Graph Distance from a Topological View of Nonbacktracking Cycles

Leo Torres
leo@leotrs.com
Pablo Suárez Serrato and Tina Eliassi-Rad

Instituto de Matemáticas UNAM
netsi
Northeastern University
Network Science Institute

In This Talk

1. The Length Spectrum
2. Modifying the Length Spectrum
3. Graph Distance
4. Properties
5. Examples

The Length Spectrum

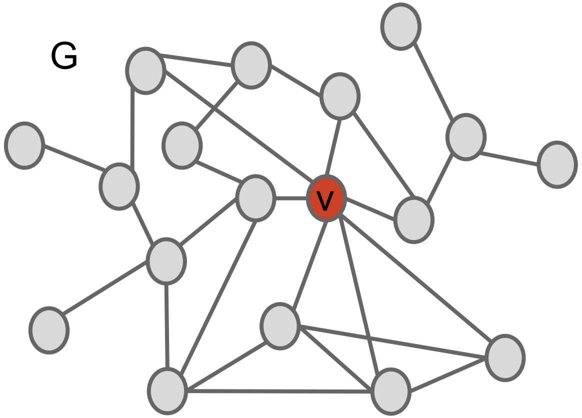
The Length Spectrum

The Length Spectrum of a graph characterizes it uniquely* up to isomorphism.[1]

[1] Constantine, David, and Jean-François Lafont. **“Marked Length Rigidity for One-Dimensional Spaces.”** Journal of Topology and Analysis, 2018. doi:10.1142/s1793525319500250.

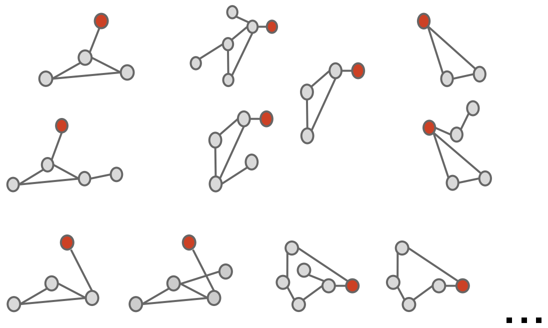
The Length Spectrum

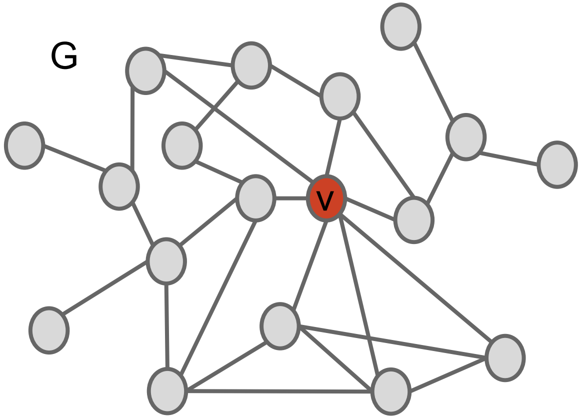
1. Given a graph G = (V, E) and a node v,



The Length Spectrum

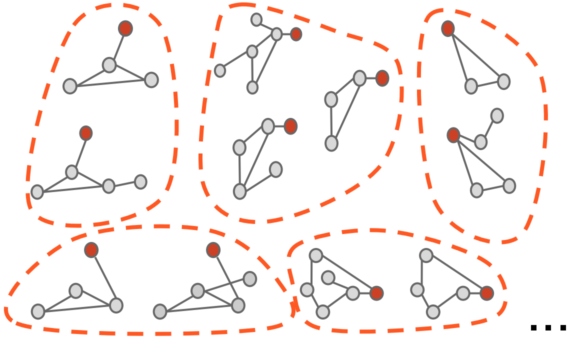
1. Given a graph G = (V, E) and a node v, consider the set of all closed walks that start and end at v.

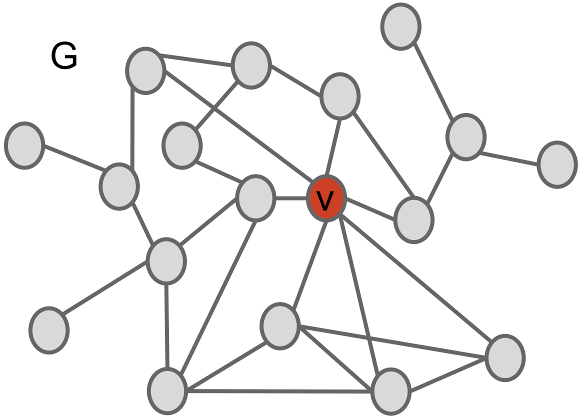




The Length Spectrum

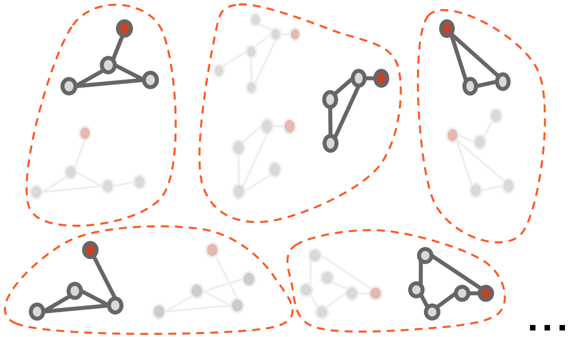
2. Walks are equivalent if they are equal save for **tree-like parts that don't go through the basepoint...**

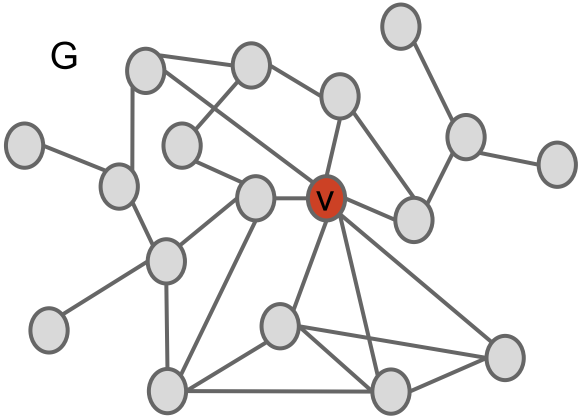




The Length Spectrum

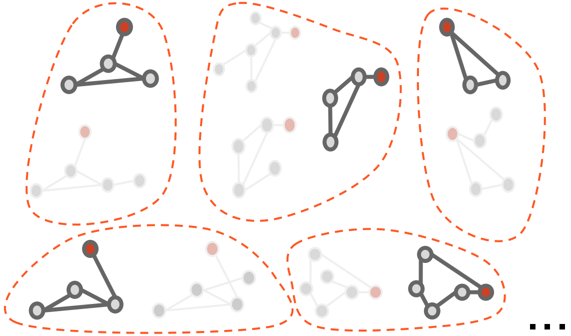
2. … and retain the shortest walk in each subset.



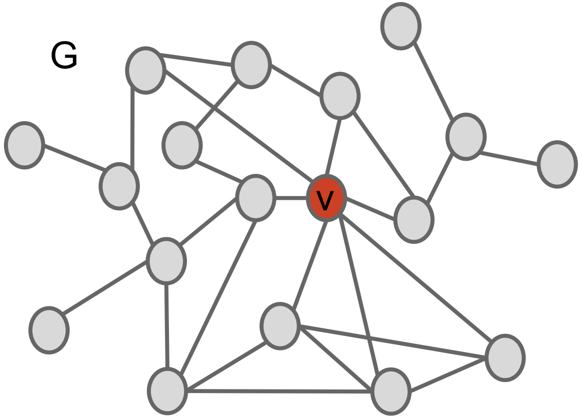


The Length Spectrum

2. This set is the fundamental group of G with basepoint v.

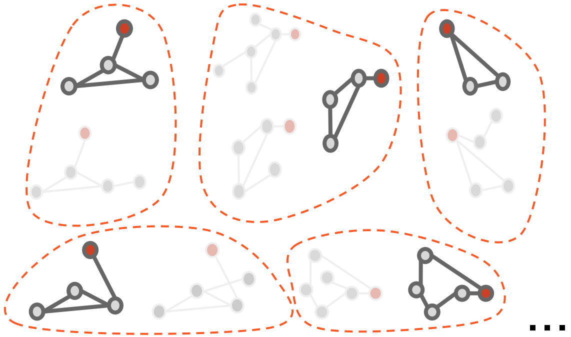


$$\pi_1(G, v)$$

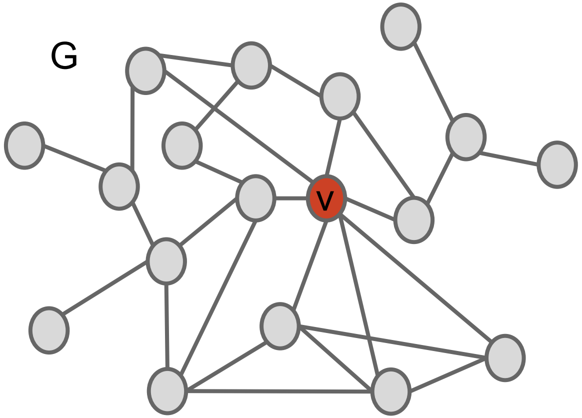


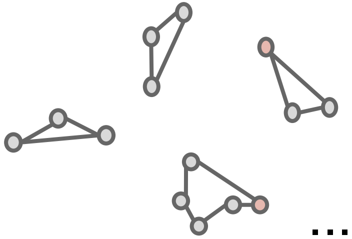
The Length Spectrum

3. Walks are equivalent if they are equal **save for tree-like parts** ~~that don't go through the basepoint~~.



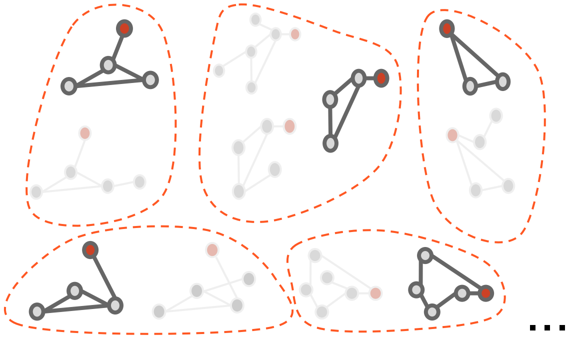
$$\pi_1(G, v)$$



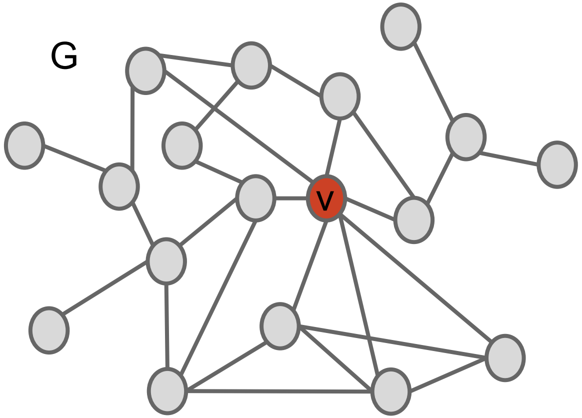


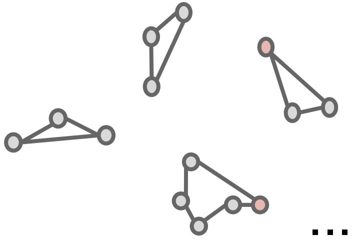
The Length Spectrum

3. This is the set of nonbacktracking cycles (NBCs) of G.



$\pi_1(G, v)$

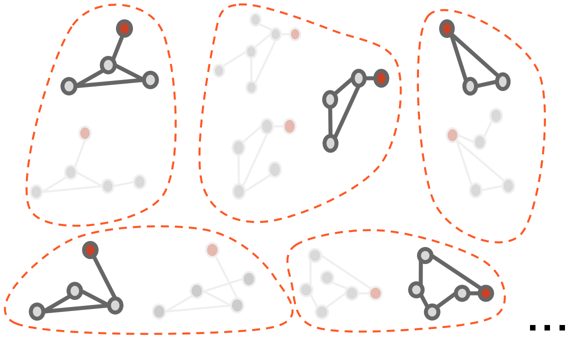




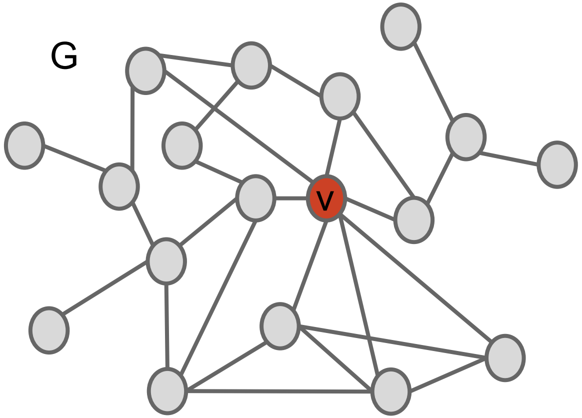
$NBCs$

The Length Spectrum

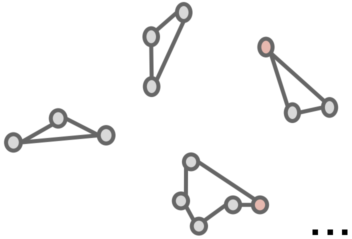
4. $\mathcal{L}$ is defined on $\pi_1(G, v)$ and assigns each walk the length of its "shaved" version.



$\pi_1(G, v)$



$$\mathcal{L} : \pi_1(G, v) \to \mathbb{R}$$



$NBCs$

The Length Spectrum

The Length Spectrum of a graph characterizes it uniquely* up to isomorphism.[1]

[1] Constantine, David, and Jean-François Lafont. **“Marked Length Rigidity for One-Dimensional Spaces.”** Journal of Topology and Analysis, 2018. doi:10.1142/s1793525319500250.

Modifying the Length Spectrum

$$d(G, H) = d(\mathcal{L}_G, \mathcal{L}_H)$$

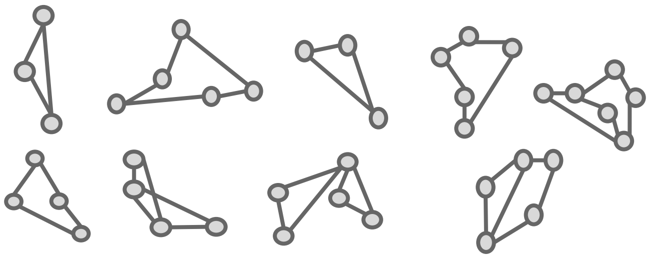
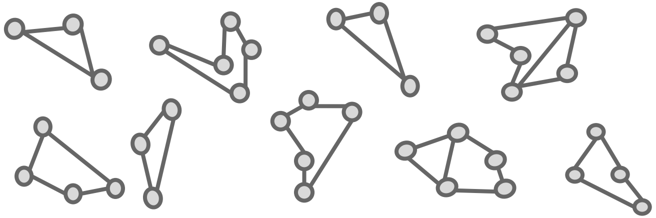
Modifying the Length Spectrum

$$d(G, H) = d(\mathcal{L}_G, \mathcal{L}_H)$$

Two assumptions	Two problems	Two solutions
$G \to \mathcal{L}_G$	How to compute?	Image instead of domain
$d(\mathcal{L}_G, \mathcal{L}_H)$	How to compare?	Partition the image

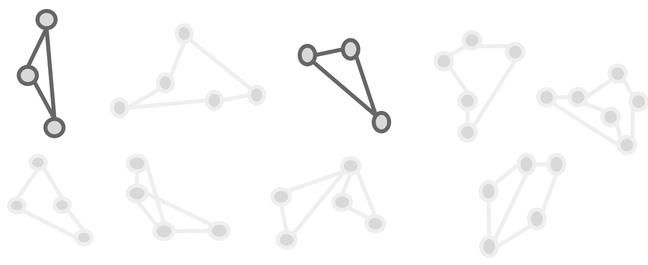
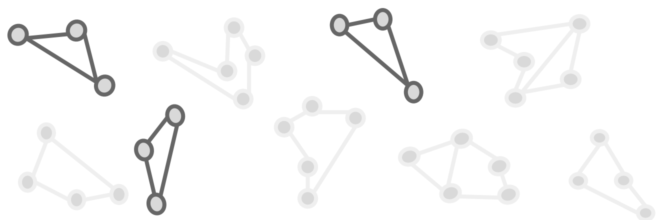
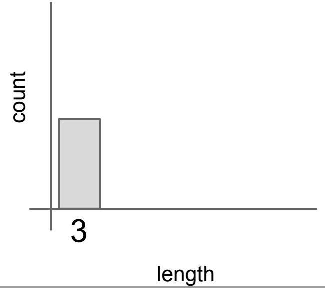
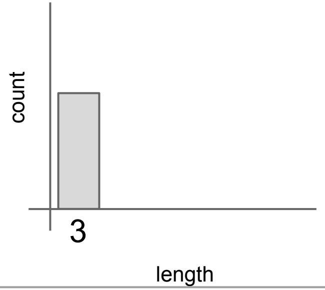
Modifying the Length Spectrum

Partition the image

	G	H
Domain		
Image		

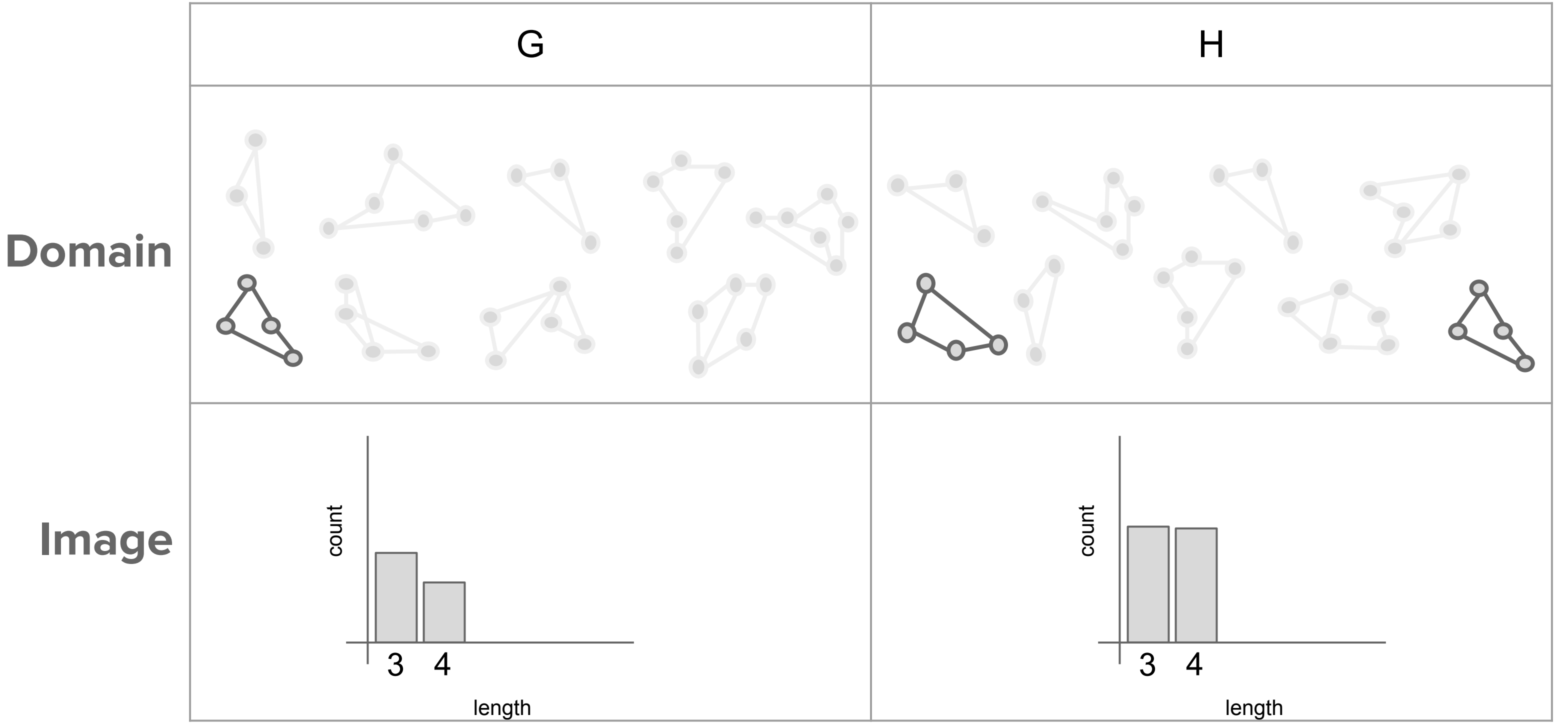
Modifying the Length Spectrum

Partition the image

	G	H
Domain		
Image		

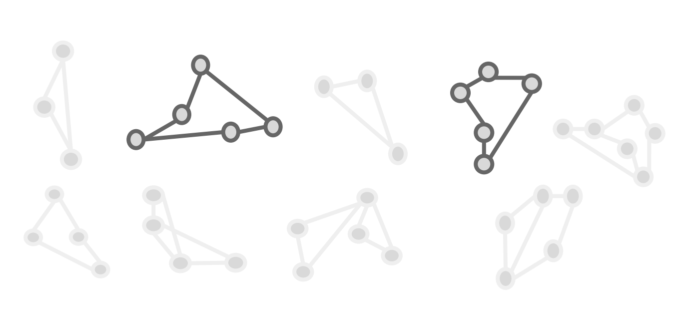
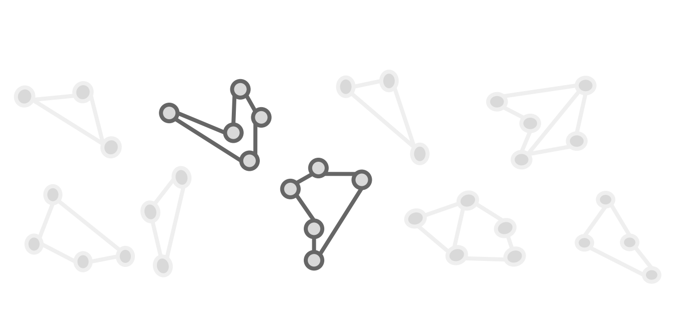
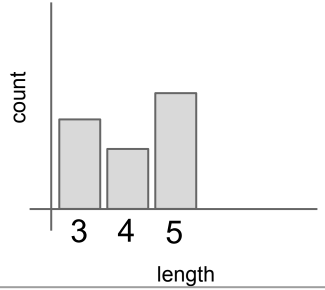
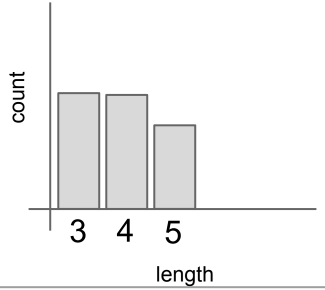
Modifying the Length Spectrum

Partition the image



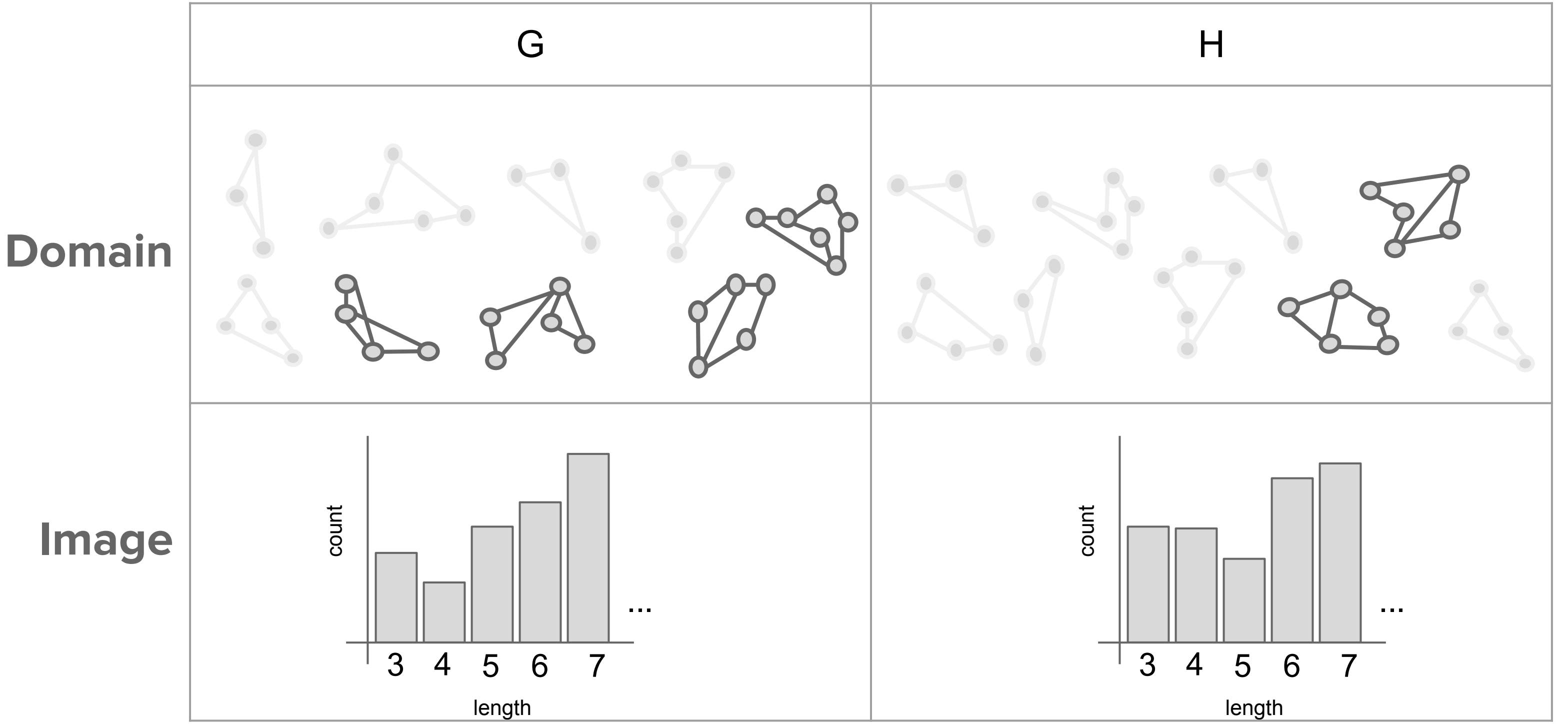
Modifying the Length Spectrum

Partition the image

	G	H
Domain		
Image		

Modifying the Length Spectrum

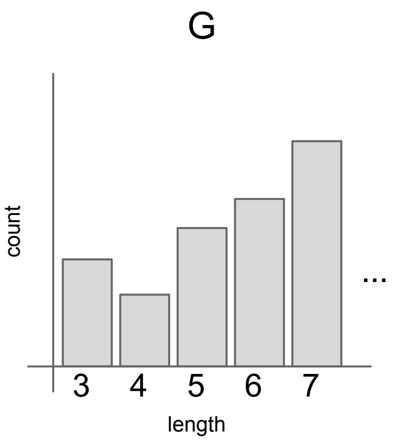
Partition the image

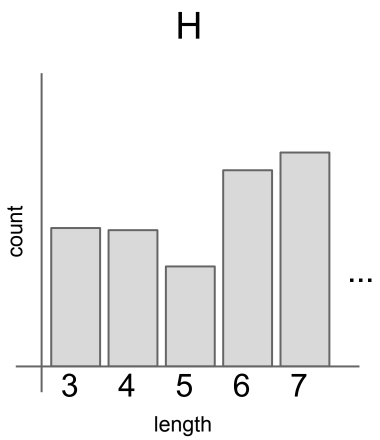


Graph Distance

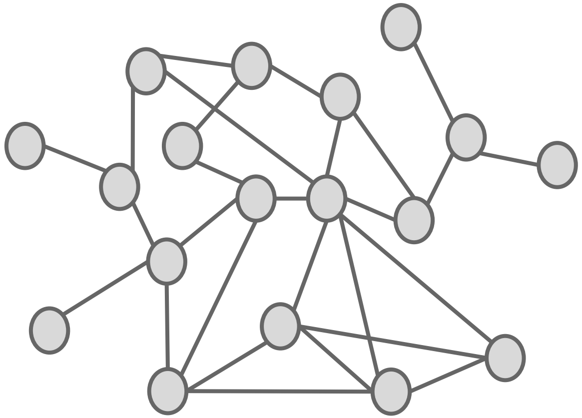
How to compare these two histograms?

- Observe that the height of each bar is the number of NBCs of a certain length.
- We can compute this using the *nonbacktracking matrix* **B**.



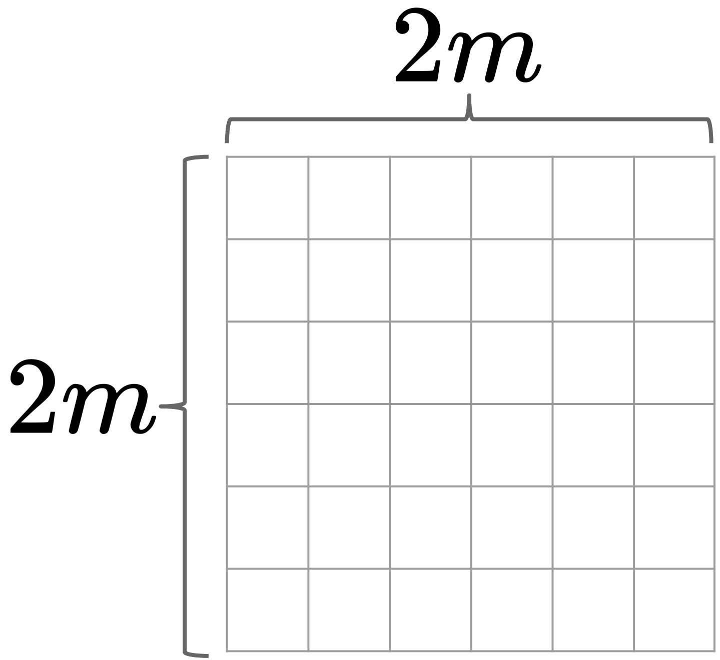


Detour: nonbacktracking matrix B



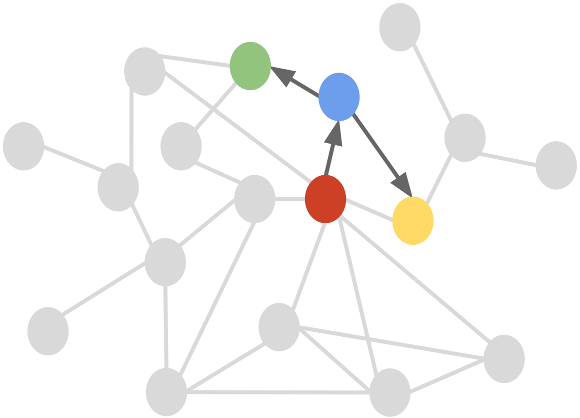
$G = (V, E)$

$|E| = m$

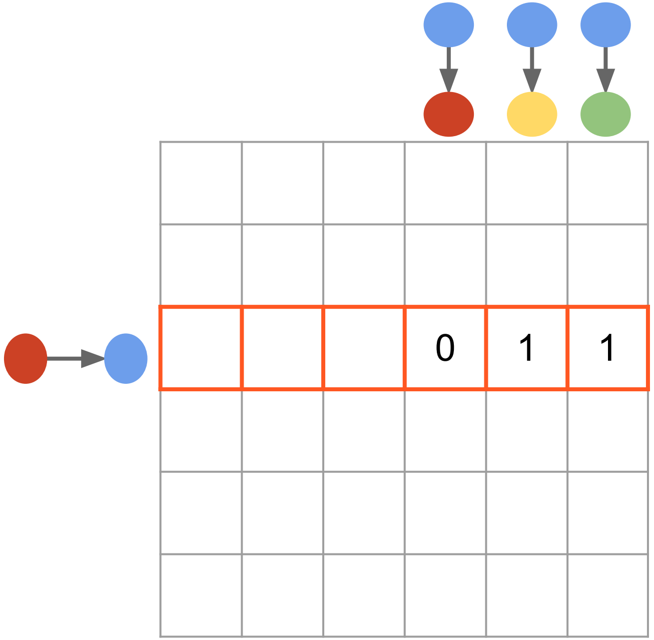


B

Detour: nonbacktracking matrix B



$G = (V, E)$
$|E| = m$

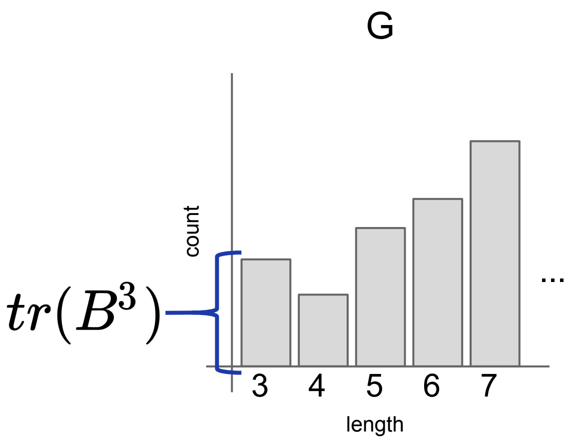


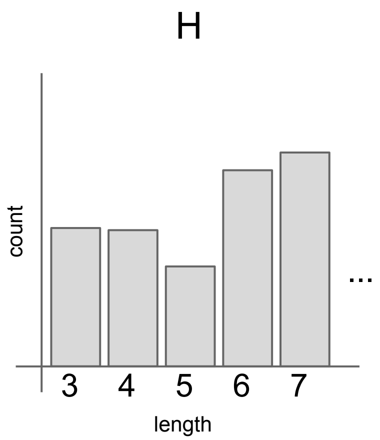
B

Graph Distance

How to compare these two histograms?

- Observe that the height of each bar is the number of NBCs of a certain length.
- We can compute this using the *nonbacktracking matrix* **B**.

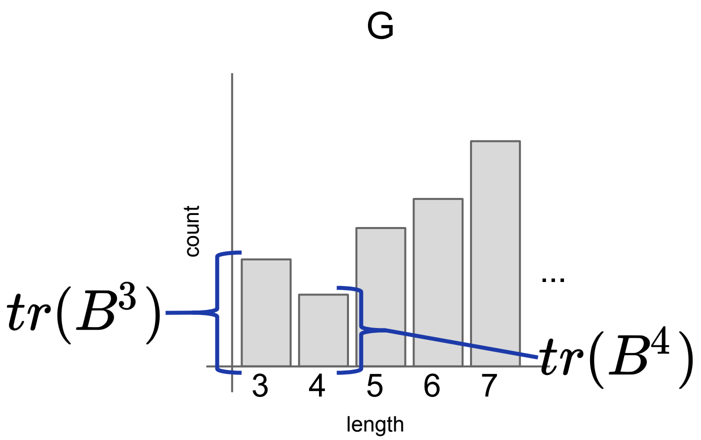


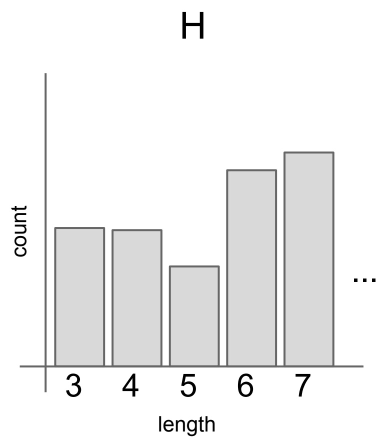


Graph Distance

How to compare these two histograms?

- Observe that the height of each bar is the number of NBCs of a certain length.
- We can compute this using the *nonbacktracking matrix* **B**.

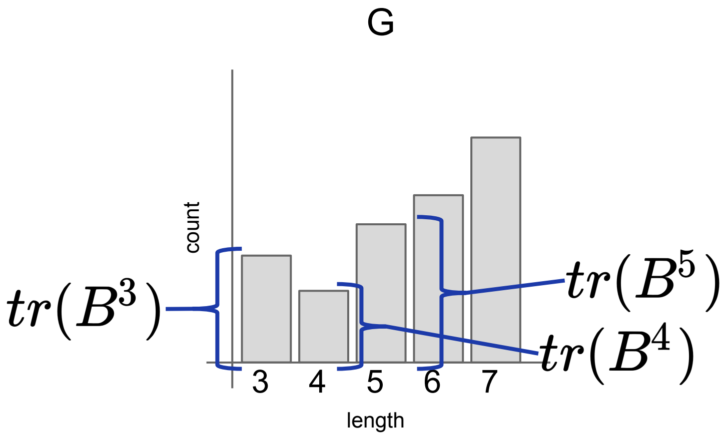


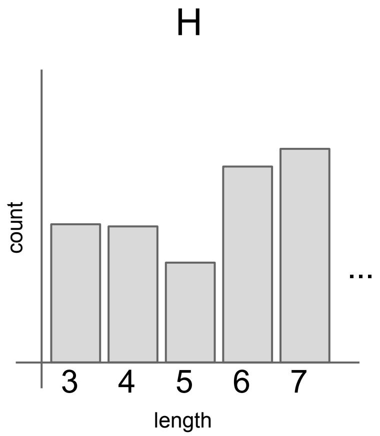


Graph Distance

How to compare these two histograms?

- Observe that the height of each bar is the number of NBCs of a certain length.
- We can compute this using the *nonbacktracking matrix* **B**.

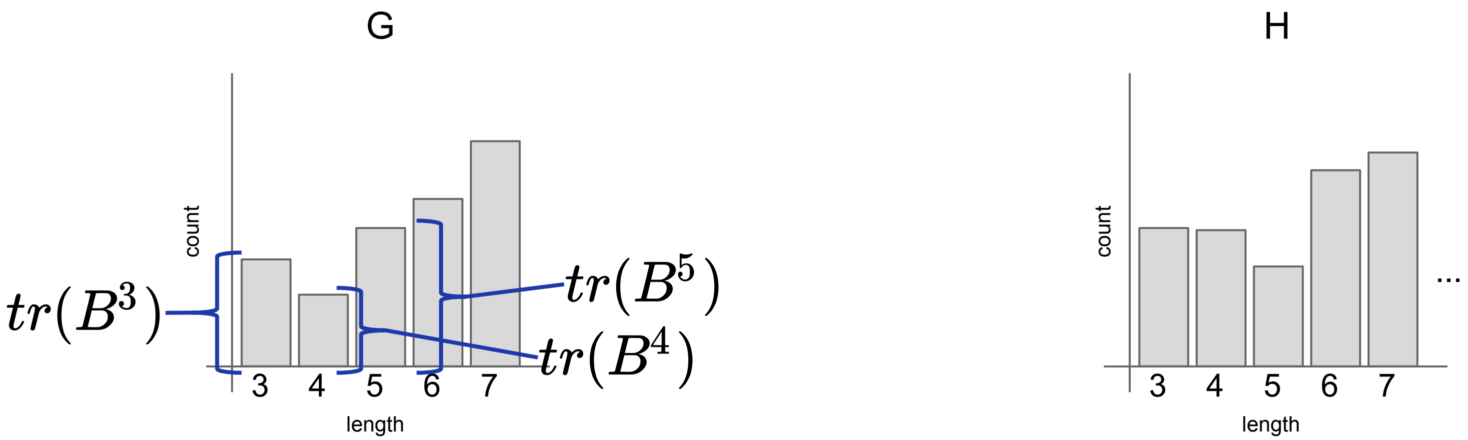




Graph Distance

How to compare these two histograms?

- Observe that the height of each bar is the number of NBCs of a certain length.
- We can compute this using the *nonbacktracking matrix* **B**.
- Thus, the histograms can be generated using only the eigenvalues of **B**.



Graph Distance

Given two graphs G, H *and an integer* r*, write* λ_k, μ_k *for the eigenvalues of their corresponding nonbacktracking matrices,* $k = 1, 2, \ldots, r$*, such that*

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_r| \qquad |\mu_1| \geq |\mu_2| \geq \ldots \geq |\mu_r|$$

Define their distance by

$$d(G, H) = \sqrt{\sum_{k=1}^{r} |\lambda_k - \mu_k|^2}$$

Graph Distance

[A1] **Identity:** $d(G,G) = 0$

[A2] **Symmetry:** $d(G,H) = d(H,G)$

[A3] **Triangle Inequality:** $d(G,H) \leq d(G,F) + d(F,H)$

~~[A4]~~ **Id. of indiscernibles:** $d(G,H) = 0 \Longrightarrow G = H$

[A5] **Divergence[2]:** $d(K_n, \bar{K}_n) \to \infty$ as $n \to \infty$

$$d(G,H) = \sqrt{\sum_{k=1}^{r} |\lambda_k - \mu_k|^2}$$

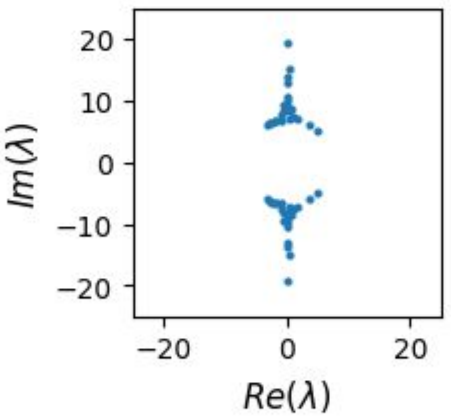
[2] Koutra, Danai, et al. **"DeltaCon: A Principled Massive-Graph Similarity Function."** Proceedings of the 2013 SIAM International Conference on Data Mining, 2013, doi:10.1137/1.9781611972832.18.

In This Talk

1. The Length Spectrum
2. Modifying the Length Spectrum
3. Graph Distance
4. Properties
5. Examples

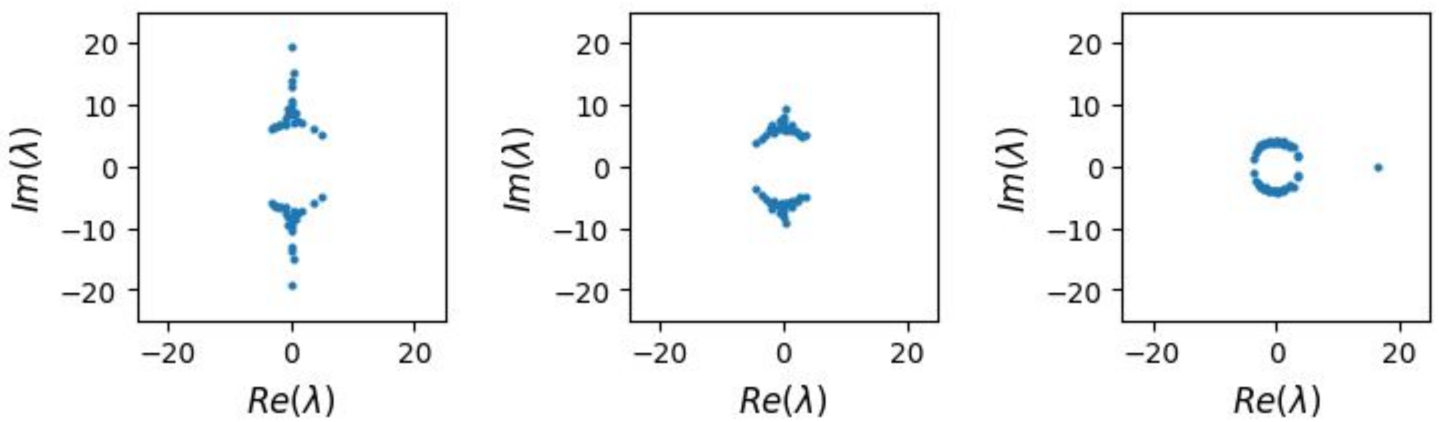
Properties: hubs

Configuration model ($n = 10k, \langle k \rangle = 10, \gamma = 2.1$)



Properties: hubs

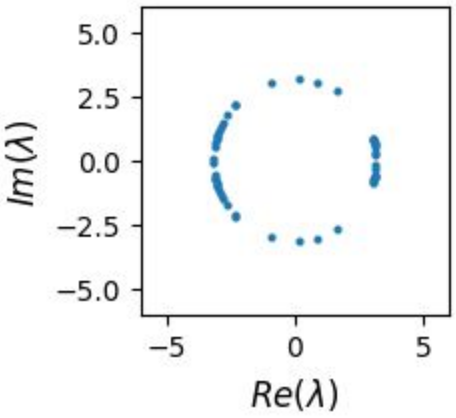
Configuration model (n = 10k,⟨k⟩= 10, γ = 2.1)



fewer hubs

Properties: triangles

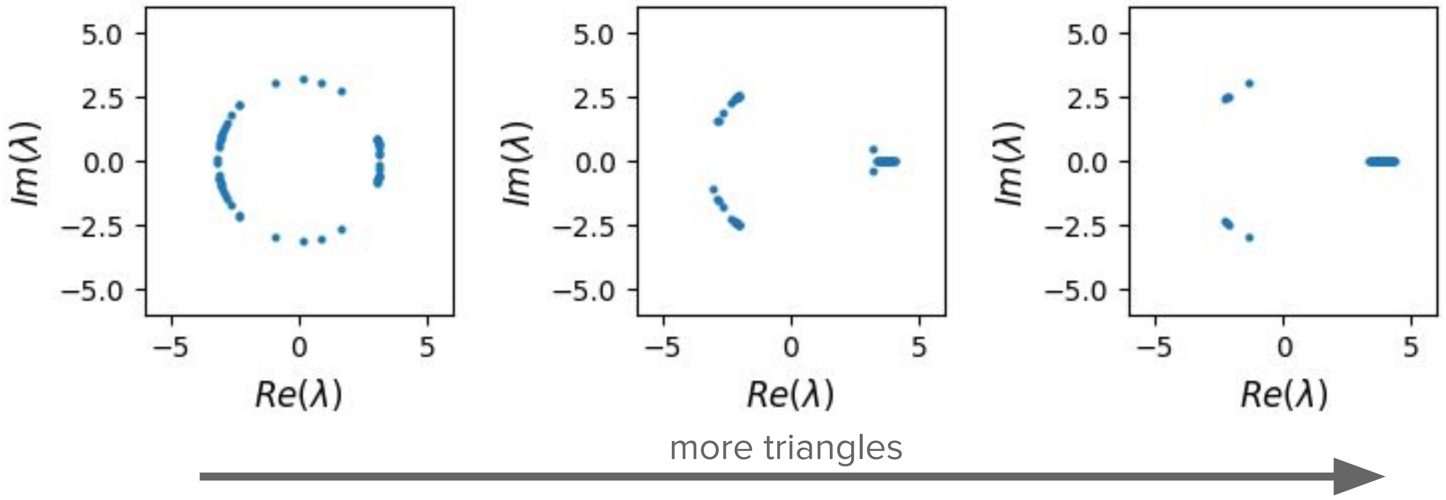
ER graph (n = 10k, p = 0.001)



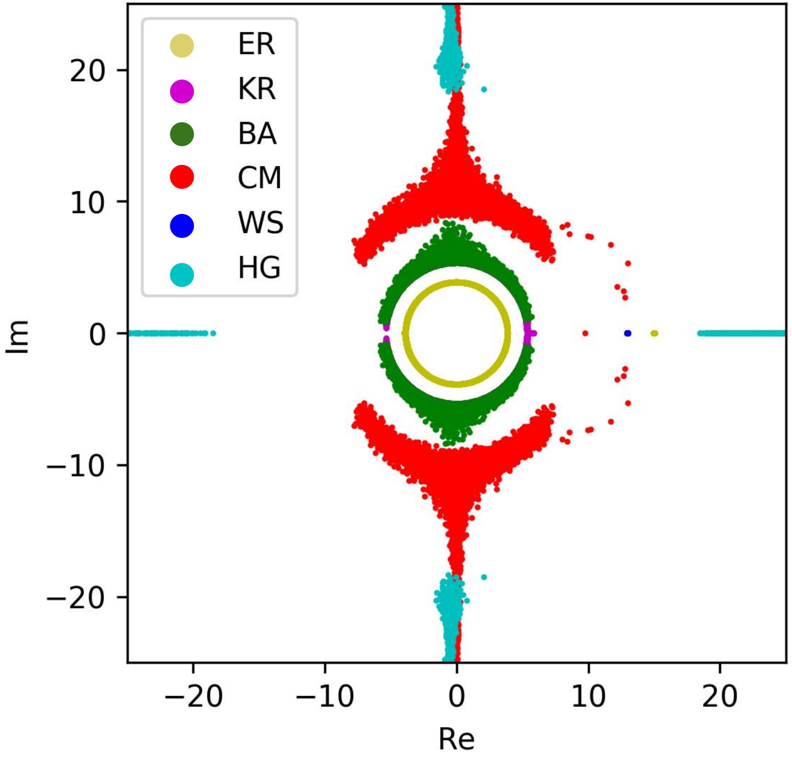
more triangles

Properties: triangles

ER graph (n = 10k, p = 0.001)

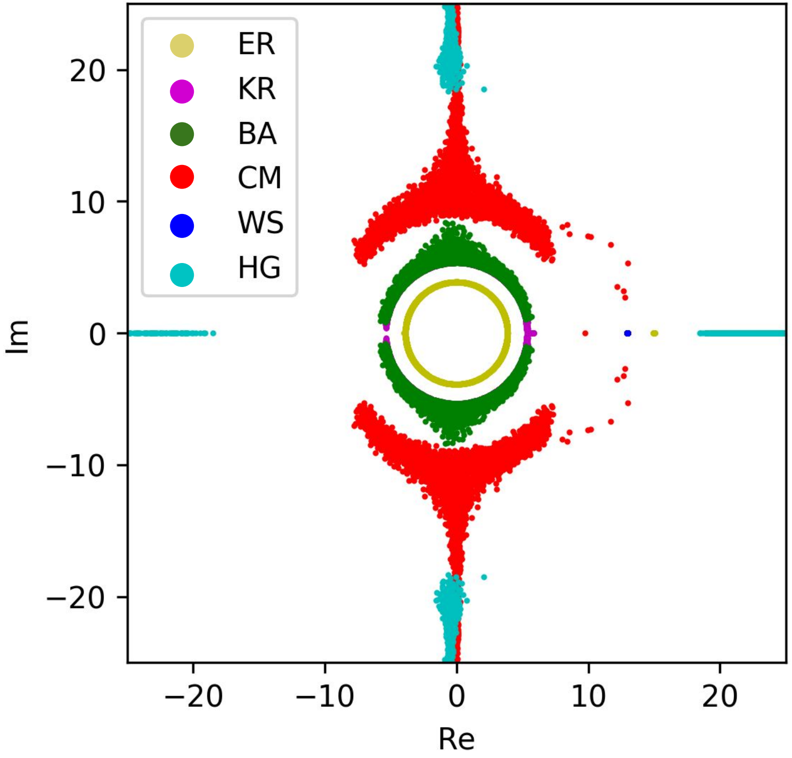


Examples: clustering

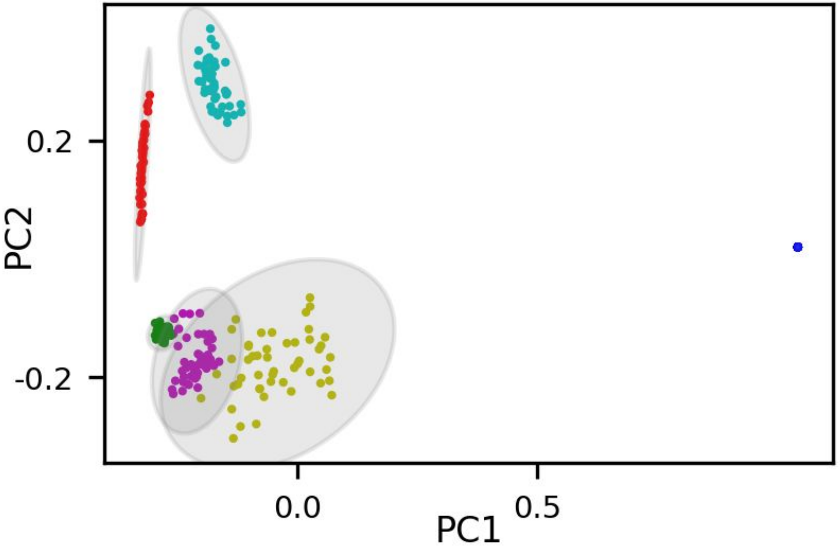


1 dot = 1 eigenvalue

Examples: clustering

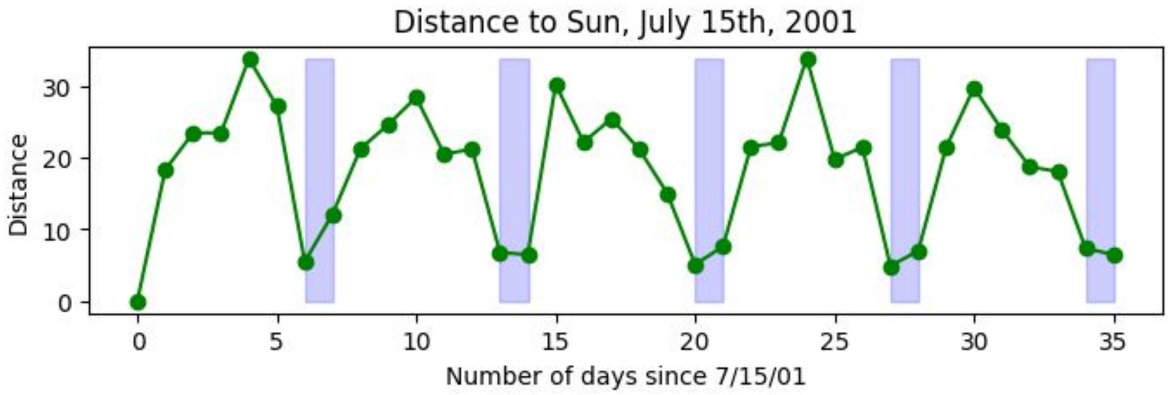


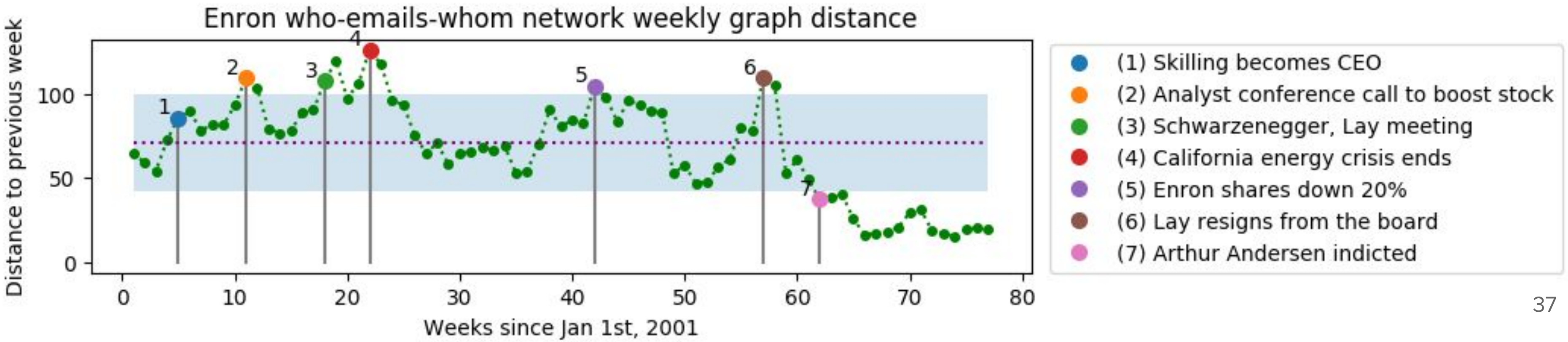
1 dot = 1 eigenvalue



1 dot = 1 graph

Examples: Enron emails



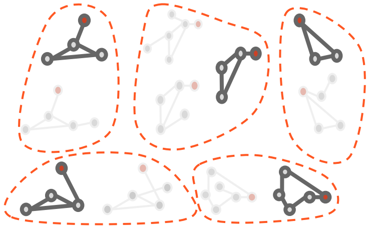


Thank You!

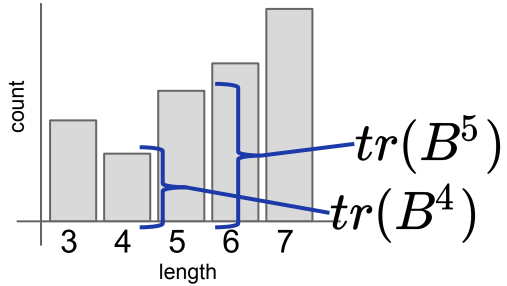
1. The length spectrum characterizes a graph uniquely*.
2. The eigenvalues of **B** account for the image of **L**.
3. We define a pseudometric, which can be interpreted in terms of triangles, degrees.
4. It can cluster random graphs well and detect anomalies.

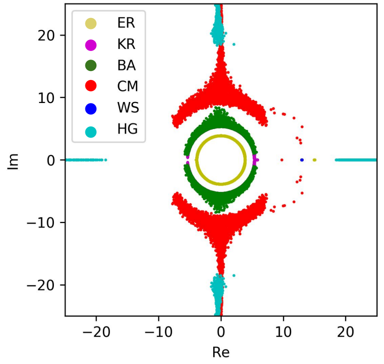
leotrs.com/science.html

$\pi_1(G, v)$



$\mathcal{L} : \pi_1(G, v) \to \mathbb{R}$





Supported by **NSF** CNS-1314603, **NSF** IIS-1741197, and **DTRA** HDTRA1-10-1-0120.