special case of Appendix B where $r = b$.

C.1 DISTMULT

Since DISTMULT is symmetric, the gradient for symmetry clauses is zero, *i.e.*, all relations already satisfy symmetry.

C.2 COMPLEX

We want to solve the following maximisation problem:

$$\begin{aligned} \mathcal{J}^{\max} &= \max_{\mathbf{h}_1,\mathbf{h}_2\in\mathcal{U}} \big(\phi_r(\mathbf{h}_1,\mathbf{h}_2) - \phi_r(\mathbf{h}_2,\mathbf{h}_1)\big) \\ &= \max_{\mathbf{h}_1,\mathbf{h}_2\in\mathcal{U}} \langle \boldsymbol{\theta}_r - \overline{\boldsymbol{\theta}_r}, \mathbf{h}_1, \overline{\mathbf{h}_2}\rangle^{\mathrm{R}} \end{aligned}$$

C.2.1 Unit sphere

From Eq. (4) with $\boldsymbol{\theta}^{\mathrm{R}}_{r,i} = \boldsymbol{\theta}^{\mathrm{R}}_{b,i}$ and $\boldsymbol{\theta}^{\mathrm{I}}_{r,i} = \boldsymbol{\theta}^{\mathrm{I}}_{b,i}$ we get:

$$\boxed{\mathcal{J}^{\max} = \max_i 2|\boldsymbol{\theta}^{\mathrm{I}}_{r,i}|}$$

C.2.2 Unit cube

Similarly, with $\boldsymbol{\theta}^{\mathrm{R}}_{r,i} = \boldsymbol{\theta}^{\mathrm{R}}_{b,i}$ and $\boldsymbol{\theta}^{\mathrm{I}}_{r,i} = \boldsymbol{\theta}^{\mathrm{I}}_{b,i}$ we get:

$$\boxed{\mathcal{J}^{\max} = 2\sum_i |\boldsymbol{\theta}^{\mathrm{I}}_{r,i}|}$$

D Link Prediction Results

Table 8: Link prediction results on the Test-I, Test-II and and Test-ALL on FB122, filtered setting.

		Test-I				Test-II				Test-ALL			
		Hits@N (%)			MRR	Hits@N (%)			MRR	Hits@N (%)			MRR
		3	5	10		3	5	10		3	5	10	
FB122	TRANSE [Bordes et al., 2013]	36.0	41.5	48.1	0.296	77.5	82.8	88.4	0.630	58.9	64.2	70.2	0.480
	KALE-PRE [Guo et al., 2016]	35.8	41.9	49.8	0.291	82.9	86.1	89.9	0.713	61.7	66.2	71.8	0.523
	KALE-JOINT [Guo et al., 2016]	**38.4**	**44.7**	**52.2**	0.325	79.7	84.1	89.6	0.684	61.2	66.4	72.8	0.523
	DISTMULT [Yang et al., 2015]	36.0	40.3	45.3	0.313	92.3	93.8	94.7	0.874	67.4	70.2	72.9	0.628
	ASR-DISTMULT	36.3	40.3	44.9	0.330	98.0	99.0	99.2	0.948	70.7	73.1	75.2	0.675
	cASR-DISTMULT	37.0	40.4	45.1	0.337	96.7	98.6	99.3	0.933	70.1	72.7	75.1	0.669
	COMPLEX [Trouillon et al., 2016]	37.0	41.3	46.2	0.329	91.4	91.9	92.4	0.887	67.3	69.5	71.9	0.641
	ASR-COMPLEX	37.3	41.0	45.9	0.338	**99.2**	**99.3**	**99.4**	**0.984**	**71.7**	**73.6**	**75.7**	**0.698**
	cASR-COMPLEX	37.9	41.7	46.2	**0.339**	97.7	**99.3**	**99.4**	0.954	71.1	**73.6**	75.6	0.680
WN18	TRANSE [Bordes et al., 2013]	57.4	72.3	80.1	0.306	87.5	95.6	98.7	0.511	79.1	89.1	93.6	0.453
	KALE-PRE [Guo et al., 2016]	60.6	74.5	81.1	0.322	96.4	98.6	99.6	0.612	86.4	91.9	94.4	0.532
	KALE-JOINT [Guo et al., 2016]	65.5	76.3	82.1	0.338	93.3	95.4	97.2	0.787	85.5	90.1	93.0	0.662
	DISTMULT [Yang et al., 2015]	80.6	81.6	82.6	0.796	96.7	98.4	99.5	0.872	92.3	93.7	94.9	0.850
	ASR-DISTMULT	**81.4**	**82.0**	**82.9**	0.801	96.7	98.4	99.5	0.869	92.4	93.8	94.9	0.851
	COMPLEX [Trouillon et al., 2016]	81.0	81.8	82.5	**0.803**	**99.9**	**100**	**100**	**0.996**	**94.7**	**95.0**	**95.1**	**0.942**
	ASR-COMPLEX	81.0	81.8	82.5	**0.803**	**99.9**	**100**	**100**	**0.996**	**94.7**	**95.0**	**95.1**	**0.942**

Supplementary References

[sup1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004. ISBN 0521833787.