

# PACE 2021-Cluster Editing Optimización Estocástica

Ivonne Monter Aldana  
ivonne.monter@cimat.mx

CIMAT

27 de mayo de 2021



Centro de  
Investigación  
en Matemáticas, A.C.

# Tabla de Contenido

## Introducción

- Cluster Editing

- Complejidad del problema

## Reto

## Método implementado

- Representación de las soluciones

- Data Reduction

- Casos especiales: Gráficas fuertemente conexas

- Búsqueda local con escala estocástica y evaluación incremental



Centro de  
Investigación  
en Matemáticas, A.C.

CIMAT

# Tabla de Contenido

## Introducción

Cluster Editing

Complejidad del problema

## Reto

## Método implementado

Representación de las soluciones

Data Reduction

Casos especiales: Gráficas fuertemente conexas

Búsqueda local con escala estocástica y evaluación incremental

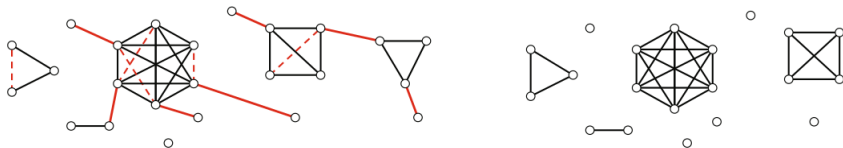


Centro de  
Investigación  
en Matemáticas, A.C.

CIMAT

# Cluster Editing

El problema de Cluster Editing consiste en transformar una gráfica en una unión disjunta de gráficas completas usando un mínimo número de modificaciones de aristas.



# Aplicaciones

- ▶ Detección a gran escala de genes homólogos.
- ▶ Modelar las interacciones entre proteínas.
- ▶ Detección de componentes patógenos.
- ▶ Análisis de redes sociales.



# Complejidad del problema

- ▶ En 1986 los matemáticos Mirko Křivánek y Jaroslav Morávek probaron que el problema de Cluster Editing es NP-hard.
- ▶ Para gráficas de a lo más grado dos, es decir caminos y círculos el problema se puede resolver en tiempo polinomial.



# Tabla de Contenido

## Introducción

Cluster Editing

Complejidad del problema

## Reto

## Método implementado

Representación de las soluciones

Data Reduction

Casos especiales: Gráficas fuertemente conexas

Búsqueda local con escala estocástica y evaluación incremental

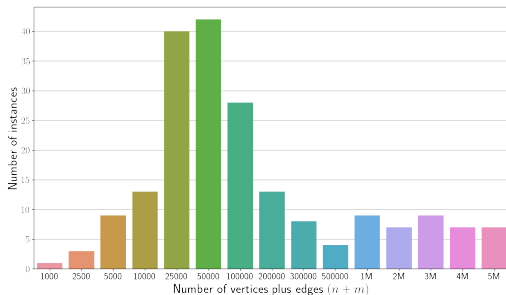


Centro de  
Investigación  
en Matemáticas, A.C.

CIMAT

# Objetivo

La tarea es encontrar la mejor solución para cada instancia en un tiempo límite de 10 minutos e imprimir la lista de aristas modificadas en menos de 30 segundos



CIMAT

Centro de  
Investigación  
en Matemáticas, A.C.



# Tabla de Contenido

## Introducción

Cluster Editing

Complejidad del problema

## Reto

## Método implementado

Representación de las soluciones

Data Reduction

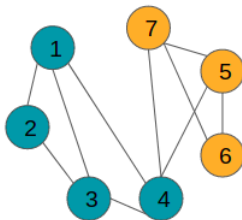
Casos especiales: Gráficas fuertemente conexas

Búsqueda local con escala estocástica y evaluación incremental



## Representación de las soluciones

- ▶ El número de clusters es a lo más el número de vértices de la gráfica.
- ▶ Dada una gráfica  $G(V, E)$  no dirigida, las soluciones al problema se representan como vectores en  $\mathbb{N}^{|V|}$ , donde cada componente pertenece al conjunto  $\{1, 2, \dots, |V|\}$ .
- ▶ La dimensión del espacio de búsqueda es  $V^{|V|}$ .



$s = (3, 3, 3, 3, 1, 1, 1)$

$(2, 4)$   
 $(4, 7)$   
 $(4, 5)$

## Data Reduction

- ▶ Dos vértices  $u$  y  $v$  son gemelos si  $v \sim u$  y  $N(v) = N(u)$ .
- ▶ La propiedad de ser gemelos es transitiva.
- ▶ La reducción consiste en compactar conjuntos de vértices gemelos entre sí a un solo vértice.

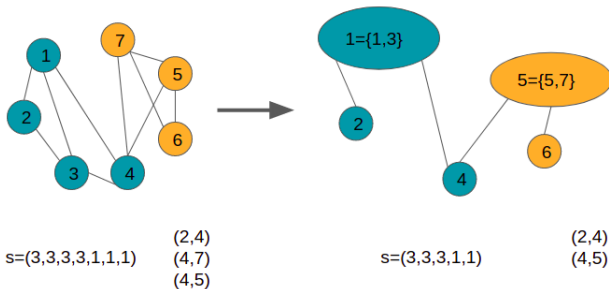
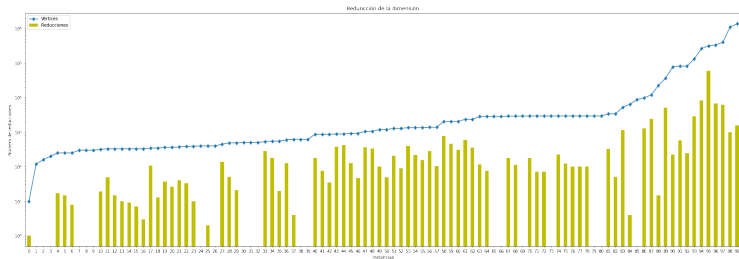


Figure: 5 y 7 son gemelos y también 1 y 3

# Gráfica de la reducción de dimensión



CIMAT

Centro de  
Investigación  
en Matemáticas, A.C.

## Casos especiales: Gráficas fuertemente conexas

- Algunas instancias corresponden a gráficas en el que el número de aristas es cercano al número de aristas de un gráfica completa.
- Si le faltan menos del 10% de aristas para ser una gráfica completa, entonces todos los vértices se colocan en un mismo cluster.

Instancia	$ E $	$ V  * ( V  - 1)/2$	Costo	Costo*
53	70027	79800	9773	8908
79	163863	191890	28027	27073
131	3353846	4005865	652019	652019
133	4005420	4005865	445	445
139	3515629	4238416	722787	722787
147	4160168	4308580	148412	148412
159	3533156	4317391	784235	784235
161	4295123	4317391	22268	22268

Table: Instancias fuertemente conexas



CIMAT

Centro de  
Investigación  
en Matemáticas, A.C.

## Búsqueda local

- ▶ Para cada componente conexa de la gráfica se implementó una búsqueda local con escalada estocástica y evaluación incremental.
- ▶ La vecindad de una solución consiste en todas las soluciones que difieren en un solo vértice.
- ▶ Para poder hacer la evaluación incremental, se observó que al cambiar un  $v$  vértice de cluster  $C1$  a un cluster  $C2$  se deben sumar y restar ciertas aristas:
  1. - Cada arista que se sumo al agregar  $v$  a  $C1$  debe ser restada del costo actual.
  2. + Cada arista original de  $v$  con el resto de vértices de  $C1$  debe ser sumada.
  3. - Por cada vértice en  $C2$  no adyacente a  $v$  se debe sumar una arista.
  4. + Cada arista original entre  $v$  y los vértices de  $C2$  debe ser restada.



## Resultados

