

Node Similarities under Random Projections: Limits and Pathological Cases



Tvrtko Tadić
Cassiano Becker
Jennifer Neville



Content

- Multiple Edge Connections and Computational Challenges
- Dimension Reduction and Random Projections
- Low and High Degree Nodes
- Ranking and the Curse of Popularity
- Further Work and Conclusion

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

$$\text{rel}_{uv} = P_{u*} P_{v*}^T,$$

and

$$\text{rel}_{uv}^R = (P_{u*} R^T)(R P_{v*}^T)^T$$

Cosine similarity

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

and

$$\text{rel}_{uv}^R = \cos(P_{u*} R^T, P_{v*} R^T)$$

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.

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

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 $\text{rel}_{uu} > \text{rel}_{uv} = 0$. Then

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

Why?

Dot product

Relevance not **normalized**

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

Estimate promotes irrelevant nodes over relevant ones with **significant probability**.

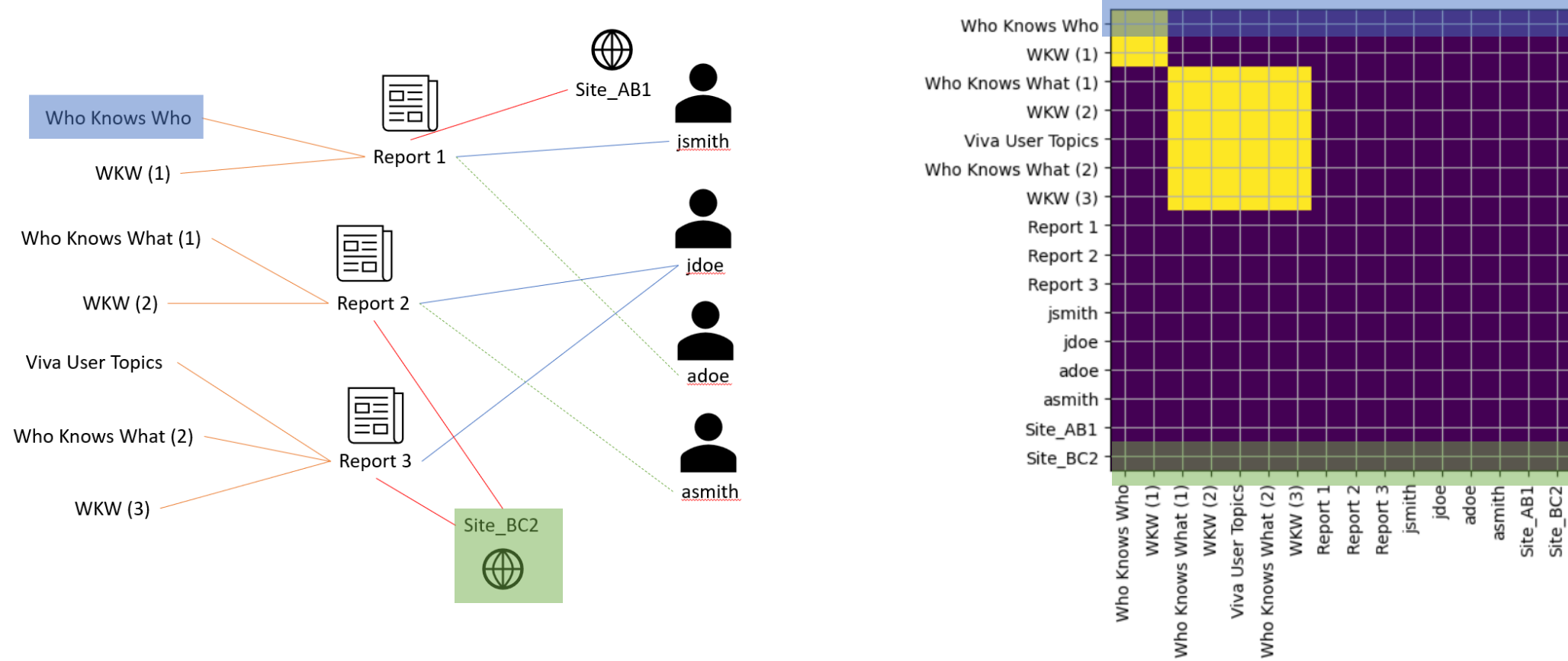
Cosine similarity

Relevance **normalized**

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

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

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

(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

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

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 $\text{rel}_{uv} = 0$, then

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

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

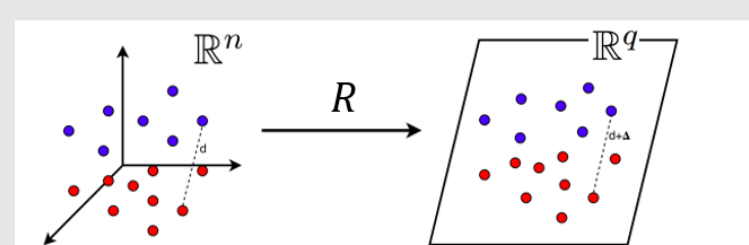
$$\text{if } \text{rel}_{uv} - \text{rel}_{wh} > 2\varepsilon, \text{ then } \mathbb{P}(\text{rel}_{uv}^R > \text{rel}_{wh}^R) \geq 1 - \delta.$$

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

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

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

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

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 $A_{ij} = A_{ji} = 1$ if page i links to page j , and 0 otherwise

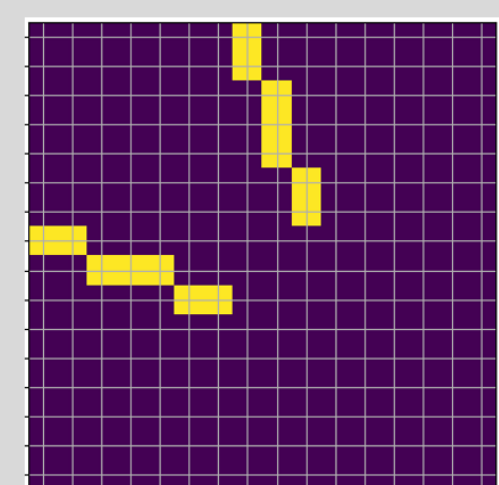
Computational Advantages

If P is a polynomial of matrices, it will be easier to compute PR^T .

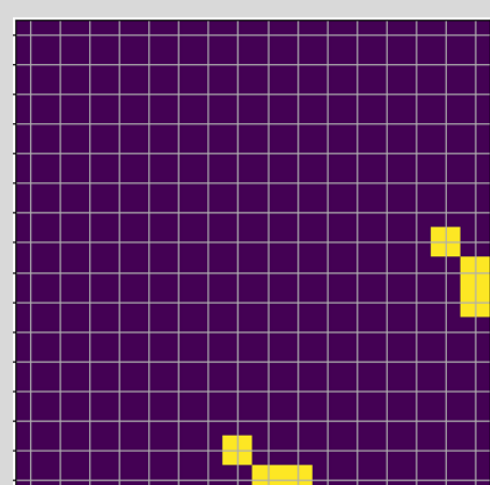
Example. $P = A_1 A_2$

Step 1: $X := A_2 R^T$

Step 2: $X := A_1 X$



$A_{\text{mentioned in}}$



$A_{\text{contained in}}$

Low and high degree nodes

For a small positive integer c we define

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

and

$$H := \{u \in V : d_u \geq \gamma^2 c q\}$$

where γ is a technical constant and q the projection dimension

Will H have nodes?
Power law considerations

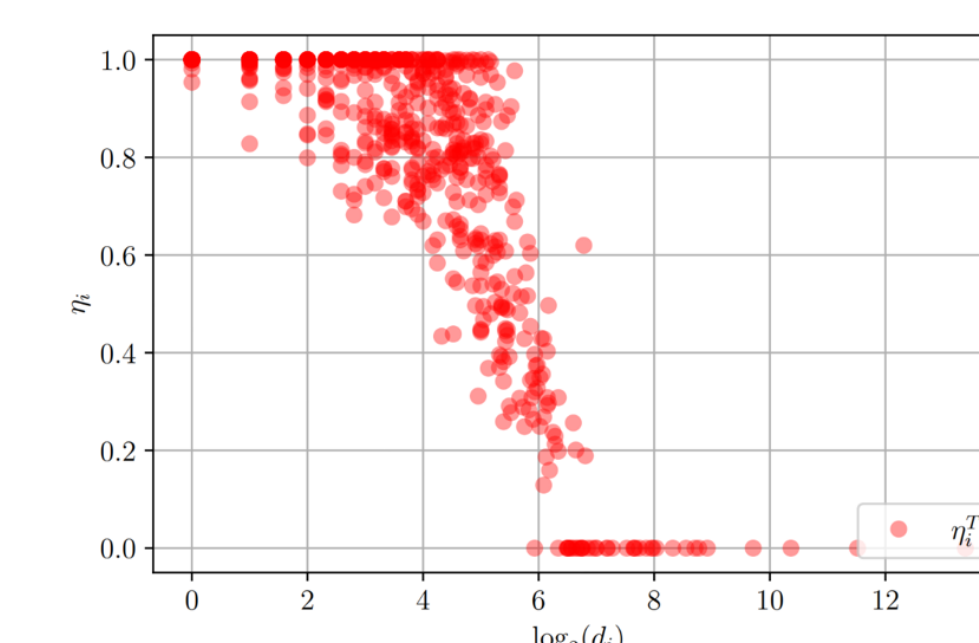
JL Lemma: $q = O(\log n)$

Estimate:

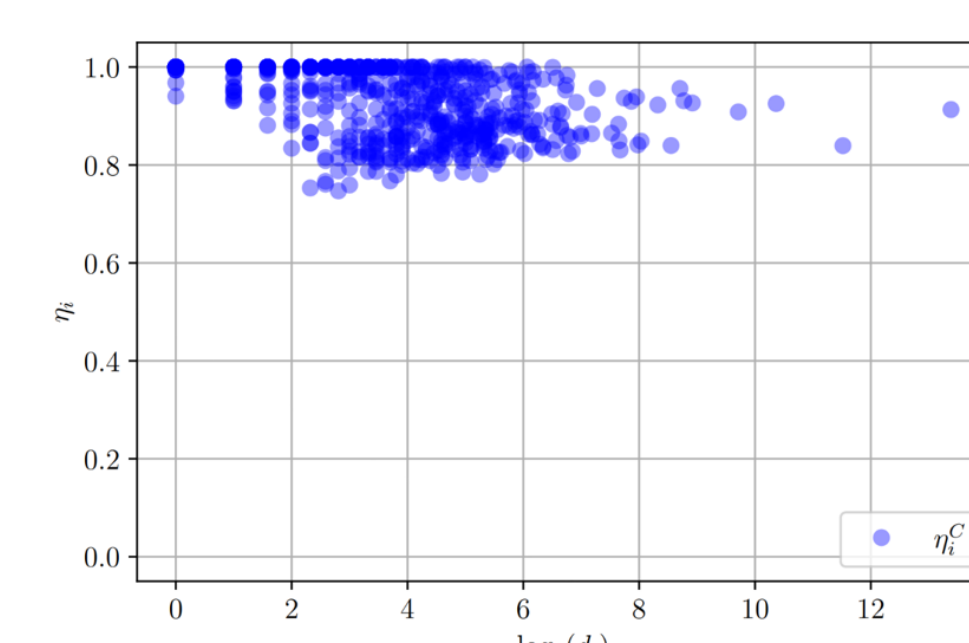
$$\frac{\#H}{n} \sim C[\log n]^{-\alpha}$$

NDCG@10 vs Node Degree

Dot product similarity



Cosine similarity



$P = T$

Published as a conference paper at ICLR 2025

NODE SIMILARITIES UNDER RANDOM PROJECTIONS: LIMITS AND PATHOLOGICAL CASES

Tvrtko Tadić
Microsoft Search, Assistant and Intelligence
tvrtkotad@microsoft.com

Cassiano O. Becker
Microsoft Search, Assistant and Intelligence
casabecker@microsoft.com

Jennifer Neville
Microsoft Research
jennneville@microsoft.com

ABSTRACT

Random Projections have been widely used to generate embeddings for various graph learning tasks due to their computational efficiency. The majority of applications have been justified through the Johnson-Lindenstrauss Lemma. In this paper, we take a step further and investigate how well dot product and cosine similarity are preserved by random projections when these are applied over the rows of the graph matrix. Our analysis provides new asymptotic and finite-sample results, identifies pathological cases, and tests them with numerical experiments. We specialize our fundamental results to a ranking application by computing the probability of random projections flipping the node ordering induced by their embeddings. We find that, depending on the degree distribution, the method produces especially unreliable embeddings for the dot product, regardless of whether the adjacency or the normalized transition matrix is used. With respect to the statistical noise introduced by random projections, we show that cosine similarity produces remarkably more precise approximations.

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