The Journal of Open Source Software

Graphem - JAX: Node Influence Maximization via Geometric Embeddings

Alexander Kolpakov [1*¶] **and Igor Rivin** [2*]

1 University of Austin, Austin TX, USA; akolpakov@uaustin.org **2** Temple University, Philadelphia PA, USA; rivin@temple.edu ¶ Corresponding author * These authors contributed equally.

DOI: 10.21105/joss.08855

Software

- Review
- Repository
- Archive

Editor: Daniel S. Katz

Reviewers:

- @leotrs
- @giorgospanay

Submitted: 18 August 2025
Published: 25 September 2025

Summary

Computing classical centrality measures such as betweenness and closeness is computationally expensive on large graphs. Graphem-JAX provides an efficient force layout algorithm that embeds a graph into a low-dimensional space in such a way that the radial distance from the origin serves as a proxy for various centrality measures.

Our method shows strong correlations with degree, PageRank, and paths-based centralities on multiple common graph families such as Erdős–Rényi random graphs, Watts-Strogatz "small world" graphs, and a few other. We also provide benchmarks on real world datasets such as Social Circles (Facebook) and Wikipedia Vote Network. Moreover in practice, Graphem-JAX allows one to find high-influence nodes in a network, and provides a fast and scalable alternative to the standard greedy algorithm.

Statement of need

Graph centrality measures provide crucial insights into network structure and influence. However, the computation of combinatorial measures such as betweenness or closeness is often infeasible for graphs with a large number of vertices n, as it often growth as $O(n^2)$ in practice, and cannot be parallelized. In contrast, spectral and force-directed methods are inherently parallelizable, and thus offer scalable alternatives.

This paper proposes a force layout algorithm that leverages a Laplacian-based initialization followed by iterative force updates to produce an embedding where the radial distance reflects node importance. We further explore potential applications of this embedding, in particular to finding high-importance communities, and compare our embedding to the baseline greedy algorithm using random cascades.

Benchmarks

We use Spearman's ρ correlation instead of Pearson's correlation as the relationship between radial ordering and centrality is not necessarily linear (as the force layout is highly non-linear), and because what matters most is the ordering, not the actual distance or centrality values.

Confidence intervals and p-values are obtained by boostrapping with $N = 1000$ replicates. More benchmarks for other graph families and embedding dimensions are available in the whitepaper (Kolpakov & Rivin, 2025).

Synthetic datasets

Centrality Measure	ρ	95% CI	p-value
Degree	0.829	[0.803, 0.854]	$< 10^{-6}$
Betweenness	0.845	[0.817, 0.867]	$< 10^{-6}$
Eigenvector	0.806	[0.778, 0.833]	$< 10^{-6}$
PageRank	0.835	[0.807, 0.859]	$< 10^{-6}$
Closeness	0.830	[0.802, 0.855]	$< 10^{-6}$
Node Load	0.845	[0.818, 0.866]	$< 10^{-6}$

Table: Spearman correlations of centrality measures with the radial distance in graph embeddings for Erdős–Rényi graphs. Embedding dimension 2.

Centrality Measure	ρ	95% CI	p-value
Degree	0.896	[0.877, 0.912]	$< 10^{-6}$
Betweenness	0.748	[0.718, 0.776]	$< 10^{-6}$
Eigenvector	0.646	[0.605, 0.682]	$< 10^{-6}$
PageRank	0.897	[0.878, 0.912]	$< 10^{-6}$
Closeness	0.594	[0.549, 0.633]	$< 10^{-6}$
Node Load	0.743	[0.711, 0.771]	$< 10^{-6}$

Table: Spearman correlations of centrality measures with the radial distance in graph embeddings for Watts–Strogatz graphs. Embedding dimension 2.

Real world datasets

Centrality Measure	ρ	95% CI	p-value
Degree	0.864	[0.851, 0.877]	$< 10^{-5}$
Betweenness	0.721	[0.704, 0.740]	$< 10^{-5}$
Eigenvector	0.537	[0.513, 0.560]	$< 10^{-5}$
PageRank	0.746	[0.730, 0.763]	$< 10^{-5}$
Closeness	0.592	[0.571, 0.610]	$< 10^{-5}$
Node Load	0.718	[0.698, 0.736]	$< 10^{-5}$

Table: Spearman correlations of centrality measures with the radial distance in a graph embedding for the SNAP *"Social circles: Facebook"* dataset (Leskovec & McAuley, 2012). Embedding dimension 4.

Centrality Measure	ρ	95% CI	p-value
Degree	0.955	[0.950, 0.959]	$< 10^{-5}$
Betweenness	0.934	[0.928, 0.939]	$< 10^{-5}$
Eigenvector	0.852	[0.840, 0.863]	$< 10^{-5}$
PageRank	0.952	[0.947, 0.956]	$< 10^{-5}$
Closeness	0.839	[0.827, 0.850]	$< 10^{-5}$
Node Load	0.933	[0.928, 0.938]	$< 10^{-5}$

Table: Spearman correlations of centrality measures with the radial distance in a graph embedding for the SNAP *"Wikipedia vote network"* dataset (Leskovec et al., 2008). Embedding dimension 3. We subsampled 5250 vertices to reduce computational load for combinatorial centrality measures.

Node influence maximization

We benchmark Graphem-JAX against the greedy maximization algorithm using Independent Cascades (IC) with adjacent node activation probability $p_{ic} = 0.1$ and $k = 10$ seed vertices. The IC algorithm realization used here is supplied by the NDlib library (Rossetti & others, 2016). The benchmark was repeated 50 time to collect a statistical sample. The outcomes of using the graph embedding as opposed to the greedy seed selection are given below.

Method	Influence	Iterations	Time (s)
Embedding	24.6 $\pm$ 6.9	200	0.26 $\pm$ 0.48
Greedy	23.7 $\pm$ 5.1	247,200	15.97 $\pm$ 0.08

Table: Influence spread, number of NDlib simulation iterations, and runtime for embedding-based vs. greedy method. Synthetic dataset: a random Erdős–Rényi graph on 128 nodes with edge probability $p = 0.05$.

Method	Influence	Iterations	Time (s)
Embedding	23.9 $\pm$ 6.0	200	0.19 $\pm$ 0.01
Greedy	22.9 $\pm$ 5.7	247,200	15.95 $\pm$ 0.07

Table: Influence spread, number of NDlib simulation iterations, and runtime for embedding-based vs. greedy method. Real-world dataset: *"General Relativity and Quantum Cosmology collaboration network"* (Leskovec et al., 2007).

Code availability

The Graphem-JAX repository is available on GitHub. An installable package is available on PyPI.

Whitepaper

The Graphem-JAX whitepaper is available on the arXiv preprint server.

Acknowledgements

This work is supported by the Google Cloud Research Award number GCP19980904.

References

Kolpakov, A., & Rivin, I. (2025). Fast geometric embedding for node influence maximization. *arXiv:2506.07435*. https://doi.org/10.48550/arXiv.2506.07435

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2008). *Wikipedia vote network*. http://snap.stanford.edu/data/wiki-Vote.html. https://doi.org/10.1145/1753326.1753532

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). *Graph of coauthorships in the e-print arXiv general relativity and quantum cosmology category*. http://snap.stanford.edu/data/ca-GrQc.html. https://doi.org/10.1145/1217299.1217301

Leskovec, J., & McAuley, J. J. (2012). *Facebook social circles network dataset*. http://snap.stanford.edu/data/egonets-Facebook.html.

Rossetti, G., & others. (2016). *NDlib: A Python Library to Model and Analyze Diffusion Processes on Complex Networks*. https://github.com/GiulioRossetti/ndlib. https://doi.org/10.1007/s41060-017-0086-6