Network Alignment by Discrete Ollivier-Ricci Flow

Chien-Chun Ni [1,3] Yu-Yao Lin [2,3] Jie Gao [3] Xianfeng Gu [3]
Speaker: Kin Sum Liu[3]

September 28th 2018.

[1]Yahoo! Research

[2]Intel Inc.

[3]Stony Brook University

An Example

Consider the same group of people who participate in two social network platforms:

- Private network: identities not revealed, e.g., Facebook .

An Example

Consider the same group of people who participate in two social network platforms:

- Private network: identities not revealed, e.g., Facebook .
- Public network: identities shown in public, e.g., LinkedIn.

An Example

Consider the same group of people who participate in two social network platforms:

- Private network: identities not revealed, e.g., Facebook .
- Public network: identities shown in public, e.g., LinkedIn.

Assume these two networks have **almost the same** topology.

An Example

Consider the same group of people who participate in two social network platforms:

- Private network: identities not revealed, e.g., Facebook .
- Public network: identities shown in public, e.g., LinkedIn.

Assume these two networks have **almost the same** topology.

Goal: align the two networks by vertex correspondences, hence reveals the identifies of the private network.

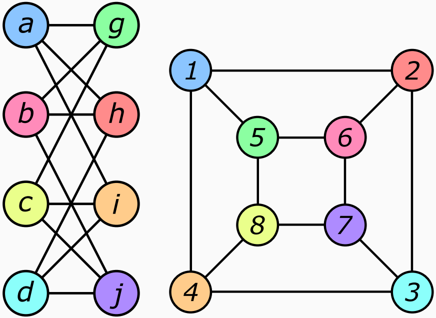
Graph Isomorphism

Given a pair of graphs G_1, G_2, find a one-to-one correspondence of the vertices in G_1 to vertices in G_2 such that (u, v) is an edge in G_1 if and only if their corresponding nodes $f(u), f(v)$ are connected in G_2[1] .

[1] credit: wikipedia

Graph Isomorphism

Given a pair of graphs G_1, G_2, find a one-to-one correspondence of the vertices in G_1 to vertices in G_2 such that (u, v) is an edge in G_1 if and only if their corresponding nodes $f(u), f(v)$ are connected in G_2[1] .



[1] credit: wikipedia

Graph Isomorphism

One of the most fundamental problems in theoretical computer science.

- In NP.

Graph Isomorphism

One of the most fundamental problems in theoretical computer science.

- In NP.
- Nov 2015/Jan 2017, László Babai claimed quasi-polynomial time algorithm: $O(\exp(\log^{O(1)} n))$.

Graph Isomorphism

One of the most fundamental problems in theoretical computer science.

- In NP.
- Nov 2015/Jan 2017, László Babai claimed quasi-polynomial time algorithm: $O(\exp(\log^{O(1)} n))$.
- Many practical algorithms: e.g., NAUTY.

Graph Isomorphism

One of the most fundamental problems in theoretical computer science.

- In NP.
- Nov 2015/Jan 2017, László Babai claimed quasi-polynomial time algorithm: $O(\exp(\log^{O(1)} n))$.
- Many practical algorithms: e.g., NAUTY.
- Subgraph isomorphism is NP-complete.

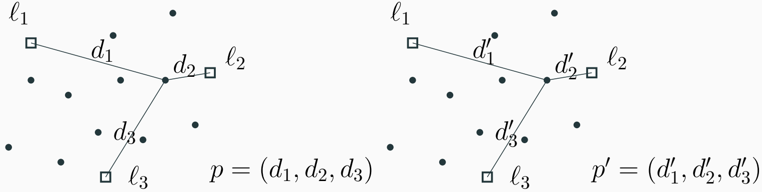
Graph Isomorphism

One of the most fundamental problems in theoretical computer science.

- In NP.
- Nov 2015/Jan 2017, László Babai claimed quasi-polynomial time algorithm: $O(\exp(\log^{O(1)} n))$.
- Many practical algorithms: e.g., NAUTY.
- Subgraph isomorphism is NP-complete.
- Approximate graph isomorphism: find the best correspondence between vertices in G_1 and G_2 s.t. if u, v are connected in G_1 their corresponding nodes are likely connected in G_2.

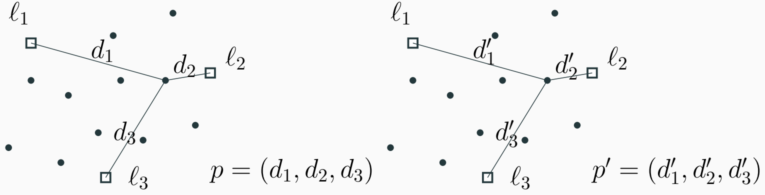
Our Solution: A Geometric Embedding Approach

How to align two sets of points in some embedding plane, assuming that some landmarks ℓ_i are already aligned?



Our Solution: A Geometric Embedding Approach

How to align two sets of points in some embedding plane, assuming that some landmarks ℓ_i are already aligned?



- Any point p can be represented by the barycentric coordinates (d_1, d_2, d_3), d_i is distance to ℓ_i.
- If the barycentric coordinates of p and p' are similar, we match p and p'.

Quantify the ‘Position’ of a Node in a Network

In a social network there are often nodes that can be easily identified as *landmarks*. Define the position of a node wrt landmarks.

Quantify the 'Position' of a Node in a Network

In a social network there are often nodes that can be easily identified as *landmarks*. Define the position of a node wrt landmarks.

Q: What distance to use?

Quantify the 'Position' of a Node in a Network

In a social network there are often nodes that can be easily identified as *landmarks*. Define the position of a node wrt landmarks.

Q: What distance to use?

- Tie strength – Trouble: not easy to measure.

Quantify the 'Position' of a Node in a Network

In a social network there are often nodes that can be easily identified as *landmarks*. Define the position of a node wrt landmarks.

Q: What distance to use?

- Tie strength – Trouble: not easy to measure.
- Count # hops to these landmarks – Trouble: small world property;

Quantify the 'Position' of a Node in a Network

In a social network there are often nodes that can be easily identified as *landmarks*. Define the position of a node wrt landmarks.

Q: What distance to use?

- Tie strength – Trouble: not easy to measure.
- Count # hops to these landmarks – Trouble: small world property;
- Distances from some geometric embedding (spectral embedding, Tutte embedding).

Quantify the 'Position' of a Node in a Network

In a social network there are often nodes that can be easily identified as *landmarks*. Define the position of a node wrt landmarks.

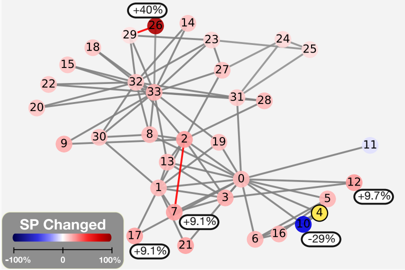
Q: What distance to use?

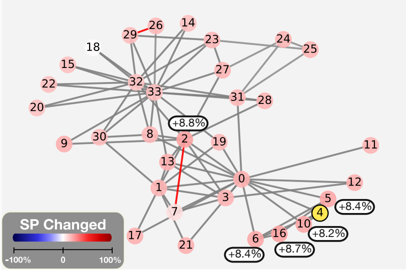
- Tie strength – Trouble: not easy to measure.
- Count # hops to these landmarks – Trouble: small world property;
- Distances from some geometric embedding (spectral embedding, Tutte embedding).

Q: Robust to noises (edge insertion/deletion)?

Robustness: Remove Two Edges

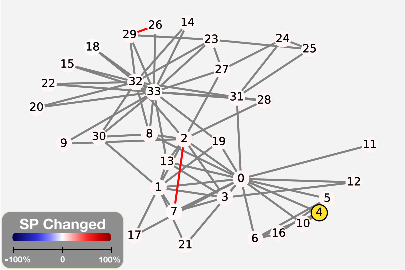
Left: Spectral embedding; Right: Tutte/Spring embedding.

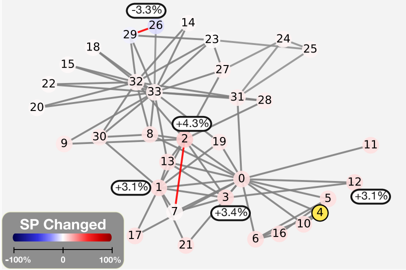




Robustness: Remove Two Edges

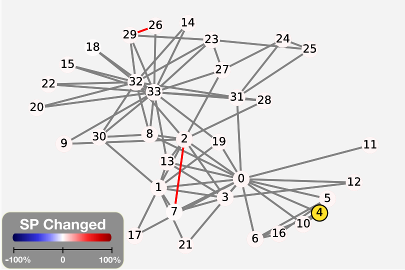
Left: Hop count; Right: our metric.

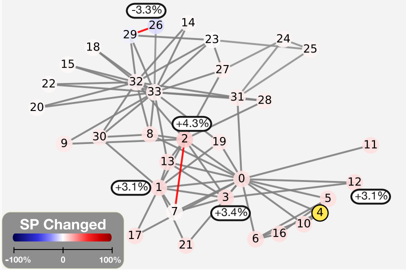




Robustness: Remove Two Edges

Left: Hop count; Right: our metric.



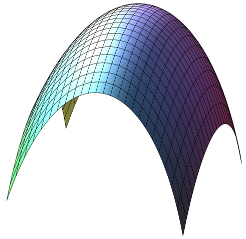
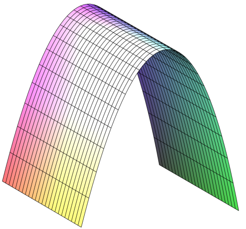
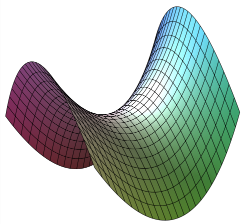


Q: How is our metric defined?

Discrete Ricci Curvature & Ricci Flow

Curvature in Geometry

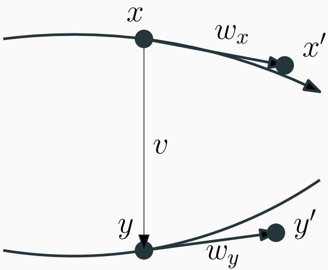
- Sphere: positive curvature;
- Plane: zero curvature;
- Hyperbolic plane: negatie curvature.

Sectional Curvature

Consider a tangent vector $v = xy$. Take another tangent vector w_x and transport it along v to be a tangent vector w_y at y.

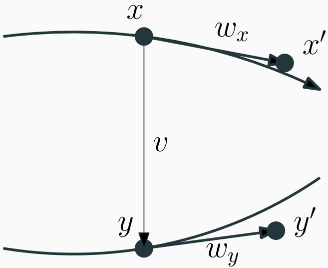
If $|x'y'| < |xy|$ the sectional curvature is positive.



Sectional Curvature

Consider a tangent vector $v = xy$. Take another tangent vector w_x and transport it along v to be a tangent vector w_y at y.

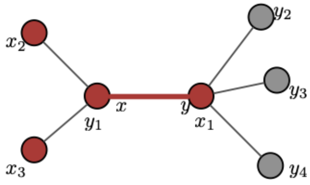
If $|x'y'| < |xy|$ the sectional curvature is positive.



- Ricci Curvature: averaging over all direction w.

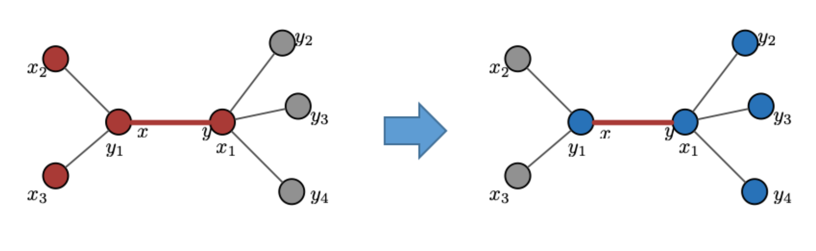
Discrete Ricci Curvature

Take the analog: for an edge xy, consider the distances from x**'s neighbors** to y**'s neighbors** and compare it with the length of xy.



Discrete Ricci Curvature

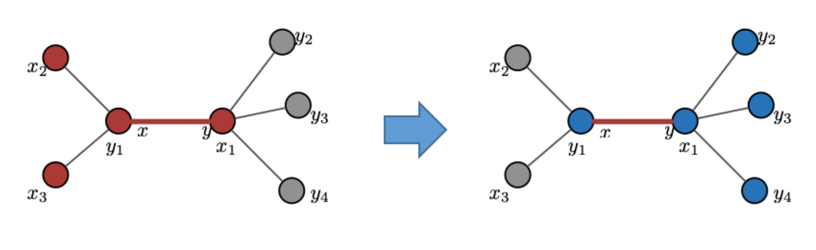
Take the analog: for an edge xy, consider the distances from x**'s neighbors** to y**'s neighbors** and compare it with the length of xy.



- Issue: how to match x's neighbors to y's neighbors?

Discrete Ricci Curvature

Take the analog: for an edge xy, consider the distances from x**'s neighbors** to y**'s neighbors** and compare it with the length of xy.



- Issue: how to match x's neighbors to y's neighbors?
- Assign uniform distribution μ_1, μ_2 on x' and y's neighbors.
- Use optimal transportation distance (earth-mover distance) from μ_1 to μ_2: the matching that minimize the total transport distance.

Definition (Ollivier)
Let (X, d) be a metric space and let m_1, m_2 be two probability measures on X. For any two distinct points $x, y \in X$, the (Ollivier-) Ricci curvature along xy is defined as

$$\kappa(x, y) := 1 - \frac{W_1(m_x, m_y)}{d(x, y)},$$

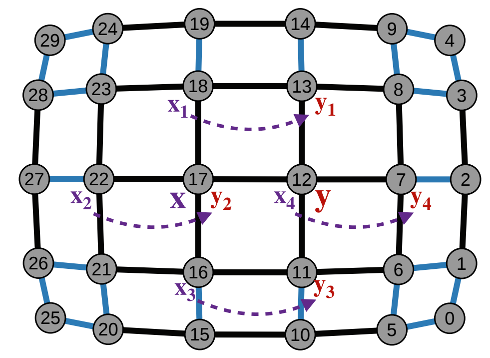
where m_x (m_y) is a probability distribution defined on x (y) and its neighbors, $W_1(\mu_1, \mu_2)$ is the L_1 **optimal transportation distance** between two probability measure μ_1 and μ_2 on X:

$$W_1(\mu_1, \mu_2) := \inf_{\psi \in \Pi(\mu_1, \mu_2)} \int_{(u,v)} d(u, v) d\psi(u, v)$$

For a node w with k neighbors, we define
$m_w(w) = \alpha; m_w(v) = (1 - \alpha)/k$. We choose $\alpha = 1/2$.

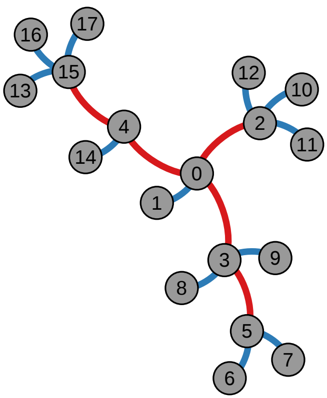
Examples

Zero curvature: 2D grid.



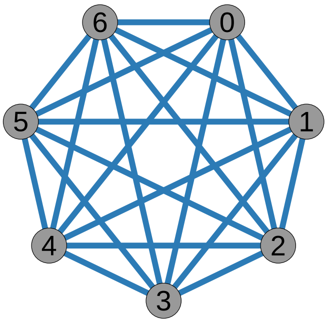
Examples

Negative curvature: tree: $\kappa(x, y) = 1/d_x + 1/d_y - 1$, d_x is degree of x.



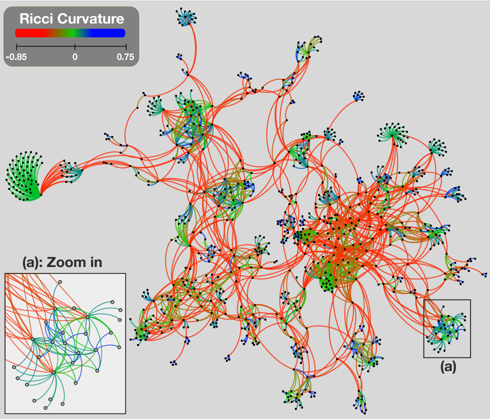
Examples

Positive curvature: complete graph.



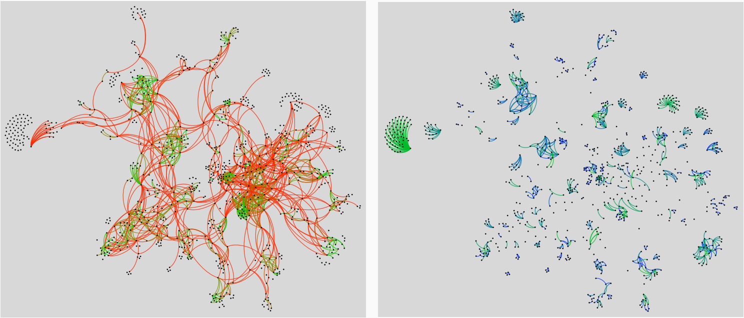
Example: Ricci Curvature

Negatively curved edges are like “backbones”, maintaining the connectivity of clusters, in which edges are mostly positively curved.



Curvature Distribution

Left: Negative curvature edges. Right: Positive curvature edges.[2]



[2]ForceAtlas layout by Gephi

Edge Weights Generated by Ricci flow

Given a graph G in which $d(x, y)$ is the weight of the edge xy and $\kappa(x, y)$ is the discrete Ricci curvature, we run

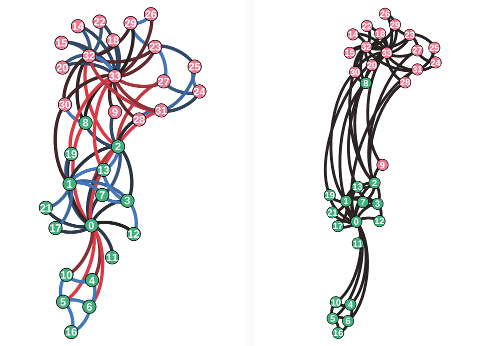
$$d_{i+1}(x, y) = (d_i(x, y) - \varepsilon \cdot \kappa_i(x, y) \cdot d_i(x, y)) \cdot N$$

Until convergence, where N is to rescale to make sure total edge weights remain the same.

At the limit, $W(x, y)/d(x, y)$ is the same for all edges.

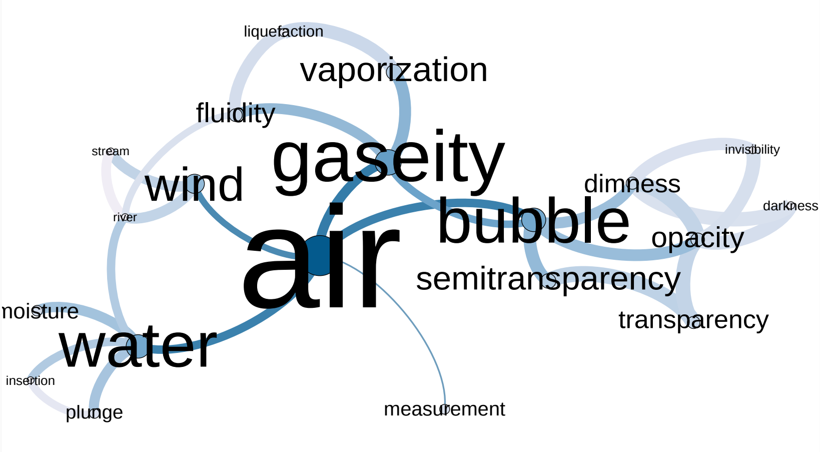
Ricci Flow Metric

Intuition: flatten the network – shrink an edge if it is within a well connected community; stretch an edge if otherwise, s.t., the network curvature is uniform everywhere.[3]



[3] Karate Club by Gephi ForceAtlas layout

As similarity metric: On wordnet, edges between similar words are shrank s.t. similar words are closer with Ricci Flow Metric.



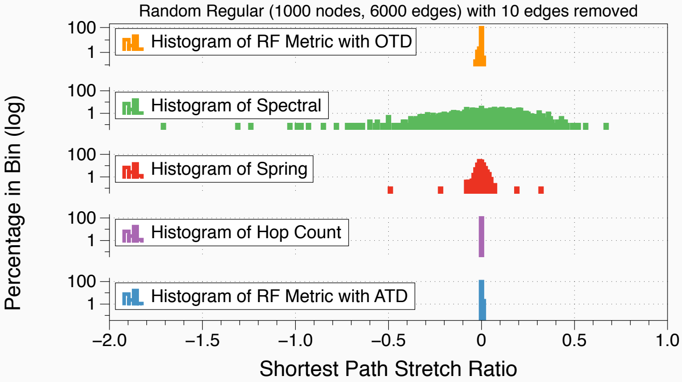
Ricci Flow Metric on Semantic Wordnet

Table 1: Node similarity: Word distance by RF-Metric and hop count

Word	RF-Metric	Hop		Word	RF-Metric	Hop
air	0	0		**heaven**	0	0
gaseity	1.084512	1		hell	0.476738	1
bubble	1.233986	1		pleasure	0.673406	1
water	1.241377	1		pleasurableness	0.786310	1
wind	1.560098	1		hope	0.920200	1
vaporization	1.854184	2		pain	1.104767	2
semitransparency	1.900589	2		cheerfulness	1.253568	2
opacity	1.993095	2		content	1.254039	2
fluidity	2.032685	2		restoration	1.391618	1
transparency	2.077700	3		sweetness	1.432170	2
dimness	2.084738	2		physical pleasure	1.450673	2
moisture	2.204766	2		feeling	1.471766	2

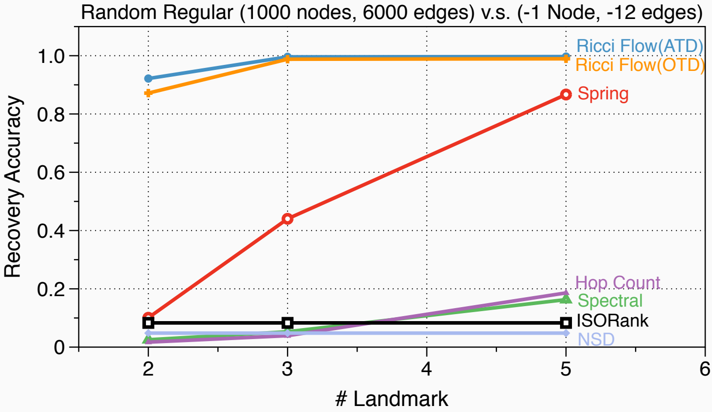
Evaluation on Resilience

Randomly remove 10 edges in a random regular graph.



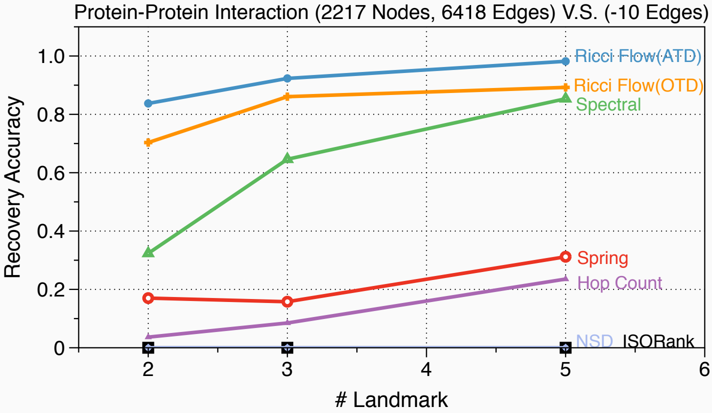
Evaluation on Matching Performance

- Randomly remove one node in a random regular graph w/ degree 12.



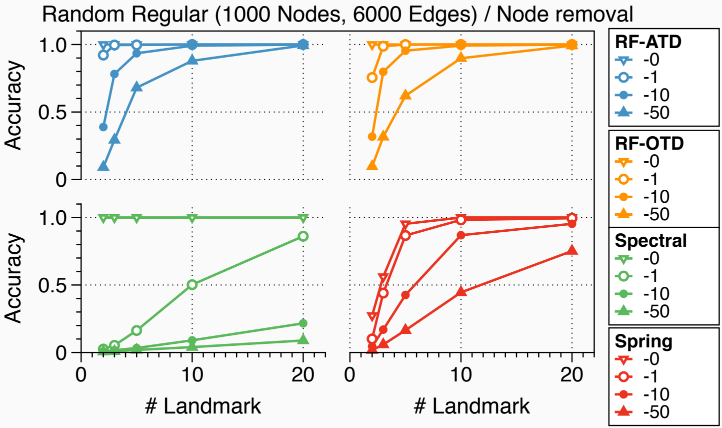
Evaluation on Matching Performance

- Randomly remove 10 edges in a protein protein network.



Evaluation on Matching Performance

- Random Regular Graph - remove Nodes



Conclusions

Ricci flow metric on graph:

- A geometric metric that is robust to noises.
- Only require topology information to compute.
- Highly related to node similarity.

Ricci Curvature & Ricci Flow Source code Available:
https://github.com/saibalmars/GraphRicciCurvature

Contact: Chien-Chun Ni(chien-chun.ni@oath.com)