A Study of Cycle Length Spectra: Connecting Homotopy to Network Science

Northeastern University
Network Science Institute

Leo Torres
leo@leotrs.com

Tina Eliassi-Rad
tina@eliassi.org

Abstract

- We highlight the connection between **nonbacktracking walks (NBWs)** and **homotopy theory** from algebraic topology.
- We show how **NBWs track structural graph measures** such as the degree distribution and clustering coefficient.
- We propose a **graph distance measure based on NBWs**.

Nonbacktracking Matrix

- Given a graph G with m edges, the **nonbacktracking matrix B** is a $2m \times 2m$ matrix.
- Each edge in G is represented by two rows in **B**, one per orientation: $u \to v$ and $v \to u$.
- For two edges $u \to v$ and $k \to l$, **B** is given by

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

where δ_{ij} is the Kronecker delta.

- Example: 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.

(1) Computing B and its Properties

- We compute **B** in two steps

Step 1: compute

$M^{+}_{x, u \to v} = \delta_{xu}$	$O(m)$
$M^{-}_{x, u \to v} = \delta_{xv}$	$O(m)$
$C = (M^{+})^{T} M^{-}$	$O(n \langle k^2 \rangle)$

Step 2: compute **B** entrywise:

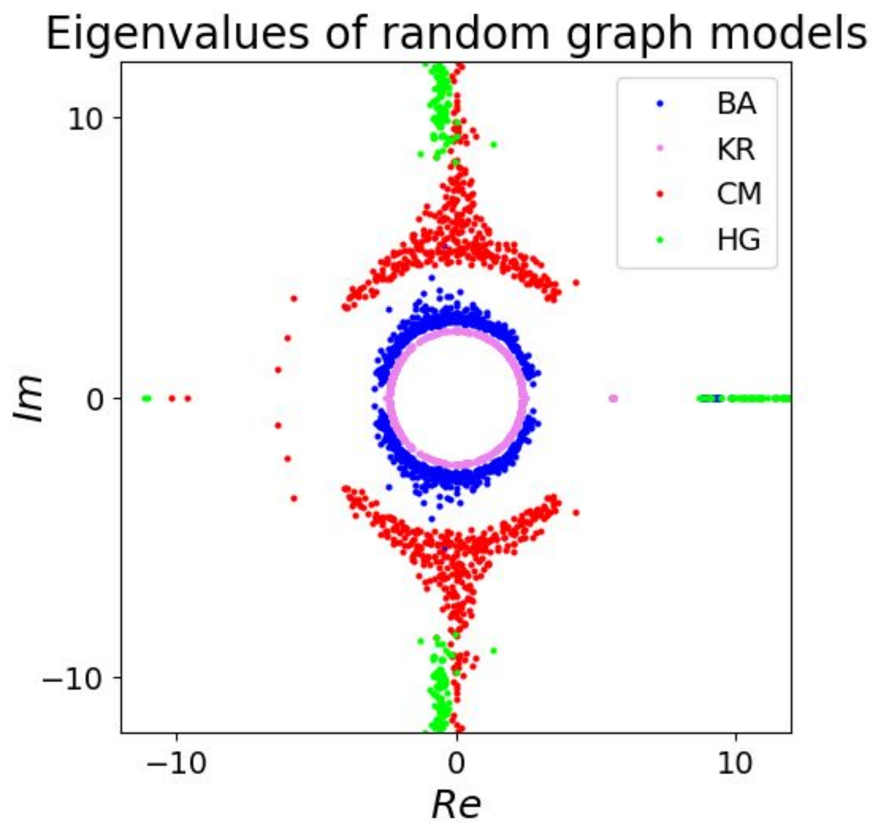
$$C_{k \to l, u \to v} = \delta_{vk}$$

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

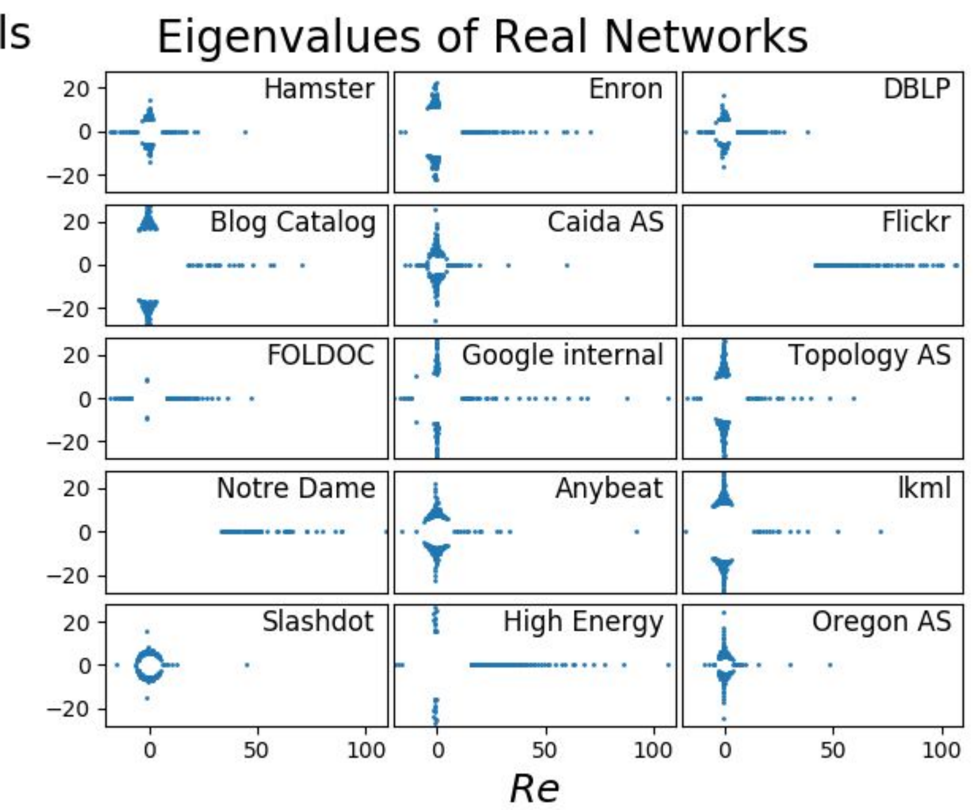
- **Time complexity** $O(m + n\langle k^2 \rangle)$.
- For homogeneous networks: $O(m + n)$.
- For power-law degree distributions: between $O(m + n)$ and $O(n^2)$.
- The number of non-zero elements of **B** is related to the **the degree distribution**: $nnz(\mathbf{B}) = n\langle k^2 \rangle - n\langle k \rangle$.
- If $\lambda_k = a_k + i\beta_k$ are the **eigenvalues** of **B**, then the **number of triangles** is

$$tr(B^3) = \sum_k a_k (a_k^2 - 3b_k^2)$$

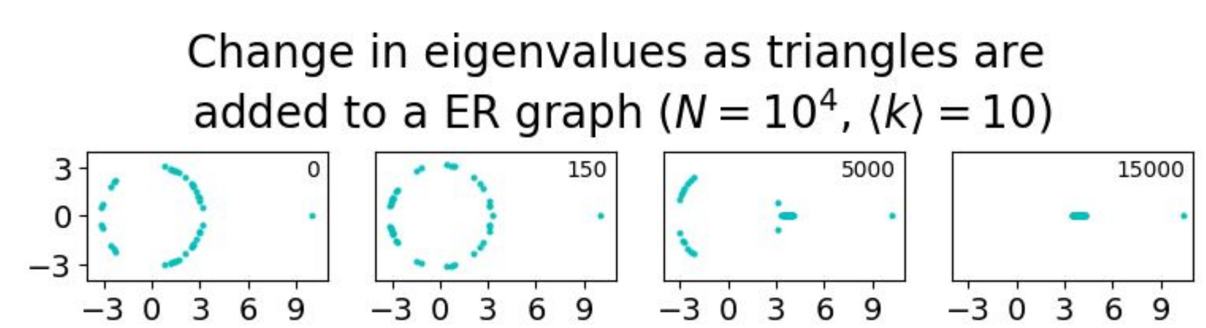
- If $(\sum_k a_k^2)$ is large and $(\sum_k b_k^2)$ is small, then number of triangles is large.



Shown are the largest 40 eigenvalues of B of 20 random graphs of each model. BA: Barabási-Albert, KR: Kronecker Graph, CM: Configuration Model, HG: Hyperbolic Graph. CM and HG generated with power law degree distribution with exponent $\gamma = 2.1$. All graphs have $n = 10^4$ nodes and approximately $\langle k \rangle = 10$.



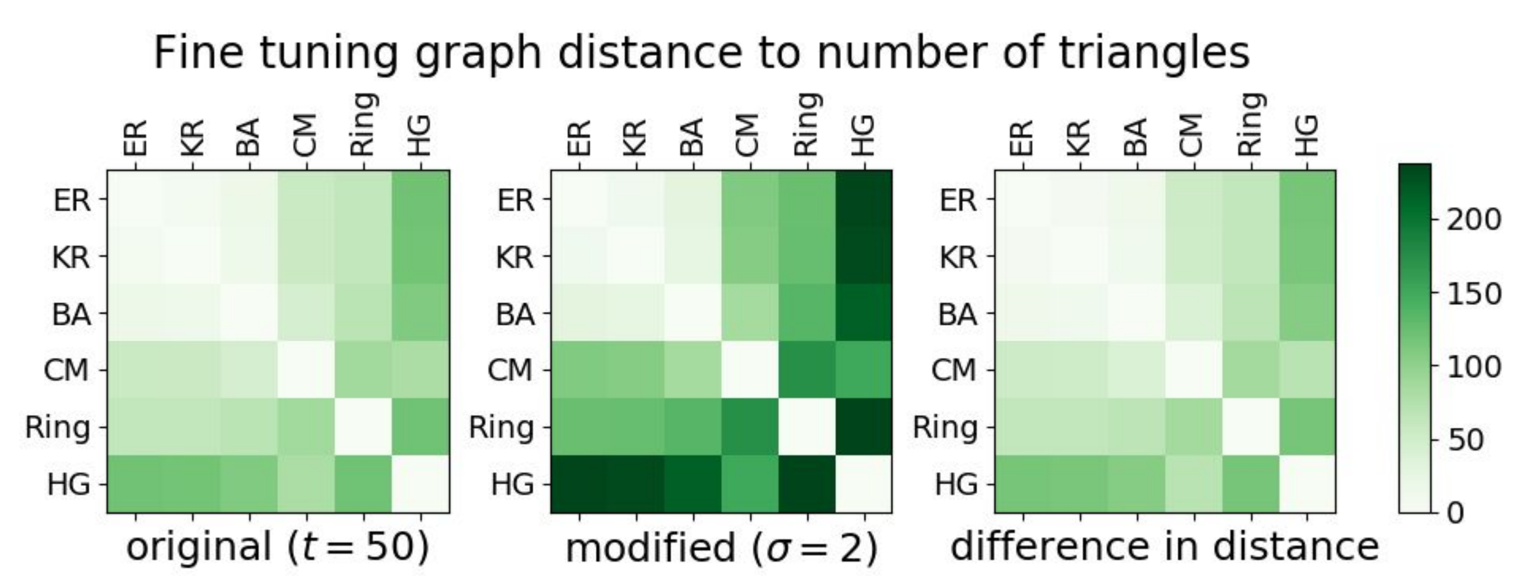
Top 100 eigenvalues of B for real networks of diverse domains: social (Hamster, Blog Catalog, Anybeat, Flickr), communications (Enron, Ikml), scientific citations (DBLP, High Energy), autonomous systems of the internet (Caida AS, Oregon AS, Topology AS), web graphs (Notre Dame, Google Internal, FOLDOC).



As more triangles are added to the network, B's eigenvaluesmove further right and closer to the horizontal axis. If $\lambda_k = a_k + ib_k$, then a_k^2 increases and b_k^2 decreases.

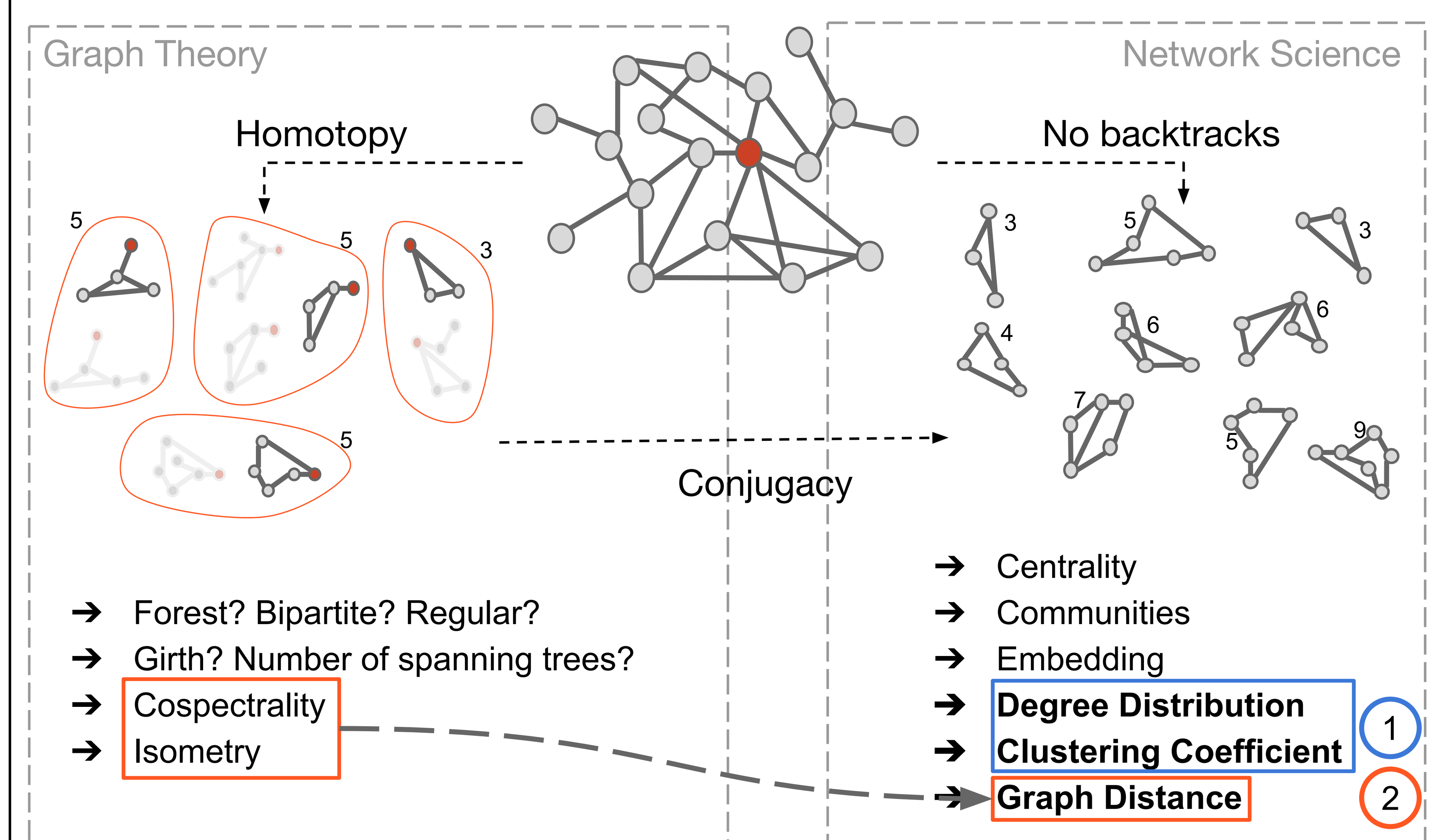
(2) A Measure of Graph Distance

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	$(\sigma a, b/\sigma)$	$(\sigma\alpha, \beta/\sigma)$



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. ER: Erdos-Renyi, KR: Kronecker Graph, BA: Barabasi-Albert, CM: Configuration Model, Ring: ring lattice, HG: Hyperbolic Graph.

Homotopy Theory



Discussion

We propose **B as a better alternative for computing graph distance** because it:

I. **carries information about important features** such as degrees and triangles,

II. can be **fine tuned** to be more or less sensitive to such features,

III. is backed by the **cospectrality and isometry results from homotopy theory**.

To our knowledge, the **connection between NBWs and homotopy theory** hasn't been fully realized in the Network Science literature.

Open Questions and Future Research

1. How best to take advantage of **homotopy theory in the study of complex networks?**
2. What other **structural and dynamic graph measures** are stored in NBWs and B?
3. Can **algebraic-topological features** be used to define a metric on the space of all graphs?
4. What other **graph mining tasks** can we improve by the use of NBWs and B?

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.