Digital Division of labor and informational magnetism: Mapping participation in Wikipedia.

**Submitted by – Abhishek Bhardwaj**

The paper “Digital Division of Labor and Informational Magnetism” discusses the struggles over the ways people and organizations try to control information produced, reproduced, and used. The paper focuses on Wikipedia, which is the world’s largest and most used repository of user-generated content. The paper argues that digital mediations of spatial knowledge have compounded the subordination of local voices, erasing the positionality of the user and social contents under which knowledge is produced. The paper takes this situation as a starting point to investigate whether the Internet, with its mass participation, affords a potentially disruptive role in breaking the digital divisions of labor. The authors describe the data used in a study of Wikipedia’s geographies, which sought to explore the geographic distribution of Wikipedia content and contributors. Three types of data were used here: data about the location of Wikipedia articles, data about the origins of edits, and data about the geographic focus of those edits. Further in the paper it is explained that how the data were processed, and any limitations of the data. It also discusses the challenges of analyzing anonymous edits and the edits by registered editors and the methods used to geolocate those edits. After that the paper discusses the geographic distribution of edits committed to all Wikipediaes from different countries as measured by Wikimedia. The study found that countries in North America, Europe, and Asia had the highest number of edits, while Sub-Sharan Africa had the lowest number. The article also analyzed the propensity of people in any country to commit edits and found that North America, Europe, and much of Oceania stood out strongly compared with regions having medium participation levels and regions with very low levels found mostly in Sub-Saharan Africa. It further explores the factors that covary with the geography of participation in Wikipedia and offers insight into the geographies of participation and voice on the English-language version of Wikipedia. The article demonstrates that the availability of broadband connectivity is a central predictor of the spatial unevenness of participation. The authors use geocoded articles and origin locations of anonymous and registered edits to analyze the source locations, target location, and respective editing volumes. They investigate the geographies of local voice by looking at the volume of autochthonous content by region, analyzing the within- region-edits, and looking at trajectories or networks of editing over space. The authors conclude that large amounts of geospatial content show no sign of deterring people from further contributions and editing, and that North America and Europe commit more than they receive into their territories.

Now based on the above paper, there are certain patterns that we can expect in a dataset of translation between different Wikipediaes. I would be stating these patterns in the form of hypothesis as follows:

* Broadband connectivity will predict the curve of translations between different Wikipediaes up to and extent: In the paper itself it is mentioned that - countries with very small and high numbers of broadband Internet connections commit more edits to Wikipediaes than one would expect assuming a linear trend. On average, countries with medium numbers of broadband Internet connections commit fewer edits than expected.

The same trend is expected when we think of translations the reason being that the languages which are being spoken in countries and regions having better broadband connectivity have more people interested in consuming information. So, the information in languages spoken in areas having better broadband connectivity will be in more demand than in the languages spoken in areas having lesser broadband internet connections.

* Number of people speaking a language will also impact the translation dataset: As the number of people speaking a specific language are more there will be higher number people demanding or needing information in that language. The same happens in case of Wikipedia edits. Population is a baseline variable that is related to the pool of people that could participate in editing Wikipedia. Population has a medium (rPearson = 0.45) correlation with the number of edits. So it is expected that Population will have a medium correlation with number of translations.
* Gross Enrollment Ratio (GER): Gross Enrolment Ratio (GER) calculates the number of those enrolled in school in relation to a country’s total population of 5–17-year-olds. GER may thus serve as a proxy for a baseline amount of literacy and schooling necessary to engage with a text-based resource like Wikipedia. The correlation of GER with the number of edits is rPearson = 0.59.

But in case of number of translations I believe GER will not be directly impacting the data but if we could find the language specific GER data (i.e. Gross Enrollment Ration in a specific language education) then we might be able to get an insight of relation between GER in a language and Translations in that language.

* Content in languages used in countries that are home to large blocks of editors will get translated into languages which are used in countries that are home to lesser number of editors:

As the paper also suggests that, countries that are home to large blocks of editors have the ability to dominate the production of knowledge about smaller countries the same goes for translation of Wikipediaes. The Wikipediaes in languages that are spoken in countries with more number of editors and content writers will have more content in their languages which will be translated to other languages.