

# Desarrollo de una herramienta de visualización de redes sociales mediante un enfoque difuso

Natalia Marín Pérez

Tutor: Carlos González Alvarado

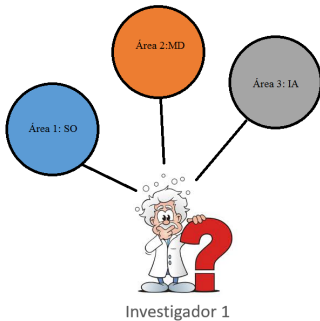
Instituto Tecnológico de Costa Rica

Noviembre 27,2017

# Tabla de contenidos

- 1 Introducción
  - Objetivos
- 2 Antecedentes
  - Clustering: fuzzy c-means
  - Análisis de redes sociales
- 3 Metodología
  - Análisis y pre-procesamiento de datos
  - Análisis de distancias
  - Algoritmo difuso
  - Características de la visualización
- 4 Validación
  - Visualización
- 5 Conclusión
- 6 Material de apoyo

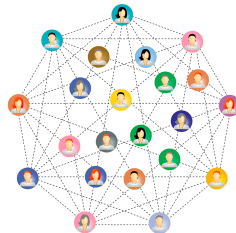
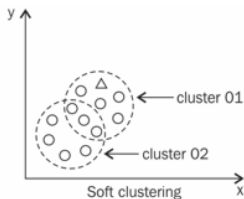
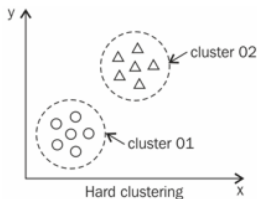
# Introducción



Áreas de investigación: Investigador 1



# Importancia de la difusidad en redes sociales



# Objetivos

Desarrollar una herramienta de visualización de redes sociales utilizando un enfoque difuso. Evaluando los datos de migración en Costa Rica, así como la representación de cómo interactúan las personas con la tecnología.

- Analizar y comparar diferentes distancias con el fin de escoger cuál se adapta mejor al problema
- Desarrollar una metodología que permita el análisis de clústeres difusos en una red social
- Implementar una herramienta que represente, de forma difusa, la conexión entre objetos en diferentes particiones
- Validar la herramienta con diferentes conjuntos de datos para medir su efectividad

# Algoritmo difuso: fuzzy c-means

## Fórmula fuzzy c-means

$$J_m(U, V) = \sum_{k=1}^N \sum_{i=1}^C u_{ij}^m \|Y_k - v_i\|^2$$

$Y = y_1, y_2, \dots, y_n \subset R^n = \text{datos},$

$c = \text{numero de clusteres en } Y; 2 \leq c < n,$

$m = \text{exponente en peso}; 1 \leq m < \infty,$

$U = \text{particion difusa de } Y; U \in M_{fc},$

$v = v_1, v_2, \dots, v_c = \text{vectores de centros},$

$v_i = (v_{i1}, v_{i2}, \dots, v_{in}) = \text{centro de cluster } i$

# Concepto de redes sociales y difusidad

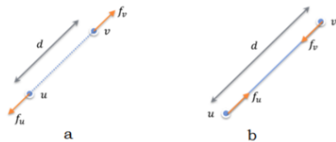


Figure: a. Fuerza de atracción nodo a nodo b. Fuerza de repulsión nodo a nodo

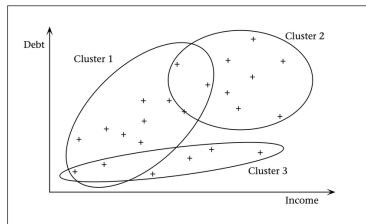


Figure: Ejemplo de representación de cluster difuso







## Caso 1: Datos de migración en Costa Rica (INEC)

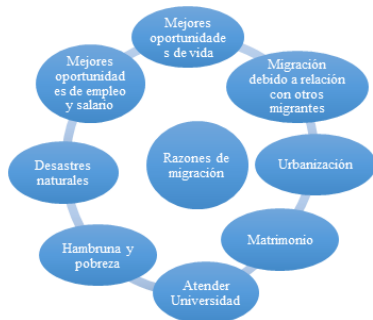
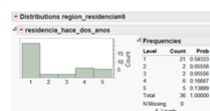
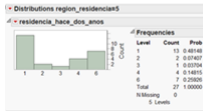
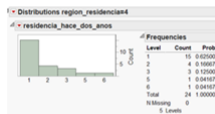


Table: Atributos a analizar

Atributos	
Región de residencia actual	Condición de actividad
Región residencia hace dos años	Título
Mantiene familia	Estado conyugal
Ingreso por persona neto	Ocupación

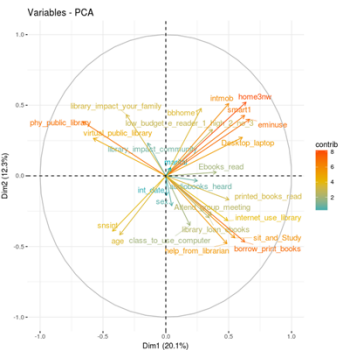
# Caso 1: Análisis de datos de migración en Costa Rica



## Caso 2: Análisis de datos de interacción de personas con la tecnología

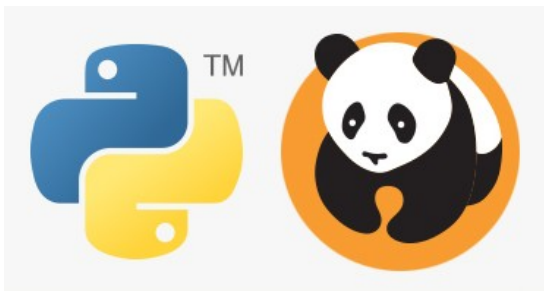
Table: Atributos a analizar

Atributos	
Uso de internet (general)	Tiene teléfono inteligente
Uso de internet en dispositivo móvil	Tenencia de computadora
Uso de internet en el lugar de residencia	Frecuencia en redes sociales
Nivel de conexión a internet	



# Pre-procesamiento de datos

- Limpieza de datos
- Transformación de los datos: normalización, discretización
- Obtención de muestra: solo personas migrantes (interna), mayor de edad.



# Análisis y selección de distancia

- Distancia Ahmad and Dey: comparación de dos datos categóricos

## Análisis y selección de distancia

- Distancia Ahmad and Dey: comparación de dos datos categóricos
- Distancia Euclideana: no aplica para datos categóricos

## Análisis y selección de distancia

- Distancia Ahmad and Dey: comparación de dos datos categóricos
- Distancia Euclideana: no aplica para datos categóricos
- Distancia Hamming: recomendada para análisis de tipo binario





# Distancia de gower

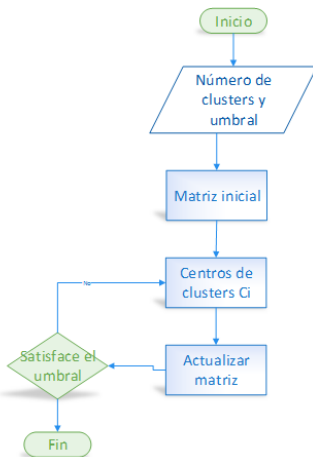
---

**Algorithm 1** Pseudocódigo distancia de gower

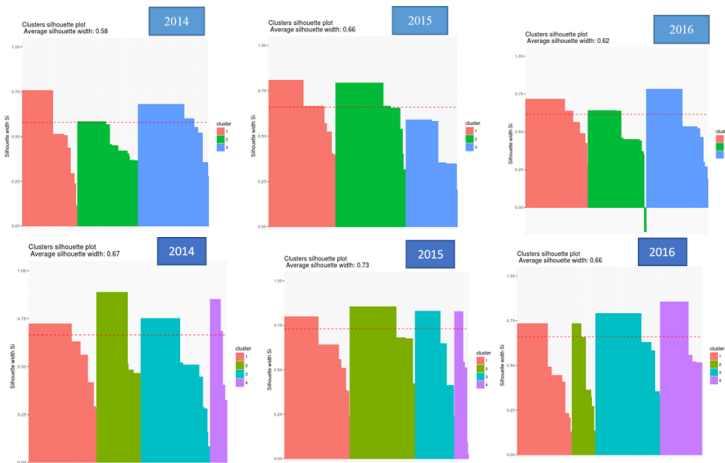
---

```
1: suma_elemento_ij = 0
2: suma_peso_ij = 0
3: for c en columnas do
4:   objeto_ij = 0.0
5:   peso_ij = 0.0
6:   if tipo_de_datos[c] es numero then
7:     objeto_ij = absoluto(vi[c] -  
       vj[c])/(datos_mixtos[c])
8:     peso_ij = peso[c]
9:   else
10:    suma_elemento_ij = [1,0][vi[c] ==  
      vj[c]]
11:    peso_ij = (peso[c])
12:   end if
13:   suma_elemento_ij + = (peso_ij*objeto_ij)
14:   suma_peso_ij + = peso_ij
15: end for
16: return suma_elemento_ij/suma_peso_ij
```

## Proceso algoritmo fuzzy c-means

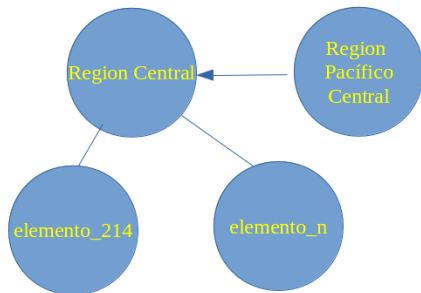


# Evaluación del agrupamiento



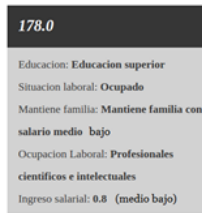
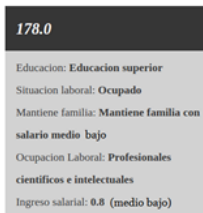
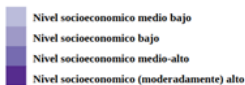
# Estructura de archivo para generar visualización

```
{
  "nodes": [{"id": "Region Central",
    "inmigracion_total": 71,
    "emigracion_total": 99,
    "children": [
      {
        "id": "element_214",
        "titulo": "secundaria_completada",
        "ingreso_por_persona": "medio_bajo",
        "condicion_laboral": "Ocupado",
        "membership": [ {
          "percent": 0.06953217343210001
        }, {
          "percent": 0.88512507829
        }, {
          "percent": 0.045342748278
        }
        ]
      }
    ]
  },
  ...
],
  "links": [
    {
      "source": "Region Pacifico Central",
      "length": 500,
      "target": "Region Central",
      "linkDistance": "strong",
      "preferredCluster": 0,
      "width": 13
    }
  ]
}
```



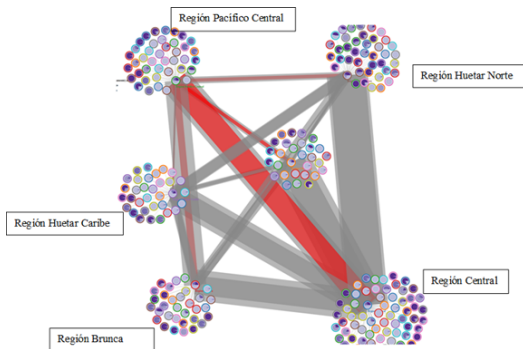
## Comparación: Enfoque difuso(soft) y k-means (hard)

**Figure:** Comparación de análisis de agrupación utilizando algoritmo k-means con respecto al algoritmo de fuzzy c-means

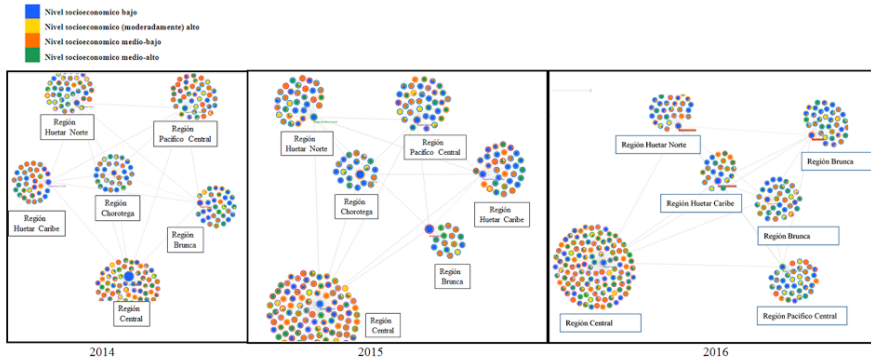


# Visualización de enlace a partir de regiones específicas

**Figure:** Movimientos y densidad de migración a partir de la región Pacífico Central

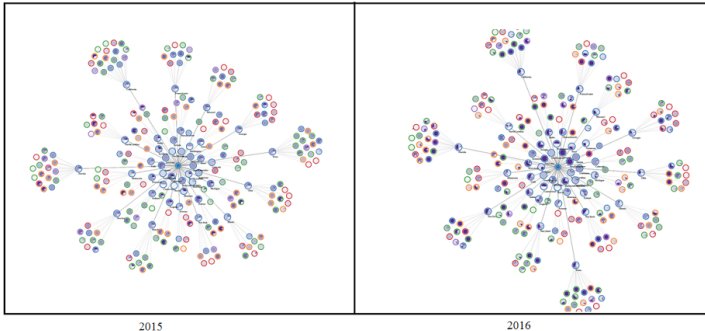
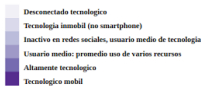


## Caso 1: Migración interna en Costa Rica





## Caso 2: Interacción de personas con la tecnología en Estados Unidos



# Conclusión

- Evolución en el tiempo de datos agrupados con un enfoque difuso para una red social
- Comparación del análisis difuso con respecto a uno binario
- Visualización propuesta permite una manera intuitiva de inferir la membresía y relaciones de manera dinámica
- El algoritmo *force-directed* para agrupar los nodos de acuerdo a su membresía
- Casos analizados: Migraciones internas en Costa Rica y encuesta sobre la interacción de personas con la tecnología.

# ¡Muchas gracias por su atención!

¿Preguntas?

# Material de apoyo

Diapositivas de apoyo

## Trabajo futuro

- Probar la metodología con otros estudios como:
  - el movimiento de las personas a través del tiempo influido por las nuevas tecnologías
  - La influencia de las relaciones en la migración de personas tanto de manera interna como externa
- Se puede mejorar la transición de la visualización para que no se note de una manera más natural a través del tiempo

## Similitudes en el conjunto de datos

**Table:** Par de elementos más similares del conjunto de datos Migración Interna en Costa Rica

Título	Condición laboral	Ingreso	Estado conyugal	Calidad vivienda	Mantiene familia
0.7	0	1	1	0	1
0.7	0	1	1	0	0.75



