

**Universidad Nacional de San Antonio Abad del Cusco  
Facultad de Ingeniería Eléctrica, Electrónica, Informática  
y Mecánica**

**Escuela Profesional de  
Ingeniería Informática y de Sistemas**



**Tesis**

---

**”Algoritmo de Optimización Basado en el Comportamiento  
Social de Arañas para Clustering”**

---

Tesis Presentada por:

**Br. Harley Vera Olivera**

**Br. José Luis Soncco Álvarez**

Para Optar al Título Profesional de:

**Ingeniero Informático y de Sistemas**

Asesor:

**Msc. Lauro Enciso Rodas**

Cusco-Perú

2016

## Resumen

*Clustering* es una técnica popular de análisis de datos para identificar grupos homogéneos de objetos basado en los valores de sus atributos, utilizado en muchas disciplinas y aplicaciones. En este trabajo utilizamos el algoritmo de optimización basado en el comportamiento social de las arañas (SSO) para optimizar grupos de datos tomando como métrica la suma de distancias euclidianas. El algoritmo SSO, propuesto por Cuevas et al., se basa en la simulación del comportamiento social cooperativo de las arañas. Los individuos usados en este algoritmo son arañas (macho y hembra) que interactúan entre sí basados en el comportamiento cooperativo de una colonia de arañas; este comportamiento es la directriz para el funcionamiento de este algoritmo. En este trabajo el algoritmo SSO fue analizado, adaptado e implementado para optimizar el problema de *Clustering*. Para efectos de comparación fueron implementados otros algoritmos importantes de la literatura: el algoritmo K-means y un algoritmo genético (AG) adaptado a el problema de *Clustering*. Experimentos fueron realizados usando 5 datasets tomados del repositorio *UCI Machine Learning Repository*, cada algoritmo fue ejecutado varias veces y después fueron calculadas varias medidas como: promedio, mediana, mínimo, y máximo valor de los resultados. Estos experimentos mostraron que el algoritmo SSO supera al algoritmo *K-means* y obtiene resultados igual de competitivos que el AG. Todos estos resultados fueron confirmados por la prueba estadística no paramétrica de *Wilcoxon* realizada sobre las salidas de los algoritmos.

**Palabras Clave:** Clustering, optimización de clusterings, inteligencia de enjambre, algoritmos de optimización, algoritmos bio-inspirados.