

RESEARCH ARTICLE

# Closed trail distance in a biconnected graph

Vaclav Snasel, Pavla Drazdilova, Jan Platos\*

Department of Computer Science, Faculty of Electrical Engineering and Computer Science, VŠB - Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

\* [jan.platos@vsb.cz](mailto:jan.platos@vsb.cz)



## Abstract

Graphs describe and represent many complex structures in the field of social networks, biological, chemical, industrial and transport systems, and others. These graphs are not only connected but often also  $k$ -connected (or at least part of them). Different metrics are used to determine the distance between two nodes in the graph. In this article, we propose a novel metric that takes into account the higher degree of connectivity on the part of the graph (for example, biconnected fullerene graphs and fullerooids). Designed metric reflects the cyclical interdependencies among the nodes of the graph. Moreover, a new component model is derived, and the examples of various types of graphs are presented.

## OPEN ACCESS

**Citation:** Snasel V, Drazdilova P, Platos J (2018) Closed trail distance in a biconnected graph. PLoS ONE 13(8): e0202181. <https://doi.org/10.1371/journal.pone.0202181>

**Editor:** Baruch Barzel, Bar-Ilan University, ISRAEL

**Received:** October 17, 2017

**Accepted:** July 30, 2018

**Published:** August 31, 2018

**Copyright:** © 2018 Snasel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data for this study are third-party data from Zachary's Karate Club network. The minimal data set for this study can be accessed via <http://konect.uni-koblenz.de/networks/ucidata-zachary>. The authors accessed the data in the same way as the instructions provided and they did not have any special access privileges that others would not have.

**Funding:** This work was supported by the project SP2018/126 "Parallel processing of Big Data V" received by P. Drazdilova of the Student Grant System, VSB-Technical University of Ostrava.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

More interconnected parts of graphs play an essential role in the social and natural sciences. The formalization of the term "more connected part" can be defined in many ways. In this article, we focus on generalizing biconnected components of a graph and we define a novel metric that considers higher degree of connectivity on the part of the graph. Biconnected components of the graph do not allow good scalability, and their definition is complicated for weighted graphs. Our approach is based on the cycle length limit in the definition of biconnected components. The first work devoted to the study of cycles of limited length is [1, 2].

Topological data analysis often uses tools developed in algebraic topology [3–5]. Cycles play a significant role in algebraic topology. For example, Poincaré's theorem of duality [6] shows that the homology group  $H_n(M, Z_2)$  is a space with the inner product above the field  $Z_2$  where the inner product is defined as an intersection index. In the case that  $M$  is a continuous manifold, then any homology class  $x \in H_1(M, Z_2)$  can be represented by a closed curve  $\gamma \subset M$ . In this case, the intersection index  $x \cdot x$  becomes zero when and only when a small surrounding of the curve  $\gamma$  is orientable.

Cycles with limited length play an essential role in the application of algebraic topology [3, 4]. When calculating the topological properties of data, it is necessary to look for cycles of limited length [7, 8]. Cyclical structures are also very often found in materials research. For example, fullerenes form long cycles [9], in which topological properties play an important role. Complex and social networks are another field in which cyclic structures appear [10].

The partitioning of large complex networks is a challenging task. Such networks are used as a representation of proteins, chemical compounds, co-author networks, social networks, etc.

The classical partitioning methods have problems with densely connected subgraphs that cannot be partitioned easily. A list of the largest biconnected component in the selected network was published by Leskovec [11].

In the case of protein interaction networks in computational biology, the authors in [12] found vertices that are articulation points (determined by the computing of biconnected components), but they have a low degree and, therefore, they are unlikely to be essential to the network. In [13] the authors found the biconnected components that enabled further analysis.

Molecular topology [14] is another area where topological and metric distances are used in graphs representing molecules. The fullerenes are cage-like, hollow molecules of pseudo-spherical symmetry consisting of pentagons and hexagons only, resulting in a trivalent polyhedron with precisely three edges (bonds) joining every vertex occupied by carbon [9]. In graph theoretical terms, fullerenes belong to the class of cubic, planar, three-connected, and simple graphs, see Fig 1. The authors of [15] give an overview of some graph invariants that can possibly correlate with the stability of a fullerene molecule.

Fullerenes have been the subject of intense research for their unique physical, chemical, and biological properties and for their technological applications, especially in materials science, electronics, nanotechnology, and medicine [16–19].

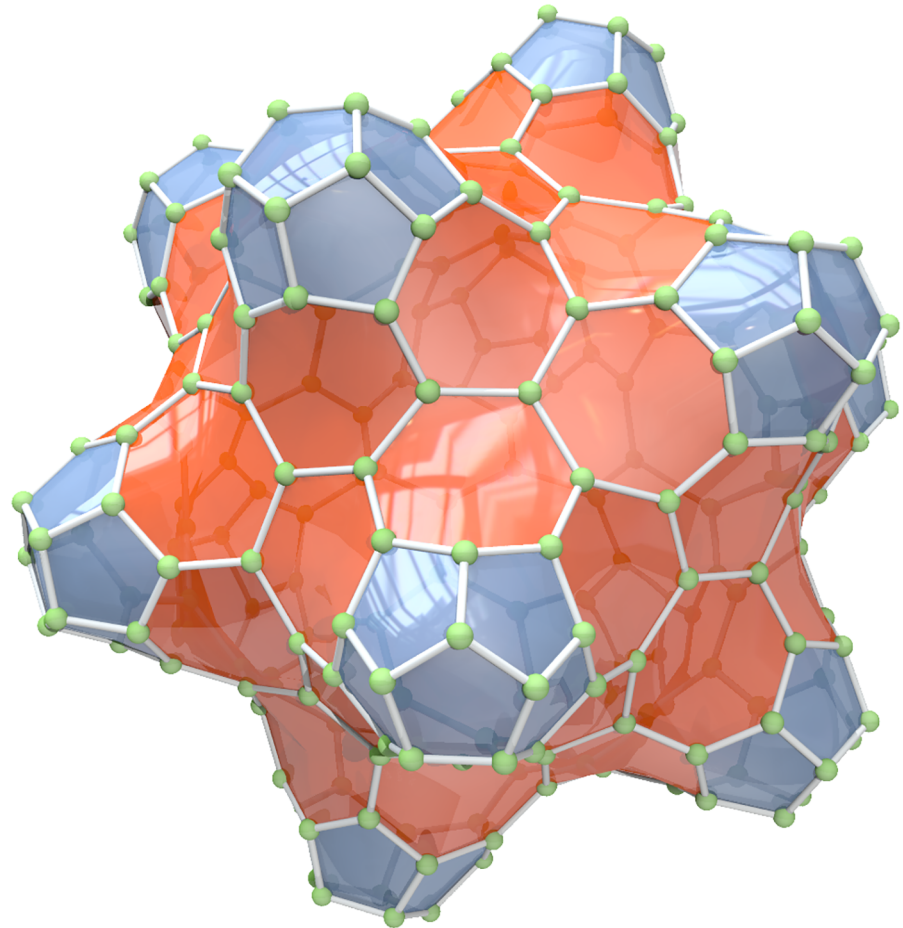
The measuring of the distances between two nodes in a graph is a difficult task. The standard measure for this distance is the shortest path between two nodes in a graph [20, 21]. Another way is the expected lengths of random walks on the graph, which can be used to derive the commute time distance [22]. The authors of [23] examine generalized distances on graphs that interpolate, depending on a defined parameter, between the shortest path distance and the commute time or resistance distance. Variants of node distances are described in detail in [24–27].

In addition to the node distance measure, the quality of the components (partitions) is also measured by means of several approaches. In 2014, a metric that measured the quality of communities according to the number of 3-cycles split across the communities was published [28]. The idea is based on the four types of directed triangles that contain cycles. These triangles are used to identify communities in directed networks. In [29], a measure that integrates both the concept of closed walks and clustering coefficients to replace the edge betweenness in the divisive hierarchical clustering algorithm (the Girvan and Newman method) was published. Levoranto et al. [30] used the strongly  $p$ -connected components for community detection in oriented networks. Community detection in undirected graphs is different. One of the major and well-known approaches uses the union of cliques to define a community [31]. Edaschery et al. [32] defined *distance- $k$  clique* of a graph  $G = (V, E)$  as a subgraph of  $G$  with a diameter  $k$ . The authors use these *distance- $k$  cliques* for a clustering.

Our approach defines a new type of metric in a graph based on “cyclical distances”. This metric is based on the definition of a biconnected component. The distance between two vertices in the graph is defined as the length of the shortest closed trail that contains these two vertices. The distance defined in this way allows straightforward generalization for weighted graphs and also allows scalability.

In this article, we define a new measure on an undirected connected graph without bridges for the measurement of distances using cyclic subgraphs. This measure satisfies the metric properties. Our innovative measure may be used to define a new type of components that highlight the locally connected subgraphs. Moreover, these components are not based on the biconnectivity property and, therefore, are able to partition densely connected biconnected components easily.

We will first introduce the terminology and the notation which we use in the article. In the next section, we define the new distance in biconnected undirected graphs, and we describe



**Fig 1. Graph of a fulleroid  $C_{260} - I[5, 7]$**  (The image was provided by Peter Schwerdtfeger, Lukas N. Wirz and James Avery under a CC BY license).

<https://doi.org/10.1371/journal.pone.0202181.g001>

some properties of this distance. In conclusion, we discuss the advantages and limitations of the defined distance.

### Terminology and notation

In this section, knowledge of graph theory will be required. The definitions of the following terms were taken from [33]:

A *loop* is an edge (directed or undirected) that joins a single endpoint to itself. A *walk* on a graph is an alternating series of vertices and edges

$$W(v_{(0)}, v_{(k)}) = v_{(0)}e_{(1)}v_{(1)}e_{(2)} \dots v_{(k-1)}e_{(k)}v_{(k)}$$

such that for  $j = 1, \dots, k$  the vertices  $v_{(j-1)}$  and  $v_{(j)}$  are the end points of the edge  $e_{(j)}$ . A *closed walk* is a walk where the initial vertex is also the final vertex. The *length of a walk* is the number of edges in this walk. We will denote the length of a walk  $W(u, v)$  as  $|W(u, v)|$ . A *trail* is a walk in which no edge occurs more than once. A *closed trail (circuit)* is a closed walk with no repeating edges. We will denote a closed trail which contains the vertices  $u, v$  as

$$CT(u, v) = ue_{(1)}v_{(1)}e_{(2)} \dots v_{(k)} \dots e_{(k)}u.$$

A *path* is a walk in which no edge or internal vertex occurs more than once (a trail in which all the internal vertices are distinct). We will denote a path with an initial vertex  $u$  and a final vertex  $v$  as  $P(u, v)$ . A *cycle* is a closed path with a length at least one. We will denote the closed path containing  $u, v$  as  $CP(u, v)$ . A *clique* is a subgraph where each node is adjacent to every other node. A *planar graph* is a graph that can be drawn on a sphere or a plane with no edge crossings.

A *connected* graph is a graph such that between every pair of vertices there exists a walk. A graph is called *k-connected* if the removal of fewer than  $k$  vertices leaves neither a non-connected graph nor a trivial one. A *component* of a graph is a maximal connected subgraph. An edge  $e$  is a *bridge* of the connected graph  $G$  if  $\{e\}$  is a disconnecting edge-set of  $G$ . An *articulation* is a vertex of a graph whose removal increases the number of components. A *biconnected* graph is a connected and “nonseparable” graph, meaning that if any vertex were to be removed, the graph would remain connected. Therefore a biconnected graph has no articulation vertices.

The property of being *2-connected* is equivalent to biconnectivity, with the caveat that the complete graph of two vertices is sometimes regarded as biconnected but not 2-connected. This property is especially useful in maintaining a graph with a two-fold redundancy, to prevent disconnection upon the removal of a single edge (or connection).

A *biconnected component* (or 2-connected component) is a maximal biconnected subgraph.

### Equivalent characterizations of biconnectivity

Let  $G = (V, E)$  be a simple undirected graph (loop-less, no multiple edges) that contains at least three points. Each of the following statements is equivalent to that  $G$  is biconnected

1. for every  $v_1, v_2 \in V$  there is a circuit of  $G$  containing  $v_1$  and  $v_2$ ,
2.  $E \neq \emptyset$  and for every  $v \in V$  and  $e \in E$  there is a cycle of  $G$  containing  $v$  and  $e$ ,
3.  $G$  has no isolated vertices and for every  $e_1, e_2 \in E$  there is a cycle containing  $e_1$  and  $e_2$ ,
4. for every  $v_1, v_2, v_3 \in V$  there is a path from  $v_1$  to  $v_2$  containing  $v_3$ ,
5. for every  $v_1, v_2, v_3 \in V$  there is a path from  $v_1$  to  $v_2$  not containing  $v_3$ ,
6.  $E \neq \emptyset$  and for every  $v_1, v_2 \in V$  and  $e \in E$  there is a path from  $v_1$  to  $v_2$  containing  $e$ .

The shortest path metric ( $d_{sp}$ ) [25] is the one most commonly used to determine the distance between vertices of the graph. It is a metric on the vertex-set  $V$  of a connected graph  $G = (V, E)$ , defined,  $\forall u, v \in V$ , as the length of the shortest path ( $P(u, v)$ ) in  $G$ . This metric does not affect greater coherence between vertices in the biconnected graph. Our goals in this article are to define a new metric on the undirected biconnected graph for the measurement of distances using cyclic subgraphs and to use higher connectivity among the vertices in the biconnected graph.

### Closed trail distance in a biconnected graph without loops

In this section, a metric between the vertices in a biconnected graph without loops via a closed trail (circuit) will be defined.

### Closed trail distance in a undirected graph

**Definition 1.** A graph is a *k-closed trail connected graph (k-CT)* if every two vertices lie on the closed trail (circuit) with a length  $\leq k$ . A *k-CT component* of the graph is a maximal *k-CT* subgraph.

**Definition 2.** A  $\infty$ -CT graph is a graph where every two vertices lie on a closed trail of any length.

**Definition 3.** Let  $G = (V, E)$  be a graph. Let  $d_{ct} : V \times V \rightarrow \mathbb{R}_0^+$  be defined by the equation

$$d_{ct}(u, v) = \min_{CT(u,v) \subseteq G} |CT(u, v)|,$$

where  $CT(u, v)$  is a closed trail that contains the vertices  $u, v$ . Then the function  $d_{ct}$  is called the closed trail distance (CT-distance).

**Theorem 1.** The CT-distance is a metric on the set  $V$ .

*Proof.* We verify the properties of the metrics.

From Definition 3 it follows that  $d_{ct}(v, v) = \min_{CT(v,v) \subseteq G} |CT(v, v)|$ . The length of the shortest closed trail that started in a vertex  $v$  and ended in the same vertex is equal to 0. Then  $\forall v \in V$   $d_{ct}(v, v) = 0$ . And from the definition it follows that  $\forall u, v \in V$   $d_{ct}(u, v) \geq 0$  because the distance between  $u$  and  $v$  is the number of edges in the closed trail containing  $u$  and  $v$ .

Symmetry of the distance is obvious:  $\forall u, v \in V$   $d_{ct}(u, v) = \min_{CT(u,v) \subseteq G} |CT(u, v)| = d_{ct}(v, u)$  because  $G = (V, E)$  is a undirected graph.

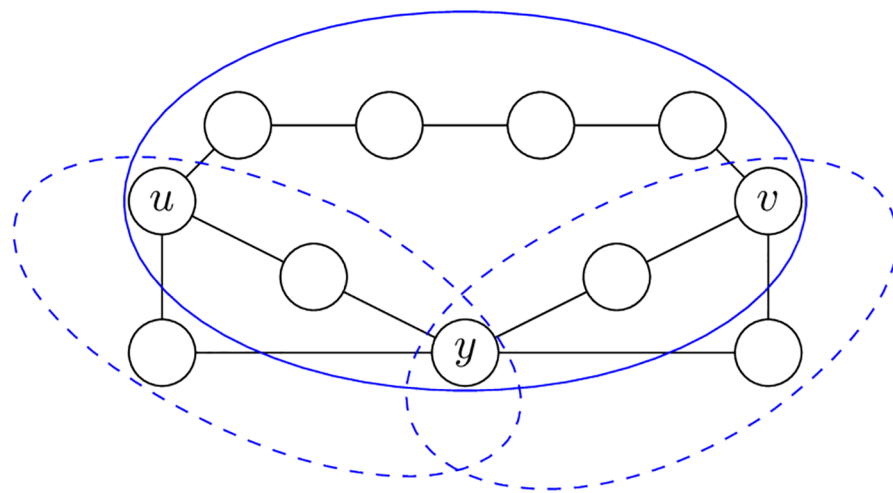
The triangle inequality:  $\forall u, v, z \in V$   $d_{ct}(u, v) + d_{ct}(v, z) \geq d_{ct}(u, z)$  is proved in a constructive way. Let  $CT(u, v)$  and  $CT(v, z)$  exist and let  $|CT(u, v)| = d_{ct}(u, v)$  and  $|CT(v, z)| = d_{ct}(v, z)$ . In the biconnected graph  $\forall u, v \in V$  a closed path ( $CP(u, v)$ ) exists. Then we have two possibilities for the closed trail which satisfy  $|CT(u, z)| = d_{ct}(u, z)$ . First:  $d_{ct}(u, v) + d_{ct}(v, z) = |CP(u, v)| + |CP(v, z)| \geq |CP(u, z)| = d_{ct}(u, z)$  which does not violate the triangle inequality. Second: if  $|CP(u, v)| + |CP(v, z)| < |CP(u, z)|$  then we can join  $CP(u, v)$  and  $CP(v, z)$  in the vertex  $v$  to one a closed trail  $CT(u, z)$  and  $d_{ct}(u, v) + d_{ct}(v, z) = |CP(u, v)| + |CP(v, z)| = |CT(u, z)| \geq d_{ct}(u, z)$  which does not violate the triangle inequality either.

**Example 1.** Let graph  $G$  be defined as shown in Fig 2.

Let  $d_{sc}(u, v)$  be the length of the shortest closed cycle in the biconnected graph that goes through the nodes  $u$  and  $v$ . The  $d_{sc}(u, v)$  is not a metric, because it does not satisfy the triangle inequality:

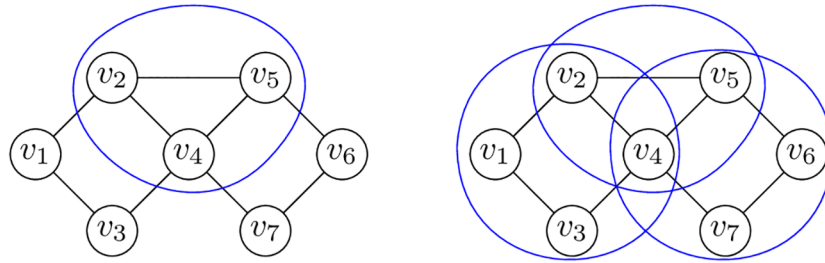
$$d_{sc}(u, y) + d_{sc}(y, v) < d_{sc}(u, v).$$

For the graph in Fig 2 it is true that  $d_{sc}(u, y) = 4$ ,  $d_{sc}(y, v) = 4$  and  $d_{sc}(u, v) = 9$ .



**Fig 2.** Example of biconnected graph where  $d_{sc}(u, y) + d_{sc}(y, v) < d_{sc}(u, v)$ .

<https://doi.org/10.1371/journal.pone.0202181.g002>



**Fig 3. 3-CT and 4-CT components of the graph.**

<https://doi.org/10.1371/journal.pone.0202181.g003>

The violation of the triangle inequality by the cycles ( $d_{sc}$ ) may be solved using the closed trails ( $d_{ct}$ ).

In a similar way to that in which we defined the  $k$ -CT component, a  $k$ -SC component may be defined using cycles of a length up to  $k$ . The graph is a  $k$ -shortest cycle connected graph ( $k$ -SC) if every two vertices lie on a cycle of the length  $\leq k$ . The  $k$ -SC component of the graph is a maximal  $k$ -SC subgraph.

**Example 2.** Figs 3 and 4 demonstrate maximal and different  $k$ -CT components in the undirected biconnected graph without loops.

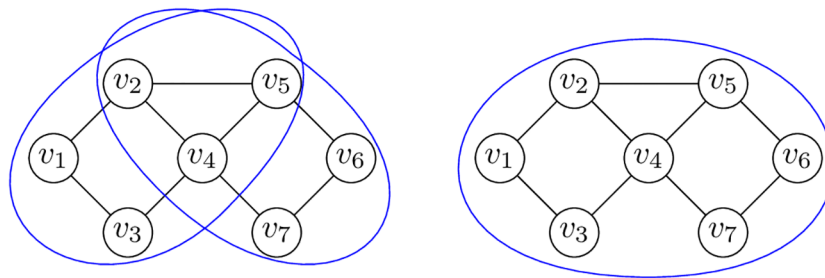
**Lemma 1.** Every 3-CT component is a clique.

*Proof.* Let  $Co3$  be a 3-CT component. According to Definition 3:  $\forall u, v \in Co3 |CT(u, v)| \leq 3$ . It is obvious that this  $CT(u, v)$  contains only three vertices and three edges and there exists the edge  $(u, v)$ . Therefore, all the vertices in  $Co3$  are adjacent and create a clique.

**Example 3.** Fig 5 shows the differences between the closed trail distance and the shortest path distance. We chose a subgraph of the fullerene graph for the comparison between distances.

**Lemma 2.** Every closed trail with a length 4 or 5 is a cycle with the same length.

*Proof.* Proof by contradiction. Let there exist a closed trail with a length of 4 ( $CT4$ ) or a closed trail with a length of 5 ( $CT5$ ) which is not a cycle. Let  $CP3$  be a cycle (closed path) with a length of 3. It is obvious that it is not possible to create  $CT4$  ( $CT5$ ) on the set with one, two or three vertices because  $CT4$  ( $CT5$ ) contains four (five) different edges. We have at least four vertices and  $CT4$  ( $CT5$ ) has to contain the vertex  $u$  which appears at least three times in the sequence of the trail (the initial, final and inner vertex of the closed trail). Vertex  $u$  is initial and inner vertex of the trail if  $CT4$  ( $CT5$ ) contain  $CP3$ . When we add one (two) edge(s) to the  $CP3$  we have a trail with a length of 4 (5) but it not possible to create a closed trail because the edge between the last—but—one vertex and initial (final) vertex is already on the trail (Fig 6). The only closed trails with a length of 4 (5) are therefore cycles with a length of 4 (5).



**Fig 4. 5-CT and 7-CT components of the graph.**

<https://doi.org/10.1371/journal.pone.0202181.g004>

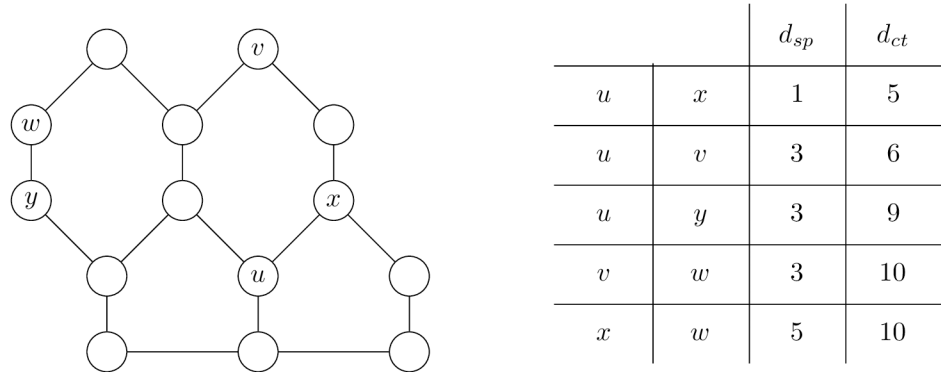


Fig 5. Selected vertices with the shortest path distance and the closed trail distance.

<https://doi.org/10.1371/journal.pone.0202181.g005>

The demonstration of the impossibility of creating a CT4 or CT5 which is not a cycle and the vertex  $u$  is the initial, inner but not final vertex of the closed trail with a length 4 (5):

$$CT4 = ue_1xe_2ye_3ue_4z, \quad CT5 = ue_1xe_2ye_3ue_4ze_3x(y).$$

The first closed trail that is not a cycle is CT6 (Fig 6, third picture).

**Theorem 2.** *The 3,4,5-CT components are biconnected. They are the same as the 3,4,5-SC (shortest cycle) components.*

*Proof.* It is obvious for 3-CT components. The proof follows from Lemma 1. 3-CT components are cliques and therefore 3-CT components are biconnected.

We suppose that for 4(5)-CT components it is not true. There exists a connected 4(5)-CT component which is not biconnected. Then the component contains the articulation  $x$ —see Fig 7. Then there exist vertices  $u, v$ , where  $d_{ct}(u, v) = 4(5)$ . From Definition 3 it follows that there exist  $CT(u, v)$  and  $|CT(u, v)| = 4(5)$  ( $CT(u, v)$  is blue in Fig 7). From Lemma 2 it follows that all closed trails with a length of 4(5) are cycles and then the vertex  $x$  is not an articulation. This is a contradiction.

**Example 4.** *The  $k$ -CT component is not sometime a biconnected subgraph for  $k \geq 6$ . The 6-CT component can be the smallest closed trail-connected component which is not biconnected (Fig 6).*

**Lemma 3.**  $d_{ct}(u, v)$  is a metric in any connected graph without bridges and defines the distances between two nodes  $u$  and  $v$ .

*Proof.* The graph is connected without bridges. There is a path between any pair of nodes  $u$  and  $v$ . Let  $P_1(u, v) = ue_{(1)} \dots e_{(i)} xe_{(i+1)} \dots e_{(k)}v$  be a path between any pair of nodes  $u$  and  $v$ . Because the graph has no bridges (it is not 1-edge connected), a path  $P_2(u, v)$  that has no common edge with  $P_1(u, v)$  exists (see Fig 8). The joining of  $P_1(u, v)$  and  $P_2(u, v)$  creates a closed trail on which the nodes  $u$  and  $v$  lie ( $CT(u, v)$ ). Therefore, a closed trail between any two nodes

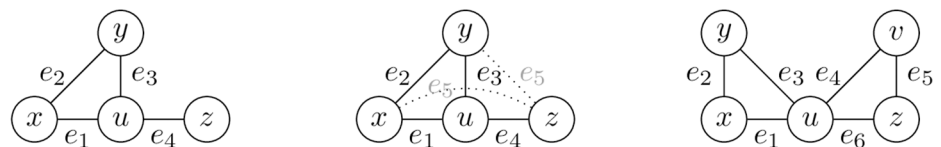
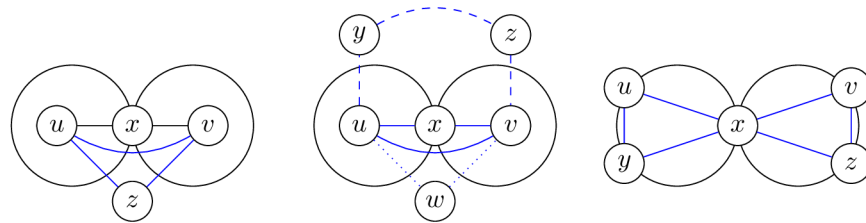


Fig 6.  $u$  is an initial and inner vertex in the trails T4 and T5. For the last closed trail CT6 is the vertex  $u$  initial, inner and final.

<https://doi.org/10.1371/journal.pone.0202181.g006>



**Fig 7.  $k$ -CT components.** The first picture:  $x \notin CT(u, v)$  and the component is biconnected. The second picture:  $x \in CT(u, v)$  where  $|CT(u, v)| = 3(4, 5)$  and the component is biconnected. The third picture:  $x \in CT(u, v)$  where this CT is the shortest closed trail which is not a cycle and the component is not biconnected.

<https://doi.org/10.1371/journal.pone.0202181.g007>

$u$  and  $v$  with a length of  $d_{ct}(u, v)$  may be created in any connected graph without bridges. The connected graph without bridges is a  $\infty$ -CT component.

**Corollary 1.** Graph and component hierarchy is defined as follows:

$$\text{biconnected graph} \subset \text{connected graph without bridges} = \infty\text{-CT component}.$$

**Corollary 2.** The ordering of type of components according to its cardinality is defined as follows:

$$3\text{-CT} \subset 4\text{-CT} \subset 5\text{-CT} \subset \dots \subset \infty\text{-CT}.$$

**Example 5.** Fig 9 shows the biggest 4-CT component (blue vertices) in the biconnected subgraph (in the blue ellipse) of the Zachary’s karate club graph (Fig 9). This 4-CT component is biconnected. The Table 1 contains numbers of  $k$ -CT components which are maximal for selected  $k$ . The Fig 10 shows the sample of  $k$ -CT components on the graph of fulleroid  $C_{260} - I[5, 7]$ .

### Closed trail distance in a weighted undirected graph

The definition of the CT-distance may be extended for the weighted graph  $G = (V, E, w)$  where  $w$  is a mapping  $w: E \rightarrow R^+$ .

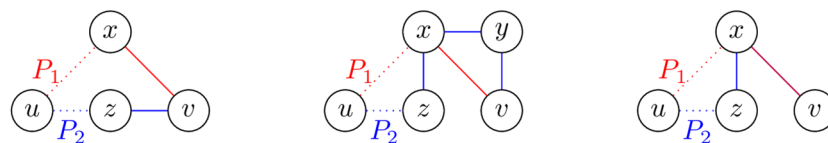
**Definition 4.** Let  $G = (V, E, w)$  be a weighted graph and let the mapping  $d_{wct}: V \times V \rightarrow R_0^+$  be defined by the equation

$$d_{wct}(u, v) = \min_{CT(u,v) \subseteq G} \left( \sum_{e \in CT(u,v)} \frac{1}{w(e)} \right).$$

Then the function  $d_{wct}$  is called the weighted closed trail distance (wCT-distance).

**Theorem 3.** The wCT-distance is a metric on the set  $V$ .

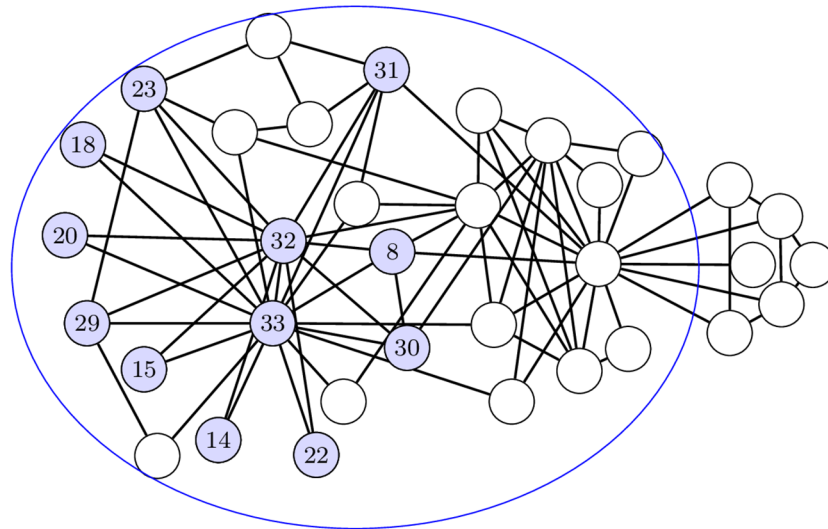
*Proof.* We verify the properties of the metrics.



**Fig 8. Path  $P_1, P_2$  between vertices  $u, v$ .** First graph is biconnected ( $P_1 \cap P_2 = \{u, v\}$ ), second graph is connected with articulation and without bridge ( $P_1 \cap P_2 = \{u, v, x\}$ ) and third graph is connected with bridge ( $P_1 \cap P_2 = \{u, v, x, (x, v)\}$ ).

<https://doi.org/10.1371/journal.pone.0202181.g008>





**Fig 9.** The largest 4-CT component (blue vertices) in the biconnected subgraph (in the blue ellipse) of Zachary’s karate club graph.

<https://doi.org/10.1371/journal.pone.0202181.g009>

From Definition 4 it follows that  $\forall v \in V$   
 $d_{wct}(v, v) = \min_{CT(v,v) \in G} (\sum_{e \in CT(v,v)} \frac{1}{w(e)}) = 0$  and  $\forall u, v \in V$   
 $d_{tw}(u, v) = \min_{CT(u,v) \in G} (\sum_{e \in CT(u,v)} \frac{1}{w(e)}) \geq 0$ .

Symmetry of the distance is obvious:  
 $d_{wct}(u, v) = \min_{CT(u,v) \in G} (\sum_{e \in CT(u,v)} \frac{1}{w(e)}) = d_{wct}(v, u)$  because  $G = (V, E)$  is an undirected graph.  
 A triangle inequality  $\forall u, v, z \in V$   $d_{wct}(u, v) + d_{wct}(v, z) \geq d_{wct}(u, z)$ , is proved in a constructive way. Let  $CT(u, v)$  be a closed trail with the property that  $\sum_{e \in CT(u,v)} \frac{1}{w(e)} = d_{wct}(u, v)$  and let  $CT(v, z)$  be a closed trail with the property that  $\sum_{e \in CT(v,z)} \frac{1}{w(e)} = d_{wct}(v, z)$ . Suppose that a closed trail  $CT(u, v, z)$  formed by the merged closed trail  $CT(u, v)$  and  $CT(v, z)$  exists. This closed trail has a weight of  $w(CT(u, v, z)) = \sum_{e \in CT(u,v)} \frac{1}{w(e)} + \sum_{e \in CT(v,z)} \frac{1}{w(e)}$ . This equality implies that  $d_{wct}(u, z) \leq d_{wct}(u, v) + d_{wct}(v, z)$

**Definition 5.** The  $k$ -wCT component is a maximal subgraph  $H$  which satisfies  $\forall u, v \in V$   
 $d_{wct}(u, v) \leq k$ .

**Example 6.** Fig 11 demonstrate a different wCT components. A closed trail marked in blue  $CT = v_1 - (v_1, v_2) - v_2 - (v_2, v_4) - v_4 - (v_4, v_3) - v_3 - (v_3, v_1) - v_1$  has the weight  $\frac{1}{2} + \frac{1}{10} + \frac{1}{1} + \frac{1}{1} = \frac{13}{5} = 2.6$ . It is the 4-CT component with the biggest weight. There exist 4-CT or 5-CT components with smaller weight (red— $\{v_4, v_5, v_6, v_7\}$  and olive— $\{v_4, v_2, v_5, v_6, v_7\}$ ). The red marked closed trail contains edges with bigger weight than a blue marked closed trail. The value of the wCT component is useful for better scaling of components.

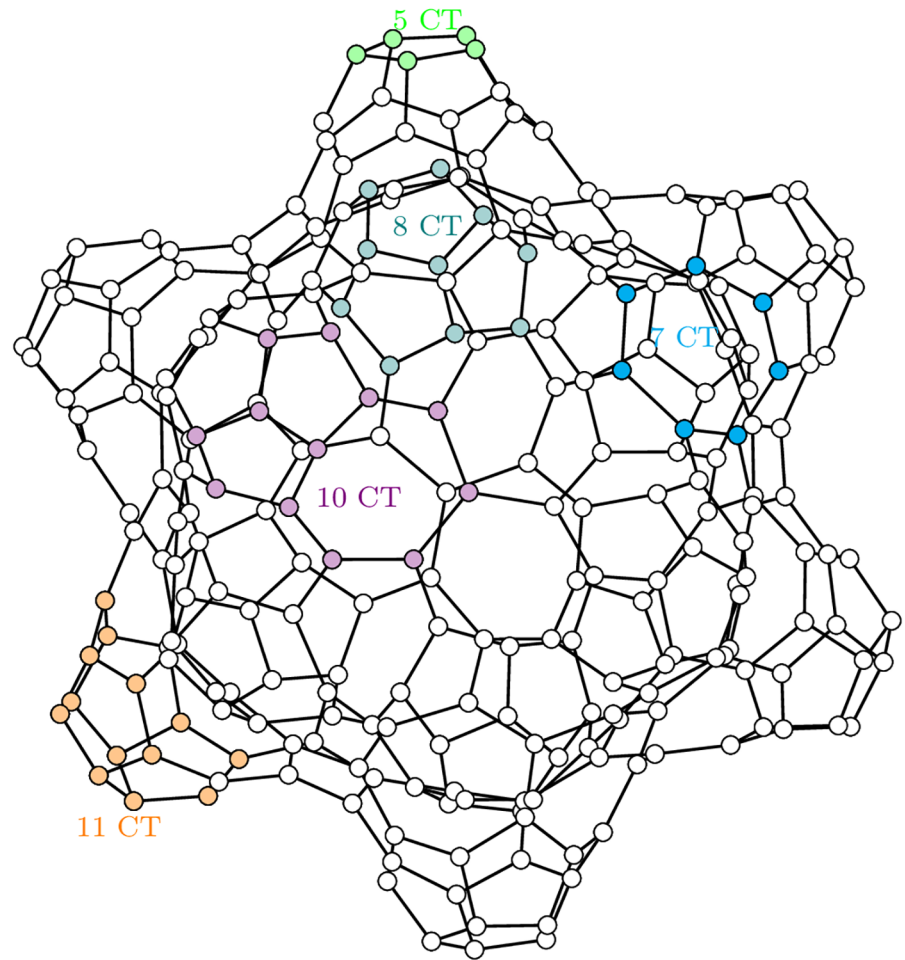
### Conclusion

In this article, we defined a new metric for measuring the distance between two nodes of a biconnected graph. Moreover, the defined method holds the properties of a metric on the

**Table 1.** The number of different  $k$ -CT components in the largest biconnected subgraph of Zachary’s karate club graph.

	3-CT	4-CT	5-CT	6-CT	7-CT	8-CT	9-CT
no. of components	20	17	10	6	3	2	1

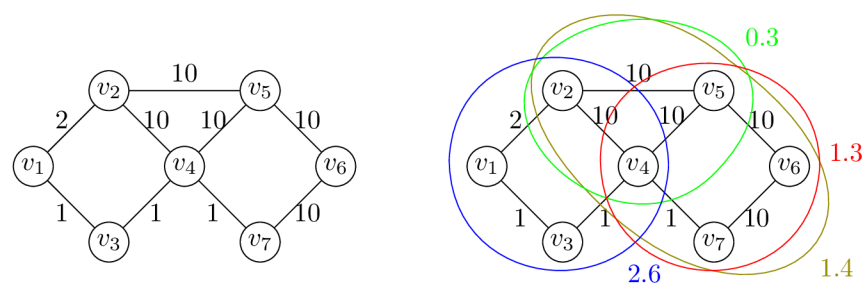
<https://doi.org/10.1371/journal.pone.0202181.t001>



**Fig 10. Graph of a fulleroid with a highlighted sample of the  $k$ -CT components.**

<https://doi.org/10.1371/journal.pone.0202181.g010>

connected graphs without bridges. The same algorithm works on any connected graph, but it does not hold the metric properties. The metric reflects the cyclical interdependencies among the vertices of the graph. The metric derives components, called  $k$ -CT, the sets of vertices that have a distance between each pair of vertices less than or equal to  $k$ . The 3,4,5-CT components are biconnected because the closed trails with the length 3, 4, and 5 are cycles. The defined metric is applicable for both unweighted and weighted graphs.



**Fig 11. Weighted graph and selected wCT components of the graph.**

<https://doi.org/10.1371/journal.pone.0202181.g011>

The paper [34] contains a method for community detection based on network decomposition. Another approach to detection of overlapping communities is used by Palla et al. [35]. Both approaches use a clique percolation method [36] for community detection, more precisely, they use cliques for specifying part of a community. A clique is a 3-CT component according to our approach. Our proposed measure and a component definition decompose the graph into overlapping components using different  $k$ -CT components.

## Acknowledgments

We would like to thank to the Peter Schwerdtfeger, Lukas N. Wirz and James Avery for the providing of the fulleroid data and images to be used in our article.

## Author Contributions

**Conceptualization:** Vaclav Snašel, Pavla Drazdilova, Jan Platos.

**Formal analysis:** Vaclav Snašel, Pavla Drazdilova.

**Methodology:** Vaclav Snašel.

**Software:** Jan Platos.

**Validation:** Jan Platos.

**Visualization:** Pavla Drazdilova.

**Writing – original draft:** Vaclav Snašel, Pavla Drazdilova, Jan Platos.

**Writing – review & editing:** Pavla Drazdilova, Jan Platos.

## References

1. Borůvka O. O jistém problému minimálním [About a certain minimal problem]. 1926; p. 37–58.
2. Nešetřil J, Milková E, Nešetřilová H. Otakar Borůvka on minimum spanning tree problem, translation of both papers the 1926, comments, history. *Discrete Mathematics*. 2001; 233(1-3):3–36. [https://doi.org/10.1016/S0012-365X\(00\)00224-7](https://doi.org/10.1016/S0012-365X(00)00224-7)
3. Carlsson G. Topology and data. *Bulletin of the American Mathematical Society*. 2009; 46(2):255–308. <https://doi.org/10.1090/S0273-0979-09-01249-X>
4. Carlsson G. Topological pattern recognition for point cloud data. *Acta Numerica*. 2014; 23:289. <https://doi.org/10.1017/S0962492914000051>
5. Snášel V, Nowaková J, Xhafa F, Barolli L. Geometrical and topological approaches to Big Data. *Future Generation Computer Systems*. 2017; 67:286–296. <https://doi.org/10.1016/j.future.2016.06.005>
6. Spanier EH. *Algebraic topology*. vol. 55. Springer Science & Business Media; 1994.
7. De Silva V, Ghrist R. Homological sensor networks. *Notices of the American mathematical society*. 2007; 54(1).
8. De Silva V, Ghrist R, Muhammad A. Blind Swarms for Coverage in 2-D. In: *Robotics: Science and Systems*; 2005. p. 335–342.
9. Schwerdtfeger P, Wirz LN, Avery J. The topology of fullerenes. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2015; 5(1):96–145. <https://doi.org/10.1002/wcms.1207> PMID: 25678935
10. Vázquez A, Oliveira JG, Barabási AL. Inhomogeneous evolution of subgraphs and cycles in complex networks. *Physical Review E*. 2005; 71(2):025103. <https://doi.org/10.1103/PhysRevE.71.025103>
11. Leskovec J, Lang KJ, Dasgupta A, Mahoney MW. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*. 2009; 6(1):29–123. <https://doi.org/10.1080/15427951.2009.10129177>
12. Bader DA, Madduri K. A graph-theoretic analysis of the human protein-interaction network using multi-core parallel algorithms. *Parallel Computing*. 2008; 34(11):627–639. <https://doi.org/10.1016/j.parco.2008.04.002>

13. Bader DA, Madduri K. SNAP, Small-world Network Analysis and Partitioning: an open-source parallel graph framework for the exploration of large-scale networks. In: Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE; 2008. p. 1–12.
14. Balaban AT. From chemical topology to three-dimensional geometry. Springer Science & Business Media; 2006.
15. Andova V, Kardoš F, Škrekovski R. Mathematical aspects of fullerenes. *Ars Mathematica Contemporanea*. 2016; 11(2):353–379. <https://doi.org/10.26493/1855-3974.834.b02>
16. Bakry R, Vallant RM, Najam-ul Haq M, Rainer M, Szabo Z, Huck CW, et al. Medicinal applications of fullerenes. *International Journal of Nanomedicine*. 2007; 2(4):639. PMID: [18203430](https://pubmed.ncbi.nlm.nih.gov/18203430/)
17. Hirsch A. The era of carbon allotropes. *Nature Materials*. 2010; 9(11):868–871. <https://doi.org/10.1038/nmat2885> PMID: [20966925](https://pubmed.ncbi.nlm.nih.gov/20966925/)
18. Jariwala D, Sangwan VK, Lauhon LJ, Marks TJ, Hersam MC. Carbon nanomaterials for electronics, optoelectronics, photovoltaics, and sensing. *Chemical Society Reviews*. 2013; 42(7):2824–2860. <https://doi.org/10.1039/c2cs35335k> PMID: [23124307](https://pubmed.ncbi.nlm.nih.gov/23124307/)
19. Sattler KD. Handbook of Nanophysics: Clusters and Fullerenes. CRC press; 2010.
20. Goddard W, Oellermann OR. Distance in graphs. In: Structural Analysis of Complex Networks. Springer; 2011. p. 49–72.
21. Ashrafi AR, Diudea MV. Distance, Symmetry, and Topology in Carbon Nanomaterials. vol. 9. Springer; 2016.
22. Göbel F, Jagers A. Random walks on graphs. *Stochastic processes and their applications*. 1974; 2(4):311–336. [https://doi.org/10.1016/0304-4149\(74\)90001-5](https://doi.org/10.1016/0304-4149(74)90001-5)
23. Kivimäki I, Shimbo M, Saelens M. Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica A: Statistical Mechanics and its Applications*. 2014; 393:600–616. <https://doi.org/10.1016/j.physa.2013.09.016>
24. Chebotarev P. The walk distances in graphs. *Discrete Applied Mathematics*. 2012; 160(10):1484–1500. <https://doi.org/10.1016/j.dam.2012.02.015>
25. Deza MM, Deza E. Encyclopedia of distances. In: Encyclopedia of Distances. Springer; 2009. p. 1–583.
26. Estrada E. The communicability distance in graphs. *Linear Algebra and its Applications*. 2012; 436(11):4317–4328. <https://doi.org/10.1016/j.laa.2012.01.017>
27. Luxburg UV, Radl A, Hein M. Getting lost in space: Large sample analysis of the resistance distance. In: Advances in Neural Information Processing Systems; 2010. p. 2622–2630.
28. Klymko C, Gleich D, Kolda TG. Using triangles to improve community detection in directed networks. arXiv preprint arXiv:14045874. 2014;.
29. Yang Y, Sun PG, Hu X, Li ZJ. Closed walks for community detection. *Physica A: Statistical Mechanics and its Applications*. 2014; 397:129–143. <http://dx.doi.org/10.1016/j.physa.2013.11.034>.
30. Levorato V, Petermann C. Detection of communities in directed networks based on strongly p-connected components. In: Computational Aspects of Social Networks (CASoN), 2011 International Conference on. IEEE; 2011. p. 211–216.
31. Everett MG, Borgatti SP. Analyzing clique overlap. *Connections*. 1998; 21(1):49–61.
32. Edachery J, Sen A, Brandenburg FJ. Graph clustering using distance-k cliques. In: International Symposium on Graph Drawing. Springer; 1999. p. 98–106.
33. Gross JL, Yellen J. Handbook of graph theory. Gross JL, Yellen J, editors. CRC Press; 2004. Available from: [http://books.google.com/?id=mKklGlea\\_BkC](http://books.google.com/?id=mKklGlea_BkC).
34. Ding Z, Zhang X, Sun D, Luo B. Overlapping community detection based on network decomposition. *Scientific reports*. 2016; 6:24115. <https://doi.org/10.1038/srep24115> PMID: [27066904](https://pubmed.ncbi.nlm.nih.gov/27066904/)
35. Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *nature*. 2005; 435(7043):814. <https://doi.org/10.1038/nature03607> PMID: [15944704](https://pubmed.ncbi.nlm.nih.gov/15944704/)
36. Derényi I, Palla G, Vicsek T. Clique percolation in random networks. *Physical review letters*. 2005; 94(16):160202. <https://doi.org/10.1103/PhysRevLett.94.160202> PMID: [15904198](https://pubmed.ncbi.nlm.nih.gov/15904198/)