Northeastern University
Network Science Institute

Graph Distance from the Topological Perspective of Nonbacktracking Cycles

Leo Torres
leo@leotrs.com

Tina Eliassi-Rad
tina@eliassi.org

Abstract

- Measuring **distances between two or more graphs** have many applications in machine learning and data mining -- e.g., transfer learning, graph clustering, and anomaly detection.
- A **nonbacktracking cycle (NBC)** is a closed walk which does not retrace an edge immediately after traversing it.
- **NBCs track graph features** such as the degrees and triangles.
- We propose a **graph distance measure based on NBCs**, backed by results from **homotopy theory** (a branch of **algebraic topology**).

Nonbacktracking Matrix B

- For a graph with m edges, the **nonbacktracking matrix B** is $2m \times 2m$.
- Each edge is represented by **two rows** and **two columns**, one per orientation.
- There is a 1 in the entry indexed by row $u \to v$ and column $k \to l$ when $u \neq l$ and $v = k$; and a 0 otherwise.
- **B** is the transition matrix of a random walker with one-step memory that never traces an edge immediately after traversing it.
- **B** is a good alternative for computing **graph distance** because it:
  I. **tracks information about features** such as degrees and triangles,
  II. can be **fine tuned** to be more or less sensitive to such features,
  III. is backed by **results from homotopy theory**.
- We give an efficient algorithm to compute the **nonbacktracking matrix.**
- To our knowledge, the **connection between NBCs and homotopy theory** hasn't been fully realized.

Future Work

1. Can **algebraic-topological features** (e.g. lengths of NBCs) be used to define a **metric** on the space of all graphs?
2. What other **structural and dynamic graph measures** are stored in **NBCs and B**?
3. What other **learning and mining tasks** can be improved by the use of **NBCs and B**?
4. Which results from **homotopy theory** are **computationally tractable?**

Computing B and its Properties

Step 1. Compute the $n \times 2m$ matrices $M^{+}_{x,u\to v} = \delta_{xu} \quad M^{-}_{x,u\to v} = \delta_{xv}$, and their product $C = (M^{+})^{T} M^{-}$.

Step 2. Observe that $C_{k\to l,u\to v} = \delta_{vk}$ while $B_{k\to l,u\to v} = \delta_{kv}(1-\delta_{ul})$

Thus, we can compute ***B*** entrywise

$$B_{k\to l,u\to v} = C_{k\to l,u\to v}(1 - C_{u\to v,k\to l})$$

Here, δ_{ij} equals 1 when $i = j$.

Time complexity:

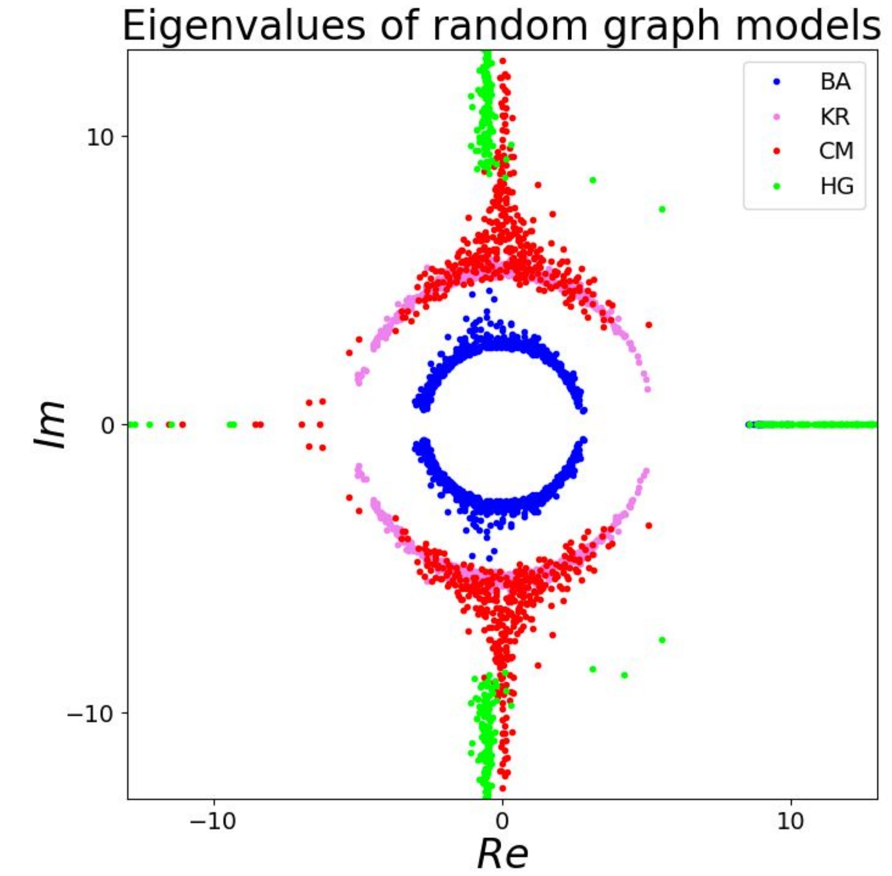
- $O(m + n\langle k^2\rangle)$ in general
- $O(m + n)$ in graphs with homogeneous degree distributions
- Between $O(m + n)$ and $O(n^2)$ in graphs with power-law degree distributions

B and the **degree distribution**

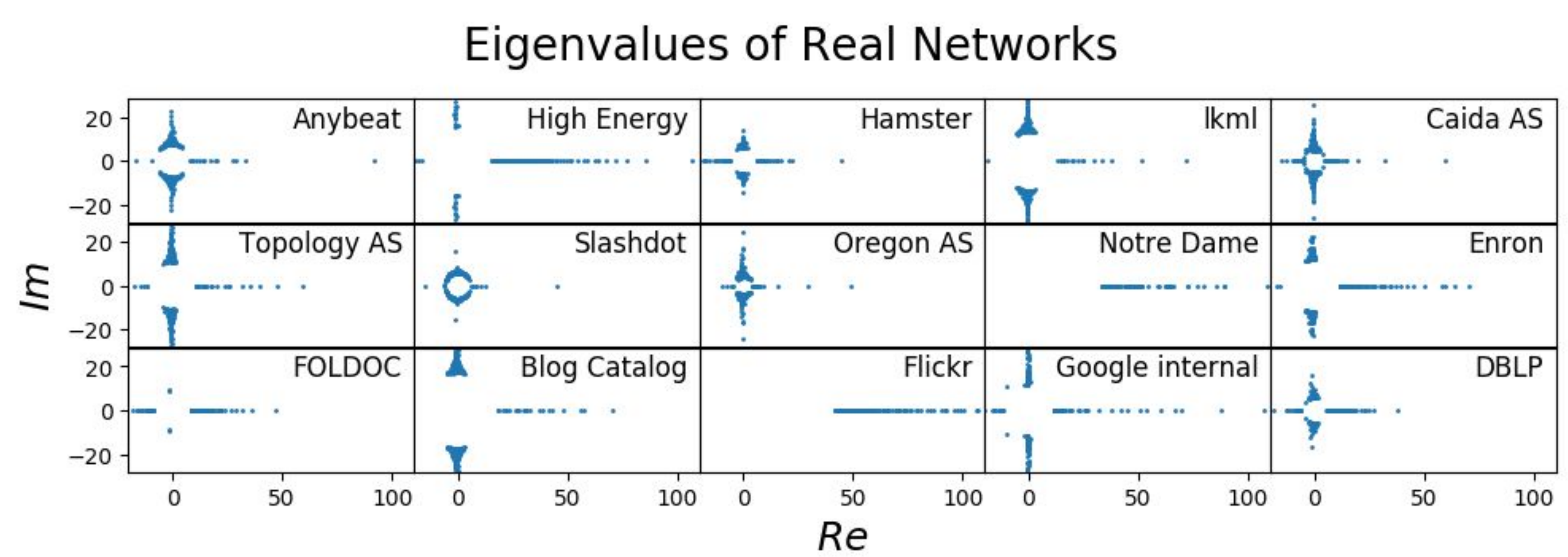
$nnz(\mathbf{B}) = n\langle k^2\rangle - n\langle k\rangle$

B and **triangles**

$tr(B^3) = \Sigma_k \alpha_k(\alpha_k^2 - 3\beta_k^2)$ for $\lambda_k = \alpha_k + i\beta_k$.



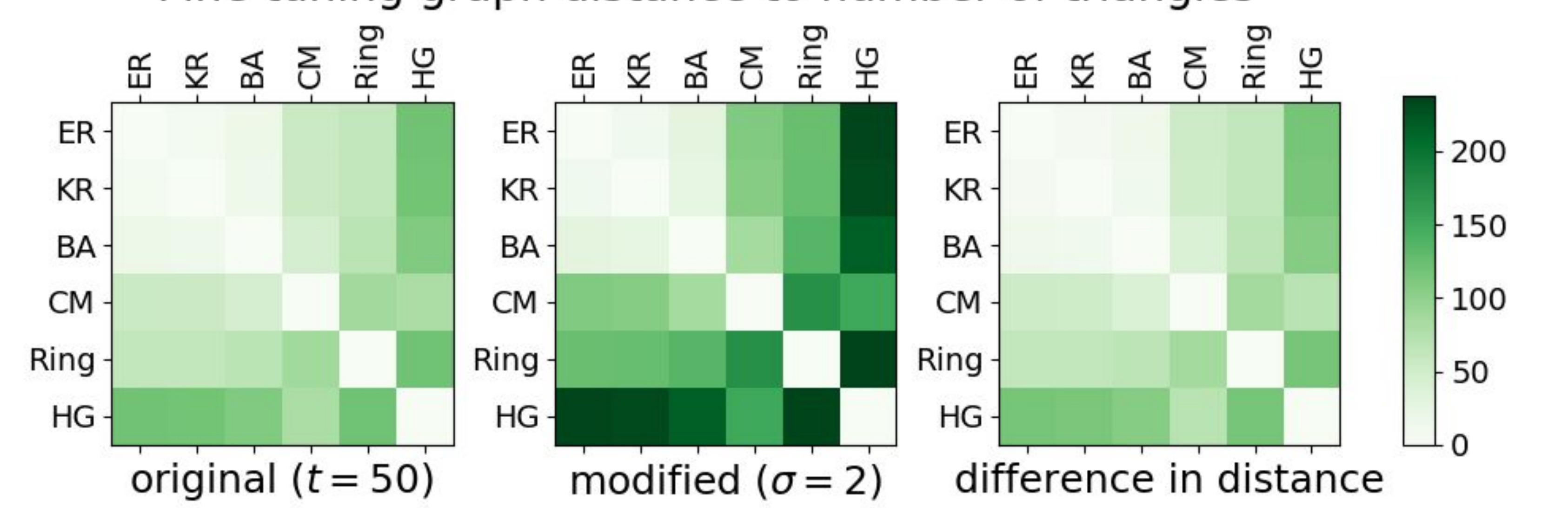
Top 100 eigenvalues for random graphs of different models: Barabasi-Albert (BA), KR (Kronecker Graphs), Configuration Model (CM), Hyperbolic Graphs (HG). All graphs have 11,000 nodes and approximately⟨k⟩ = 10.



Top 100 eigenvalues for real networks of diverse domains: social, communications, scientific citations, autonomous systems of the internet, and Web graphs.

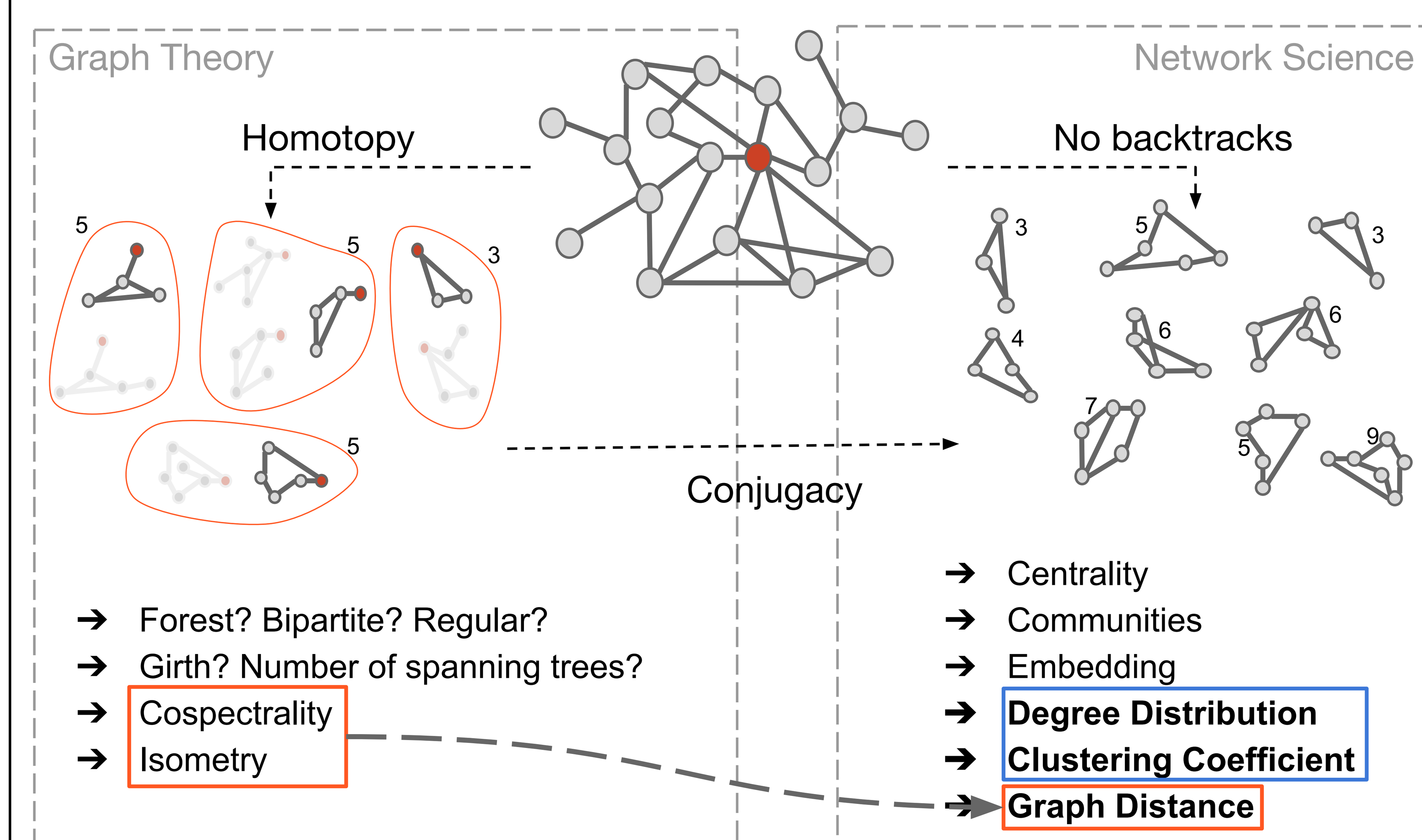
A Measure of Graph Distance

Graph	G_1	G_2
Eigenvalues	$\lambda_k = a_k + ib_k,\ k = 1..t$	$\mu_k = \alpha_k + i\beta_k,\ k = 1..t$
Features	$(a_1,..,a_t,b_1,..,b_t) = (a,b)$	(α,β)
Fine-tuned	$(\sigma a, b/\sigma)$	$(\sigma\alpha, \beta/\sigma)$



- ER: Erdos-Renyi, KR: Kronecker Graph, BA: Barabasi-Albert, CM: Configuration Model, Ring: ring lattice, HG: Hyperbolic Graph.
- **Left:** distance between the mean feature vector of several random graph models. Models are ordered in increasing number of triangles from left to right.
- **Center:** distance between mean modified feature vector.
- **Right:** difference in previous two distances.
- Observe that elements away from the diagonal have a larger difference in number of triangles, which are amplified accordingly.

Relationship to Homotopy Theory



We thank **Evimaria Terzi** and **Pablo Suárez Serrato** for their contributions to this work. Torres and Eliassi-Rad were supported by **NSF** CNS-1314603, **NSF** IIS-1741197, and **DTRA** HDTRA1-10-1-0120.