

Modular Graphs with High Cheeger Constant

Krunal K Patel (krunalkachadia@gmail.com)

Abstract- Modularity and centrality properties in graphs is an area of research that is becoming increasingly important with growing applications in analysing big data, especially social network data, computation for biology, communication network optimization and design, VLSI, etc. to name a few. We want the network to have community structure. At the same time we also want a network to be decently connected so that it can withstand certain amount of failures. We can measure how well a network is connected by its Cheeger constant. In this work in our search for ways to test the modularity properties in networks we focused on possible relationships between the Cheeger constant of a graph and modularity. Although modularity is a local graph property and Cheeger constant is a global graph property, it turns out that there is a relation between them. Low Cheeger constant means that the graph is weakly connected. Such graphs are more likely to have community structure. A paper by Gregory Gauthier on "Graph Fortresses and Cheeger Values" proves that any network having Cheeger constant less than 1 has positive modularity. They introduced graph fortress and cellular automaton to prove the result. First they proved that any graph having Cheeger constant less than 1 will have a double fortress in it. Subsequently, another result of theirs states that any graph with non-trivial double fortress will have a partition with positive modularity. In this document we first prove that any minimal Cheeger set is always internally connected. This gives us a lower bound on number of edges in a minimal Cheeger set with respect to its cardinality. We then prove that if a graph has two disjoint weak fortress then the graph will also have a double weak fortress. Using these two results we prove that if all minimal Cheeger set having Cheeger value ϵ have cardinality at least $2/2-\epsilon$ then the graph has a non-trivial double weak fortress. Finally we prove that the presence of a non-trivial double weak fortress gives us a partition of vertices of graph with non-negative modularity.

Index Terms- Cheeger Constant, Cheeger set, Cheeger value, Graphs, Graph fortress, Modularity, Connectivity

1 INTRODUCTION

Modularity and centrality properties in graphs is an area of research that is becoming increasingly important with growing applications in analysing big data, especially social network data, computation for biology, communication network optimization and design, VLSI, etc. to name a few.

We want the network to have community structure. At the same time we also want a network to be decently connected so that it can withstand certain amount of failures. We can measure how well a network is connected by its Cheeger constant (see Section 2).

In this work in our search for ways to test the modularity properties in networks we focused on possible relationships between the Cheeger constant of a graph and modularity. Although modularity is a local graph property and Cheeger constant is a global graph property, it turns out that there is a relation between them. Low Cheeger constant means that the graph is weakly connected. Such graphs are more likely to have community structure.

A paper by Gregory Gauthier on "Graph Fortresses and Cheeger Values"[2] proves that any network having Cheeger constant less than 1 has positive modularity. They introduced graph fortress and cellular automaton to prove the result(see the Section \ref{Definitionsec}). First they proved that any graph having Cheeger constant less than 1 will have a double fortress in it.

Theorem 1.1 Let G be a finite graph with $h_G < 1$ or $h_G = 1$ and $d_v \geq 2$ for all $v \in V(G)$ then G has a non-trivial double fortress.

Subsequently, another result of theirs states that any graph with non-trivial double fortress will have a partition with positive modularity.

Theorem 1.2 Let G be a finite graph, and $F \subseteq V(G)$ a double fortress. Then the modularity of the G based on F is nonnegative, with the modularity being zero only if F is trivial or if each vertex has the same number of neighbors in its own partition as in the other partition.

We observed that in the proof of second result, the conditions are stronger than needed. A double fortress is same as strong community. Even a graph possessing double weak fortress (weak community)

will also have a partition with positive modularity and hence the community structure. For example the graph in the Figure 1 has double weak fortress but no double fortress in it.

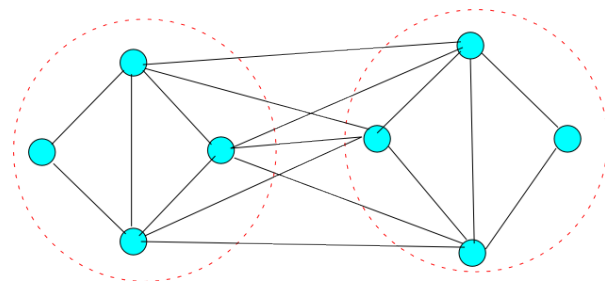


Fig1. A small network with weak community and no strong community

Since we weakened the conditions in the second result, we attempted to get better conditions in first result. i.e. we now only need the graph to have a weak double fortress instead of double fortress. It turns out that a graph with Cheeger constant up to 2 with certain cardinality conditions on minimal Cheeger set(see the definitions section Section 2) will have a double weak fortress. Also we don't need to use Cellular Automaton to prove our results. Our results are mainly based on the properties of weak fortress (weak community) and minimal Cheeger sets.

In this document we first prove that any minimal Cheeger set is always internally connected. This gives us a lower bound on number of edges in a minimal Cheeger set with respect to its cardinality. We then prove that if a graph has two disjoint weak fortress then the graph will also have a double weak fortress. Using these two results we prove that if all minimal Cheeger set having Cheeger value ϵ have cardinality at least $\frac{2}{2-\epsilon}$ then the graph has a non-trivial double weak fortress. modularity.

IJSER

Finally we prove that the presence of a non-trivial double weak fortress gives us a partition of vertices of graph with non-negative modularity.

2. NOTATIONS AND DEFINITIONS

2.1. NOTATIONS

We will use the following notations for this document.

- $G \rightarrow$ Finite Simple Graph
- $V(G) \rightarrow$ Vertex set of graph G
- $E(G) \rightarrow$ Edge set of graph G
- $A^c \rightarrow$ Complement of set A
- $N \rightarrow |V(G)|$
- $\delta(A, B) \rightarrow$ Number of edges with one end point in A and the other in B .
Where A and B are disjoint subsets of $V(G)$
- $d_v \rightarrow$ Degree of vertex v

2.2. DEFINITIONS

Among many similar or equivalent sets of definitions, we have mainly followed [1].

Definition If A is a nonempty proper subset of $V(G)$, then we define the *Cheeger Value* of A as

$$h_G(A) = \frac{\delta(A, A^c)}{\min(|A|, |A^c|)} \quad (1)$$

We note that $h_G(A) = h_G(A^c)$.

Definition *Cheeger Constant* of G is defined as $h_G = \min h_G(A)$, ($\emptyset \subset A \subset V(G)$)

Definition If A is a nonempty proper subset of $V(G)$, then A is *minimal cheeger set* if for all $\emptyset \subset B \subset A$, $h_G(B) > h_G(A)$.

Modularity[3]

Modularity is a concept in network theory that describes how well a graph can be broken into individual strongly-connected modules. M. E. J. Newman quantifies the modularity Q as follows:

$$Q = \frac{1}{4m} \sum_{i,j \in V(G)} \left(A_{ij} - \frac{d_i d_j}{2m} \right) s_i s_j \quad (2)$$

where A is the adjacency matrix and s_i is +1 or -1 depending on the sign of vertex i . m is the number of edges in G .

The motivation is to indicate partitions of graphs in which there are more intramodular edges and fewer intermodular edges than would be expected. A positive value indicates a clear break between modules as indicated by the partition, while a negative value indicates a bad choice of partition. Further, the trivial partition where all the vertices are in one partition has zero modularity, which allows zero to be the baseline.

Definition If A is a subset of $V(G)$, then A is a *fortress (strong Community)* if for all $v \in A$, $\delta(\{v\}, A - \{v\}) \geq \delta(\{v\}, A^c)$. A is a *double fortress* if both A and A^c are fortresses.

Definition

If A is a subset of $V(G)$, then A is a *weak fortress (weak Community)* if $\sum_{v \in A} \delta(\{v\}, A - \{v\}) \geq \sum_{v \in A} \delta(\{v\}, A^c) = \delta(A, A^c)$. A is a *double weak fortress* if both A and A^c are weak fortresses.

IJSER

Part 2:

Given that A and B are two disjoint weak fortresses in G, let

$$\delta(A, A^c) = e_1 \text{ and } \delta(B, B^c) = e_2.$$

Since $A \cap B = \emptyset \Rightarrow A \subset B^c$.

Without loss of generality let $e_1 \geq e_2$.

$$\begin{aligned} & \sum_{v \in B^c} \delta(\{v\}, B^c - \{v\}) \\ & \geq \sum_{v \in A} \delta(\{v\}, A - \{v\}) \geq \delta(A, A^c) = e_1 \\ & > e_2 \\ & = \delta(B^c, B) \end{aligned}$$

$\Rightarrow B^c$ is a weak fortress and B is a required non-trivial double weak fortress.

Theorem 3.3 If G is a finite graph with cheeger constant $\epsilon < 2$ with all minimal cheeger sets having cheeger value ϵ have cardinality at least $\frac{2}{2-\epsilon}$ then the graph has a non-trivial weak fortress and a non-trivial double weak fortress.

Proof:

Let A be a minimal cheeger set with smallest cardinality having cheeger value ϵ .

Let $|V(G)| = N$
 $|A| = n \Rightarrow n \leq N/2$.

$$h_G(A) = \frac{\delta(A, A^c)}{n} = \epsilon$$

By Theorem 3.1 A is internally connected.

\Rightarrow Number of internal edges in A $\geq n - 1$

$$\Rightarrow \sum_{v \in A} \delta(\{v\}, A - \{v\}) \geq 2n - 2$$

and $\delta(A, A^c) = n\epsilon$

also $n \geq \frac{2}{2-\epsilon}$

$$\begin{aligned} & \Rightarrow 2n - n\epsilon \geq 2 \\ & \Rightarrow 2n - 2 \geq n\epsilon \end{aligned}$$

$$\Rightarrow \sum_{v \in A} \delta(\{v\}, A - \{v\}) \geq \delta(A, A^c)$$

Hence A is a non-trivial weak fortress.

Also note that $h_G(A) = h_G(A^c) = \epsilon$. Hence $\exists B \subseteq A^c$ such that $h_G(B) = \epsilon$ and B is a minimal cheeger set. Let $|B| = m$.

Case 1:

$$m \geq N/2 \geq n$$

$$\begin{aligned} & \sum_{v \in A^c} \delta(\{v\}, A^c - \{v\}) \\ & \geq \sum_{v \in B} \delta(\{v\}, B - \{v\}) \geq 2m - 2 \geq 2n - 2 \\ & \geq n\epsilon \\ & = \delta(A^c, A) \end{aligned}$$

$\Rightarrow A^c$ is a weak fortress and A is a required non-trivial double weak fortress.

Case 2:

$m < N/2$

\Rightarrow Number of internal edges in B $\geq m - 1$

$$\Rightarrow \sum_{v \in B} \delta(\{v\}, B - \{v\}) \geq 2m - 2$$

and $\delta(B, B^c) = m\epsilon$

also $m \geq \frac{2}{2-\epsilon}$

$$\begin{aligned} & \Rightarrow 2m - m\epsilon \geq 2 \\ & \Rightarrow 2m - 2 \geq m\epsilon \end{aligned}$$

$$\Rightarrow \sum_{v \in B} \delta(\{v\}, B - \{v\}) \geq \delta(B, B^c)$$

Hence B is a non-trivial weak fortress and clearly $A \cap B = \emptyset$. Hence by Lemma 3.2 G has a non-trivial double weak fortress.

4. THE RELATION TO MODULARITY

From the definition, weak fortresses are strongly connected subsets of vertices. It should come as no surprise that double weak fortresses are a good choice for partitioning vertices into modules. The following theorem illustrates this property.

Theorem 4.1 Let G is a finite graph and $F \subseteq V(G)$ be a double weak fortress. Then the modularity of G based on F is nonnegative with modularity being zero if and only if F is trivial or sum of all internal degrees is equal to sum of all external degrees in F or F^c .

Proof:

For any $v \in A \subseteq V(G)$ define advantage of v in A by $adv(v) = \delta(\{v\}, A - \{v\}) - \delta(\{v\}, A^c)$. Further if $B \subseteq A$ then $adv(B) = \sum_{v \in B} adv(v)$. Note that if A is a weak fortress then $adv(A) \geq 0$.

Rewriting equation (2)

$$Q = \frac{1}{4m} \sum_{i,j \in V(G)} \left(A_{ij} S_i S_j - \frac{d_i d_j S_i S_j}{2m} \right) \quad (10)$$

Fix an arbitrary $i \in V(G)$. Let $i \in F$ (alternatively $i \in F^c$). Then the first term corresponds to how many more neighbors of i are in F than F^c (alternatively, in F^c than F). But this is just $adv(i)$. Summing over all $i \in F$ and $i \in F^c$ gives

$$Q = \frac{1}{4m} \left(adv(F) + adv(F^c) - \frac{d_i d_j S_i S_j}{2m} \right) \quad (11)$$

Now let $X = \sum_{v \in F} d_v$ and $Y = \sum_{u \in F^c} d_u$. Hence $X + Y = 2m$ and $X - Y = \sum_{v \in F} d_v - \sum_{u \in F^c} d_u$.

Hence $(X-Y)^2 = \sum_{i,j \in V(G)} d_i d_j S_i S_j$

$$Q = \frac{1}{4m} \left(adv(F) + adv(F^c) - \frac{(X - Y)^2}{2m} \right) \quad (12)$$

Let $Z = \delta(F, F^c)$ and for each $v \in F$

$$d_v = \delta(\{v\}, F - \{v\}) + \delta(\{v\}, F^c)$$

$$adv(v) = \delta(\{v\}, F - \{v\}) - \delta(\{v\}, F^c)$$

$$\Rightarrow \delta(\{v\}, F^c) = \frac{1}{2} (d_v - adv(v))$$

sum over all $v \in F$

$$\Rightarrow Z = \frac{1}{2} (X - adv(F))$$

similarly for all $v \in F^c$

$$\delta(\{v\}, F) = \frac{1}{2} (d_v - adv(v))$$

sum over all $v \in F^c$

$$\Rightarrow Z = \frac{1}{2} (Y - adv(F^c))$$

$$\Rightarrow Z = \frac{1}{2} (X - adv(F)) = \frac{1}{2} (Y - adv(F^c))$$

$$\Rightarrow X - Y = adv(F) - adv(F^c)$$

Since F and F^c are weak fortresses, $adv(F)$ and $adv(F^c)$ are non negative.

$$adv(F) + adv(F^c) \geq |adv(F) - adv(F^c)| \quad (13)$$

Equality holds if and only if either advantage is zero i.e. F is trivial or sum of all internal degrees is equal to sum of all external degrees in F or F^c

Also

$$2m = X + Y \geq |X - Y| = |adv(F) - adv(F^c)| \quad (14)$$

with equality holding if and only if X or Y is zero.

by equation (13) and (14)

$$2m(adv(F) + adv(F^c)) \geq (adv(F) - adv(F^c))^2$$

$$= (X - Y)^2 \quad (15)$$

hence by equation (12)

$Q \geq 0$ with equality holding if and only if F is trivial or sum of all internal degrees is equal to sum of all external degrees in F or F^c .

5. CONCLUSION

When study of connectivity and expansion in a sparse graph is combined with study of modularity in a graph we get some rewarding results. Apparently, the two properties, though not completely orthogonal nor diametrically opposite, are at variance with each other. However, surprising connections are exposed once we delve deeper into their structure and dynamics. It turns out that the Cheeger constant is too much of an aggregate global parameter, compressing too much information into a single number, losing structural information in the process.

Work by Sitabhra Sinha et al.[4] indicates that modularity is more ubiquitous in naturally arising networks than previously recognized. Our results strengthen this work by showing that indeed, graphs with high Cheeger constant also may be modular. Our results can be extended and developed in several directions. Investigating whether there is an absolute or parametric upper bound on the Cheeger constant of the graph for it to have modularity will be an interesting abstract problem. Our first result, that nontrivial minimal Cheeger sets must be internally connected, may have wider applications than study of modularity. Moreover, the particular technique is a contribution. What is the quantitative relation between bounds on the Cheeger constant and on the minimum cardinality of a minimal Cheeger set for modularity is also an interesting question.

The techniques used in this work are mainly combinatorial, but the use of finiteness and discreteness is minimal, and exploring how or when these results are applicable to infinite graphs or even uncountable relations (connecting the Cheeger numbers and inequalities to their original form -- isoperimetric inequalities on Riemannian manifolds) can be very productive. To some extent, Gauthier[2] explores this.

We have not explored this here, though our techniques should be easily extensible to the infinite case.

ACKNOWLEDGEMENT

I would like to thank the IJSER team for publishing my work. I would also like to thank Mr. Ramprasad Joshi for his guidance.

REFERENCES

- [1] F. R. K. Chung. Spectral Graph Theory. American Mathematical Society, 1997.
- [2] Gregory Gauthier. "Graph Fortresses and Cheeger Values". 2010. url: <http://www3.canisius.edu/~kahngb/reu/papggauthier.pdf>.
- [3] M. E. J. Newman. "Modularity and Community Structure in Networks". In: Proceedings of the National Academy of Sciences of the United States of America 103 (2006), pp. 8577–8582.
- [4] S. Sinha. "The importance of community: How modular organization of social networks affects their collective dynamics". In: Studies in Microeconomics 2 (2014), pp. 49–61.

-----End of Paper-----