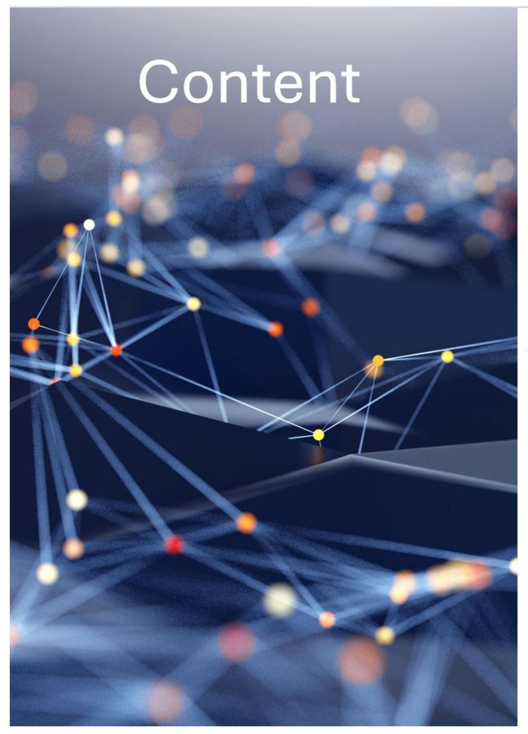
Node Similarities under Random Projections: Limits and Pathological Cases

Tvrtko Tadić
Cassiano Becker
Jennifer Neville

Microsoft

ICLR

Content



Takeaway:

Quality of embeddings, obtained by random projection, can vary depending on the node degree and the similarity measure used.

Relevance Measures and Estimates

Dot product similarity

$$\mathrm{rel}_{uv} = P_{u*}P_{v*}^{\top},$$

and

$$\mathrm{rel}_{uv}^{R} = (P_{u*}R^{\top})(RP_{v*}^{\top})^{\top}$$

Cosine similarity

$$\mathrm{rel}_{uv} = \cos(P_{u*}, P_{v*}),$$

and

$$\mathrm{rel}_{uv}^{R} = \cos(P_{u*}R^{\top}, P_{u*}R^{\top})$$

Dot Product Similarity Issues

NEW RESULT

Corollary 3.3. *If v is low-degree and u high-degree node, then the standard deviation in asymptotic estimate is greater than its expectation.*

$$\mathrm{rel}_{uv}^{R} \overset{a}{\sim} \mathcal{N}\left(\mathrm{rel}_{uv}, \frac{1}{q}\left[\mathrm{rel}_{uv}^{2} + \mathrm{rel}_{uu}\mathrm{rel}_{vv}\right]\right)$$

NEW RESULT

Corollary 3.4. *Let u be a high-degree vertex and v a low-degree vertex with no common neighbors, i.e., such that $\mathrm{rel}_{uu} > \mathrm{rel}_{uv} = 0$. Then*

$$\mathbb{P}(\mathrm{rel}_{uu}^{R} < \mathrm{rel}_{uv}^{R}) > \mathbb{P}(\mathcal{N}(0,1) > 1) \approx 15.8\%.$$

Why?

Dot product

Relevance not normalized

$$\mathrm{rel}_{uv}^{R} \overset{a}{\sim} \mathcal{N}\left(\mathrm{rel}_{uv}, \frac{1}{q}\left[\mathrm{rel}_{uv}^{2} + \mathrm{rel}_{uu}\mathrm{rel}_{vv}\right]\right)$$

Estimate promotes irrelevant nodes over relevant ones with significant probability.

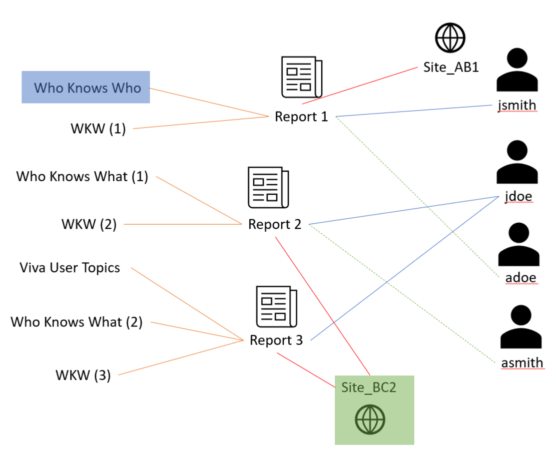
Cosine similarity

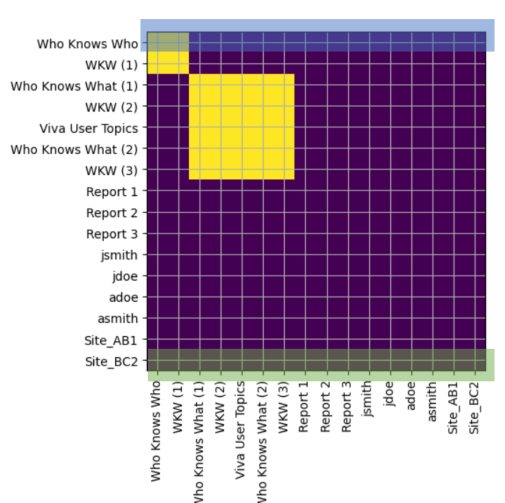
Relevance normalized

$$\mathrm{rel}_{uv}^{R} \overset{a}{\sim} \mathcal{N}\left(\mathrm{rel}_{uv}, \frac{1}{q}\left(1 - \mathrm{rel}_{uv}^{2}\right)^{2}\right)$$

Estimate unlikely to promote irrelevant node over relevant.

Graph *Connectivity* Through Matrix





Dot Product Similarity

Theorem 2.1. *The following statements hold:*
(a) Asymptotic result. For $u, v \in V$ and large q

$$\mathrm{rel}_{uv}^{R} \overset{a}{\sim} \mathcal{N}\left(\mathrm{rel}_{uv}, \frac{1}{q}\left[\mathrm{rel}_{uv}^{2} + \mathrm{rel}_{uu}\mathrm{rel}_{vv}\right]\right)$$

(b) Finite sample result. For $\varepsilon \in (0,1)$ and $\delta \in (0,1)$, if $q \geq 4\frac{1+\varepsilon}{\varepsilon^2}\log\left[\frac{n(n-1)}{\delta}\right]$, then

$$|\mathrm{rel}_{uv}^{R} - \mathrm{rel}_{uv}| < \varepsilon\sqrt{\mathrm{rel}_{uu}\mathrm{rel}_{vv}}$$

for all $u, v \in V$ holds with probability at least $1-\delta$.

Cosine Similarity Stability

NEW RESULT

Corollary 3.7. *For $u, v, w, h \in V$, the following holds.*

(a) If $\mathrm{rel}_{uv} = 0$, then

$$\mathrm{rel}_{uu} = 1 > 0 = \mathrm{rel}_{uv} \text{ and } \mathrm{rel}_{uu}^{R} > \mathrm{rel}_{uv}^{R} \text{ with probability } 1.$$

(b) Under the conditions of Theorem 2.4 (b),

$$\text{if } \mathrm{rel}_{uv} - \mathrm{rel}_{wh} > 2\varepsilon, \text{ then } \mathbb{P}(\mathrm{rel}_{uv}^{R} > \mathrm{rel}_{wh}^{R}) \geq 1-\delta.$$

Further work

Matrix R was Gaussian and as such dense.

FAST RP method uses the sparse matrix where for $s \geq 1$:

$$\mathbb{P}\left(R_{ij} = \frac{k\sqrt{s}}{\sqrt{q}}\right) = \begin{cases} \frac{1}{2s} & \text{for } k=-1 \\ (1-\frac{1}{s}) & \text{for } k=0 \\ \frac{1}{2s} & \text{for } k=1 \end{cases}$$

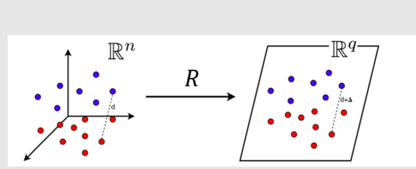
We believe that the results shown for Gaussian matrices will qualitatively hold in a modified form for Fast RP.

Dimension Reduction using J-L Lemma

Random projection method is based on using a $q \times n$ random matrix :

$$R = \begin{bmatrix} N(0,1/q) & \cdots & N(0,1/q) \\ \vdots & \ddots & \vdots \\ N(0,1/q) & \cdots & N(0,1/q) \end{bmatrix}$$

Johnson- Lindenstrauss Lemma guarantees low relative error, with high probability, when $q = O(\log n)$.



Cosine Similarity

NEW RESULT

Theorem 2.4. *The following claims hold:*
(a) Asymptotic result. For $u, v \in V$ and large q

$$\mathrm{rel}_{uv}^{R} \overset{a}{\sim} \mathcal{N}\left(\mathrm{rel}_{uv}, \frac{1}{q}\left(1 - \mathrm{rel}_{uv}^{2}\right)^{2}\right)$$

(b) Finite-sample result. Let $\varepsilon \in (0, 0.05]$ and $\delta \in (0,1)$. If $q \geq \frac{2\ln\left[\frac{2n(n-1)\left(1+\frac{\varepsilon^2}{4}\right)}{\delta}\right]}{\ln\left[1+\frac{\varepsilon^2}{2(1+\varepsilon\sqrt{2})}\right]}$, then

$$\left|\mathrm{rel}_{uv}^{R} - \mathrm{rel}_{uv}\right| \leq \varepsilon\left(1 - \mathrm{rel}_{uv}^{2}\right)$$

holds for all $u, v \in V$ with probability at least $1-\delta$.

Impact on Wikipedia graph Ranking

Gleich/wikipedia-20060925 dataset:

- **2,983,494** vertices (web pages)
- **37,269,096** edges (web links), wherein $A_{ji} = 1$ if page i links to page j, and 0 otherwise

- Computing relevance / connectivity between nodes in a graph can be time-costly.
- Random projections can speed up the process and help us estimate the relevance
- Depending on the type of relevance, quality of the estimation might be degree-dependent
- High degree nodes are scarce, and detecting problems using experimental methods might be difficult

Conclusion

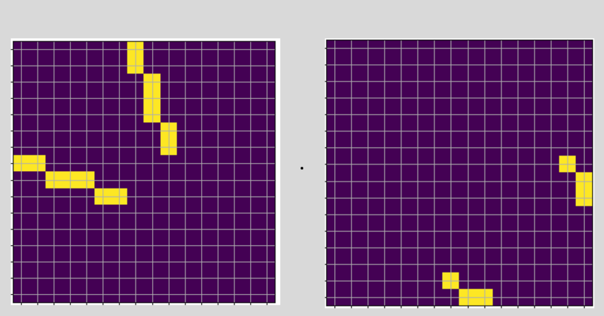
Computational Advantages

If P is a polynomial of matrices, it will be easier to compute $PR^{\top}$.

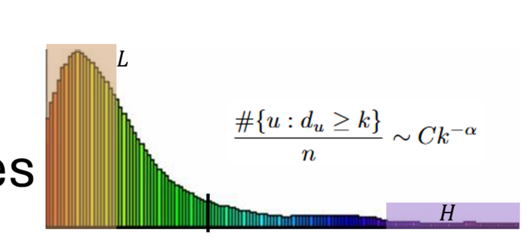
Example. $P = A_1A_2$

Step 1: $X \coloneqq A_2R^{\top}$

Step 2: $X \coloneqq A_1X$



Low and high degree nodes



For a small positive integer c we define

$$L \coloneqq \{v \in V : d_v \leq c\}$$

and

$$H \coloneqq \{u \in V \ : \ d_u \geq \gamma^2 cq\}$$

where γ is a technical constant and q the projection dimension

Will H have nodes?

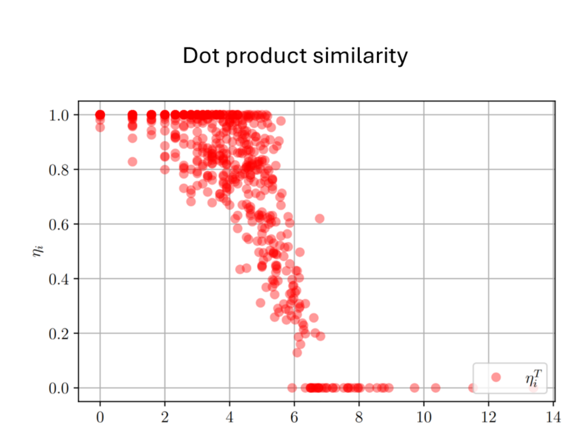
Power law considerations

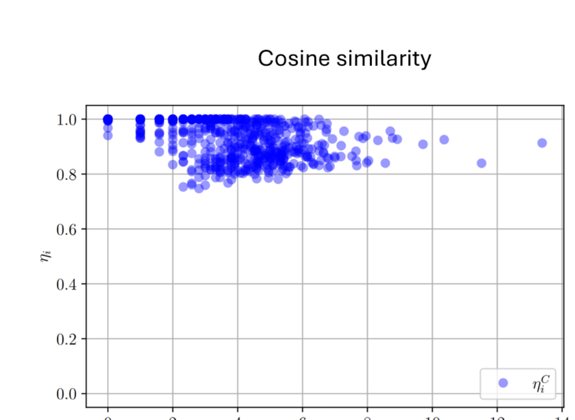
JL Lemma: $q = \mathcal{O}(\log n)$

Estimate:

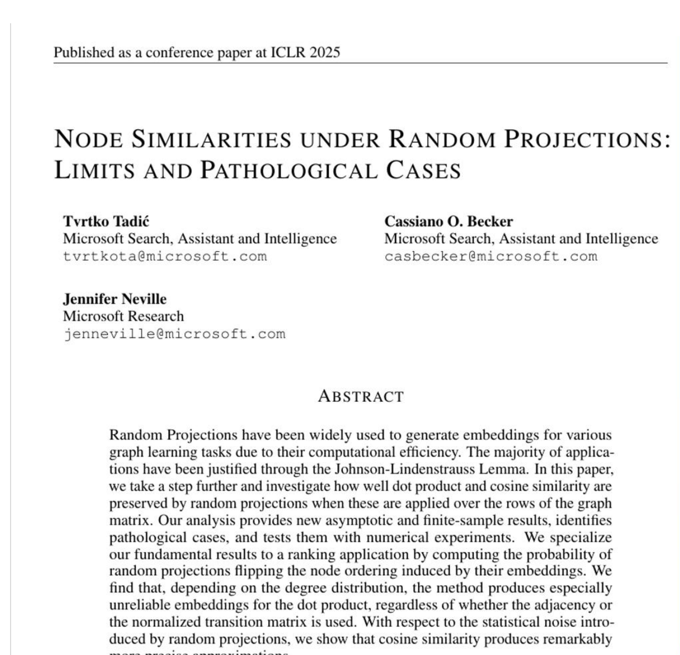
$$\frac{\#H}{n} \sim C[\log n]^{-\alpha}$$

NDCG@10 vs Node Degree





$P = T$



Acknowledgments

- Expanded on the experimentation work done by the Graph Intelligence Sciences Team at Microsoft
- Thanks to management for hacking weeks & learning days used for this work
- Thanks to the University of Washington Math Department for hosting the first author