"Vasile Alecsandri" University of Bacău
Faculty of Sciences
Scientific Studies and Research
Series Mathematics and Informatics
Vol. 19 (2009), No. 2, 493 - 498

COMBINATORIAL OPTIMIZATION ALGORITHMS FOR POLAR GRAPHS AND THEIR APPLICATIONS IN FINANCE

D. PACURARI, M. MUNTEANU,M.TALMACIU

Abstract. Many natural problems in finance involve partitioning assets into natural groups or identifying assets with similar properties. Building a diversified portfolio is somehow dual to clustering. An approach to clustering constructs an similarity graph, where elements i and j are connected by an edge if and only if i and j are similar that they should/can be in the same cluster. If the similarity measure is totally correct and consistent, the graph will consist of disjoint cliques, one per cluster. A graph is (s, k)-polar if there exists a partition A,B of its vertex set such that A induces a complete s-partite graph and B a disjoint union of at most k cliques. Recognizing a polar graph is known to be NP-complete. In this paper we determine the density and the stability number for (s,k)-polar graphs with algorithms that are comparable, while respect to computing time, with the existing ones and we give some applications in finance.

1. Introduction

Throughout this paper, $G = (V, E)$ is a connected, finite and undirected graph ([1]), without loops and multiple edges, having $V = V(G)$ as the vertex set and $E = E(G)$ as the set of edges.

Keywords and phrases: Recognition algorithm, polar graph, weakly decomposition, time-series, clustering, data-mining.
(2000)Mathematics Subject Classification: 05C85, 05C90.

$\overline{G}$ is the complement of G. If $U \subseteq V$, by $G(U)$ we denote the subgraph of G induced by U. By $G - X$ we mean the subgraph $G(V - X)$, whenever $X \subseteq V$, but we simply write $G - v$, when $X = \{v\}$. If $e = xy$ is an edge of the graph G, then x and y are adjacent, while x and e as well as y and e are incident. If $xy \in E$, we also use $x \sim y$, and $x \not\sim y$ whenever x, y are not adjacent in G. If $A, B \subset V$ are disjoint and $ab \in E$ for every $a \in A$ and $b \in B$, we say that A, B are *totally adjacent* and we denote by $A \sim B$, while by $A \not\sim B$ we mean that no edge of G joins some vertex of A to a vertex of B and, in this case, we say A and B are *non-adjacent*.

The *neighborhood* of the vertex $v \in V$ is the set $N_G(v) = \{u \in V : uv \in E\}$, while $N_G[v] = N_G(v) \cup \{v\}$; we denote $N(v)$ and $N[v]$, when G appears clearly from the context. The *degree* of v in G is $d_G(v) = |N_G(v)|$. The neighborhood of the vertex v in the complement of G will be denoted by $\overline{N}(v)$.

The neighborhood of $S \subset V$ is the set $N(S) = \cup_{v \in S} N(v) - S$ and $N[S] = S \cup N(S)$. A graph is complete if every pair of distinct vertices is adjacent. A *clique* is a subset Q of V with the property that $G(Q)$ is complete. The *clique number* of G, denoted by $\omega(G)$, is the size of the maximum clique. A clique cover is a partition of the vertex set such that each part is a clique. $\theta(G)$ is the size of the smallest possible clique cover of G; it is called the *clique cover number* of G. A stable set is a subset X of vertices where every two vertices are not adjacent. $\alpha(G)$ is the number of vertices of a stable set of maximum cardinality; it is called the *stability number* of G. $\chi(G) = \omega(\overline{G})$ and it is called the *chromatic number* of G.

By P_n, C_n, K_n we mean a chordless path on $n \geq 3$ vertices, a chordless cycle on $n \geq 3$ vertices, and a complete graph on $n \geq 1$ vertices, respectively.

Let $\mathcal{F}$ denote a family of graphs. A graph G is called *$\mathcal{F}$-free* if none of its subgraphs are in $\mathcal{F}$.

The *Zykov sum* of the graphs G_1, G_2 is the graph $G = G_1 + G_2$ having:

$$V(G) = V(G_1) \cup V(G_2),$$
$$E(G) = E(G_1) \cup E(G_2) \cup \{uv : u \in V(G_1), v \in V(G_2)\}.$$

When searching for recognition algorithms, it frequently appears a type of partition for the set of vertices in three classes A, B, C, which we call a *weak decomposition*([2], [3]), such that: A induces a connected subgraph, C is totally adjacent to B, while C and A are totally nonadjacent.

Further on, we present the concepts of data-mining, time-series and clustering.

Data-mining represents the extraction, from existing data, through non-everyday methods, of potential information, unknown previously and possibly useful ([4]).

A continuous sequence of real values is known as time series.

The notion of clustering here is similar to that of conventional clustering of discrete objects. Given a set of individual time series data, the objective is group similar time-series into the same cluster.

2. Polar graphs

In this paper we determine the density and the stability number for (s,k)-polar graphs with algorithms that are comparable, while respect to computing time.

We give, in Theorem 1 below, a necessary and sufficient condition for a connected non-complete graph to be polar.

Theorem 1. *Let $G = (V, E)$ be a connected incomplete graph and, also, a cograph. Let (A, N, R) be a weak decomposition with (*[6], [7]*) with $G(A)$ the weak component. G is (s, k)-polar complete if and only if A is the clique and $G(V - A)$ is $(s, k - 1)$-polar complete.*

Proof.

We take $G = (V, E)$ (s, k)-polar complete, with $V = S \cup Q$, $S = \cup_{i=1}^{s} S_i$, $Q = \cup_{j=1}^{k} Q_j$. We apply the weak decomposition procedure and we obtain (A, N, R) with $G(A)$ the weak component. If we initially take $A = \{a_1\}$, where $a_1 \in S_1$, then, because $\{a_1\} \not\sim S_1 - \{a_1\}$, $(S - S_1) \cup Q \sim S_1 - \{a_1\}$ and $\{a_1\} \sim (S - S_1) \cup Q$ it follows that at the end of the applied procedure of weak decomposition, we have: $A = \{a_1\}$, $N = (S - S_1) \cup Q$, $R = S_1 - \{a_1\}$. If we initially take $A = \{b_1\}$, where $b_1 \in Q_1$, then, because $Q_1 \not\sim Q - Q_1$, $S \sim Q - Q_1$ si $Q_1 \sim S$, it follows that at the end of the applied procedure of weak decomposition, we have: $A = Q_1$, $N = S$, $R = Q - Q_1$. Moreover, we consider A with the propriety that $|A|=max_{j=1,\ldots,k}|Q_j|$. Then $R = Q - A$. Because $max_{j=1,\ldots,k}|Q_j| \geq 1$ then the last variant is the appropriate one. We have: A clique and $G(V - A)$ graph $(s, k - 1)$-polar complete. Reverse, let A be the clique and $G(V - A) = G(N \cup R)$ graph $(s, k - 1)$-polar complete. Since G is co-graph, we have $A \sim N \sim R$. Since A a clique and $A \not\sim N$ and $G(N \cup R)$ $(s, k-1)$-polar complete it follows that $G = G(A \cup (V - A)) = G(A \cup N \cup R)$ is (s, k)-polar complete.

Consequence 1. *If G is (s, k)-polar complete and co-graph then:*
(i) $\alpha(G)$=$max\{max_{i=1,...,s}|S_i|, k\}$;
(ii) $\omega(G) = max\{|A| + 1, s + 1\}$;
(iii) $\nu(G)$=$min_{i=1,...,s}|S_i|$.

Proof. We know ([6]) that
$\alpha(G)$=$max\{\alpha(G(A \cup N)), \alpha(G(A)) + \alpha(G(R))\}$.
Since $A \sim N$ and A clique, it follows that
$\alpha(G(A \cup N))$=$max_{i=1,...,s}|S_i|$ and
$\alpha(G(A)) + \alpha(G(R)) = k$.
So,
$\alpha(G) = max\{max_{i=1,...,s}|S_i|, k\}$.
$\omega(G) = max\{|A| + 1, s + 1\}$.
Because $A \sim N \sim R$ and $N = S = \cup_{i=1}^{s} S_i$, $S_i \sim S_j$, $\forall i, j = 1, ..., s$, it follows that a dominant set of minimum cardinal is $min_{i=1,...,s}|S_i|$. So, $\nu(G) = min_{i=1,..,s}|S_i|$.

3. The recognition algorithm.

Theorem 1 leads to the following recognition algorithm.

Input: A connected non-complete cograph $G = (V, E)$.
Output: An answer to the question: "Is G polar"?
begin
1. Generate L_G, the family of the weak components of G as follows:
    $L_G \leftarrow \emptyset$
    while $V \neq \emptyset$ *do*
      determine the weak component A with the weak
      decomposition algorithm
2. If A not clique
    $then$ $Return$: "G is not polar"
  $else$
    if $\exists i \in S, \exists j \in Q$ such as $ij \notin E$
      $then$ $Return$: "G is not complet polar"
      $else$ Introdu $G(N \cup R)$ in L
    "G is polar complet"

The Complexity of the Algorithm. We determine the degrees of the vertices of graph G. The determination of a weak decomposition with A the weak component takes $O(n+m)$ time. The fact that $S \sim Q$ does not take place, takes $O(n^2)$. Because $A \sim N \sim R$ and $A \not\sim R$ it follows that, if $\exists a \in A$ such that $d_G(a) \neq (|A|-1)+|N|$ then A is not a clique. So, the complexity of the algorithm is $O(n^3)$.

4. Some applications in finance

A lot of finance issues implicate the partition of activities in natural groups or the identification of activities with similar proprieties (e.g. we can partition the supplies in logical groups, based on time-series and other date?). Building a portfolio that can allow the selection of a group of supplies that are not interrelated is dual to clustering. An alternative type of approach to clustering builds a similar graph, where the elements i and j are joined through an edge if and only if they are similar enough so that they can be in the same cluster. If the measuring of similarity is correct then the graph will consist of a disjoint area of cliques and then the similarity will reduce itself to finding the clique. We have just determined the size of the clique for the polar co-graphs class and it is $max\{max_{j=1,\dots,k}|Q_j|, s+1\}$.

In [5]) the authors have studied different types of measuring the distance time-series, at the price of supplies, to see the result from the clusters of the best pairs of industrial groups. They have parameterized the distance measurements on 3 dimensions (representation, normalization and reduction of dimensions) The first dimension was much more informative then the original series, the third dimension leads to clusters better than the original series.

5. Conclusions and future work

In this paper we have characterized the polar graphs, characterization which has led to a recognition algorithm. We have determined the density, the stability number and the domination number for this class of graphs. We have shown the possibility of applying the polar graphs in finance. In the future, a case study can be conducted in financial time-series clustering.

References

[1] C. Berge, *Graphs*, North-Holland, Amsterdam, 1985.

[2] C. Croitoru, E. Olaru, M. Talmaciu, *Confidentially connected graphs*, The annals of the University "Dunarea de Jos" of Galati, Suppliment to Tome XVIII (XXII) 2000, Proceedings of the international conference "The risk in contemporany economy".

[3] C. Croitoru, M. Talmaciu, *On Confidentially connected graphs*, Bul. Stiint. Univ. Baia Mare, Ser B, Matematica - Informatica, Vol. XVI (2000), Nr. 1, 13-16.

[4] W Frawley, G. Piotetsky-Shapiro, C. Matheus, *Knowledge discovery in databases: An overview*, All Magazine, Vol. 13, nr. 3(57-70),1992.

[5] M. Gavrilov, D. Anguelov, P. Indyk, and R. Motwani, *Mining the Stock Market: Which Measure is Best ?*, Proc. Sixth Int. Conf. Knowledge Discovery and Data Mining, 487-496, 2000.

[6] M. Talmaciu, *Decomposition Problems in the Graph Theory with Applications in Combinatorial Optimization - Ph. D. Thesis,* University "Al. I. Cuza" Iasi, Romania, 2002.

[7] M. Talmaciu, E. Nechita, *Recognition algorithm for diamond-free graphs*, Informatica, 2007, vol. 18,no. 3, 457-462 .

D. PACURARI, M. MUNTEANU,
Economics Studies Faculty
M.TALMACIU,
Faculty of Science
University of Bacău
8 Spiru Haret str., 600114
Bacău, ROMANIA
email: doinap_ro@yahoo.com
mircea.muntean.bc@mfinante.ro