Family of graphs without 2-community structure

Cristina Bazgan — Université Paris-Dauphine, CNRS, LAMSADE, France
Janka Chlebíková — University of Portsmouth,UK
Clément Dallard — University of Portsmouth, UK
Thomas Pontoizeau — Université Paris-Dauphine, LAMSADE, France

Abstract

In the context of community structure detection, we study the existence of a partition of the vertex set of a graph into two parts such that each part is a community, namely a 2-*community structure*. We use the definition of a community where each vertex of the graph has a larger proportion of neighbors in its community than in the other community. There are few results about this problem, and it was not known if there exist graphs without 2-community structure, except stars. In this paper, we present a class of graphs without 2-community structure and leave some interesting open questions about the problem.

1 Introduction

Social networks can be represented as graphs, with members as nodes and relationships as edges. In these graphs, a natural problem consists in partitioning the graph into dense subgraphs, also called *communities*. Partition problems about community detection have been widely studied in the literature, see *e.g.* [4] for an overview. In this paper, we are interested in the community structure definition recently introduced by Olsen [8]. A *community structure* is described as a partition of the vertices into communities, where a part is a *community* if and only if each vertex respects a condition. This condition captures the proportion of neighbors a vertex has within the communities with regard to their sizes, and must be greater in its own community. We focus on community structures with a fixed number of communities, specifically with 2 communities. Firstly, we formally define the notions of communities and community structures. Secondly, we outline the known results about the 2-community structure problem, *i.e.* the problem of deciding if a 2-community structure exists in a given graph. Thirdly, we construct an infinite family of graphs without 2-community structure. As far as we know, these are the first negative results regarding the existence of a 2-community structure.

2 Communities and community structures

Although the concept of community is intuitively clear, many theoretical definitions for a partition into communities have been studied in the literature. For instance, minimizing the number of links between communities is the problem of finding a cut of minimum size [7]. On the other hand, maximizing the number of links within communities corresponds to a partition into cliques. The problem of finding a partition into cliques minimizing the number of cliques, also called clique cover, is well known and has been proved NP-hard in [5]. Yet, a clique is a very restrictive definition and may not be appropriate to describe a community. Closely related to the 2-community structures, Gerber and Kobler defined the problem of deciding if a graph has a vertex partition into two nonempty parts such that each vertex has at least as many neighbors in its part as in the other part [6]. Bazgan *et al.* proved that this problem is NP-hard [1]. Olsen refined the concept of community by using the notion of proportion of neighbors rather than the degree only [8]. More

specifically, a community structure is a partition in which each vertex has a greater proportion of neighbors in its community than in any other community.

Definition 1. A community structure of a graph $G = (V, E)$ is a partition $\Pi = \{C_1, \ldots, C_k\}$, $k \geq 2$, of V such that $|C| \geq 2$ for each $C \in \Pi$ and

$$\forall C_i, C_j \in \Pi,\ \forall u \in C_i,\ \frac{d_{C_i}(u)}{|C_i| - 1} \geq \frac{d_{C_j}(u)}{|C_j|}\,.$$

The elements of Π are called communities. Given $C \in \Pi$, we say a vertex $u \in C$ is *satisfied* (in C) if it respects the above inequality.

According to Definition 1, graphs that are isomorphic to stars do not have a community structure. Indeed, since each part of a community structure has to have at least 2 vertices, the vertices that are not in the same part as the center of the star cannot satisfy the inequality. In order to generalize the definition also for stars, we can relax the inequality without compromising the notion of community structure.

Definition 2. A community structure of a graph $G = (V, E)$ is a partition $\Pi = \{C_1, \ldots, C_k\}$, $k \geq 2$, of V such that

$$\forall C_i, C_j \in \Pi,\ \forall u \in C_i,\ |C_j| \cdot d_{C_i}(u) \geq (|C_i| - 1) \cdot d_{C_j}(u)\,.$$

Hence, Definition 1 and Definition 2 are completely equivalent when all communities have at least 2 vertices, but the latter is more general. Notice that if a graph is disconnected, the problem becomes trivial, hence we consider that the graphs are connected. Therefore, if a community contains a single vertex, then all the neighbors of this vertex must be universal in the graph, otherwise they cannot be satisfied. In [8], the author gives a polynomial time algorithm to compute a community structure. Unfortunately, the number of communities can be large. On the other hand, when some vertices are asked to belong to the same community, the problem is proved to be NP-hard. However, the complexity of the problem remains unknown when the number of communities is fixed. In this paper, we discuss the problem of finding a community structure with exactly 2 communities, namely a 2-community structure.

3 Known results for 2-community structure

So far, only few results are known about the complexity of finding a 2-community structure in a graph. It has been proved in [3] that deciding if a graph contains a 2-community structure with both communities of the same size is NP-complete. In a *connected 2-community structure*, each community must induce a connected subgraph. It has been proved that the connected 2-community structure problem is polynomial time solvable on trees [2, 3], and graphs with maximum degree 3 or minimum degree $n-3$ for n is the order of the graph [2]. The proofs and algorithms extensively use the connectedness of the communities. For instance, to find a connected 2-community structure in a tree, one can prove that there exists an edge such that its removal yields two connected components, and their vertex sets are connected communities. If the graph has a maximum degree at most 3, a greedy algorithm keeps decreasing the size of a cut under some constraints. The final cut results in two connected components whose vertex sets are communities.

To the best of our knowledge, no graphs were known to not have a (connected) 2-community structure. In the next section, we present an infinite family of graphs without 2-community structure.

4 Graphs without 2-community structure

The question if there exists a graph without a 2-community structure was originally introduced in [2]. In this section we present the infinite class $\mathcal{G}$ (see Definition 3) of graphs with even number of vertices without a 2-community structure. Based on our computer based verification, it seems that the graphs of minimum order without 2-community structure have 10 vertices – we are aware of 4 such (connected) graphs and they all belong to the class $\mathcal{G}$. Moreover, with the exception of these 4 graphs, all graphs with at most 10 vertices have a connected 2-community structure. On the other hand, we are aware of graphs with 11 vertices without a connected 2-community structure, but with a (disconnected) 2-community structure.

In the following we describe an infinite family of graphs without a 2-community structure.

Definition 3. Let $\mathcal{G}$ be the class of the graphs such that for $G = (V, E) \in \mathcal{G}$:

- $V = W_1 \cup W_2 \cup \{w, x, y, z\}$, where W_1, W_2 are cliques of the same size k, $k \geq 3$, and $\{x, y, w\}$ is a clique of size 3;
- w is adjacent to all vertices in $W_1 \cup W_2$;
- z is a pendant vertex adjacent to the vertex y;
- $1 \leq d_{W_1}(x) = d_{W_2}(x) \leq k-1$ and $2 \leq d_{W_1}(y) = d_{W_2}(y) \leq k-1$;
- $|W_i \cap (N(x) \cup N(y))| > \frac{3k}{k+3}$ for each $i \in \{1, 2\}$, and furthermore there exist vertices $\alpha, \beta \in W_1 \cup W_2$ such that $\alpha \in N(y) \setminus N(x)$, and $\beta \in N(x) \cap N(y)$;
- there is no edge between the vertex sets W_1 and W_2.

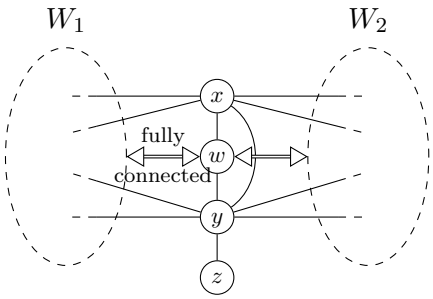


Figure 1: A schematic representation of a graph in $\mathcal{G}$.

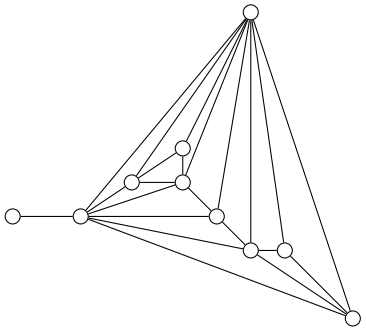


Figure 2: A planar graph with 10 vertices without a 2-community structure.

Theorem 1. *Each graph in $\mathcal{G}$ does not have a 2-community structure.*

Interestingly, there exists a planar graph in $\mathcal{G}$, hence without a 2-community structure, (see Figure 2).

5 Conclusion and further work

The definition we use to describe a community structure is interesting because each vertex has to satisfy a condition depending not only on its degree but also on the size of the communities. The

removal or addition of a vertex in a part of the partition affects all the vertices of the two parts, as the sizes of both parts are updated. This makes the problem particularly interesting to work with, and complex from an algorithmic point of view.

Thanks to our counterexamples, we get a new insight into the problem, and we hope to find some necessary structural properties to characterize graphs that do not have a 2-community structure. The main open question is the complexity of the 2-community structure problem which is unknown for both connected and disconnected variants. Other community related problems could also be studied, and the results may yield to a better understanding of the inherent complexity of using such a definition.

References

[1] C. Bazgan, Z. Tuza, and D. Vanderpooten. The satisfactory partition problem. *Discrete Applied Mathematics*, 154(8):1236–1245, 2006.

[2] C. Bazgan, J. Chlebíková, and T. Pontoizeau. Structural and algorithmic properties of 2-community structures. *Algorithmica*, 80(6):1890–1908, 2018.

[3] V. Estivill-Castro and M. Parsa. On connected two communities. In *Proceedings of the 36th Australasian Computer Science Conference (ACSC 2013)*, pages 23–30, 2013.

[4] S. Fortunato. Community detection in graphs. *Physics Reports*, 486:75–174, 2010.

[5] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1979.

[6] M. Gerber and D. Kobler. Partitioning a graph to satisfy all vertices. *Technical report,Swiss Federal Institute of Technology*, 1998.

[7] O. Goldschmidt and D. S. Hochbaum. Polynomial algorithm for the k-cut problem. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science (FOCS 1988)*, pages 444–451, 1988.

[8] M. Olsen. A general view on computing communities. *Mathematical Social Sciences*, 66(3): 331–336, 2013.