Discrete Mathematics 341 (2018) 119–125

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Note

Large values of the clustering coefficient

Michael Gentner [a], Irene Heinrich [b], Simon Jäger [a], Dieter Rautenbach [a,*]

[a] Institute of Optimization and Operations Research, Ulm University, Ulm, Germany
[b] Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany

ARTICLE INFO

Article history:
Received 18 November 2016
Received in revised form 11 August 2017
Accepted 12 August 2017
Available online 14 September 2017

Keywords:
Clustering coefficient
Connected caveman graph
Cliquishness

ABSTRACT

A prominent parameter in the context of network analysis, originally proposed by Watts and Strogatz (1998), is the clustering coefficient of a graph G. It is defined as the arithmetic mean of the clustering coefficients of its vertices, where the clustering coefficient of a vertex u of G is the relative density $m(G[N_G(u)])/\binom{d_G(u)}{2}$ of its neighborhood if $d_G(u)$ is at least 2, and 0 otherwise. It is unknown which graphs maximize the clustering coefficient among all connected graphs of given order and size.

We determine the maximum clustering coefficients among all connected regular graphs of a given order, as well as among all connected subcubic graphs of a given order. In both cases, we characterize all extremal graphs. Furthermore, we determine the maximum increase of the clustering coefficient caused by adding a single edge.

1. Introduction

Watts and Strogatz [8] proposed the clustering coefficient of a graph in order to quantify the corresponding property of networks. For a vertex u of a simple, finite, and undirected graph G, let the *clustering coefficient* of u in G be

$$C_u(G) = \begin{cases} \frac{m(G[N_G(u)])}{\binom{d_G(u)}{2}}, & \text{if } d_G(u) \geq 2, \text{ and} \\ 0, & \text{otherwise,} \end{cases}$$

where $N_G(u)$ denotes the neighborhood $\{v \in V(G) : uv \in E(G)\}$ of u in the graph G whose vertex set is $V(G)$ and whose edge set is $E(G)$, $d_G(u)$ denotes the degree $|N_G(u)|$ of u in G, $G[N_G(u)]$ denotes the subgraph of G induced by $N_G(u)$, and $m(G[N_G(u)])$ denotes the size of this subgraph, that is, $m(G[N_G(u)])$ equals exactly the number of triangles of G that contain the vertex u.

Furthermore, let the *clustering coefficient* of G be the average

$$C(G) = \frac{1}{n(G)} \sum_{u \in V(G)} C_u(G)$$

of the clustering coefficients of its $n(G) = |V(G)|$ vertices.

For an integer ℓ, let $[\ell]$ denote the set of all positive integers at most ℓ.

While the clustering coefficient received a lot of attention within social network analysis [1,5–7], some fundamental mathematical problems related to it are still open. It is unknown [4,7], for instance, which graphs maximize the clustering coefficient among all connected graphs of a given order and size.

* Corresponding author.
E-mail addresses: michael.gentner@uni-ulm.de (M. Gentner), heinrich@mathematik.uni-kl.de (I. Heinrich), simon.jaeger@uni-ulm.de (S. Jäger), dieter.rautenbach@uni-ulm.de (D. Rautenbach).

http://dx.doi.org/10.1016/j.disc.2017.08.020
0012-365X/

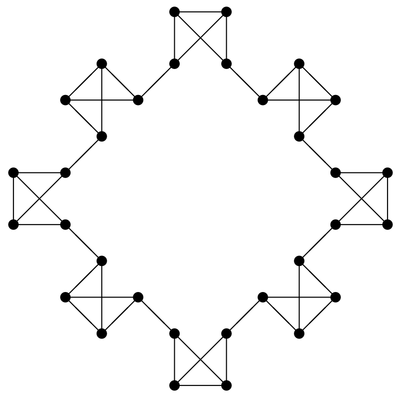

Fig. 1. The unique graph $G(3,8)$ with the largest clustering coefficient among all connected 3-regular graphs of order 24.

Watts [6,7] suggested the so-called *connected caveman* graphs as a possible extremal construction. For integers k and ℓ at least 2, these arise from ℓ disjoint copies $G_1,\dots,G_\ell$ of $K_{k+1}-e$, the complete graph of order $k+1$ minus one edge, arranged cyclically by adding, for every i in $[\ell]$, an edge between one of the two vertices of degree $k-1$ in G_i and one of the $k-1$ vertices of degree k in G_{i+1}, where the indices are identified modulo ℓ. Actually, it is rather obvious that these graphs do not have the largest clustering coefficient among all connected graphs of given order and size, because removing the edge between G_1 and G_2, and adding a new edge between the two vertices of degree $k-1$ in G_1, increases the clustering coefficient.

Fukami and Takahashi [2,3] considered *clustering coefficient locally maximizing graphs* whose clustering coefficient cannot be increased by some local operations such as an edge swap.

In the present paper we determine the maximum clustering coefficients among all connected regular graphs of a given order, as well as among all connected subcubic graphs of a given order. In both cases, we characterize all extremal graphs. Furthermore, we determine the maximum increase of the clustering coefficient caused by adding a single edge.

2. Results

We introduce a slightly modified version of the connected caveman graphs. For integers k and ℓ with $k \geq 3$ and $\ell \geq 2$, let $G(k,\ell)$ be the k-regular connected graph that arises from ℓ disjoint copies $G_1,\dots,G_\ell$ of $K_{k+1}-e$ arranged cyclically by adding, for every i in $[\ell]$, an edge between a vertex in G_i and a vertex in G_{i+1}, where the indices are identified modulo ℓ. Note that $G(k,\ell)$ is uniquely determined up to isomorphism by the requirement of k-regularity. See Fig. 1 for an illustration.

Theorem 1. *Let k and n be integers with $n \geq k+2$ and $k \geq 3$. If G is a connected k-regular graph of order n, then*

$$C(G) \leq 1 - \frac{6}{k(k+1)}$$

with equality if and only if $n/(k+1)$ is an integer and G equals $G(k, n/(k+1))$.

Proof. Let G be a connected k-regular graph of order n. For a non-negative integer i, let V_i be the set of vertices u of G with $m(G[N_G(u)]) = \binom{k}{2} - i$. Since G is connected and has order at least $k+2$, no vertex has a complete neighborhood, that is, V_0 is empty. For a set U of vertices of G, let $\sigma(U) = \sum_{u\in U} C_u(G)$. In order to obtain a useful decomposition of G, we consider some special graphs. For $k \leq 4$, one such graph suffices, while for $k \geq 5$, two more are needed.

Let $G_1,\dots,G_r$ be a maximal collection of disjoint subgraphs of G that are all copies of $K_{k+1}-e$. Let $A = V(G_1)\cup\cdots\cup V(G_r)$ and $R = V(G)\setminus A$. Note that every vertex in A has at most one neighbor in R. Suppose that R contains a vertex u from V_1. Since every vertex in $N_G(u)$ has at least two neighbors in the closed neighborhood $N_G[u]$ of u, the subgraph G_{r+1} of G induced by $N_G[u]$ does not intersect A and is a copy of $K_{k+1}-e$. Now, $G_1,\dots,G_r,G_{r+1}$ contradicts the maximality of the above collection, which implies that R does not intersect V_1.

Since each G_i contains $k-1$ vertices from V_1 and two vertices whose neighborhood induces $K_1\cup K_{k-1}$, and $|A| = r(k+1)$, we have

$$\sigma(A) = \sum_{i\in[r]}\sum_{u\in V(G_i)} C_u(G) = r\left((k-1)\frac{\binom{k}{2}-1}{\binom{k}{2}} + 2\frac{\binom{k-1}{2}}{\binom{k}{2}}\right) = |A|\left(1-\frac{6}{k(k+1)}\right).$$