Progress Report of the project “MAPPING OF HIGHER EDUCATION, UNIVERSITY AND RESEARCH INSTITUTIONS IN THE COMIFAC REGION AND INTERNATIONAL PARTNERS”

« CARTOGRAPHIE DES INSTITUTIONS D’ENSEIGNEMENT SUPÉRIEUR, UNIVERSITAIRE ET DE RECHERCHE DE L’ESPACE COMIFAC ET DE PARTENAIRES INTERNATIONALS »

by
Honoré Tekeu, PhD
Brunhel N'tambu Vambi, PhD student
Damase P. Khasa, Professor
Figure showing the historical context of each institution, highlighting the range from older, well-established entities to newer ones in the ecosystem. This diversity in foundation years reflects the depth and breadth of experience and expertise brought into the collaborative efforts within the Congo Basin.
Distribution of Establishment years of Institutions

- Year Established
  - 1650: 1
  - 1750: 0
  - 1850: 1
  - 1950: 5
  - 2000: 2

Number of Institutions
Sentiment Analysis Using NLP method

- A. Data Collection
- B. Pre-Processing
  1. Removing punctuations, numbers & symbols.
  2. Converting into lowercase.
  3. Tokenization.
  4. Removing stopwords.
  5. Lemmatizing.
- C. Vectorization
- D. Applying ML Classification Algorithms
  → Evaluation
The word cloud visualizing the mission of each institution, generated using NLP techniques to analyze the data. This approach highlights key themes and objectives as described in their mission statements, focusing on the frequency of terms used across missions.

NB = Each word in the cloud represents a part of their mission, with the size of the words reflecting their frequency of appearance in mission descriptions.
Network analysis

In order to illustrate how institutions are interconnected through their partnerships, with node sizes reflecting their degree of connectivity, we used NetworkX method in Python to build an artificial network undirected graph. It is a type of network used to represent relationships where the connections between nodes (in this case, institutions) have no direction. In an undirected graph, an edge between two nodes implies a mutual relationship, meaning the relationship is bidirectional or does not inherently flow from one node to another. This type of network is particularly useful for illustrating connections such as partnerships, collaborations, or associations where the direction of the relationship is not relevant to the analysis.
Involvement of Institution in the Congo basin

- **Green** = institutions that have indicated their involvement in the Congo Basin.
- **Red** = institutions that are not involved or where no information was provided, but it appears all extracted institutions are involved based on the given criteria.

This graph highlights the connectivity and focus of institutions towards the Congo Basin, showcasing those that are actively engaged in the region. The absence of red nodes implies a strong commitment or interest among the surveyed institutions towards to the Congo Basin.
Here, we illustrated a wide range of involvement durations among the institutions, from longstanding participants like Ulg (since 1817), indicating a deep historical connection to the Congo Basin, to more recent entrants such as R2FAC (2018). Each institution has a node in the graph, where the node size correlates with the length of involvement in the Congo Basin.
As shown, our analysis offers a more interconnected view of the institutions involved in the CBFP Partnership, emphasizing both the duration of their membership and the potential for collaboration among institutions with similar lengths of engagement.
Partnerships among institutions in the Congo Basin

- **ERAIFT** emerges as a highly connected institution with the highest degree centrality (0.217), indicating it has the most direct connections to other institutions in the network. This suggests ERAIFT is a central player in the CBFP partnerships, potentially engaging in various collaborative efforts. Its betweenness centrality (0.239) also indicates a significant role in connecting different parts of the network.

- **RIFFEAC** has a notable betweenness centrality (0.410), the highest among the institutions, despite a lower centrality degree (0.144) compared to ERAIFT. This suggests RIFFEAC plays a critical role as a bridge in the network, facilitating interactions between institutions that might not be directly connected.

- Other institutions like **FASAE_UNIKIN**, **FFGG ULVAL**, and **EFG.1** show varying levels of centrality, indicating their roles in the network. Although their direct connections might be fewer, their strategic position could influence the network’s dynamics, especially for those with higher betweenness centrality.

- The clustering coefficient for the top institutions is 0, indicating that their partners are not directly connected to each other. This low clustering suggests that these institutions may connect disparate parts of the network, bridging communities within the partnership ecosystem.

Overall, the network properties highlight the diversity of roles that institutions play within the partnerships in the Congo Basin. Some act as central hubs with numerous partnerships, while others serve as critical connectors that bridge different parts of the network.
Network properties of each institution and partners

<table>
<thead>
<tr>
<th>No</th>
<th>Institution</th>
<th>Degree</th>
<th>Betweenness Centrality</th>
<th>Closeness Centrality</th>
<th>Clustering Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>ERAIFT</td>
<td>0.217391</td>
<td>0.239344</td>
<td>0.246468</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>RIFFAC</td>
<td>0.144928</td>
<td>0.410806</td>
<td>0.30857</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>FASAE_UNIKIN</td>
<td>0.086957</td>
<td>0.006394</td>
<td>0.086957</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>FFGG, ULAVAL</td>
<td>0.072464</td>
<td>0.084399</td>
<td>0.23052</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>EFG.1</td>
<td>0.072464</td>
<td>0.136331</td>
<td>0.275975</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>65</td>
<td>UY2</td>
<td>0.014493</td>
<td>0</td>
<td>0.191163</td>
<td>0</td>
</tr>
<tr>
<td>66</td>
<td>UY2, UDouala, UDischang, IRAD, IRET, USTM, INER...</td>
<td>0.014493</td>
<td>0</td>
<td>0.133749</td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>CENAREST</td>
<td>0.014493</td>
<td>0</td>
<td>0.194967</td>
<td>0</td>
</tr>
<tr>
<td>68</td>
<td>CEDAMM</td>
<td>0.014493</td>
<td>0</td>
<td>0.203049</td>
<td>0</td>
</tr>
<tr>
<td>69</td>
<td>GIE_SCEVN</td>
<td>0.014493</td>
<td>0</td>
<td>0.047431</td>
<td>0</td>
</tr>
</tbody>
</table>

Interpreting the network graph of institutions and their partnerships in the Congo Basin involves analyzing various network properties that can offer insights into the structure and dynamics of the network. As presented in the Table above, key properties include degree centrality, betweenness centrality, closeness centrality, and clustering coefficient. These metrics can help identify the most influential or central institutions within the network, their role in facilitating connectivity, and the overall cohesion of the network.

➢ **Degree Centrality**: Indicates the number of direct connections an institution has with others. Institutions with higher degree centrality are directly connected to many other institutions, suggesting they have numerous partnerships and potentially play a significant role in the network.

➢ **Betweenness Centrality**: Measures the extent to which an institution lies on paths between other institutions. Institutions with high betweenness centrality can be seen as important bridges or intermediaries within the network, facilitating the flow of information or resources.

➢ **Closeness Centrality**: Reflects how close an institution is to all other institutions in the network. A higher closeness centrality indicates an institution can quickly interact with all other institutions, suggesting efficiency in communication or influence spreading.

➢ **Clustering Coefficient**: Measures the degree to which an institution's partners are also partners with each other. A higher clustering coefficient for the network indicates a tightly knit group of institutions, where partnerships tend to form a cohesive community.
Number of administrative scientific staff per Institution

NB: Data from some institutions have not yet been verified and validated. For this reason, we did not generate their distribution histogram.
Number of students per Institution

NB: Data from some institutions have not yet been verified and validated. For this reason, we did not generate their distribution histogram.
Number of scientific and teaching staff per Institution

NB: Data from some institutions have not yet been verified and validated. For this reason, we did not generate their distribution histogram.
Thank you!