Detecting the Coarse Geometry of Networks

Melanie Weber, Princeton University
Jürgen Jost, Max Planck Institute for Mathematics in the Sciences
Emil Saucan, ORT Braude & Technion

Download Paper as PDF:

Contact:
mw25@math.princeton.edu
https://web.math.princeton.edu/~mw25

Abstract

Clustering and sampling are key methods for the study of relational data. Learning efficient representations of such data relies on the identification of major geometric and topological features and therefore a characterization of its coarse geometry.

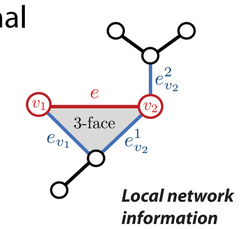
Here, we introduce an efficient sampling method for identifying crucial structural features using a discrete notion of Ricci curvature. The introduced approach gives rise to a complexity reduction tool that allows for reducing large relational structures (e.g., networks) to a concise core structure on which to focus further, computationally expensive analysis and hypothesis testing.

Theory

Curvature-based sampling chooses points whose metric density is inversely proportional to curvature. Existing approaches are based on extrinsic curvature, thus requiring the -- potentially expensive -- construction of isometric embeddings. We take an intrinsic approach, asking whether we can construct a coarse embedding of a weighted graph viewed as metric measure space.

The operator Ric_F is symmetric in the nodes v_1, v_2 of an edge $e(v_1, v_2)$, i.e. defines a kernel k_F, which can be related to a positive notion by setting $k_F^* = e^{-k_F}$. By a direct application of classic results, this gives a map from the ambient space to a real Hilbert space: The existence of a coarse embedding of a graph into a Hilbert space follows from
(i) being a positive kernel ;
(ii) the edges with curvature bounded above generating the coarse structure of the network.

Discrete Ricci Curvature
Simple, scalable network characteristic that evaluates local information. It is defined as a function on the edges, allowing us to introduce the concept of ***edge-based sampling***: We understand relational data as determined not by its members, but by the relations between them, suggesting an edge-based approach to structural analysis.



Local network information

Forman-Ricci Curvature
Networks $G = \{V, E\}$ form regular, 1-dim. cell complexes in which case the following curvature function can be defined:

$$\mathrm{Ric}_F(e) = \omega(v_1) + \omega(v_2) - \sum_{\substack{e_{v_1} \sim e \\ e_{v_2} \sim e}} \left[\omega(v_1) \sqrt{\frac{\omega(e)}{\omega(e_{v_1})}} + \omega(v_2) \sqrt{\frac{\omega(e)}{\omega(e_{v_2})}} \right]$$

It is defined for edges $e(v_1, v_2)$ connecting vertices v_1, v_2 ; ω denotes the weights of edges and vertices.

Methods

Informally, coarse geometry denotes the study of the geometric (or sometimes topological) properties, without considering fine-grained, small-scale features.

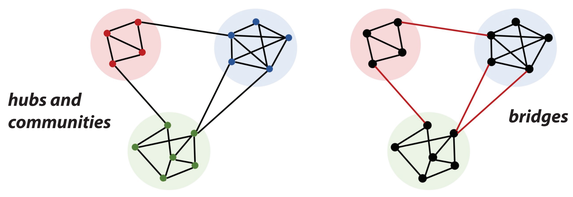
(1) ***Network Backbone***
Captures essential structural properties, such as clusters or communities and the "long-range" connections between distinct network regions.

(2) ***Sampling***
A "good" sample is representative of the crucial features of the full data set, i.e. it resembles its core structure and coarse geometric features.

Network backbone
We call a subnetwork $G' = \{V', E'\}$ $(V' \subseteq V, E' \subseteq E)$ that captures structurally important nodes (hubs) and edges (bridges), a ***backbone*** of $G = \{V, E\}$. It is structure-preserving, i.e., its structural features (e.g., node degree distribution, community structure) are representative of G.

hub: node with high degree and a high betweenness centrality

bridges: edges that govern the mesoscale structure of G (e.g., long-range connections between communities)



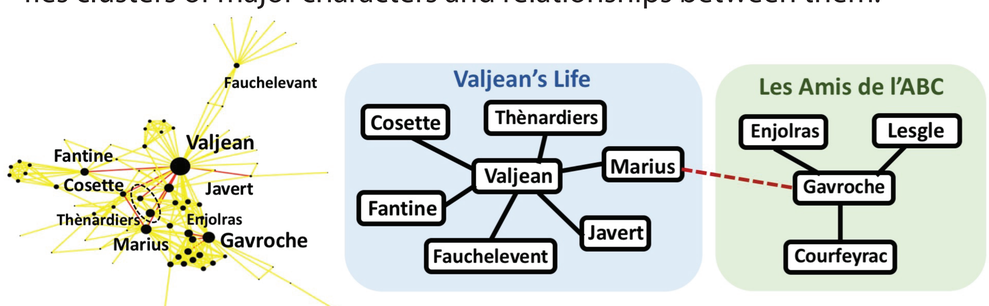
Algorithm
High curvature can be linked to high structural importance: Sample nodes or edges with high curvature to identify the coarse geometry of the network.

Algorithm 1 Curvature-based sampling

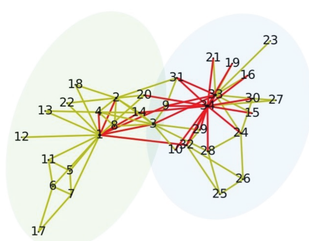
Input: $G = \{V, E, w_v, w_e\}$; (i) k, (ii) r
for $u, v \in V,\ u \sim v$ **do**
  $k_F(u, v) \leftarrow Ric_F(e = (u, v))$
end for
(i) $S \leftarrow \{e = (u, v) \,|\, k(e) < k < 0\}$
(ii) $k_F^* \leftarrow e^{-k_F},\ \hat{k}_F^* \leftarrow reconstruct(kernelPCA(k_F^*), r)$
  $S \leftarrow \{e = (u, v) \,|\, \hat{k}_F^*(e) \neq 0\}$
Output: $G' = \{V|_S, E|_S, w_v|_S, w_e|_S\}$

Experiments

Identify major structural features
Sampling on weighted network of character co-occurrences in Victor Hugo's Les Miserables (backbone threshold t=5%, marked in red) identifies clusters of major characters and relationships between them.



Preserve community structure
Comparison of known community structure of Zachary's karate club with sampled backbone. Keeping the >30% highest absolute curvature edges, preserves the major structural features.


threshold: 30%

Larger-scale Experiments
Identification of backbones in two larger data sets demonstrating potential application as complexity reduction tool.