PLOS ONE

RESEARCH ARTICLE

On the limit value of compactness of some graph classes

Tatiana Lokot[1☯], **Alexander Mehler**[2☯], **Olga Abramov**[3☯]*

1 Retired, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany, **2** Affiliation Faculty of Computer Science and Mathematics, Goethe-University Frankfurt, Frankfurt, Germany, **3** Affiliation CITEC/Faculty of Technology, Bielefeld University, Bielefeld, Germany

☯ These authors contributed equally to this work.
* oabramov@techfak.uni-bielefeld.de

Citation: Lokot T, Mehler A, Abramov O (2018) On the limit value of compactness of some graph classes. PLoS ONE 13(11): e0207536. https://doi.org/10.1371/journal.pone.0207536

Editor: Siamak Yassemi, University of Tehran, ISLAMIC REPUBLIC OF IRAN

Received: June 28, 2018

Accepted: November 1, 2018

Published: November 20, 2018

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by BMBF (https://www.bmbf.de/) funded project "Linguistic Networks" (http://www.linguistic-networks.net/) and DFG (http://www.dfg.de/) funded project "EcoGest" (https://scs.techfak.uni-bielefeld.de/ecogest/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Abstract

In this paper, we study the limit of compactness which is a graph index originally introduced for measuring structural characteristics of hypermedia. Applying compactness to large scale small-world graphs (Mehler, 2008) observed its limit behaviour to be equal 1. The striking question concerning this finding was whether this limit behaviour resulted from the specifics of small-world graphs or was simply an artefact. In this paper, we determine the necessary and sufficient conditions for any sequence of connected graphs resulting in a limit value of $C_B = 1$ which can be generalized with some consideration for the case of disconnected graph classes (Theorem 3). This result can be applied to many well-known classes of connected graphs. Here, we illustrate it by considering four examples. In fact, our proof-theoretical approach allows for quickly obtaining the limit value of compactness for many graph classes sparing computational costs.

Introduction

Evidently, a hypertext forms a network of documents mostly linked on the basis of content-related connections. There is a range of studies applying the compactness measure itroduced in [2] in order to answer questions concerning the structure of hypermedia [1–8]. All these studies addressed compactness computing the values of particular graph invariants which implies high computational costs. In this paper, we take a different perspective considering the limit of compactness for different classes of connected graphs proof-theoretically. Our approach allows for omitting the computational step when the conditions below hold.

The paper is organized as follows. Section starts with repeating graph-theoretical notions used throughout the paper. Section outlines our main findings regarding the limit value of compactness. Section illustrates the application of our tool on four selected graph classes. Finally, Section summarizes our mathematical findings and gives an outlook on results obtained which are part of a subsequent publication. More specifically, in Section we give an overview of those graph classes for which compactness can be easily obtained applying our mathemaical tool (in fact, we have studied about 30 well-known graph classes, there are presumably more than those mentioned here for which our tool can be applied).

Preliminaries

In this Section, we recall some definitions from graph theory to be used throughout this paper. Let G be a simple undirected graph with the *vertex set* $V = V(G)$ and the *edge set* $E = E(G)$. The *order* n of G is the number of its vertices ($n = |V|$). The *size* of G is the number of its edges.

Definition 1. *The degree (or valency)* $\deg(v)$ *of a vertex v of a graph G is the number of edges incident to v in G.*

Definition 2. *The geodesic distance* $\delta(v, w)$ *of two vertices u and v in graph G is the number of edges of the shortest path in G connecting them.*

Definition 3. *The diameter* $D(G)$ *of a graph G is the maximum of geodesic distances in G.*

By $L(G)$ we denote the average geodesic distance in graph $G = (V, E)$ [9]:

$$L(G) = \frac{\sum_{\{v,w\}\in[V]^2} \delta(v, w)}{n(n-1)} \tag{1}$$

Further, we denote the numerator of the fraction in (1) by $\Sigma(G)$, that is:

$$\Sigma(G) = \sum_{\{v,w\}\in[V]^2} \delta(v, w) \tag{2}$$

Thus, (1) can be rewritten as:

$$L(G) = \frac{\Sigma(G)}{n(n-1)} \tag{3}$$

Further, for every vertex $c \in V$ we denote by $\Sigma(c, G)$ the sum of $n - 1$ geodesic distances from c to vertices in $V \setminus \{c\}$. That is:

$$\Sigma(c, G) = \sum_{u\in V} \delta(c, u) \tag{4}$$

and using this notation we write

$$\Sigma(G) = \sum_{u\in V} \Sigma(u, G) \tag{5}$$

For example, for the *path graph* P_2 on two vertices u and v connected by an edge we get

$$\Sigma(u, P_2) = \Sigma(v, P_2) = 1$$

and

$$\Sigma(P_2) = \Sigma(u, P_2) + \Sigma(v, P_2) = 2$$

We repeat the definition of the compactness $C_B(G)$ of a graph $G = (V, E)$, $|V| = n > 1$, as introduced in [2] in a version obtained from [1]:

$$C_B(G) = \frac{K(n-1)n - \left(\sum_{\{v,w\}\in[V]^2} \delta(v, w) + K\sum_{G'\in Com(G)} |G'|(|V| - |G'|)\right)}{(K-1)n(n-1)} \tag{6}$$

where K is "the maximum value an entry in the converted distance matrix [of a graph] can assume" [2, p. 161], $Com(G)$ is the set of connected components of G and $|G'|$ is the order of the graph G' (connected component of the graph G). In what follows, we set $K = n$. Here, we consider only connected graphs, so we can obviously write the following:

$$C_B(G) = \frac{n}{n-1} - \frac{L(G)}{n-1}. \tag{7}$$

On can easily see that $C_B(G) \leq 1$. Further, since $\forall \{v, w\} \in [V]^2$: $D(G) \geq \delta(v, w)$, we have:

$$0 \leq L(G) \leq D(G). \tag{8}$$

Thus, with (7) and (8) we get for every connected graph G:

$$1 \geq C_B(G) = \frac{n}{n-1} - \frac{L(G)}{n-1} = 1 - \frac{L(G)-1}{n-1} \geq 1 - \frac{D(G)-1}{n-1} \tag{9}$$

Definition 4. *The path graph P_m, $m \geq 2$, is a simple connected undirected graph with two vertices of degree 1 (called terminal vertices) and $m - 2$ vertices of degree 2 (called internal vertices).*

The order n of P_m is equal to m and its diameter $D(P_m) = m - 1$. The vertices of P_m can be labeled by the consecutive integers $\{1, 2, \ldots, m\}$ in such a way that the terminal vertices are labeled by 1 and m, respectively, and for every integer i, $1 \leq i \leq m - 1$, the consecutive vertices with labels i and $i + 1$ are adjacent.

Further we need the following formula the proof of which one can easily get with the straightforward calculation:

$$\Sigma(P_m) = \frac{m(m^2-1)}{3} \tag{10}$$

Hence in view of (3) we have

$$L(P_m) = \frac{m(m^2-1)}{3m(m-1)} = \frac{m+1}{3}$$

and

$$C_B(P_m) = \frac{m}{m-1} - \frac{m+1}{3(m-1)} = \frac{2}{3} + \frac{1}{3(m-1)} \tag{11}$$

Definition 5. *The Cartesian product $G_1 \Box G_2$ of two graphs G_1, G_2 is a graph with vertex set $V(G_1) \times V(G_2)$ such that any two vertices (v, u), $(w, z) \in V(G_1 \Box G_2)$ are adjacent iff $v = w$ and u and z are adjacent in G_2 or v and w are adjacent in G_1 and $u = z$.*

Remark 1. *If $G = G_1 \Box G_2$ is the Cartesian product of two graphs G_1 of order n_1 and G_2 of order n_2, then the following properties (referred to below) hold*:

- *G is connected iff both G_1 and G_2 are connected*;
- *the diameter of G is the sum of the diameters of G_1 and G_2*:

$$D(G_1 \Box G_2) = D(G_1) + D(G_2)$$

- *the order n of G is the product $n_1 n_2$ of the order n_1 of G_1 and the order n_2 of G_2.*

Example 1. *Let us consider the Cartesian product $G(m)$ of two copies of the path graph P_m, that is, $G(m) = P_m \Box P_m$ (a so-called square lattice graph whose compactness C_B is investigated below). According to Remark 1, $G(m)$ is connected (since P_m is connected), its order is m^2 and its diameter $D(G(m))$ is $2(m - 1)$.*

Main results

Throughout the present paper we deal with sequences $\{G(m)|m = 1, 2, \ldots\}$ of connected graphs that satisfy the following "natural" condition

$$\lim_{m\to\infty} n(m) = \infty \tag{12}$$

where n is the order of the graph $G(m)$.

Theorem 1. *Let $\{G(m)|m = 1, 2, \ldots\}$ be a sequence of simple undirected connected graphs $G(m)$ such that the order $n = n(m) \to \infty$ for $m \to \infty$. Assume that the following holds:*

$$lim_{m\to\infty} \frac{D(G(m))}{n} = 0, \tag{13}$$

where $D(G(m))$ is the diameter of the graph $G(m)$. Then, the compactness $C_B(G(m))$ tends to 1 for $m \to \infty$.

Proof. In view of (9) we have

$$1 \geq C_B(G(m)) \geq 1 - \frac{D(G(m)) - 1}{n - 1}$$

which implies with our assumptions that

$$\lim_{m\to\infty} C_B(G(m)) = 1$$

Theorem 2. *Let $\{G(m)|m = 1, 2, \ldots\}$ be a sequence of simple undirected connected graphs $G(m)$ such that the order $n = n(m)\to\infty$ for $m \to \infty$. Then, $L(G(m))/n \to 0$ for $m \to \infty$ $(n \to \infty)$ iff $D(G(m))/n \to 0$ for $m \to \infty$ $(n \to \infty)$.*

Proof. In view of (8), we can easily see that we only need to prove that if $D(G)/n \nrightarrow 0$ for $m \to \infty$ then $L(G)/n \nrightarrow 0$ for $m \to \infty$.

Without loss of generality we assume that

$$\lim_{m\to\infty} D(G)/n = c \neq 0 \quad (0 < c \leq 1)$$

Hence, if we take any number a, $0 < a < c$, then for all sufficiently large numbers m we have

$$D(G) > an$$

which implies that there is a geodesic path (subgraph $P_{k(n)}$) in G of length $k(n)$ where $k(n)$ is the integer part of the number an. So we have $an = k(n) + \varepsilon_n$ with ε_n $(0 \leq \varepsilon_n < 1)$ being the fractional part of the number an. Therefore, in view of $\Sigma(G) > \Sigma(P_{k(n)})$ and with (10) we have for all sufficiently large m

$$L(G)/n > \Sigma(P_{k(n)})/(n^2(n-1)) = (1/3)k(n)(k(n)^2 - 1)/(n^2(n-1)).$$

Thus, with $k(n) = an - \varepsilon_n$ we get

$$L(G)/n > \frac{(an - \varepsilon_n)((an - \varepsilon_n)^2 - 1)}{3n^2(n-1)} = \frac{(a - \varepsilon_n/n)((a - \varepsilon_n/n)^2 - 1/n^2)}{3(1 - 1/n)}$$

which implies in view of $\lim_{m\to\infty} n = \infty$ that for all sufficiently large numbers m we have

$$\frac{L(G)}{n} \geq a^3/4 > 0.$$

Hence, $L(G)/n \nrightarrow 0$.

From these two theorems obviously follows:

Corollary 1. *For any sequence of simple undirected connected graphs $G(m)$ for which the order $n = n(m)$ of $G(m)$ tends to ∞ whenever $m \to \infty$, $C_B(G(m)) \to 1$ for $m \to \infty$ iff for $m \to \infty$ $L(G(m))/n \to 0$ ($D(G(m))/n \to 0$).*

And what about the case of disconnected graphs? It turns out that Corollary 1 can be easily generalized with some consideration for the case of disconnected graph classes. That is, the following statement holds:

Theorem 3. *Let $\{G(m)|m = 1, 2, \ldots\}$ be a sequence of simple undirected not necessarily connected graphs $G(m)$ such that the order $n = n(m) \to \infty$ for $m \to \infty$. Then, the compactness C_B of $G(m)$ tends to 1 iff both of the following equalities hold*:

1. *$L(G(m))/n \to 0$ (or $D(G(m))/n \to 0$) for $m \to \infty$*
2. $\lim_{m\to\infty} n_1/n = 1$ *(equivalently*, $\lim_{m\to\infty}(n - n_1)/n = 0$), *where n_1 is the order of the largest connected component of $G(m)$.*

These results give an answer to the question in which case $C_B(G(m))$ tends to 1 for $m \to \infty$ ($n \to \infty$).

Some simple applications

In this section, we consider four simple classes of undirected connected graphs and examine their compactness C_B in the limit of their order (i.e., $n \to \infty$). Sometimes, C_B is easily estimated as in the case of complete graphs. In most cases, however, it is difficult to calculate the exact value of C_B or to give a good estimation of it. Here, we refer to Corollary 1 in order to do this.

The examples of graphs considered here have the following properties. Their diameter $D(G)$ is either constant or grows slower than its order n in such a way that $D(G)/n$ tends to 0 whenever n tends to ∞.

Complete graphs

A *complete graph* K_n of order n is a simple undirected graph with n vertices such that each pair of distinct vertices is connected by a unique edge. That is, the average geodesic distance $L(K_n)$ and the diameter $D(K_n)$ both are equal to 1. Using (7), we get the exact value of compactness $C_B(K_n)$ of K_n:

$$C_B(K_n) = \frac{n}{n-1} - \frac{L(G)}{n-1} = \frac{n}{n-1} - \frac{1}{n-1} = 1$$

It is worth noting that the complete graph K_n is the only graph for which $C_B(K_n)$ equals 1. This trivially results in the following equality:

$$\lim_{n\to\infty} C_B(K_n) = 1$$

The same result is obtained by directly applying Corollary 1.

Star graphs

A *star graph* S_m on m vertices ($m > 2$) is a simple undirected connected graph in which one vertex called *central* vertex has degree $m - 1$ and another $m - 1$ vertices have degree 1.

Consider a sequence $\{S_m | m = 1, 2, \ldots\}$ of connected star graphs of order $n = m$ where diameter $D(S_m)$ is obviously equal to 2 $\forall m \geq 3$. By using Corollary 1 we immediately obtain (seemingly counter-intuitively to what we expect should be measured by compactness):

$$\lim_{m\to\infty} C_B(S_m) = 1$$

It is worth noting that for S_m ($m > 2$) it is easy to calculate the value of $L(S_m)$ and of $C_B(S_m)$. Indeed, we have

$$\Sigma(S_m) = 2(m-1)^2$$

where $\Sigma(G)$ is defined by (2). Next, using (3), $L(S_m)$ can be computed as follows:

$$L(S_m) = 2(m-1)^2/(m(m-1)) = 2(m-1)/m$$

So with (7) we clearly have

$$C_B(S_m) = \frac{m}{m-1} - \frac{2}{m}$$

Hence, it follows that $C_B(S_m) \to 1$ as $m \to \infty$. Thus, we get the same result as in the case of complete connected graphs by calculating $C_B(S_m)$ without Corollary 1.

Lattice graphs

We consider a simple undirected graph $G(m)$ whose vertices can be associated with the points in the plane with the integer x and y coordinates being both in the range $1, 2, \ldots m$. Two vertices are connected by an edge if and only if the distance between them is equal to 1. Such a graph is called a *lattice graph* or a square grid graph and can be viewed as the Cartesian product of two copies of the path graph P_m, that is, $G(m) = P_m \square P_m$ (see Example 1). So, we only repeat that $G(m)$ is connected, has the order $n = m^2$ and the diameter $D(G(m)) = 2(m - 1)$.

Let us consider a sequence $\{G(m) | m = 1, 2, \ldots\}$ of lattice graphs. What is the limit value of $C_B(G(m))$? With $n = m^2$ and $D(G(m)) = 2(m - 1)$ we have

$$\lim_{m\to\infty} n = \infty$$

and

$$\lim_{m\to\infty} \frac{D(G(m))}{n} = \lim_{m\to\infty} \frac{2(m-1)}{m^2} = 0$$

Hence, with Corollary 1 we immediately have $\lim_{m\to\infty} C_B(G(m)) = 1$.

Hypercube graphs

A *hypercube graph* Q_m is a simple undirected connected graph on 2^m vertices labeled by the numbers $0, 1, \ldots, 2^m - 1$. Two vertices are connected by an edge if and only if the binary representations of their labels differ exactly in one position. Q_m can also be defined as the Cartesian product of m copies of the path graph P_2:

$$Q_m = \underbrace{P_2 \Box P_2 \Box \ldots \Box P_2}_{m}$$

In view of Remark 1 of Section, we see that Q_m is connected (because P_2 is connected), its diameter $D(Q_m)$ equals m and its order n is 2^m. We easily see that the fraction $D(Q_m)/n = m/2^m$ tends to zero for $m \to \infty$, so with Corollary 1 we obtain:

$$\lim_{m\to\infty} C_B(Q_m) = 1.$$

Concluding remarks

We confined us here to providing only four simple examples of the graph classes, for which our tool can be easily applied. Actually, we have found more than 30 well-known graphs classes for which our tool is applicable. So, the compactness of these graphs tends to 1 whenever their order tends to ∞.

First, among these graph classes there are those whose diameter does not depend on the order n. These are *book graphs*, *complete r-partite graphs*, *crown graphs*, *Hadamard graphs*, *Keller graphs*, *lating square graphs*, *Paley graphs*, *strongly regular graphs*, *Turán graphs*, *wheel graphs*, *windmill graphs* and some others.

Next, we found some graph classes for each of which the estimation of its diameter as a function of the order n allows the application of our tool. Among those graph classes are the following: *de Bruijn graphs*, *cube-connected cycles*, *Fibonacci cube graphs*, *folded cube graphs*, *Hamming graphs*, *Johnson graphs*, *king's graphs*, *Kneser graphs*, *knight's graphs*, *perfect undirected binary trees*, *self-complementary graphs*, *Ramanujan graphs* and others.

Further, we have seen in Section that the complete graph K_m has the largest possible value of compactness which is 1. So we can say that the graph K_m is the most compact graph among all the graphs of the same order m. If we try now to get the limit value of compactness of the path graph P_m using our tool, we see that this is not possible because $D(P_m)/m \nrightarrow 0$ for $m \to \infty$. Indeed,

$$\lim_{m\to\infty} \frac{D(P_m)}{m} = \lim_{m\to\infty} \frac{m-1}{m} = 1$$

but with (11) we easily get $\lim_{m\to\infty} C_B(P(m)) = 2/3$. We can prove (to appear) that the path graph P_m is the least compact among all the simple connected undirected graphs of the same order m. That is, for each such graph G of order m the following holds:

$$\frac{2}{3} < C_B(P_m) \leq C_B(G)) \leq C_B(K_m) = 1.$$

This finding defines the range of possible values of compactness for connected graphs. Hence, the limit value of compactness for any sequence of simple connected undirected graphs lies within the interval [2/3; 1]. Moreover, we can prove (to appear) that for any number α in the interval [2/3; 1] a graph family can be constructed for which the limit value is exactly α.

It is worth noting that in the case of not necessarily connected graphs the value of compactness lies within the interval [0, 1]. Our future work will consider, amongst others, an extended set of graph classes and the study of a range of invariants including weighted and unweighted ones.

Acknowledgments

This work has been funded by German Federal Ministry of Education (BMBF) in the framework of the research project *Linguistic Networks: Text Technological Representation, Computational Linguistic Synthesis and Physical Modeling* and by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the research project EcoGest (https://scs.techfak.uni-bielefeld.de/ecogest/). Financial support by the BMBF and the DFG is gratefully acknowledged.

Author Contributions

Conceptualization: Tatiana Lokot, Alexander Mehler, Olga Abramov.

Formal analysis: Tatiana Lokot.

Funding acquisition: Alexander Mehler.

Methodology: Alexander Mehler, Olga Abramov.

Project administration: Alexander Mehler.

Supervision: Alexander Mehler.

Validation: Tatiana Lokot.

Writing – original draft: Olga Abramov.

Writing – review & editing: Tatiana Lokot, Alexander Mehler, Olga Abramov.

References

1. Mehler A. Structural Similarities of Complex Networks: A Computational Model by Example of Wiki Graphs. Applied Artificial Intelligence. 2008; 22(7-8):619–683. https://doi.org/10.1080/08839510802164085
2. Botafogo RA, Rivlin E, Shneiderman B. Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. ACM Transactions on Information Systems. 1992; 10(2):142–180. https://doi.org/10.1145/146802.146826
3. Mendes E, Counsell S, Mosley N. Measurement and Effort Prediction for Web Applications. In: Web Engineering, Software Engineering and Web Application Development. London, UK, UK: Springer-Verlag; 2001. p. 295–310. Available from: http://dl.acm.org/citation.cfm?id=647063.714744.
4. Egghe L, Rousseau R. A measure for the cohesion of weighted networks. Journal of the American Society for Information Science and Technology. 2003; 54(3):193–202. https://doi.org/10.1002/asi.10155
5. Smeaton AF, Morrissey PJ. Experiments On The Automatic Construction Of Hypertext From Texts; 1995.
6. Smeaton AF. Building Hypertexts under the Influence of Topology Metrics. In: Fraïssé S, Garzotto F, Isakowitz T, Nanard J, Nanard M, editors. Hypermedia Design. London: Springer London; 1996. p. 105–106.
7. Abramov O, Mehler A. Automatic Language Classification by Means of Syntactic Dependency Networks. Journal of Quantitative Linguistics. 2011; 18(4):291–336. https://doi.org/10.1080/09296174.2011.608602
8. Mehler A, vor der Brück T, Gleim T Rüdiger und Geelhaar. In: Towards a Network Model of the Coreness of Texts: An Experiment in Classifying Latin Texts using the TTLab Latin Tagger. Springer; 2015. p. 87–112.
9. Newman MEJ. The structure and function of complex networks. SIAM Review. 2003; 45:167–256. https://doi.org/10.1137/S003614450342480