Using Extremal Combinatorics to Bound the Differential of a Graph by Its Order

S. Bermudo [1,2]

Department of Economy, Quantitative Methods and Economic History, Pablo de Olavide University, Carretera de Utrera Km. 1, 41013-Sevilla, Spain

H. Fernau [3]

Fachbereich 4—Abteilung Informatik, Universität Trier, D-54286 Trier, Germany

Keywords: differential of a graph, lower bound results

1 Introduction

1.1 Definition

Let $\Gamma = (V, E)$ be a graph of order n and let $B(D)$ be the set of vertices in $V \setminus D$ that have a neighbor in the vertex set D. The *differential* of D is defined as $\partial(D) = |B(D)| - |D|$ and the differential of a graph is equal to $\max\{\partial(D) : D \subseteq V\}$. The graph parameter ∂ was introduced in [9]. There, also several basic properties were derived. Notice that for a graph Γ of order

[1] Partially supported by Ministerio de Ciencia y Tecnología, (MTM 2009-09501) and by the Junta de Andalucía (FQM-260) and (P06-FQM-02225).

[2] Email: sbernav@upo.es

[3] Email: fernau@uni-trier.de

n, $0 \leq \partial(\Gamma) \leq n-2$. For every graph Γ with connected components $\Gamma_1, ..., \Gamma_k$, $\partial(\Gamma) = \partial(\Gamma_1) + \cdots + \partial(\Gamma_k)$. Therefore, we will only consider connected graphs.

The study of the differential in graphs was continued in [1,2,12,14].

1.2 Motivation

Social networks, such a Facebook or Twitter, have served as an important medium for communication and information disseminating. As a result of their massive popularity, social networks now have wide applications in the viral marketing of products and political campaigns. Motivated by its wide applications in these topics some authors proposed some influence maximization problems [4,6,8] as a fundamental algorithmic problem for information diffusion in social networks. This problem consists in determining the best group of nodes to influence the rest. The study of the graph parameter $\partial(\Gamma)$, called the differential of Γ, could be motivated from such scenarios.

The differential of a set could act as a measure of how this set can influence the rest of the vertices. Suppose that we have a political party and we are interested in giving some political talks in some cities of a country to influence the people. A natural problem would be to find "the best cities" to organize those talks, in the sense that we want to give the talks in the cities where we can influence most people, assuming a certain bound on the number of cities that we are able to visit. We could see the map of the country as a graph and, to avoid weights, we could consider all the cities having the same population and the same importance, and all roads between cities having the same length. We also assume that people might go to a meeting if this takes place in their city or in a neighboring city. In such a particular case, if we want to influence everybody, it looks logical to choose the cities which belong to the dominating set but, sometimes, the dominating set contains some vertices which do not dominate anybody but themselves and possibly one single neighbor. From the economical point of view, it might not be interesting to give a talk in a city if virtually nobody is supposed to attend it. In this example, the best choice could be the cities which belong to the minimum differential set, although we do not influence every city in the country.

More generally, the idea of viral marketing (as explained in [4,6,8]) tries to (ab)use customers acquired by specific marketing offers as multiplicators, influencing their immediate neighborhood to buy certain products. This model is a stochastic one from the start, but can be simplified to lead to the graph problem studied in this paper.

2 Preliminaries

2.1 Connections with Domination and the Enclaveless Number

As explained in [9], the differential of a graph is related to the well-known parameter $\gamma(\Gamma)$ denoting the minimum size of a dominating vertex set in Γ. Namely [5,15],

$$\Psi(\Gamma) := \max\{|B(D)| : D \subseteq V\} = n - \gamma(\Gamma),$$

where the parameter Ψ is known as the *enclaveless number* of a graph and, for a B with $|B(D)| = \Psi(\Gamma)$, $B(D)$ is also known as a *nonblocker* set; see [5,7].

Moreover, for any graph without isolated vertices,

$$\Psi(\Gamma) - \gamma(\Gamma) = n - 2\gamma(\Gamma) \leq \partial(\Gamma) \leq \Psi(\Gamma) - 1, \tag{1}$$

see [9]. We have shown in [1] that computing $\partial(\Gamma)$ is of a complexity similar to computing $\Psi(\Gamma)$, being NP-complete on rather restricted graph classes but solvable using parameterized algorithms (with a standard parameterization); confer to [5].

In this paper, we are studying lower bounds on the differential of a graph, obtaining results that nicely complement what is known about the enclaveless number. For that parameter, the following is known:

- [11] For any connected graph Γ of order $n \geq 2$, $\Psi(\Gamma) \geq n/2$.
- [3,10] For any connected graph Γ of order $n \geq 8$ and minimum degree $\delta(\Gamma) \geq 2$, $\Psi(\Gamma) \geq \frac{3n}{5}$. Moreover, there are seven exceptional connected graphs at all that violate this bound.
- [13] For any graph Γ of order n satisfying $\delta(\Gamma) \geq 3$, $\Psi(\Gamma) \geq \frac{5n}{8}$.

The second item immediately implies, when combined with Eq. (1), that $\partial(\Gamma) \geq \frac{n}{5}$ for any connnected graph Γ of order $n \geq 8$ satisfying $\delta(\Gamma) \geq 2$.

2.2 Statement of Main Results

Here, we derive the following main results, improving this immediate bound:

Theorem 1 *For any connected graph Γ of order $n \geq 3$, $\partial(\Gamma) \geq n/5$.*

Theorem 2 *For any connected graph Γ of order n that has minimum degree two, $\partial(\Gamma) \geq \frac{3n}{11}$ apart from five exceptional graphs.*

An alternative way of defining the differential of a graph is the following, which is based on the notion of a *big star*, i.e., some star S_d with $d \geq 2$. Given the graph $\Gamma = (V, E)$, a *big star packing* is given by a vertex-disjoint collection $\mathcal{S} = \{X_i \mid 1 \leq i \leq k\}$ of (not necessarily induced) big stars $X_i \subseteq V$, i.e., $\Gamma[X_i]$ contains some S_d with $d = |X_i| - 1 \geq 2$. If $\mathcal{S}$ is a big star packing of Γ, we also denote this property by $SP(\Gamma, \mathcal{S})$.

Proposition 1 $\partial(\Gamma) = \max\{\sum_{S\in\mathcal{S}}(|S| - 2) : SP(\Gamma, \mathcal{S})\}$.

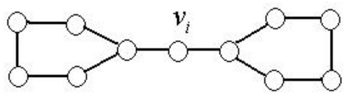
Proposition 2 *There exists an infinite family of connected graphs Γ'_k, $k \geq 1$, of order n with a differential of $n/5$.*

We will say that a vertex $v \in V$ is a *critical vertex* if $v \in D \cup B(D)$ for every set $D \subseteq V$ such that $\partial(D) = \partial(\Gamma)$.

Lemma 1 *Let $\Gamma_i = (V_i, E_i)$ be a graph which has a critical vertex u_i, for $i = 1, 2$. If $\Gamma = (V, E)$ is a graph such that $V = V_1 \cup V_2$ and $E = E_1 \cup E_2 \cup \{u_1u_2\}$, then u_1 and u_2 are critical vertices of Γ and $\partial(\Gamma) = \partial(\Gamma_1) + \partial(\Gamma_2)$.*

Proposition 3 *There exists an infinite family of connected graphs Γ_k, $k \geq 1$, of minimum degree two and of order $11k$ with a differential of $3k$.*

Proof. Consider the following graph $\Gamma^i = (V^i, E^i)$ of order 11.



The differential of this graph is 3 and v_i is the only critical vertex. Now, we consider a connected graph $\Gamma_k = (V_k, E_k)$ such that $V_k = \bigcup_{i=1}^k V^i$ and $E_k = \bigcup_{i=1}^k E^i \cup \{v_1v_2, v_2v_3, ..., v_{k-1}v_k\}$. By Lemma 1, we know that $\partial(\Gamma) = 3k$. □

This also shows that the bounds given in the Main Results are best possible.

3 Towards Proving the Main Results

As the proof of Theorem 1 is pretty straightforward, we will sketch the proof of Theorem 2 only in the following. We will use methods from Extremal Combinatorics.

If our theorem was false, then there should exist an example $\Gamma = (V, E)$ with $|V| = n$, $\delta(\Gamma) \geq 2$ and $\partial(\Gamma) < \frac{3n}{11}$. If such a counterexample exists, we could also ask for a proof, i.e., we are also given a big star packing $\mathcal{S}(D)$ such that the set D of its star centers satisfies $\partial(D) = \partial(\Gamma)$. Let $\mathcal{D}_1(\Gamma) = \{D \subseteq V \mid D$ is the set of star centers of a big star packing and $\partial(D) = \partial(\Gamma)\}$. Define $C(D) = V \setminus (D \cup B(D))$.

Lemma 2 *If $D \in \mathcal{D}_1(\Gamma)$, then the induced graph $\Gamma[C(D)]$ decomposes into K_1- and K_2-components.*

Recall that, apart from $D \in \mathcal{D}_1(\Gamma)$, we are also given a corresponding big star packing $\mathcal{S}(D)$. Hence, $\partial(D) = \sum_{S \in \mathcal{S}(D)} (|S| - 2)$ and $D \cup B(D) = \bigcup_{S \in \mathcal{S}(D)} S$. Notice that there might be several big star packings that testify the differential claimed for D, but we will fix one of these big star packings, denoted as $\mathcal{S}(D)$, in the following discussion. For every $j = 2, ..., \Delta$, we will denote by $\mathcal{S}_j(D)$ the set of all stars S in $\mathcal{S}(D)$ such that $|S| = j+1$, and $\mathcal{S}_{\geq 3}(D) = \bigcup_{j \geq 3} \mathcal{S}_j(D)$. Since there could be several differential sets that attain the differential $\partial(\Gamma)$, we will furthermore ask for a differential set D that maximizes $|D|$ among all those with $\partial(D) = \partial(\Gamma)$. Let $\mathcal{D}_2(\Gamma) = \{D \in \mathcal{D}_1(\Gamma) \mid \forall D' \in \mathcal{D}_1(\Gamma)\ (|D'| \leq |D|)\}$.

Lemma 3 *If $D \in \mathcal{D}_2(\Gamma)$, then any vertex x in $B(D)$ has at most one neighbor in $C(D)$.*

As a possible application of Lemma 2 and Lemma 3, we state:

Lemma 4 *For any connected graph Γ of order $n \leq 8$ that has minimum degree two, if $\partial(\Gamma) < \frac{3n}{11}$, then Γ is one of the five exceptional graphs.*

Due to Lemma 2, we could also establish a third priority; let $k_2(D)$ denote the number of K_2-components in $\Gamma[C(D)]$. Let $\mathcal{D}_3(\Gamma) = \{D \in \mathcal{D}_2(\Gamma) \mid \forall D' \in \mathcal{D}_2(\Gamma)\ (k_2(D') \leq k_2(D))\}$. Fix some arbitrary $D \in \mathcal{D}_3(\Gamma)$ in the following discussion. In order to explain the importance of the given sequence of priorities, we establish:

Lemma 5 *If $S \in \mathcal{S}(D)$ with $|S| \geq 4$, then no $x \in S \setminus D$ is neighbor of a K_2-component in $\Gamma[C(D)]$.*

Our previous reasoning shows that stars with more than three vertices are giving a very good bound. Bad situations arise with S_2-stars, which we are now studying in details.

A sequence of pairwise distinct adjacent vertices $v_1, \ldots, v_t$ is called an S_2 *sequence* if it obeys the following recursive definition:

- either $t = 1$ and $v_1 \in C(D)$, or
- $t = 3$, $\{v_1, v_2, v_3\} \in \mathcal{S}(D)$, and $v_2 \in D$, or
- $t > 1$, $v_t \in C(D)$, and $v_1, \ldots, v_{t-1}$ is an S_2 sequence, or
- $t > 3$, $\{v_{t-2}, v_{t-1}, v_t\} \in \mathcal{S}(D)$, $v_{t-1} \in D$, and $v_1, \ldots, v_{t-3}$ is an S_2 sequence.

An S_2 sequence s is called *maximal* if there are no vertices x (or x, y, z) that do not already occur in s, such that s, x or x, s or x, y, z, s or s, x, y, z form an S_2 sequence. An S_2 sequence $s = v_1, v_2, \ldots, v_t$ is called *maximal from* v_1 if there are no vertices x (or x, y, z) that do not already occur in s, such that s, x or s, x, y, z form an S_2 sequence. Clearly, if $s = v_1, v_2, \ldots, v_t$ is an S_2 sequence, then $s^- = v_t, v_{t-1}, \ldots, v_1$ is an S_2 sequence, as well; s is maximal if and only if s is maximal from v_1 and s^- is maximal from v_t. If v_1 is adjacent to v_t we will consider that s and s^- are equivalent. For $S \in \mathcal{S}(D)$, let $C(S, D)$ collect all vertices from $C(D)$ that are neighbors of vertices from S. For a collection $\mathcal{S} \subseteq \mathcal{S}(D)$ of stars, let $C(\mathcal{S}, D) = \bigcup_{S \in \mathcal{S}} C(S, D)$. Let $\mathcal{D}_4(\Gamma) = \{D \in \mathcal{D}_3 \mid \forall D' \in \mathcal{D}_3(|C(\mathcal{S}_{\geq 3}(D'), D')| \leq |C(\mathcal{S}_{\geq 3}(D), D)|)\}$. So, in the following, let $D \in \mathcal{D}_4(\Gamma)$.

For every $D \in \mathcal{D}_4(\Gamma)$ we denote by $s_2(D)$ the number of maximal inequivalent S_2 sequences in Γ, and $\mathcal{D}_5(\Gamma) = \{D \in \mathcal{D}_4(\Gamma) : \forall D' \in \mathcal{D}_4(\Gamma)\ (s_2(D) \leq s_2(D'))\}$. In the following, we consider $D \in \mathcal{D}_5(\Gamma)$.

According to our priorities, for the discussion of any maximal S_2 sequence s starting with x, it sufficient to distinguish three different cases:

non-C case: s contains no vertex from $C(D)$ at all.

single-C case: s contains exactly one vertex from $C(D)$, which is x.

double-C case: s contains exactly two vertices from $C(D)$, which are the first vertex x in s and the second vertex y in s.

In order to prove our bound, the first of the three cases does not harm, since it implies a better ratio (of three) on the S_2 path. Of particular danger to our counting are those $C(D)$-vertices that are not close to big stars.

In the following, we abbreviate $C_2(D) = C(D) \setminus C(\mathcal{S}_{\geq 3}(D), D)$.

We start discussing the single-C case:

Lemma 6 *Consider a maximal S_2 sequence s starting with x such that s contains exactly one vertex from $C(D)$, which is x. Assume that $x \notin N(S)$ for any $S \in \mathcal{S}_{\geq 3}(D)$. Let z be the last vertex of s and let $N(z, \notin s)$ collect all neighbors of z that are not already in s. Then, $N(z, \notin s) = \emptyset$.*

A similar statement is true for the double-C case.

We will say that a vertex $x \in C_2(D)$ has a *private S_2 star* if there exists a maximal S_2 sequence starting with x containing this star, which does not belong to a S_2 sequence starting with $x' \in C_2(D)$ with $x' \neq x$.

Lemma 7 *Every maximal S_2 sequence starting with $x \in C_2(D)$ contains more than one S_2 star.*

A similar Lemma is true for the double-C case.

Lemma 8 *If s is an S_2 sequence that starts with $x \in C_2(D)$ (or with $e = \{x, y\} \subseteq C_2(D)$), then any S_2 star in s is private for x (or for e).*

We denote by k_j the number of $S \in \mathcal{S}(D)$ such that $|S| = j + 1$. By the previous lemmas we obtain the following result.

Lemma 9 *If $|C_2(D)| = r$, then $3r \leq 2k_2$.*

This is the final cornerstone to establish a contradiction to the existence of a counterexample by a rather straightforward counting argument.

4 Concluding Remarks

Using arguments very similar to those leading to Theorem 2, we could improve the lower bound ratio from $\frac{3}{11}$ to $\frac{2}{7}$ in the cases of subcubic graphs of minimum degree (at least) two, under some mild additional conditions. Since our lower bound example graph families are of maximum degree four, this completes the study with respect to a further maximum degree bound, since 2-regular graphs have been already completely studied in [9]. Furthermore, in the long version of this paper, we also study the influence of the graph parameter "maximum number of induced P_5" in order to improve on Theorem 1.

Having established some lower bounds, some natural questions prevail: (1) Reed could establish a better bound in relation with the domination (or enclaveless) parameter for graphs with minimum degree three. Are similar achievements possible for the differential? (2) So far, we could only make use of Theorem 1 for purposes of parameterized complexity in the spirit of reference [5]; see [1]. Is there any way to employ Theorem 2 for this purpose to obtain better running times of our parameterized algorithms?

References

[1] Bermudo, S. and H. Fernau, *Computing the differential of a graph (abstract)*, in: *Cologne-Twente Workshop on Graphs and Combinatorial Optimization*, 2011.

[2] Bermudo, S., J. M. Rodríguez and J. M. Sigarreta, *On the differential in graphs*, (2011), submitted.

[3] Blank, M., *An estimate of the external stability number of a graph without suspended vertices (in Russian)*, Prikl. Math. i Programmirovanie Vyp. **10** (1973), pp. 3–11.

[4] Culotta, A., *Maximizing cascades in social networks*, Technical report, University of Massachusetts. USA (2003). URL http://www.cs.umass.edu/~culotta/pubs/culotta03maximizing.pdf

[5] Dehne, F., M. Fellows, H. Fernau, E. Prieto and F. Rosamond, NONBLOCKER*: parameterized algorithmics for* MINIMUM DOMINATING SET, in: J. Štuller, J. Wiedermann, G. Tel, J. Pokorný and M. Bielikova, editors, *Software Seminar SOFSEM*, LNCS **3831** (2006), pp. 237–245.

[6] Domingos, P. and M. Richardson, *Mining the network value of customers*, in: *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining, KDD*, 2001, pp. 57—66.

[7] Downey, R. G. and M. R. Fellows, "Parameterized Complexity," Springer, 1999.

[8] Kempe, D., J. M. Kleinberg and É. Tardos, *Influential nodes in a diffusion model for social networks*, in: L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi and M. Yung, editors, *Automata, Languages and Programming, ICALP*, LNCS **3580** (2005), pp. 1127–1138.

[9] Mashburn, J. L., T. W. Haynes, S. M. Hedetniemi, S. T. Hedetniemi and P. J. Slater, *Differentials in graphs*, Utilitas Mathematica **69** (2006), pp. 43–54.

[10] McCuaig, B. and B. Shepherd, *Domination in graphs of minimum degree two*, Journal of Graph Theory **13** (1989), pp. 749–762.

[11] Ore, O., "Theory of Graphs," Colloquium Publications **XXXVIII**, American Mathematical Society, 1962.

[12] Pushpam, P. R. L. and D. Yokesh, *Differentials in certain classes of graphs*, Tamkang Journal of Mathematics **41** (2010), pp. 129–138.

[13] Reed, B., *Paths, stars, and the number three*, Combinatorics, Probability and Computing **5** (1996), pp. 277–295.

[14] Sigarreta, J. M., *Differential in Cartesian product graphs* (2010), to appear in Ars Combinatorica.

[15] Slater, P. J., *Enclaveless sets and MK-systems*, Journal of Research of the National Bureau of Standards **82** (1977), pp. 197–202.