Characterizing Combinatorially Max-Supergeometric Trees

Amir Daneshgar[†] and Ramin Javadi[‡]

Department of Mathematical Sciences

[†] Sharif University of Technology
http://www.sharif.ir/~daneshgar
daneshgar@sharif.ir

[‡] Isfahan University of Technology
http://rjavadi.iut.ac.ir
rjavadi@cc.iut.ac.ir

Shahrivar 1392 (September, 2013)

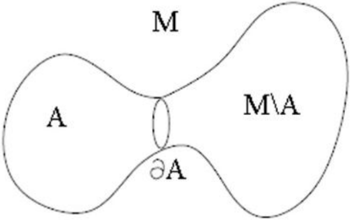
Cheeger constant of a Riemannian Manifold

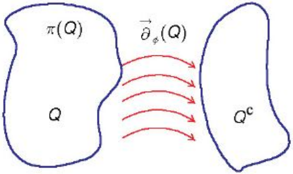
Cheeger constant of a (compact) n-dimensional Riemannian manifold G:

$$\xi_2^M(G) \stackrel{\text{def}}{=} \inf_A \max \left\{ \frac{\mu_{n-1}(\partial A)}{\mu_n(A)}, \frac{\mu_{n-1}(\partial A)}{\mu_n(A^c)} \right\}$$

A runs over open subsets of M.

μ_n: n-dimensional measure, ∂A: the boundary of A



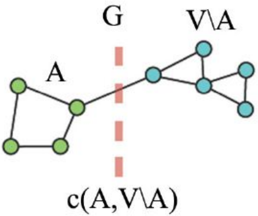


The case of simple graphs

For a simple graph $G = (V, E)$:
The max version: (Cheeger constant or edge expansion)

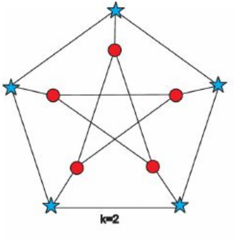
$$\varsigma_2^M(G) \stackrel{\text{def}}{=} \min_{A \subseteq V(G)} \max \left\{ \frac{|E(A, A^c)|}{|A|}, \frac{|E(A, A^c)|}{|A^c|} \right\}$$

2-Isoperimetry Problem: Finding a 2-partition (A, A^c) of $V(G)$ attaining the edge expansion of G.



Define $\overrightarrow{E}(A) \stackrel{\text{def}}{=} E(A, A^c)$.

2-isoperimetry (partitioning): examples



The Petersen graph P

Let the induced graph on A be connected and contain a cycle. Therefore, $\varsigma_2^M(P) = 1$.

Can we justify this reasoning?

Weighted graphs

Model: (A finite weighted graph) A simple graph $G = (V, E)$ together with two weight functions $w : V \to \mathbb{Q}^+$ and $c : E \to \mathbb{Q}^+$.

Notations: For every $x \in V$ and $A, B \subseteq V$,

$$deg(x) \stackrel{\text{def}}{=} \sum_{y \sim x} c(xy).$$

$$E(A, B) \stackrel{\text{def}}{=} \{e = uv \in E : \ u \in A, v \in B\},$$

$$w(A) \stackrel{\text{def}}{=} \sum_{u \in A} w(u), \quad c(A) \stackrel{\text{def}}{=} \sum_{e \in E(A, A^c)} c(e).$$

For the case of weighted graphs with potentials see [R. JAVADI PHD THESIS 2011].

A naive generalization: the normalized cut problem

Let $\mathcal{P}_k(V)$ be the set of k-partitions of V. Given a weighted graph $G = (V, E, c, w)$ and an integer k ($2 \leq k \leq |V|$), find a k-partition $(A_1, \ldots, A_k)$ that gives rise to equality in the following:

A naive generalization of Cheeger's constant (a $\|.\|_\infty$ version):

$$\tilde{\iota}_k^M(G) \overset{\text{def}}{=} \min_{\{A_i\}_1^k \in \mathcal{P}_k(V)} \max_{1 \leq i \leq k} \frac{c(A_i)}{w(A_i)}.$$

We use the acronym NCP for the corresponding problems.

The combinatorial case

The combinatorial case

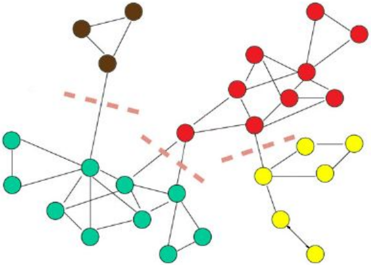
The combinatorial case is when the weight functions c and w are constant and equal to 1, i.e. when we are dealing with simple graphs.

Note that in this case,

$$\tilde{\varsigma}_k^M(G) \stackrel{\text{def}}{=} \min_{\{A_i\}_1^k \in \mathcal{P}_k(V)} \max_{1 \le i \le k} \frac{|\overrightarrow{E}(A_i)|}{|A_i|}.$$

An example

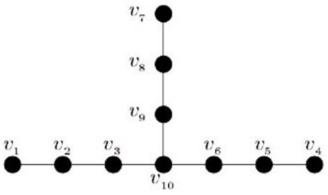
All edge and vertex weights are equal to 1, $k = 4$.



$$\tilde{\varsigma}_4^M \leq \max\{\frac{1}{3}, \frac{3}{10}, \frac{3}{9}, \frac{1}{6}\} = \frac{1}{3}.$$

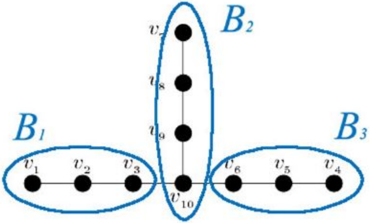
Another example

(All edge and vertex weights are equal to 1.)



Another example

(All edge and vertex weights are equal to 1.)



$$\tilde{\varsigma}_3^M(G) \leq \max(\tfrac{1}{3}, \tfrac{2}{4}, \tfrac{1}{3}) = \tfrac{1}{2}.$$

Hardness of NCP

Convention

NCP stands for the Normalized Cut Problem. Also,

Subscript k: appears when k is a constant, disappears when k is part of the input.

Example:

NCP_k^M :

CONSTANTS: An integer k.
INPUTS: A weighted graph $G = (V, E, w, c)$ and a positive integer N.
QUERY: Is it true that $\tilde{\iota}_k^M(G) \leq N$?

[J. AND D. 2010]

NCP_k^M is NP-complete for simple graphs.

NCP^M is NP-complete for simple trees.

Approximating NCP: real relaxation

What about approximations?

[Hajiabolhassan, D. 2008]

$\tilde{\iota}_k^M$ does not admit a real relaxation!

The isoperimetric constants

Relaxing the definition!

A k-subpartition consists of k nonempty and disjoint subsets of $V(G)$. Let $\mathcal{D}_k(G)$ be the class of all k-subpartitions of $V(G)$. Define the kth (Max) isoperimetric constants of G as,

$$\iota_k^M(G) \stackrel{\text{def}}{=} \min_{\{A_i\}_1^k \in \mathcal{D}_k(V)} \max_{1\leq i\leq k} \frac{c(A_i)}{w(A_i)},$$

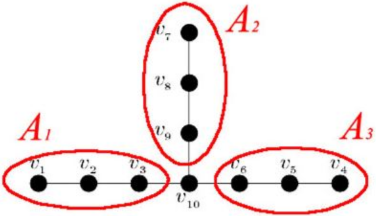
and the combinatorial version as

$$\varsigma_k^M(G) \stackrel{\text{def}}{=} \min_{\{A_i\}_1^k \in \mathcal{D}_k(V)} \max_{1\leq i\leq k} \frac{|\vec{E}(A_i)|}{|A_i|}.$$

We use the acronym IPP for the corresponding problems.

Another example

(All edge and vertex weights are equal to 1.)



$$\varsigma_3^M(G) \le \max(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}) = \tfrac{1}{3}.$$

The gradient operator

Let $\mathcal{F}_w(G)$ and $\mathcal{F}_c(G)$ be the set of all real functions on $V(G)$ and $E(G)$, respectively, equipped with the corresponding weighted inner-products. Define the gradient as

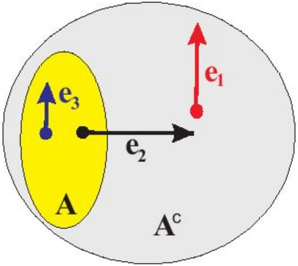
$$\nabla : \mathcal{F}_w(G) \longrightarrow \mathcal{F}_c(G), \quad \nabla f(uv) \stackrel{\text{def}}{=} f(v) - f(u).$$

Gradient of characteristic functions

If $f = \frac{1}{w(A)}\chi_A$ is the normalized characteristic function of a subset $A \subseteq V(G)$ then

$$\|\nabla f\|_{1,c} = \frac{c(A)}{w(A)} = \frac{\|\nabla \chi_A\|_{1,c}}{\|\chi_A\|_{1,w}}.$$

The gradient operator: Figure



Let f be the characteristic function of a subset $A \subset V$. Then,

$$\nabla f(e_1) = \nabla f(e_3) = 0, \quad \nabla f(e_2) = -1.$$

Also, note that in this case,

$$\|\nabla f\|_1 = |E(A, A^c)|, \text{ and } \|f\|_1 = |A|.$$

A real relaxation of parameters

Define,

$$\mathcal{O}_k^+(G) \stackrel{\text{def}}{=} \left\{ \{f_i\}_1^k \;\middle|\; \{f_i\}_1^k \text{ is positive orthonormal} \right\},$$

$$\tilde{\mathcal{O}}_k^+(G) \stackrel{\text{def}}{=} \left\{ \{f_i\}_1^k \in \mathcal{O}_k^+(G) \;\middle|\; \{\text{supp}(f_i)\}_1^k \in \mathcal{P}_k(G) \right\}.$$

and the relaxed parameters,

$$\gamma_k^M(G) \stackrel{\text{def}}{=} \inf_{\{f_i\}_1^k \in \mathcal{O}_k^+(G)} \max_i (\|\nabla f_i\|_{1,c}),$$

$$\tilde{\gamma}_k^M(G) \stackrel{\text{def}}{=} \inf_{\{f_i\}_1^k \in \tilde{\mathcal{O}}_k^+(G)} \max_i (\|\nabla f_i\|_{1,c}).$$

Justifications for definitions

[Hajiabolhassan, J. and D. 2010]

For both max and mean versions, $\gamma_k(G) = \tilde{\gamma}_k(G) = \iota_k(G)$.

The intrinsic inequality

By definitions, in general, we have $\iota_k(G) \leq \tilde{\iota}_k(G)$, where the inequality can be strict (in both maximum and mean versions)!

To the best of our knowledge, the correctness of definitions for subpartitions has been first indipendently observed in [Miclo 2007], [Hajiabolhassan, D. 2008], and [Helffer, T. Hoffmann-Ostenhof, Terracini 2008].

Justifications for definitions

Test function approximation

The equality $\gamma_k(G) = \tilde{\gamma}_k(G) = \iota_k(G)$ shows that $\iota_k(G)$ can be effectively approximated by test functions.

Subpartitions are richer

Computationally, a move from partitions to subpartitions usually makes the problem easier!
(e.g. the polynomial time algorithm for minimum k-subpartition problem [NAGAMOCHI, KAMIDOI 2007]).

Subpartition residues [SHARIATRAZAVI, J. AND D. 2011]

There is evidence supporting the fact that subpartition residues contain nontrivial information. Hence, the subpartition setup makes it possible to gain more information in an easier way!

Hardness of IPP

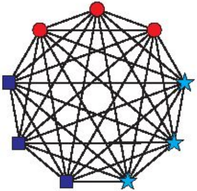
[J. and D. 2010]

IPP^M is polynomial (actually linear) time solvable for weighted trees (even with potentials)!.

[Shariatrazavi, J. and D. 2011]

There exists an algorithm that given a weighted tree with rational weights (and potentials!) on n vertices and an integer k, computes ι_k^M and a minimizer in $(n \log n)$-time.

The complete graph K_n

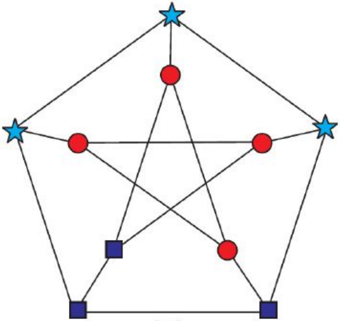


Note that for every set $A_i \subset V$,

$$\frac{|\overrightarrow{E}(A_i)|}{|A_i|} = \frac{|A_i|(n - |A_i|)}{|A_i|} = n - |A_i|.$$

Thus, $\varsigma_k^M(K_n) = \tilde{\varsigma}_k^M(K_n) = n - \lfloor \frac{n}{k} \rfloor$.

The Petersen graph P for $k = 3$



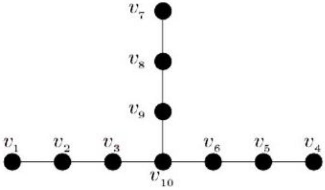
k=3

Note that at least one of the parts is acyclic. Hence,

$$\varsigma_3^M(P) = \tilde{\varsigma}_3^M(P) = 1 + \frac{2}{3}.$$

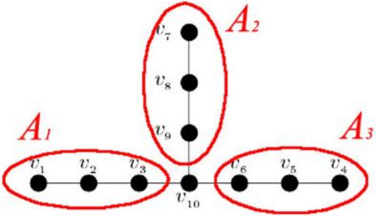
The case of trees

(All edge and vertex weights are equal to 1.)



The case of trees

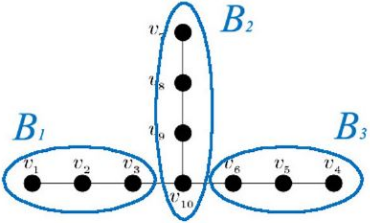
(All edge and vertex weights are equal to 1.)



$$\varsigma_3^M(G) = \max\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{1}{3}.$$

The case of trees

(All edge and vertex weights are equal to 1.)



$$\varsigma_3^M(G) = \max(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = \frac{1}{3}.$$

$$\tilde{\varsigma}_3^M(G) = \max(\frac{1}{3}, \frac{2}{4}, \frac{1}{3}) = \frac{1}{2}.$$

Definition

Combinatorially max-supergeometric graphs

A simple graph is said to be combinatorially max-supergeometric if

$$\forall\, 1 \leq k \leq |V| \quad \tilde{\varsigma}_k(G) = \varsigma_k(G).$$

Note

Hereafter, in this talk the word supergeometric stands for the phrase combinatorially max-supergeometric.

On supergeometric trees

Importance of the problem in theory

A move from a computationally hard problem NCP to a simpler one IPP is interesting by itself. Suergeometric graphs provide simple instances of the hard problem. Hence, studying such graphs may reveal properties that can simplify the original hard problem! Definitely, for the case of trees with an efficient solutions to IPP this can be considered as an effective solution to a subset of instances of the original hard problem!

On supergeometric trees

Importance of the problem in applications

Note that supergeometry is a property that depends on the weights!
Also, it is known that connectivity (and hence clustering) can be estimated through finding minimum spanning trees.
A crucial question is whether there exists an algorithm that given a weighted graph the algorithm extracts a supergeometric minimum spanning tree!
This is related to the existence of outliers and uniformity of data-set!

Small graphs

Remark

Note that $|A| = \{u\} = 1$ implies that $\frac{E(A,A^c)}{|A|} = \deg(u)$.

$|V| = 2$

K_2 is the only connected graph on two vertices and is supergeometric.

$|V| = 3$

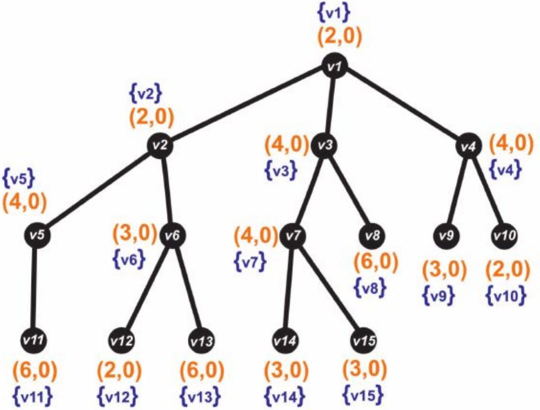
K_3 and P_3 are supergeometric.

$|V| = 4$

K_4, P_4, C_4 and S_4 (the star on four vertices) are supergeometric. The other two cases will be discussed on the board!

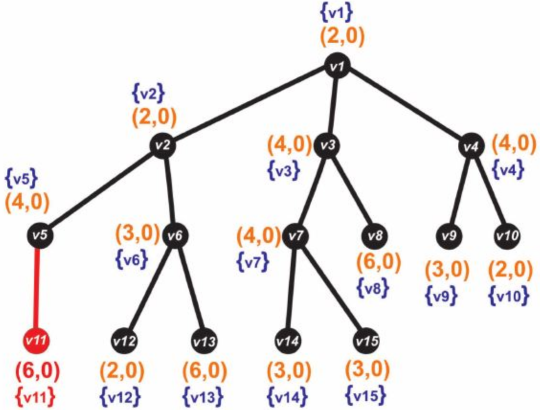
The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



The algorithm to decide IPP^M

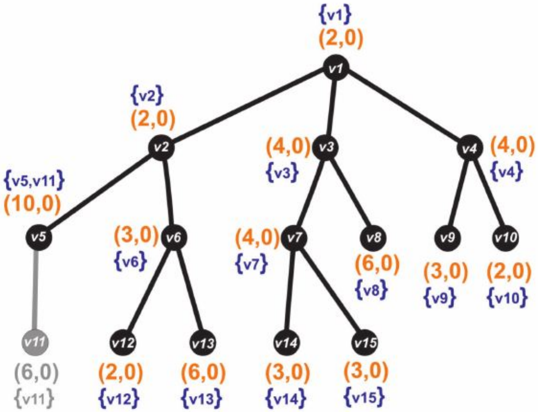
Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



$$\frac{\gamma(v_{11})+\varphi(e)}{\omega(v_{11})} = \frac{0+1}{6} > \frac{1}{10} \quad \text{but} \quad \frac{\gamma(v_{11})-\varphi(e)}{\omega(v_{11})} = \frac{0-1}{6} < \frac{1}{10}.$$

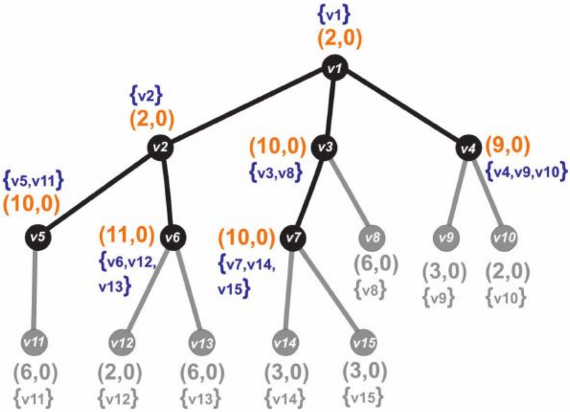
The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



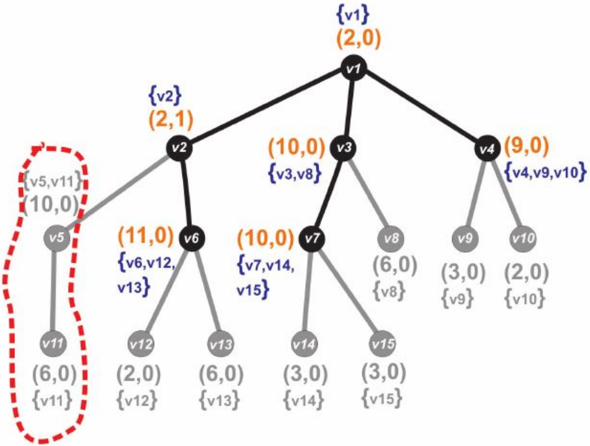
The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



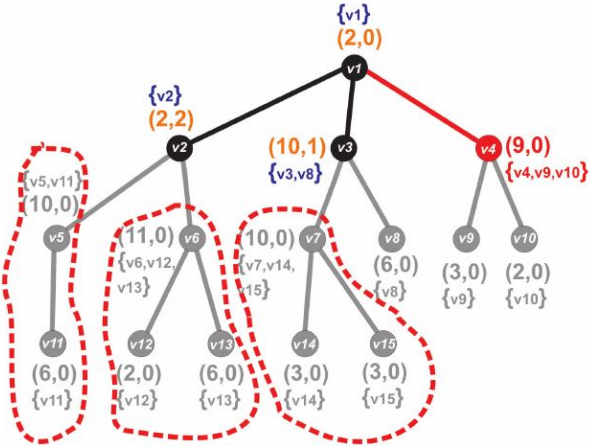
The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



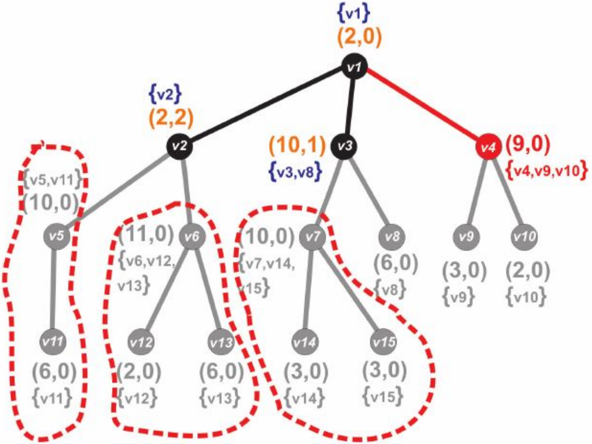
The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



The algorithm to decide IPP^M

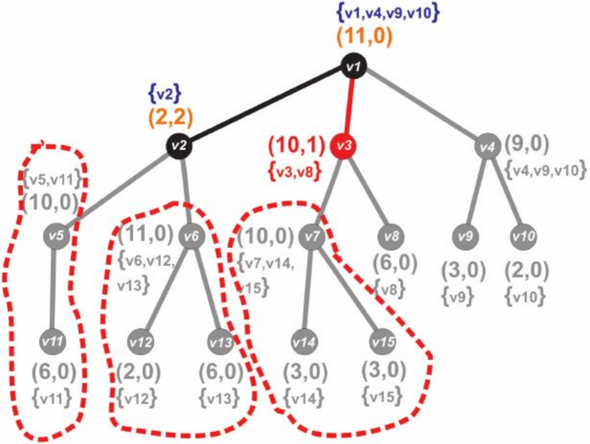
Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



$$\frac{\gamma(v_4)+\varphi(e)}{\omega(v_4)} = \frac{0+1}{9} > \frac{1}{10} \quad \text{but} \quad \frac{\gamma(v_4)-\varphi(e)}{\omega(v_4)} = \frac{0-1}{9} < \frac{1}{10}.$$

The algorithm to decide IPP^M

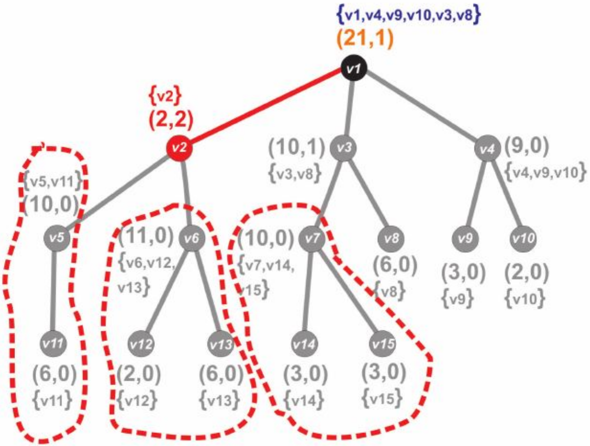
Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



$$\frac{\gamma(v_3)+\varphi(e)}{\omega(v_3)} = \frac{1+1}{10} > \frac{1}{10} \quad \text{but} \quad \frac{\gamma(v_3)-\varphi(e)}{\omega(v_3)} = \frac{1-1}{10} < \frac{1}{10}.$$

The algorithm to decide IPP^M

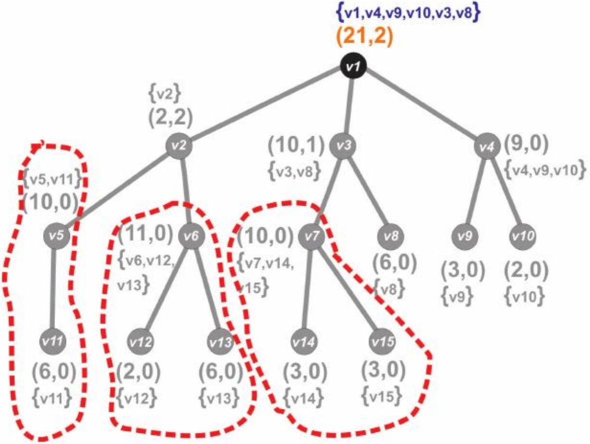
Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



$$\frac{\gamma(v_2)+\varphi(e)}{\omega(v_2)} = \frac{2+1}{2} > \frac{1}{10} \quad \text{and} \quad \frac{\gamma(v_2)-\varphi(e)}{\omega(v_2)} = \frac{2-1}{2} > \frac{1}{10}.$$

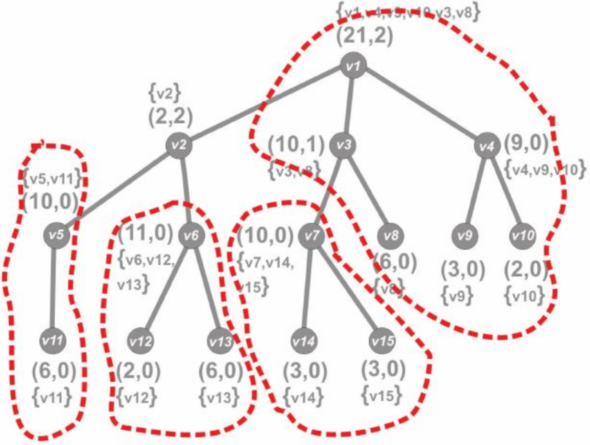
The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.



The algorithm to decide IPP^M

Edge weights=1, Vertex weights=(ω, p), $N = 1/10$ and $k = 4$.

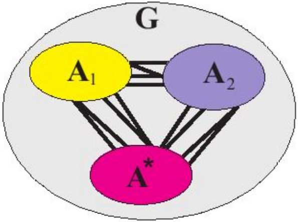


Bipartition

$\varsigma_2 = \tilde{\varsigma}_2$

Let $\{A_1, A_2\}$ be a minimizer for ς_2 and without loss of generality assume $|E(A_1, A^*)| \leq |E(A_2, A^*)|$, where $A^* \stackrel{\text{def}}{=} V - (A_1 \cup A_2)$. Then

$$\frac{|\overrightarrow{E}(A_2 \cup A^*)|}{|A_2 \cup A^*|} = \frac{|\overrightarrow{E}(A_2)| - |E(A_2, A^*)| + |E(A_1, A^*)|}{|A_2| + |A^*|} \leq \frac{|\overrightarrow{E}(A_2)|}{|A_2|}.$$



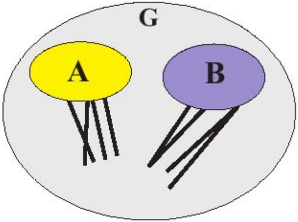
Connectedness of parts

[J. and D. 2010]

There always exists a minimizer of $\varsigma_k(G)$ such that the induced graph on any part is connected!

If there is no edge between A and B,

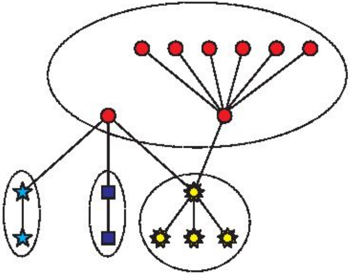
$$\min\left(\frac{|\vec{E}(A)|}{|A|}, \frac{|\vec{E}(B)|}{|B|}\right) \leq \frac{|\vec{E}(A)| + |\vec{E}(B)|}{|A| + |B|} = \frac{|\vec{E}(A \cup B)|}{|A \cup B|}.$$



An example

Remark

The above fact is also true for $\tilde{\varsigma}_3$ but not for $\tilde{\varsigma}_k$ when $K \notin \{2, 3\}$ in general!



$$\tilde{\varsigma}_4(T) = \varsigma_4(T) = \frac{1}{2}.$$

Please listen to the discussion!

An example

Relations to the degree sequence

Let $\delta \leq \cdots d \leq \Delta$ be the degree sequence of G. Then, if $\frac{\Delta+\delta}{2} > d+1$ then G is not $(|V|-1)$-geometric, and consequently, is not supergeometric.

Relmark

This show that S_5, a star on 5 vertices, is not supergeometric. Note that this is a minimal example of such a graph!

Relmark

This result can be generalized a bit further!

Supergeometric graphs: some special cases

[J.] The case of regular graphs

If G is a d-regular graph and $E(A)$ is the edge set of $G[A]$, then

$$\frac{|\vec{E}(A)|}{|A|} = \frac{d|A| - 2|E(A)|}{|A|} = d - 2\frac{|E(A)|}{|A|}.$$

Hence,

$$\varsigma_k(G) = d - 2 \max_{\{A_i\}_1^k \in \mathcal{D}_k(V)} \min_{1 \leq i \leq k} \frac{|E(A_i)|}{|A_i|}.$$

Thank you!

Comments and Criticisms are Welcomed

daneshgar@sharif.ir,
rjavadi@cc.iut.ac.ir.