PROYECTO DE INVESTIGACIÓN

I. Título (El proyecto de tesis debe llevar un título que exprese en forma sucinta su contenido, haciendo referencia en lo posible, al resultado final que se pretende lograr. Máx. palabras 25**)**

Modelo de redes complejas multicapa para identificar comunidades de rendimiento en inglés en instituciones de Educación Básica Regular de Puno, 2025

II. Justificación del proyecto(Describa el problema y su relevancia como objeto de investigación. Es relevante una clara definición y delimitación del problema que abordará la investigación, ya que temas cuya definición es difusa o amplísima son difíciles de evaluar y desarrollar**)**

El análisis del rendimiento en inglés requiere reconocer diferencias entre instituciones que una nota global puede ocultar. En la evaluación regional de Puno de 2025, las respuestas permiten distinguir lectura, vocabulario, gramática y comprensión auditiva. Sin embargo, las estadísticas del sistema de monitoreo describen resultados individuales y distribuciones por dimensión, sin identificar ni validar grupos de instituciones con perfiles semejantes. El problema consiste en determinar cómo representar esas similitudes y obtener comunidades institucionales reproducibles, considerando el grado y la cantidad de estudiantes evaluados.

Neuman (2022) mostró que los datos de evaluación pueden representarse mediante redes para identificar agrupaciones, mientras que Gadár y Abonyi (2024) estudiaron comunidades definidas por múltiples facetas de conectividad. Estos antecedentes sustentan explorar una representación que conserve las diferencias entre dimensiones, cuya utilidad debe contrastarse con métodos multivariados convencionales.

La investigación propone desarrollar y evaluar un modelo de redes complejas multicapa para instituciones de Educación Básica Regular del nivel secundaria de Puno. Cada capa representará una sección del examen y las conexiones expresarán similitud de rendimiento. Se comparará el modelo con alternativas de agrupamiento y se evaluará la estabilidad de las comunidades mediante remuestreo y validación con datos reservados.

La disponibilidad de respuestas por pregunta, información institucional y una cohorte exploratoria de 39.808 personas identificadas mediante conciliación por un ID que respalda la factibilidad del estudio. El aporte esperado es un procedimiento reproducible para caracterizar perfiles institucionales y aportar evidencia que oriente la interpretación de necesidades educativas diferenciadas. Sus conclusiones se circunscribirán a las instituciones participantes y a las dimensiones medidas por la evaluación.

III. Antecedentes del proyecto (Incluya el estado actual del conocimiento en el ámbito nacional e internacional. La revisión bibliográfica debe incluir en lo posible artículos científicos actuales, para dejar en claro el conocimiento existente y lo que agregará la Tesis propuesta. Esto es importante para el futuro artículo que resultará como producto de este trabajo**)**

Muelle (2023) analizó la segregación académica y social del rendimiento escolar en Lima Metropolitana mediante información de evaluaciones educativas e indicadores de segregación y autocorrelación espacial. El estudio identificó una estructura territorial heterogénea y destacó la importancia de considerar la ubicación de las escuelas al interpretar los resultados. Este antecedente fundamenta el análisis territorial del rendimiento en el contexto peruano y aporta criterios para examinar la distribución geográfica de las comunidades institucionales. Su alcance difiere del presente proyecto porque aborda otras áreas curriculares y no desarrolla una representación multicapa del rendimiento en inglés.

Paredes et al. (2026) estudiaron la relación entre factores socioeconómicos, recursos escolares y resultados de matemática en estudiantes de secundaria del Perú. Emplearon técnicas de análisis espacial y modelos econométricos, y reportaron dependencia espacial de los resultados y asociaciones con características del entorno. El estudio respalda la incorporación del contexto territorial en la interpretación del desempeño institucional. Para el presente proyecto, su contribución consiste en advertir que las semejanzas entre escuelas deben examinarse

considerando su localización y contexto, aunque sus resultados en matemática no permiten establecer equivalencias con las dimensiones de la evaluación de inglés.

Neuman (2022) investigó patrones conjuntos de rendimiento y contexto socioeconómico a partir de datos de PISA. Construyó una red bipartita y aplicó Infomap para identificar agrupaciones de países según las distribuciones de sus estudiantes. Su contribución fue mostrar una representación relacional de datos de evaluación que permite reconocer configuraciones conjuntas más allá de la comparación de promedios. Este antecedente orienta la transformación de resultados educativos en redes, aunque el presente proyecto empleará instituciones como unidades de análisis y similitudes entre secciones del examen, en lugar de agrupaciones de países e indicadores socioeconómicos.

Xue et al. (2023) desarrollaron un modelo multidimensional para evaluar cursos universitarios de comprensión y expresión oral en inglés con 62 estudiantes de China. Mediante análisis factorial y una red neuronal integraron antecedentes académicos, evaluación de pares, evaluación docente y autoevaluación. Identificaron cuatro fuentes de información con capacidad predictiva sobre el resultado final. Su aporte al proyecto es sustentar la definición explícita de las dimensiones y la evaluación de modelos con datos reservados. Las dimensiones estudiadas corresponden a fuentes de valoración, por lo que sus ponderaciones no son directamente aplicables a las secciones del examen regional de Puno.

Popovych y Drlik (2024) identificaron grupos de desempeño semejante en cursos de programación con actividades evaluadas automáticamente. Aplicaron CRISP-DM y k-means a seis conjuntos de datos de la plataforma Priscilla, diferenciados por actividad o periodo. Encontraron dos grupos en los casos analizados y examinaron cambios en su composición. El estudio aporta un referente para comparar el modelo multicapa con métodos convencionales y analizar la sensibilidad de los agrupamientos a las variables seleccionadas. Su unidad de análisis es el estudiante y su contexto es la programación, mientras que esta investigación se centrará en perfiles institucionales de inglés.

Panskyi et al. (2024) analizaron diferencias entre modalidades tradicionales, STEM y STEAM en cursos de ingeniería con sensores dirigidos a 379 estudiantes de secundaria de Polonia. Consideraron conocimiento, creatividad y relevancia percibida, y evaluaron la estructura de los agrupamientos mediante diferentes herramientas de validación. Los resultados respaldaron la existencia de diferencias entre las modalidades educativas examinadas. Su contribución al proyecto es metodológica: muestra la importancia de utilizar indicadores multidimensionales y contrastar la estructura de los grupos. El presente estudio abordará resultados de una evaluación regional y no los efectos de una experiencia formativa.

Roski et al. (2024) examinaron patrones de uso de recursos educativos diseñados desde el enfoque de Diseño Universal para el Aprendizaje. Combinaron análisis factorial exploratorio y k-means para reconocer perfiles de uso de una plataforma y relacionarlos con características de los estudiantes. Este antecedente aporta un procedimiento para construir perfiles interpretables y contrastarlos con información externa al agrupamiento. Para el proyecto de Puno, constituye un referente de análisis multivariado frente al cual evaluar la representación multicapa. La diferencia principal reside en que sus variables describen utilización de recursos y no respuestas a preguntas de un examen.

Reittu et al. (2023) propusieron integrar capas y atributos de nodos en una matriz y detectar comunidades mediante una extensión de la descomposición regular orientada a compresión. Ilustraron el método con redes sintéticas, sistemas autónomos de Internet y conexiones aéreas. Su aporte fue ofrecer una alternativa de integración que no depende exclusivamente de maximizar modularidad. Este antecedente fundamenta la comparación entre distintas representaciones de la misma información institucional. Aunque su aplicación no es educativa, resulta pertinente para definir cómo combinar las dimensiones del rendimiento y evaluar qué estructura se conserva al integrarlas.

Gadár y Abonyi (2024) abordaron la identificación de comunidades que comparten múltiples facetas de conectividad en redes múltiplex. Propusieron una medida de modularidad que utiliza una capa empírica como referencia y la aplicaron a redes sociales. Su contribución destaca que las relaciones entre capas y el modelo de referencia deben responder al propósito del análisis.

Para esta investigación, el antecedente orienta la construcción de capas institucionales con significado explícito y la evaluación de la información que aporta su integración, sin asumir que una representación multicapa sea superior por definición.

Liu et al. (2024) estudiaron la detección de comunidades en redes multicapa cuyos nodos y enlaces presentan diferentes significados. Desarrollaron una formulación basada en motivos y la contrastaron mediante redes sintéticas y empíricas. El trabajo muestra que la heterogeneidad de la red condiciona la elección de la función de calidad y del procedimiento de detección. Su relevancia para el proyecto consiste en fundamentar una definición coherente de nodos, capas y conexiones. En el caso de Puno, las capas representarán las mismas instituciones, por lo que la diversidad de dimensiones no implica necesariamente heterogeneidad de nodos.

Hussain et al. (2025) formularon el consenso de agrupamientos como la búsqueda de una partición mediana e incorporaron la estructura del grafo en su procedimiento. Su propuesta aborda el desacuerdo entre soluciones y permite integrar resultados de diferentes particiones. Este antecedente sustenta evaluar la consistencia de las comunidades institucionales en múltiples ejecuciones. Para el presente proyecto, se distinguirá esa variabilidad computacional de la incertidumbre asociada a estimar perfiles con distintas muestras de estudiantes. La estabilidad se examinará tanto frente a cambios del algoritmo como frente al remuestreo de los datos.

Mamani Hilasaca (2024) examinó la relación entre aprendizaje autorregulado en entornos virtuales y competencias de inglés en 115 estudiantes de secundaria de una institución de Desaguadero, Puno. Utilizó un diseño correlacional y encontró asociaciones débiles, con diferencias entre los resultados de las pruebas estadísticas aplicadas. El estudio aporta contexto regional sobre el análisis de desempeños diferenciados en inglés y la necesidad de interpretar las asociaciones con cautela. Su alcance es individual y se limita a una institución; el presente proyecto analizará comunidades de instituciones a partir de respuestas de una evaluación regional.

IV. Referencias (Listar las citas bibliográficas con el estilo adecuado a su especialidad)

Gadár, L., & Abonyi, J. (2024). Finding multifaceted communities in multiplex networks. Scientific Reports, 14, Artículo 14521. https://doi.org/10.1038/s41598-024-65049-6

Hussain, M. T., Halappanavar, M., Chatterjee, S., Radicchi, F., Fortunato, S., & Azad, A. (2025). Parallel median consensus clustering in complex networks. Scientific Reports, 15, Artículo 3788. https://doi.org/10.1038/s41598-025-87479-6

Liu, Y., Li, A., Zeng, A., Zhou, J., Fan, Y., & Di, Z. (2024). Motif-based community detection in heterogeneous multilayer networks. Scientific Reports, 14, Artículo 8769. https://doi.org/10.1038/s41598-024-59120-5

Mamani Hilasaca, M. T. (2024). Aprendizaje autorregulado en entornos virtuales y logro de competencias del inglés en estudiantes de secundaria. Comuni@cción: Revista de Investigación en Comunicación y Desarrollo, 15(4), 360–370. https://doi.org/10.33595/2226-1478.15.4.1195

Muelle, L. E. (2023). Dimensión geoespacial de la segregación académica y social del rendimiento de alumnos en las escuelas de educación secundaria en Lima Metropolitana. Apuntes, 50(93), 101–141. https://doi.org/10.21678/apuntes.93.1521

Neuman, M. (2022). PISA data clusters reveal student and school inequality that affects results. PLOS ONE, 17(5), Artículo e0267040. https://doi.org/10.1371/journal.pone.0267040

Panskyi, T., Korzeniewska, E., & Firych-Nowacka, A. (2024). Educational data clustering in secondary school sensor-based engineering courses using active learning approaches. Applied Sciences, 14(12), Artículo 5071. https://doi.org/10.3390/app14125071

Paredes, R. P., Vilca-Huayta, O. A., Arpi-Mayta, R., Sucari-Turpo, H., Huamani-Peralta, A., Yapuchura-Saico, C. R., Gonzales, M., Quispe-Lino, C. N., Rodríguez-Limachi, O., & Condori-Palomino, J. A. (2026). Assessing the impact of socioeconomic and school factors on learning outcomes in Peru: A spatial perspective. Frontiers in Education, 10, Artículo 1600944. https://doi.org/10.3389/feduc.2025.1600944

Popovych, V., & Drlik, M. (2024). Identification of students with similar performances in micro-learning programming courses with automatically evaluated student assignments. Applied Sciences, 14(9), Artículo 3615. https://doi.org/10.3390/app14093615

Reittu, H., Leskelä, L., & Räty, T. (2023). A network community detection method with integration of data from multiple layers and node attributes. Network Science, 11(3), 374–396. https://doi.org/10.1017/nws.2023.2

Roski, M., Sebastian, R., Ewerth, R., Hoppe, A., & Nehring, A. (2024). Learning analytics and the Universal Design for Learning (UDL): A clustering approach. Computers & Education, 214, Artículo 105028. https://doi.org/10.1016/j.compedu.2024.105028

Xue, S., Xue, X., Son, Y. J., Jiang, Y., Zhou, H., & Chen, S. (2023). A data-driven multidimensional assessment model for English listening and speaking courses in higher education. Frontiers in Education, 8, Artículo 1198709. https://doi.org/10.3389/feduc.2023.1198709