Physica A 417 (2015) 221–229

Contents lists available at ScienceDirect

Physica A

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

The maximum number of 3- and 4-cliques within a planar maximally filtered graph

Jenna Birch [a], Athanasios A. Pantelous [a,b,*], Konstantin Zuev [b,1]

[a] Institute for Financial and Actuarial Mathematics, Department of Mathematical Sciences, University of Liverpool, United Kingdom
[b] Institute for Risk and Uncertainty, University of Liverpool, United Kingdom

HIGHLIGHTS

- Embeddings of n-vertex triangulations in the form of maximal planar graphs.
- Applying the generating operations proposed by Eberhard to construct the maximal planar graphs.
- Any maximal planar graph can be transformed to a standard spherical triangulation.
- The standard spherical triangulation structure always contains the maximum number of 3- and 4-cliques.

ARTICLE INFO

Article history:
Received 6 February 2014
Received in revised form 18 July 2014
Available online 16 September 2014

Keywords:
Planar Maximally Filtered Graphs
Correlation based Networks
3- and 4-cliques
Eberhard's operation
Standard spherical triangulation

ABSTRACT

Planar Maximally Filtered Graphs (PMFG) are an important tool for filtering the most relevant information from correlation based networks such as stock market networks. One of the main characteristics of a PMFG is the number of its 3- and 4-cliques. Recently in a few high impact papers it was stated that, based on heuristic evidence, the maximum number of 3- and 4-cliques that can exist in a PMFG with n vertices is $3n - 8$ and $n - 4$ respectively. In this paper, we prove that this is indeed the case.

1. Introduction

In recent years there has been increasing interest in how we can model complex systems using network theory. These can include information, technological and biological systems [1], social networks [2] and financial markets [3,4]. In particular, a network based approach of studying complex systems has become very popular in econophysics [5], an interdisciplinary research field that studies economic and financial phenomena. One of the important and fundamental problems in this approach is to filter the most relevant information from financial networks. As a result traditional algorithms from network theory have been adapted and some new methods have been introduced. In 1999 Mantegna [6] introduced a method for finding a hierarchical arrangement for a portfolio of stocks by extracting the Minimum Spanning Tree (MST) from the complete network of correlations of daily closing price returns for US stocks. This work has developed to include Asset Graphs (AG) as introduced in Refs. [7,8] and Threshold Networks [9,10]. Building from the work by Mantegna [6] with the MST, one of the most recent developments was an algorithm proposed by Tumminello et al. [11] where the complete network can be

* Correspondence to: Institute for Financial and Actuarial Mathematics, Department of Mathematical Sciences and Institute for Risk and Uncertainty, University of Liverpool, Peach Street, L697ZL Liverpool, United Kingdom. Tel.: +44 0 151 794 5079.

E-mail addresses: j.l.birch@student.liv.ac.uk (J. Birch), a.pantelous@liverpool.ac.uk, aap@liv.ac.uk (A.A. Pantelous), zuev@liv.ac.uk (K. Zuev).

1 Tel.: +44 0 151 794 5263.

http://dx.doi.org/10.1016/j.physa.2014.09.011
0378-4371/

filtered at a chosen level, by varying the genus of the resulting filtered graph. So if a graph is embedded on a surface with genus $= g$, as g increases the resulting graph becomes more complex and so reveals more information about the clusters formed, while keeping the same hierarchical tree as the corresponding MST. The simplest form of this graph is the Planar Maximally Filtered Graph (PMFG), on surface $g = 0$.

The PMFG is constructed in a similar way to the MST. Specifically, for a graph with n vertices, a weighted edge is associated with all paired vertices where the value of the edge is the similarity coefficient between the two vertices. The weighted edges $u_1, u_2, \ldots, u_e$ are placed in descending order $u_{(1)}, u_{(2)}, \ldots, u_{(e)}$. The first edge $u_{(1)}$ is selected and a graph is constructed with $u_{(1)}$ and the two vertices that it connects. The ordered edges continue to be selected and added to the network structure only if the resulting network can be embedded into a plane i.e. can be drawn on a planar surface without edges crossing. (There are some tests for planarity based on Kuratowski's Theorem. For more details on these refer to Hopcroft and Tarjan [12].) The algorithm ends when all vertices $v_1, v_2, \ldots, v_n$ are connected, using $3(n-2)$ edges.

Mantegna [6] illustrates how the MST can extract the hierarchical arrangement between the vertices of a correlation network (in particular between stocks). Similar to this, one of the key properties of the AG, threshold networks and PMFG is that cliques can form between the vertices in the network which can highlight relationships. Huang et al. [13] create threshold networks to analyse the Chinese stock market using a correlation threshold value $-1 \leq \theta \leq 1$ where θ is the correlation coefficient between two stocks. They study the relationship between the maximum clique, maximum independent set (a subset $I \subseteq V$ such that the subgraph $G(I)$ has no edges) and the threshold value θ. Huang et al. [13] state that '*the financial interpretation of the clique in the stock correlation network is that it defines the set of stocks whose price fluctuations exhibit a similar behaviour*'.

The PMFG has proven to be an important tool for filtering the most relevant information from a network, particularly in correlation based networks that model the correlation between stock prices. For example, Pozzi et al. [14] consider the level of risk and the returns on portfolios selected using filtered graphs, including PMFG. Erygit et al. [15] use PMFGs (along with MSTs and clustering methods) to analyse the daily and weekly return correlations among indices from stock exchange markets of 59 countries. In general, the PMFG can tell us about the clusters (as mentioned above) that form within the dataset, regardless of the network nature, as a result of the underlying topological properties of the network. Song et al. [16] introduce a technique to extract the cluster structure and detect the hierarchical organization within a complex dataset. This method has been developed using the topological structure of the PMFG such as the separating 3-cliques which separate a graph into two disconnected parts. Aste et al. [17] discuss the benefits of studying networks in terms of their surface embeddings. By considering the topology of the PMFG we can see that the basic structure (or motif) of the PMFG is a series of 3-cliques. Consider a sphere, a surface with $g = 0$. The PMFG separates the sphere into a sequence of triangular faces, with each vertex of the network belonging to a 3-clique. For a set of vertices there are various representations that this underlying series of 3-cliques can form (see Section 2.2). A set of three 3-cliques joined by the shared edges of a fourth 3-clique will form a 4-clique between a group of four vertices. Aste et al. [18] discuss that there must be strong relations between the properties of these 4-cliques and the ones of the system from which the cliques have been generated. For the PMFG we consider the vertices that form the 3- and 4-cliques (as the maximum number of elements that can form a clique is 4). Tumminello et al. [11] state '*…normalizing quantities are $n_s - 3$ for 4-cliques and $3n_s - 8$ for 3-cliques. Although we lack a formal proof, our investigations suggest that these numbers are the maximal number of 4-cliques and 3-cliques, respectively, that can be observed in a PMFG of n_s elements*'. As well as looking at the average correlations within the cliques and whether the cliques are from one sector or cross-sector we also consider the ratio between the number of cliques that have formed to the maximum number of cliques that could form. For this, Tumminello [11] used the normalizing quantities that have been mentioned above, an approach that has also been used in Refs. [15,18,19] and used when defining the connection strength of a sector in Ref. [20]. This paper provides the missing formal proof that $3n - 8$ and $n - 4$ are indeed the maximum number of 3-cliques and 4-cliques possible in a PMFG.

This paper is organized as follows: Section 2 provides relational definitions and notations including diagonal flips (2.1) and surface triangles and separating 3-cycles (2.2). Section 3 discusses various representations of each maximal planar graph, including standard spherical triangulation form (3.1). Our main results are presented in Section 4. Section 5 concludes and gives future direction for our research.

2. Relational definitions and notations

Here we introduce some key terminology that we use throughout the paper. Consider a graph $G(V, E)$ where V is the set of vertices belonging to G and E is the set of edges belonging to G. We denote the number of vertices $|V| = n$ and the number of edges $|E| = e$.

Let G be a *planar graph*, i.e. a graph that can be embedded in the plane in such a way that the edges of G will only intersect at the end points (the vertices of G). The planar graph divides the plane into *faces*, with each face bound by a simple cycle of G. The number of edges in this boundary is the *degree* of the face. The *planar representations* of G are all possible isomorphic embeddings of G in the plane.

A *triangulation* of a closed surface is a simple graph, one that does not contain self- or multiple-edges, which is embedded into the surface so that each face is a triangle and that two faces meet along at most one edge. A planar graph is *maximal* if it is triangulated because if a face has more than 3 edges we can add a diagonal edge. A PMFG is a triangulation of a sphere. Within this paper we shall denote P_n as a *maximal planar graph* with n vertices.