Virus for graph properties

María Rosa BRITO[*] Charles DELORME[†]
Luis Enrique MARQUEZ[‡] Oscar ORDAZ[§]

February 15, 1996

Abstract

A virus is a local configuration that prevents a graph (digraph) from possessing a given property P. The virus notion finds application in the random generation of graphs (digraphs) without certain properties, and also in the approximate resolution of NP-complete decision problems whose coproblem is NP-complete. This paper presents the concepts of family of virus and saturated virus in graphs and digraphs. A survey of results in connection with these notions is given. The paper aims at demonstrating the relevance of the surveyed results in the solution of the two problems mentioned above. Finally, some open problems on virus theory and their relation with NP-complete decision problems in graph theory are presented.

Key words: virus, saturated virus, algorithms, graph property, digraph property, NP-complete problems, random generation.

---

[*] Universidad Simón Bolivar. Departamento de Matemáticas. Caracas, Venezuela

[†] Université Paris-XI, Centre d'Orsay, LRI, Bât. 490, 91405 ORSAY Cedex, France, Research partially supported by PCP-Info (CEFI-CONICIT).

[‡] Universidad Nacional Abierta

[§] Universidad Central de Venezuela. Departamento de Matemáticas and Centro de Ingeniería de Software Y Sistema ISYS. Apartado 47567 Caracas 1041-A, Venezuela. e-mail oordaz@conicit.ve

1 Introduction

Losavio, Marquez, Meza and Ordaz (1990) introduced, informally, the notions of *germ* and *virus* in the context of random generation of graphs. Given a property P, defined for graphs of all orders, a germ for P is a structure that, if present in a graph G, guarantees that G will posses P. A virus for P, on the other hand, is a structure whose presence in G implies that G does not have P.

A formal definition is introduced in [2]. In that paper, a virus is a structure composed of a graph H and a function f defined on a part T of its vertex set, with values in $\mathbb{Z}$. The virus H is present in a graph G, if G contains a proper induced subgraph $\tilde{H}$, isomorphic to H, such that, for all $x \in T$,

$$d_G(x) - d_{\tilde{H}}(x) = f(x),$$

where $d_G(x)$ and $d_{\tilde{H}}(x)$ denote the degree of vertex x in the graphs G and $\tilde{H}$, respectively. This notion, besides being applicable to the random generation problem mentioned above, can be of use in the exact or approximate solution of decision problems associated to graphs and digraphs.

In order to clarify the applicability of the virus notion, consider first that we wish to randomly generate a graph G within the family of those graphs of a given order n, that do not posses a given property P. If we have a family of viruses which characterize the absence of P, then we can randomly select a member H from the family of viruses for P among graphs of order n, and then, fill up the rest of G, by adding edges at random in $G \setminus H$, thus obtaining the desired graph. On the other hand assume that we wish to decide if a given graph G posesses property P. For certain important properties (as for instance, '*G is hamiltonian*' or '*G has a perfect matching*', such decision problems can be very hard to solve, the exact solutions sometimes requiring a computational effort more than polynomial on the order of the input graph. If a virus V for P is known, and if the presence or absence of V can be inexpensively determined (in terms of computational time), then an approximate solution to the problem can be given simply by checking the presence/absence of V in G. If V

is in G, then we know that G does not have P. If $V \not\subset G$, we can say that *probably* G has P. The validity of this probabilistic statement will depend on how close the presence of V comes to characterize the absence of P, with respect to the uniform distribution on the family of graphs of a given order. We shall see below that, at least for some important properties, viruses of small order can be found, allowing an approximate solution to the decision problem stated, at a very low computational cost, and with asymptotic probability 1 of giving the right answer.

In this paper we describe viruses by modifying slightly the role of f: f is a mapping of $V(H)$ into $\mathcal{P}(\mathbb{Z}) \setminus \emptyset$, (here $\mathcal{P}(X)$ is the set of parts of X), and for all $x \in V(H)$, $d_G(x) - d_{\tilde{H}}(x) \in f(x)$.

In Subsection 1.1 we introduce several definitions (some of them new), relevant to the discussion to follow. In Section 2, results on the characterization of some important properties via families of viruses are given. Section 3 discusses the existence of viruses of small order which come close to characterizing certain properties, in the probabilistic sense described above. Section 4 contains some final remarks and a list of open questions related to the notion of viruses.

1.1 Terminology

Let P be a property defined on all graphs or digraphs and let $\mathbb{Z}$ be the set of nonnegative integers. Let $G = (V(G), E(G))$ be a graph, let $D = (V(D), E(D))$ be a digraph and $\mathcal{I}$ the set of intervals $[0, n]$ with $n \in \mathbb{Z}$.

Definition 1 *Let P be a property of a graph. We say that (H, f) where H is a graph and $f : V(H) \to \mathcal{I}$, is a* **virus family** *for P if every graph G containing a proper induced subgraph $\tilde{H}$, isomorphic to H, such that, for all $x \in V(\tilde{H})$,*

$$d_G(x) - d_{\tilde{H}}(x) \in f(x),$$

does not have property P.

Definition 2 *Let P be a property of a digraph. We say that (H, f^+, f^-) where H is a digraph and $f^+, f^- : V(H) \rightarrow \mathcal{I}$, is a* **virus family** *for P if every digraph D containing a proper induced subgraph $\tilde{H}$, isomorphic to H, such that, for all $x \in V(\tilde{H})$,*

$$d_G^+(x) - d_{\tilde{H}}^+(x) \in f^+(x) \quad \text{and} \quad d_G^-(x) - d_{\tilde{H}}^-(x) \in f^-(x)$$

does not have property P.

Notice that a virus is a member of a family virus. The context will clarify whether the virus or virus family notion is used.

Definition 3 *Let (H, f) be a virus for a property P. The action of adding an edge to H or replacing $f(x) = [0, m(x)]$ by $f(x) = [0, m(x) + 1]$ for some $x \in V(H)$ is defined as* **positively modified** *and the action of removing an edge to H or replacing $f(x) = [0, m(x)]$ by $f(x) = [0, m(x) - 1]$ for some $x \in V(H)$ is defined as* **negatively modified.**

In a similar way we define when a virus (H, f^+, f^-) for a property P associated to a digraph is positively and negatively modified.

Graph properties are classified in two categories:

Category A: those properties P for which if (H, f) is a virus for P and is positively modified then the new pair (H^1, f^1) is not a virus for P.

The properties: *G is connected, D is strongly connected, G or D are hamiltonian* and *G has perfect matching* belong to this category.

Category B: are those for which if (H, f) is a virus for P and is negatively modified then the new pair (H^1, f^1) is not a virus for P.

In this category we have the property: *G is bipartite*

Definition 4 *Let P be a property in category A. (H, f) is a saturated virus for P, if the new pair (H^1, f^1) obtained by positively modifying (H, f) is not a virus for P. In case that P is in category B then the new pair (H^1, f^1) obtained by negatively modifying an (H, f) virus of P,is not virus for P.*

Definition 5 *Let P be a property. (H, f) is a* **minimal virus** *for P, if there is no a virus (H^1, f^1) for P with $|V(H^1)| < |V(H)|$*

2 Viruses for some important properties

In this section we present viruses which characterize the absence of certain important properties. These viruses have been studied in [2] and [5]. In this work they have been extended to include the definition of virus family, presenting also the modifications required in order to include the saturated virus definition. Observe that the viruses for *G is connected* and *D is strongly connected* properties characterize these properties, in the sense that a graph or digraph has these properties if and only if it does not have the viruses that will be described here for these properties. Nevertheless, the viruses presented here for *G is bipartite*, *G has a perfect matching*, *G or D are hamiltonian* do not characterize these properties. A characterization of graphs without perfect matching nor viruses is present in [5]. For *G is bipartite* and *D is hamiltonian* it shall be explained in what follows.

Theorem 1 *[5] (H, f) is a virus for the property P: G is connected if and only if for every vertex $x \in V(H)$, $f(x) = \{0\}$. Moreover, the viruses for property P: G is connected characterize this property. Minimal viruses correspond to those in which H is connected.*

Theorem 2 *[2] (H, f^+, f^-) is a virus for the property P: D is strongly connected if and only if either $\forall x \in V(H)$ $f^+(x) = \{0\}$ or $f^-(x) = \{0\}$. either $\forall x \in V$ $f^+(x) = \{0\}$ or $\forall x \in V$ $f^-(x) = \{0\}$. Moreover, the viruses for the property P: D is strongly connected characterize this property. Minimal viruses correspond to those in which H is strongly connected.*

The saturated viruses for property P: *G is connected* are the pairs (H, f) with H a completed graph and $f = \{0\}$ on H and G.

The saturated viruses for property P: *D is strongly connected* constituted by every pair (H, f^+, f^-) with H a complete graph and $f^+ = \{0\}$ or $f^- = \{0\}$ on H.

Remark Some saturated viruses for the property P: *G is k-connected* are made of a complete graph with at least k vertices, whit a nonempty part where $f = \{0\}$ and $f = \mathbb{Z}$ on the complementary part.

Some saturated viruses for the property P; *G is $k-$ edge connected* are made of a complete graph K with $\sum_{x \in K} m(x) < k$.

Theorem 3 *[5] (H, f) is a virus for the property P: G is bipartite if and only if H contains an odd cycle. Besides, the viruses do not characterize the property and the non-bipartite, non virus containing graphs are the odd order circuits. (H, f) is a minimal virus if and only if H is a cycle of order three.*

The saturated virus for the property P: *G is bipartite* coincide with minimal virus for bipartite property.

Notice that the digraph $\overrightarrow{ab}, \overrightarrow{bc}, \overrightarrow{ac}$ is not a circuit, and prevents bipartiteness.

Theorem 4 *[2] (H, f) is a virus for the property P: G* is hamiltonian *if and only if for every set of disjoint maximal paths $P_1, \cdots, P_r$ covering $V(H)$ there exists a path P_j such that: if P_j consists of just one vertex $\{x_j^1\}$ then $f(x_j^1) = \{0, 1\}$ and if $V(P_j) = \{x_j^1, \cdots, x_j^{q(j)}\}$ then $f(x_j^1) = \{0\}$ or $f(x^{q(j)_j}) = \{0\}$.*

The satured virus for property P: *G is hamiltonian* is formed by the virus (H, f) verifying the above Theorem and the new pair (H^1, f^1) obtaining by positively modifying (H, f) is not a virus for the property P: *G is hamiltonian.*

Some satured viruses are made of a stable S on s vertices, with $f = \{0\}$ on S, and a complete graph of less that $s + 1$ vertices, with $f = \mathbb{Z}$, all connected to all vertices in S.

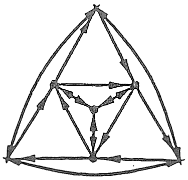

Figure 1: A non hamiltonian digraph without virus

The Petersen virus correspond to a circuit on 6 vertices, with each of them having $f = \{0\}$, then 3 other vertices, each connected by two edges to two opposite vertices of the circuit. These 3 vertices have $f = \{0, 1\}$.

Theorem 5 *[2]* (H, f^+, f^-) *is a virus for the property P: D is hamiltonian if and only if for every set of disjoint maximal paths* $P_1, \cdots, P_r$ *covering* $V(H)$ *there exists a path* P_j *consists of just one vertex* $\{x_j^1\}$ *then* $f^+(x_j^1) + f^-(x_j^1) \in [0, 1]$ *and if* $V(P_j) = \{x_j^1, \cdots, x_j^{q(j)}\}$ *then* $f^-(x_j^1) = \{0\}$ *or* $f^+(x_j^{q(j)}) = \{0\}$.

Similarly, we can define the saturated virus for the property P: D is hamiltonian

The digraph in figure 1 shows that there exist non hamiltonian digraphs without the virus described in the above theorem.

Some saturated virus for the property *P: D is hamiltonian* are made of a stable S with s vertices where $f^+ = \{0\}$ and $f^- = \mathbb{Z}$, and a complete graph K with less than s vertices, where $f^+ = f^- = \mathbb{Z}$ and all arcs between S and K in both directions.

Theorem 6 *Let* H *be graph. Then* (H, f) *is a virus for the property P: G has a perfect matching if and only if there exists* $S \subset V(H)$ *such that* $H - S$ *has at least* $|S| + 1$ *odd connected components with any vertex* x *in these components having* $f(x) = \{0\}$

The graphs with maximal matching with only one vertex, not covered by this matching, do not have a perfect matching and do not have virus, this graphs are odd order. At moment we have no knowledge even order graphs without perfect matching and without virus.

The saturated virus for property P: *G has a perfect matching* is formed by the virus (H, f) verifying the above theorem and the new pair (H^1, f^1) obtained by positively modifying (H, f) is not a virus the property P: *G has a perfect matching.*

Some saturated viruses for property P: *G has a perfect matching* are made of a stable S on s vertices, with $f = \{0\}$ on S, and a complete graph of less than $s + 1$ vertices, with $f = \mathbb{Z}$, all connected to all vertices in S.

3 Viruses of Small Order

One can produce examples that show that minimal viruses, as defined above, are not necessarily of small order. In [2] the following "metaconjecture" is presented: For many important properties there exist viruses of small order in *almost all* instances in which the property is not present. This is useful because the presence (or absence) of a virus of order k inside a graph of order n can be detected by a procedure with an execution time bounded by $O(n^{k+1})$. If k is a small number (say, less than or equal to 3) then, an *almost surely* correct answer to the question *does G have property P?* can be given with a procedure of low computational cost. Now we will show two examples in favor of this metaconjecture.

In the following theorem it is shown that the viruses of order 1 are present *almost surely* in the digraphs which are not strongly connected.

Theorem 7 *[2] Let D be a random digraph with uniform distribution over $|V(D)| = n$, and let A be the event: D has a vertex v such that $d^+(v) = 0$ or $d^-(v) = 0$. Then*

$$\frac{\Pr[A]}{\Pr[D \textit{ is not strongly connected}]}$$

tends to 1 when n tends to infinity.

An immediate consequence of the Theorem above is that, a very simple algorithm has asymptotic probability 1 of answering correctly the question *is D strongly connected?*: simply verify if the event A of the Theorem occurs in our digraph D. If A occurs, answer that D is not strongly connected, otherwise answer that D is (very likely) strongly connected. The computational complexity of this algorithm is $O(n)$, provided that the degrees of the vertices of D are given.

Among the viruses for the property *P: G has a perfect matching* the smallest is an isolated point. We say that G contains an forbidden structure for perfect matching of type (0;1) when G has an isolated vertex, and that G contains a forbidden structure for perfect matching of type (1;1,1) when $G - \{v\}$ has two isolated vertices for some vertex v.

The following Theorem explains the relative importance of these viruses in the absence of the perfect matching property.

Theorem 8 *[1]*

Let $q_0 = \Pr\{G$ contains a forbidden structure for perfect matching of type $(0;1)\}$, $q_1 = \Pr\{G$ contains a forbidden structure for perfect matching of type $(1;1,1)\}$ and $q_2 = \Pr\{G$ contains an forbidden structure for perfect matching of a type different from $(1;1,1)$ and $(0;1)\}$. Then, letting n be the order of the random graph,

$$\frac{\Pr\{G \textit{ has no perfect matching}\}}{q_0} \to 1, \quad \textit{as } n \to \infty$$

and $q_2 = o(q_1)$.

All the probabilities in the above theorem are taken with respect to the uniform distribution over the set of graphs with n vertices.

As a consequence, the following algorithm has very low probability of error even in the case of graphs with no perfect matching and no isolated vertices.

Algorithm

```
For a given graph $G$ with set of vertices $V(G)$
• Flag ← False
• Iteration: For each $v \in V(G)$
    If $v$ is an isolated vertex, Flag ← True
• End Iteration
• Iteration: For each $v \in V(G)$
    If $G - \{v\}$ contains two isolated vertices, Flag ← True
• End Iteration
• If Flag = False, then exit:"The graph $G$ contains (very likely) a perfect matching"
• If Flag = True, then exit:"The graph $G$ has no perfect matching".
```

A 'naive' implementation of the Algorithm just given has computational complexity $O(n^3)$.

4 Conclusions

The origin of virus theory concerns the requirements for computational environments [3],[4] for the assistance to researchers in graph theory, developed at the ISYS research center of the Universidad Central de Venezuela. An important functionality of these systems is a tool for automatic pseudo-random generation of digraphs, imposing given conditions on graph properties, some of which forbid a certain property.

Moreover, this theory can be also useful in the construction of algorithms for NP-complete or polynomial problems, requiring a considerable computational effort.

Natural consequences of this research are the following problems :

1. Construction of algorithms for identifying viruses for properties *P: G has a perfect matching*, *P: G is hamiltonian.*

2. There are graphs and digraphs without the virus defined here, such as *P: D is hamiltonian* or *P: G has perfect matching.* The identification or definition in such cases, of the element or "obstruction" forbidding the graph to have the required property is an open problem. An interesting issue is to find families of graphs and digraphs that do not possess the virus defined in this paper. For example the digraph that is not hamiltonian and does not possess the virus, is hypohamiltonian, that is to say a non-hamiltonian graph, but the digraph resulting eliminating any vertex is hamiltonian.

3. Let G be a graph without a property P. We want G to have property P with minimal changes, such as adding or deleting arcs. We could proceed identifying a virus for P present in G, performing minimal corrections to eradicate the virus. In a non connected graph with only two connected components and with a saturated virus, it is sufficient to connect both connected components. It is clear that the presence of graphs and digraphs without P and does not posses the virus as is mentioned in 2,has to be taken into account.

4. Consider that for property *P: G is hamiltonian* and the property *P: D is hamiltonian*, the absence of this property is related (in the sense of Erdös random graphs) to the presence of the smallest possible virus (with respect to the order of the induced subgraph). These considerations are useful to design aproximate algorithms, with asymtotically zero error probability.

References

[1] M. R. Brito. A simple, a.s correct algorithm for deciding if a graph has a perfect matching, to appear in Discrete Applied Mathematics.

[2] M. R. Brito, W. Fernández de la Vega, O. Meza. O. Ordaz. Viruses in Graphs and Digraphs. Vishwa International Journal of Graph Theory. Volume 2, Number 1, (1993) 35-55

[3] F. Losavio, L.E. Marquez, O. Meza, O. Ordaz. La Génération Aléatoire de Digraphs dans l'Environnement AMDI, Techniques et Science Informatique 3, Vol. 10, No 6 (1991)

[4] F. Losavio, A. Matteo,O. Ordaz, O. Meza, W. Gontier. An implementation of the PAC Architecture using Object-Oriented Technique. Proceedings of the 13th World Computer Congress IFIP'94. Hamburg, Alemania.1994

[5] O. Meza and O. Ordaz. Viruses in Graphs: characterization of graph families by forbidden subgraphs. Submitted to Discrete Math.