Complex networks reveal emergent interdisciplinary knowledge in Wikipedia

Luis A. Miccio[1,2], Carlos Gámez[3], Juan Luis Suarez[4] and Gustavo A. Schwartz[1,5]

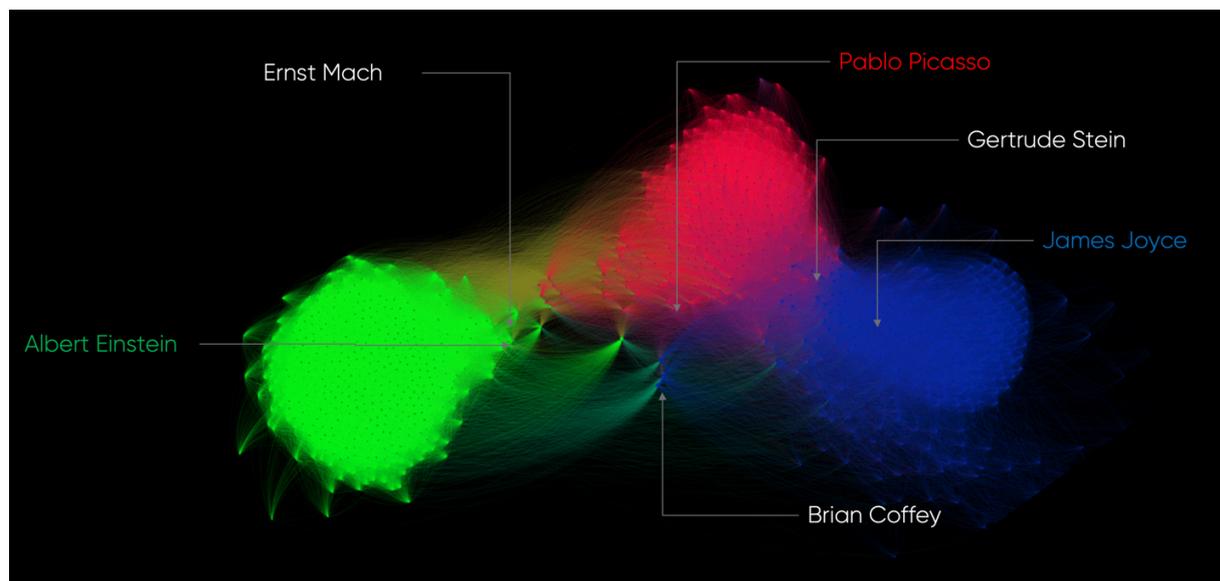[1] Donostia International Physics Center, P. M. de Lardizábal 4, 20018 San Sebastián, Spain.
[2] Institute of Materials Science and Technology (INTEMA), National Research Council (CONICET), Colón 10850, 7600 Mar del Plata, Buenos Aires, Argentina.
[3] Departament d'Educació Lingüística i Literària, i Didàctica de les Ciències Experimentals i de la Matemàtica, Universitat de Barcelona, 08035 Barcelona, Spain
[4] CulturePlex Lab, Western University, London, Ontario N6A 3K7, Canada.
[5] Centro de Física de Materiales (CSIC-UPV/EHU) - Material Physics Centre (MPC), P. M. de Lardizábal 5, 20018 San Sebastián, Spain.

In the last two decades, a great amount of work has been done on data mining and knowledge discovery using complex networks. These works have provided insightful information about the structure and evolution of scientific activity as well as important biomedical discoveries. However, interdisciplinary knowledge discovery, including disciplines other than science, is more complicated to implement because most of the available knowledge is not indexed. Here, we present a new method for mining Wikipedia to unveil implicit interdisciplinary knowledge to map and understand how different disciplines (art, science, literature) are related to and interact with each other. We use the normalized Google distance to measure the structural relatedness between each pair of Wikipedia's entries. Based on this metrics, we create an undirected graph representing an interdisciplinary knowledge map. The assortativity matrix accounts for the interaction strength between different clusters/disciplines. Furthermore, using the formalism of complex networks we characterized both individual and collective behaviour of the different *elements* (people, ideas, works) within each discipline and among them. The results obtained agree with well-established interdisciplinary knowledge and show the ability of this method to boost quantitative studies in the field of interdisciplinary cultural analytics.

Interdisciplinary knowledge map. Each dot represents an entry of Wikipedia (related to people, concepts, works). Thin lines represent links between different elements according to the relatedness defined from the normalized Google distance. Colours represent different disciplines composed of the most related nodes to the given seeds: Picasso (red), Einstein (green) and Joyce (blue). Closeness among nodes is proportional to their relatedness.