Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Applying clique-decomposition for computing Gromov hyperbolicity ☆

Nathann Cohen [a], David Coudert [b], Guillaume Ducoffe [b,*], Aurélien Lancin [c]

[a] CNRS, Univ. Paris Sud, Université Paris-Saclay, France
[b] Université Côte d'Azur, Inria, CNRS, I3S, France
[c] LAMIA Laboratory, University of the French West Indies, France

ARTICLE INFO

Article history:
Received 19 October 2016
Received in revised form 29 March 2017
Accepted 2 June 2017
Available online xxxx
Communicated by A. Marchetti-Spaccamela

Keywords:
Gromov hyperbolicity
Graph algorithms
Clique-decomposition
Outerplanar graphs

ABSTRACT

Given a graph, its *hyperbolicity* is a measure of how close its distance distribution is to the one of a tree. This parameter has gained recent attention in the analysis of some graph algorithms and the classification of complex networks. We study on practical improvements for the computation of hyperbolicity in large graphs. Precisely, we investigate on relations between the hyperbolicity of a graph G and the hyperbolicity of its *atoms*, that are the subgraphs output by the clique-decomposition invented by Tarjan [51,65]. We prove that the maximum hyperbolicity taken over the atoms is at most one unit off from the hyperbolicity of G and the bound is sharp. We also give an algorithm to slightly modify the atoms, called the "substitution method", which is at no extra cost than computing the clique-decomposition, and so that the maximum hyperbolicity taken over the resulting graphs is *exactly* the hyperbolicity of the input graph G. An experimental evaluation of our method for computing the hyperbolicity of a given graph from its atoms is provided for collaboration networks and biological networks. Finally, on a more theoretical side, we deduce from our results the first *linear-time* algorithm for computing the hyperbolicity of an outerplanar graph.

1. Introduction

In this paper we aim at improving the computation of hyperbolicity in graphs whose size ranges from thousands to tens of thousands of nodes. To this end, we establish new relations between hyperbolicity and some graph decomposition. Roughly, the hyperbolicity of a metric space is an estimate of how close it is to a metric tree (formal definitions are postponed to the technical sections of this paper). This parameter was first introduced by Gromov in the context of automatic groups [40].[1] Later on, it was applied to the study of more general metric spaces including graphs equipped with their shortest-path metric. Graph hyperbolicity is now part of the parameters in use to classify complex networks [1,3,46]. In particular, it has been proposed in [14] to consider hyperbolicity as a measure of how much a network is "democratic",

☆ This work is partially supported by ANR project Stint under reference ANR-13-BS02-0007 and ANR program "Investments for the Future" under reference ANR-11-LABX-0031-01.

* Corresponding author.
E-mail address: guillaume.ducoffe@inria.fr (G. Ducoffe).

[1] Informally speaking, elements of an automatic group are the vertices of some (Cayley) graph and it can be checked with finite-state automata whether two words represent either a same vertex or two adjacent vertices. Formal definitions and properties of automatic groups are out of our scope. We refer the reader to [35] for more information.

http://dx.doi.org/10.1016/j.tcs.2017.06.001
0304-3975/

with the level of "democracy" in a network growing with the minimum size of a *core* (subset of vertices intersecting a constant fraction of all the shortest-paths). Experiments in [2,14] suggest that the larger the hyperbolicity of a real-world graph, the more democratic it is (see also [23] for formal relationships between Gromov hyperbolicity and the existence of a core). Furthermore, the study of graphs with bounded hyperbolicity has found applications in the design and the analysis of approximation algorithms [22,29] and geometric routing schemes [12], as well as in network security [44] and bioinformatics [20,34], to name a few. As a result, hyperbolicity and its relations to other metric graph parameters has received growing attention over the last decades. The reader may refer to [1,33] for a recent survey.

Computing the hyperbolicity is useful in some of the above applications. For instance, it allows to compute an embedding of the graph into the hyperbolic plane with quasi-optimal distortion of the distances in linear time [67]. The latter is a prerequisite to many algorithms on negatively curved spaces [48]. However, the computational cost of the hyperbolicity has only recently received a bit more attention. So far, the best-known algorithm to compute the hyperbolicity [36], though it runs in polynomial-time, is impractical for large-scale graphs such as the graph of the Autonomous Systems of the Internet, road maps, etc. This comes from its challenging implementation, relying on fast square matrix multiplications, and its time complexity which is supercubic. On a more positive side, there have been recent attempts that are much more efficient than the state-of-the-art algorithm in practice, with running time dominated by the computation of the all-pairs-shortest-paths [15,26]. But on the negative side, a bunch of complexity assumptions – including the Strong Exponential Time Hypothesis [43] – implies that graph hyperbolicity cannot be computed in subquadratic-time, even for sparse graphs [16,27,36]. This motivates us to study which structural properties can help to speed-up the computation of the hyperbolicity in large graphs.

Our approach We study how the information provided by a *clique-decomposition* of a graph into *atoms* (see Section 2.2) can be used to compute the hyperbolicity. As such a decomposition is readily implemented, we strive to split the problem into the computation of the hyperbolicity of smaller graphs similar to the *atoms* themselves.

Note that there are studies supporting the existence of clique-separators in real-life graphs, such as the underlying graphs of social and biological networks [1,8,30,45], that makes our approach practical for large graphs. Furthermore, clique-decomposition has been proved useful to preprocess the graphs in the computation of many optimization problems [65] – including the computation of treelength [32] – related to hyperbolicity. Therefore, at first glance, it is not surprising that clique-decomposition can be applied to preprocess the graphs in the computation of the hyperbolicity. This being said, hyperbolicity is less robust than other metric invariants to graph modifications (e.g. it may increase through an edge-contraction [19]), and so a careful analysis is needed to prove that it is indeed the case.

Our approach in this paper has similarities with the work presented in the PHD Thesis of Soto [59]: he proved for instance that the hyperbolicity of a graph is the maximum hyperbolicity taken over the subgraphs from the split-decomposition [28] or from the modular decomposition [38]. The latter has practical applications in protein-protein interaction networks [37].

Main contributions Our first result on clique-decomposition is in sharp contrast with those obtained in [59] for modular and split decompositions. Indeed, we prove that the hyperbolicity $\delta(G)$ of a graph G *cannot* be deduced directly from the hyperbolicity of its atoms (Section 3). We prove nonetheless that it can be approximated with additive constant 1 by taking the maximum hyperbolicity $\delta^*(G)$ over the atoms (Section 4). This result requires an in-depth analysis of clique-decomposition in order to be proved. Additionally, we present necessary conditions for having that $\delta(G) > \delta^*(G)$.

Based on this characterization, we show in Section 5 how each atom can be transformed (i.e. augmented with few simplicial vertices) in order to compute exactly the hyperbolicity, and provide a complexity analysis of the procedure. Experiments in Section 7 show the benefit of our method in terms of size of the graph, when applied to some real networks from scientific communities and biology.

Finally, we apply clique-decomposition for improving the best-known complexity to compute the hyperbolicity in the class of outerplanar graphs. We detail in Section 6 the first linear-time algorithm for computing the hyperbolicity of these graphs. We find the latter result all the more interesting that under the Strong Exponential Time Hypothesis, the hyperbolicity of sparse graphs cannot be computed in subquadratic-time [16].

Definitions and notations used in this paper are introduced in Section 2.

2. Definitions and notations

We use the graph terminology of [13,31]. All graphs considered in this paper are finite, unweighted and simple. Given $G = (V, E)$, let $n = |V|$, $m = |E|$. The open neighborhood $N_G(S)$ of a set $S \subseteq V$ consists of all vertices in $V \setminus S$ with at least one neighbor in S. The closed neighborhood of S is the set $N_G[S] = S \cup N_G(S)$.

Given two vertices u and v, a *uv-path* of length $l \geq 0$ is a sequence of vertices $(u = v_0 v_1 \dots v_l = v)$, such that $\{v_i, v_{i+1}\}$ is an edge for every i. In particular, a graph G is *connected* if there exists a uv-path for all pairs $u, v \in V$, and in such a case the *distance* $\mathrm{d}_G(u, v)$ is defined as the minimum length of a uv-path in G. Note that it yields a discrete metric space (V, d_G), also known as the shortest-path metric space of G. We also denote by $\mathrm{d}(u, X) = \min_{x \in X} \mathrm{d}(u, x)$ the distance between a vertex u and a set X of vertices.

Our proofs use the notions of subgraphs, *induced* subgraphs, as well as *isometric subgraphs*, the latter denoting a subgraph H of a graph G such that $\mathrm{d}_H(u, v) = \mathrm{d}_G(u, v)$ for any two vertices $u, v \in H$.

Please cite this article in press as: N. Cohen et al., Applying clique-decomposition for computing Gromov hyperbolicity, Theoret. Comput. Sci. (2017), http://dx.doi.org/10.1016/j.tcs.2017.06.001