Final Project: Complex Networks for Circular Economy Representation Using Bipartite Graphs

Complex Networks (MAI)

Professor
GÓMEZ JIMÉNEZ, Sergio

Team members
BEJARANO SEPULVEDA, Edison Jair
edison.bejarano@estudiantat.upc.edu
Walzthöny Kreutzberg, Eric
Eric.Walzthony@estudiantat.upc.edu

Universitat Rovira i Virgili
Avinguda dels Països Catalans, 26, 43007 Tarragona, Spain
Master's degree in Artificial Intelligence
Tarragona, 2022

Contents

Abstract

The circular economy is one of the new goals that is presented to our society, this changes the paradigm of the traditional economy, where the products that are being used in some process generate other materials that are currently considered as garbage. For this reason, the circular economy is committed to the reuse of materials in other processes and thereby reducing pollution and obtaining a sustainable economy. This is a difficult task, so to delimit the problem, in this project a scenario of a chemical industry was simulated in which there are some materials that are used by different processes that have certain attributes and generate as output other materials that they could be used in turn by existing processes or create new processes.

For this approach, and thanks to the great capacity of complex networks to adapt to new environments based on the information they handle, a recommender system based on bipartite graph was developed, which in turn feeds on a database of initial data and that in the future can be enriched with new processes and materials entered by users, thus adding the possibility of designing queries such as undoing products or materials and that the network adapts to it. On the other hand, it was possible to create new processes based on certain products that are not being used, thus closing and simulating a circular economy environment that provides recommendations and in turn is validated by user feedback.

Keywords: Complex networks, circular economy, chemistry industry, process, optimization.

1 Introduction

Economic ecosystems are a type of complex network, in which the components interact amongst themselves, and the surrounding environment. In this case, the modeling of a synthetic circular economy is done based on chemical processes (materials and processes). It is known that country-product complex network can be modeled as a bipartite, binary network where the nodes of the two layers correspond to the country and the product [7]. Taking this idea, and applying it to a circular economy, this ecosystem is then modeled using a bipartite graph. The bipartite graph $G(U, V, E)$ is defined as having vertices which are divided into two disjoint and independent sets U and V, such that every edge E connects a vertex in U to one in V.

In order to have this representation, the two nodes (material and process) were populated into a database (Neo4j), and then the relationships were added for each of the corresponding processes and materials. The schema of the database can be seen in Figure 1, which shows that the Process can have two distinct relationships, namely: "RelationshipTo", and "RelationshipFrom".

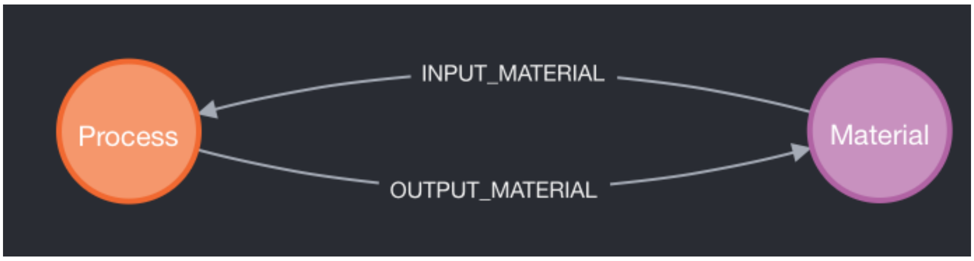

Figure 1 – Visualization of the relationships

Continuing with the representation of the circular economy, the specific processes and materials need to be generated, as well as the overall goal of the system, what is the purpose and what advantages are there with models based on complex networks with graphical structures. The main objective of the system is to be able to represent the chemical processes in a bipartite graph, having the processes as U and the materials as V, the edges E are divided into two different categories: (a) INPUT_MATERIAL (b) OUTPUT_MATERIAL, which allows to separate what each process needs as input and what is produced by it. An example is shown in Figure 2, where "M1" corresponds to Material 1, and the surrounding processes represent whether M1 is an input (green arrows) or output (red arrows), this would be an example of a sub-graph of the network.

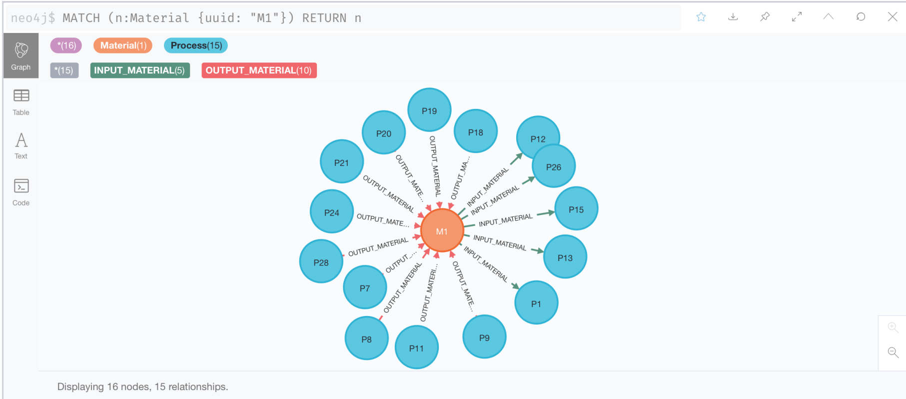

Figure 2 – Relationship of a single material (M1) to different processes as input (green) or output (red)

Furthermore, the same can be represented with the Process nodes, where the same coloring is applied, but seeing the materials in relation to the process and not vice-versa. This can be seen in Figure 3 with ”P1” being the Process 1 and the surrounding nodes the materials which are either used as input or output for the specific process.

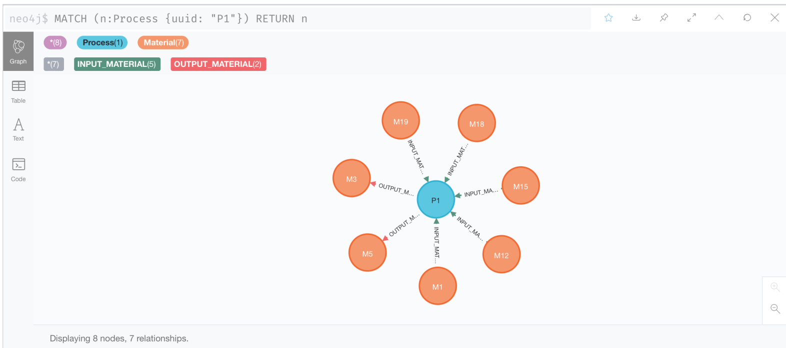

Figure 3 – Relationship of a single Process (P1) to different materials as input (green) or output (red)

In addition, the attributes of the nodes also need be defined, both for the materials and the processes. The definition of the material and process attributes are detailed below: **Material:**

- ID: Universally Unique Identifier (UUID), which serves to identify the materials.
- Name: the name of the specific material
- Quantity: the amount of the material that is used for a specific material
- Unit: the unit the quantity of the material is measured in.
- Cost: the unit cost of the material
- Description: additional information regarding the material.

Process:

- ID: Universally Unique Identifier (UUID), which serves to identify the materials.
- Name: the name of the specific material
- Input Material: which input materials are needed for the process
- Output Material: which materials are produced by this process
- Total Input Cost: total amount of materials that are needed and the total cost of the inputs

- Total Output Cost: total amount of materials that are produced and the value produced by the outputs
- Description: additional information regarding the material.

The system is designed in such a way that the attributes which can be set can be modular and free of choice for the user, this allows flexibility when wanting to model relationships with different attributes of the available nodes.

2 Methodology description

A well-defined plan is the cornerstone of a successful project, and for this project, a detailed plan has been defined, divided into different parts among the team members and implementing agile methodologies to optimize the results. Likewise, GitHub and its environment were used as complementary tools, to the version control and the manage times, tasks and processes.

For this project, the process was structured in the following parts:

- **Tools and software:**

  Github: As previously mentioned, some tools are implemented to facilitate project management, one of them was GitHub, which was used to control, share and manage code versions, which can be found in the following link: https://github.com/walzter/CircularEconomy. Also, it was used to register, manage and control the different tasks of the project, likewise, a milestone was designed to plan the development of these tasks and be able to meet the stipulated delivery date. As shown in Figure 4. The project was divided into 3 parts, first, where was allocated to think about some problem that can be solved by complex networks and investigate possible solutions. The second was to understand and create a database that can simulate a scenario where it can be applied. The last part was to design and develop some recommendation system based on complex networks that can solve the basic requirements.

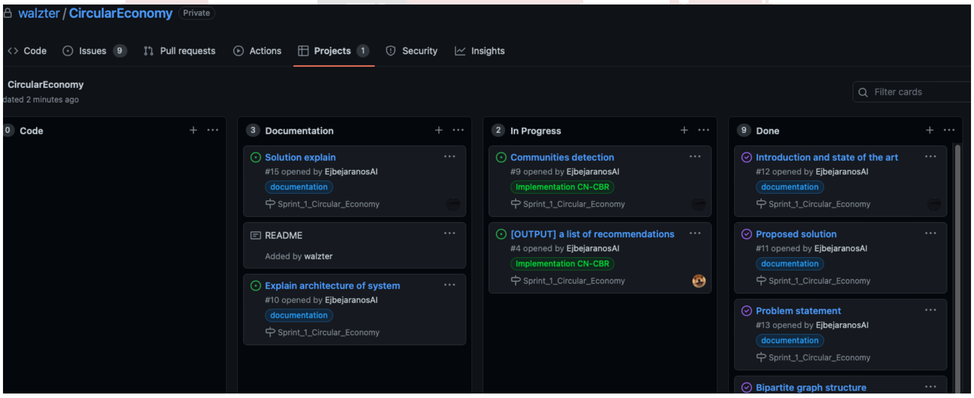

Figure 4 – GitHub Board and milestone issues

  Neo4j: This is a graph-based Database Management System (DBMS), in order to represent the bipartite network. This allows the system to make queries (CRUD), or more advanced queries (future work: Recommendations), with a simple and fast interface. This was setup both locally and on the cloud which means that the compute power is lifted from the user. Neo4j also allows for advanced visualizations of communities, and the implementation of graph-based algorithms. The GUI can be seen in Figure 5, which shows the visualization of the graph-based database, as well as a command-line based querying system in Cypher Query Language, which is a graph-based querying language.

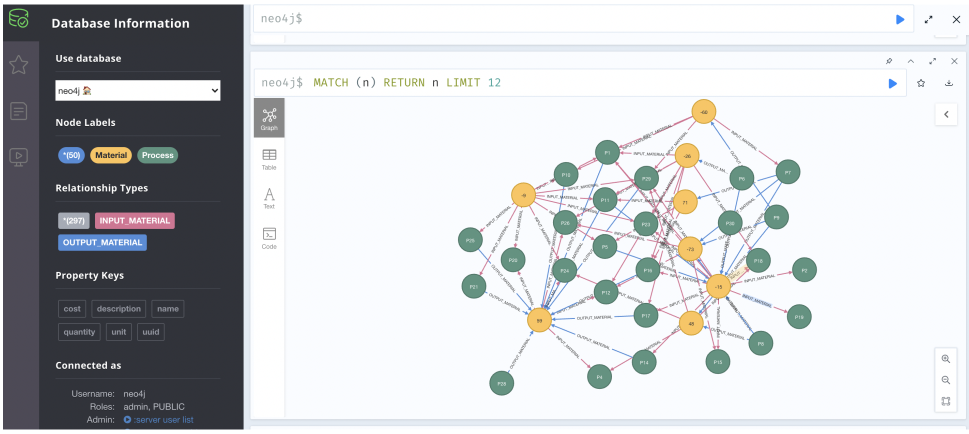

Figure 5 – Neo4j Browser user interface

Python: It was the primary language to create, model and design the software solution. The connections from and to the DBMS were done from scratch as well as the population of the database. The queries were executed in python using the Cypher Query Language, and then passed onto the local or cloud-based DB. All further analysis and exploration of the networks were also done from scratch.

Libraries: Some important libraries as Numpy to make arithmetic operations, Pandas to manage dataframes, Networkx to create, model and use complex networks properties. In addition, we used a Neo4j driver for python, called neomodel, which allows for easy connections to the database.

- **Problem Statement:**

  The circular economy is a new concept that is based on a model of production and consumption of sources. This model is a new way in which many governments and companies are trying to invest and also implement with the motivation of reduce pollution by reusing and recycling materials that were used in another process and that in the past were considered as garbage, on the contrary, in this methodology it is about establishing the recycling of materials in new processes.

  However, the problem today is that the entire economy is based on a linear economy Figure 6, where materials are taken from natural resources and made into products or services that ultimately go to waste, and this is due to that are not planned and were not taken into consideration in the manufacturing process, something very common in the last century, where economic models were not interested in sustainability.

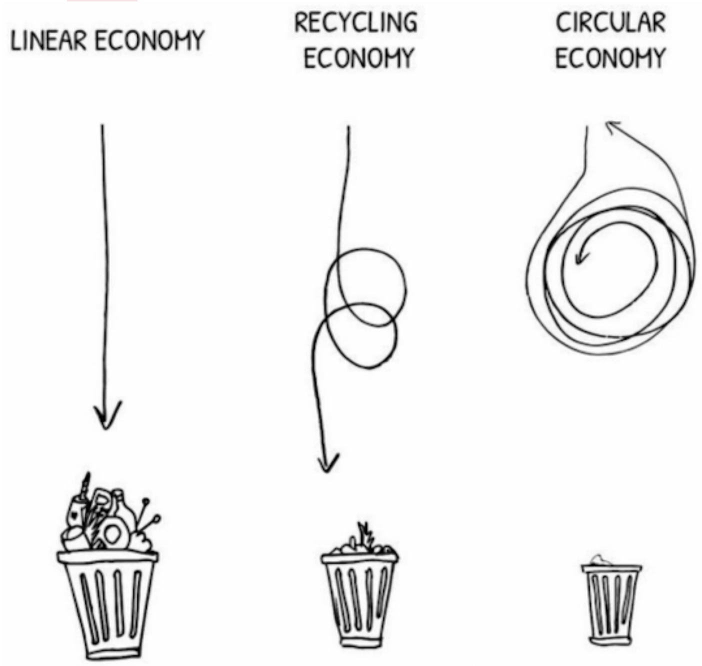

Figure 6 – Linear economy vs Circular economy

On the other hand, the circular economy aims to reduce waste by taking it as a resource to be used in the creation of processes that develop products where, at the end of their useful life, they can be used as raw material for other products, reducing waste of natural resources, reducing environmental pollution, thus avoiding further depletion of the resources we currently have, which is currently occurring with the model used, being unacceptable and incompatible with a good relationship between our society and nature.

Likewise, on the path towards the transformation of an economy from linear to circular, 3 principles are taken into account, such as the elimination of waste and pollution, the circulation of products and materials, and the regeneration of nature. Based on the above, it is understood that in practice, modeling and planning each and every one of the necessary relationships between all the actors involved, these being companies, processes and materials, coming to represent a level of complexity and level of customization much higher than in the linear economy.

That is why, in order to build a sustainable society, it is vitally important to have autonomous models that help develop and implement this type of approach, since they must have the ability to connect waste from different processes that in turn will generate materials, to new processes to be reused, that is why solutions are needed that, in addition to being efficient, have the ease of adapting to changes and that is why complex networks can become an alternative solution to this paradigm.

It is for all this that an alternative to the problem of the circular economy is proposed, which consists of designing and creating a basic recommendation system using complex bipartite networks to model and solve different principles of the circular economy.

Due to the nature of the relationships in a circular economy, a bipartite graph was used to model the complex network, such that the disjoint and independent nodes U and V represent the processes and materials in the circular economy.

- **Coordination:**

  At the time of starting the project, it was decided to carry out a division of labor, according to the main components of the project. That is why it was important to coordinate and divide the different tasks, that is why some strategies used in agile methodologies were adapted and with that roles were assigned to manage the project in the best way, that is why Eric was assigned as scrum master and Edison as product owner, and both as developers.

  However, it must be taken into account that as the project developed, the intervention of the different members in other tasks was necessary to understand, facilitate and adapt the new components in the system, so in the end an average of 80 hours per team member, in the following distribution:

- **Decisions:** As a fundamental piece for the development of the project, constant communication was necessary to be able to direct this project from scratch. One of the first decisions in the team was about how to use complex networks to solve some real world problem that could bring great value to our community. This is why we decided to venture into the circular economy since it is vital to resolve and contribute to developing this methodology to generate a positive impact on the environment and therefore on our lives.

  Another of the most significant decisions for this project was how to address this problem and from what point of view to approach it. It was discussed which industry would benefit the most from this tool, this was boiled down to the following industry: Chemical companies and their associated processes as it is a clear example where years ago a linear economy was managed and only until now has an attempt been made to begin a transformation towards a circular economy.

  However, many of these industries make these transitions manually, which prevents a scalability and possible autonomy for a machine to manage this type of problem due to the number of variables and situations that may arise.

  Moreover, the representation of the problem had to be done in such a way that it would mimic the real-world. In the beginning of this project, we mentioned the country-product ecosystem which was successfully modelled with bipartite binary networks. Thus, using the same principle, it was applied to the Circular Economy.

- **Analysis:** The analysis was centered around the descriptors of Complex Networks, in which we analyzed the most used materials by processes (both as input and output), as well as the processes and which materials they share. These were also answered using queries to the DBMS, which resulted in visual representations of the queries. We analyzed the following descriptors:

- Number of Vertices
- Number of Edges
- Maximum and Minimum Degree
- Average path length
- Degree distribution

The results of the analysis is shown in the Results section, however, this analysis is fundamental to understand the way the materials (information) is propagated through the network. Moreover, the most important descriptor that was used was the distribution of degrees in this synthetic network, since it indicates how many materials are connected to a specific process, both as input and output. Moreover, due to the bugs and unplanned limitations that were faced, the analysis was impacted due to the tools and new knowledge that was acquired to develop the application.

3 Requirements Analysis

In this part, we carry out a requirements analysis on the needs to be addressed and we model this problem in the most similar conditions with the circular economy approach, that is why to land and narrow down the problem a little more to provide an analysis and a possible solution. A scenario was selected where the chemical industry is the universe where the problem of a linear economy is represented, where many of the processes take natural resources and transform them into products and in other cases into substances that are discarded, thus leaving a clear need to optimize the source of resources, reduce pollution and generate new processes based on materials that have no use.

Based on the previous approach, some functional requirements of the system were defined, in order to define what functionalities of the system should have and what are the acceptance criteria to consider them fulfilled, being as follows:

3.1 Functional Requirements:

First, the functional requirements that the system must meet are defined. In particular, it focused on the requirements that the system must meet to be considered a circular economy based on the principles of this. Some of them are:

- **Make queries by the user**
  1. CRUD Material / Process (CRUD = Create, Replace, Update, Delete)
  2. Return quantity of materials that are not being used
  3. Incorporate a new materials into the processes available
- Make recommendations about materials that are not being used
- Ease of system to adapt to changes in processes and materials
- **Complex network representation:** As a user, it is necessary to represent all the information in a briefly way and that is why a graph representation is the best option, this is represented using neo4j as a graph-based DBMS.
- Recommend new processes based on materials that are not consumed and perform a user verification to see if it is satisfied

4 Proposed Solution Design

As a proposed solution, and as can be seen in the image Figure 7, the general structure of the system that was designed in the recommender system based on bipartite complex networks is shown. This scheme explains the proposed solution divided into 7 main components which are:

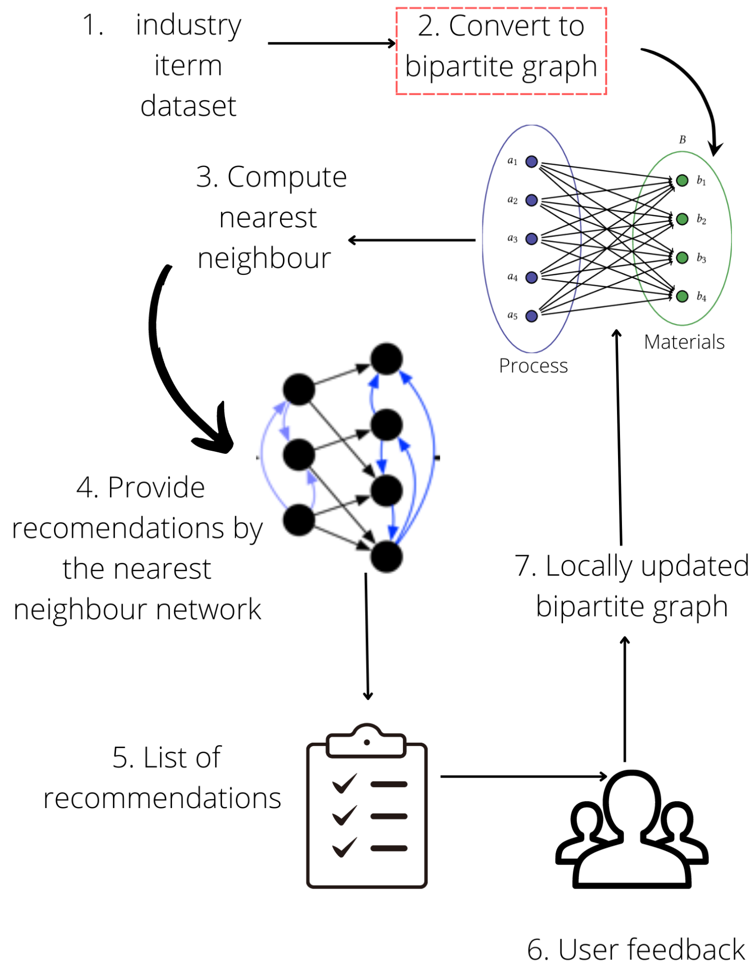

Figure 7 – Recomender system based on complex bipartities networks structure schema.

1. **Industry dataset:** As one of the most important tasks in the system is the industry data set, which is necessary to simulate the problem in an initial state, and in this project it was necessary to investigate certain processes in the chemical industry, this with the intention to limit the problem. With this, the group discussed the need to contact some experts in this industry to discuss how this problem behaves in their environment, with which important things were understood, such as the need to design a database where the process can display attributes such as the quantity and type of materials, in turn, following certain rules, a product is produced which in turn obtains an output material that in many cases is classified as waste, which is why in the dataset it was considered to create and represent a matrix as shown in the Figure 8, the processes and the materials that are necessary for it, thus determining that the positive values are those that are necessary for the process. and those that are in negative, would represent the outgoing materials of said process.

	M-001	M-002	M-003	M-004	M-005	M-006	M-007	M-008	M-009	M-010	...	M-012	M-013	M-014	M-015	M-016	M-017	M-018	M-019	M-020	Energy
Processes																					
P-001	-73.0	0.0	59.0	0.0	48.0	0.0	0.0	0.0	0.0	0.0	...	-60.0	0.0	0.0	-26.0	0.0	0.0	-9.0	-15.0	0.0	-76.0
P-002	0.0	0.0	0.0	0.0	73.0	0.0	0.0	11.0	0.0	69.0	...	0.0	36.0	-33.0	33.0	-82.0	64.0	0.0	-38.0	-92.0	112.0
P-003	0.0	0.0	0.0	58.0	26.0	0.0	0.0	92.0	43.0	0.0	...	-42.0	34.0	-4.0	5.0	0.0	0.0	0.0	0.0	0.0	212.0
P-004	0.0	71.0	-28.0	0.0	-76.0	0.0	4.0	-12.0	0.0	0.0	...	0.0	-21.0	0.0	0.0	-28.0	0.0	18.0	0.0	0.0	-83.0
P-005	0.0	-2.0	0.0	-71.0	37.0	0.0	0.0	19.0	16.0	-43.0	...	98.0	-7.0	0.0	0.0	0.0	0.0	-91.0	22.0	4.0	-27.0
P-006	0.0	-97.0	0.0	0.0	0.0	0.0	79.0	0.0	-11.0	0.0	...	0.0	8.0	0.0	16.0	0.0	0.0	0.0	20.0	22.0	37.0

Figure 8 – Circular economy dataset for chemistry industry.

As a result of the representation of the data set, it is possible to observe the energy consumption per material, as in the image (It should be noted that this was a scenario simulated by rules between the processes and materials, which can be adapted to real situations in future works).

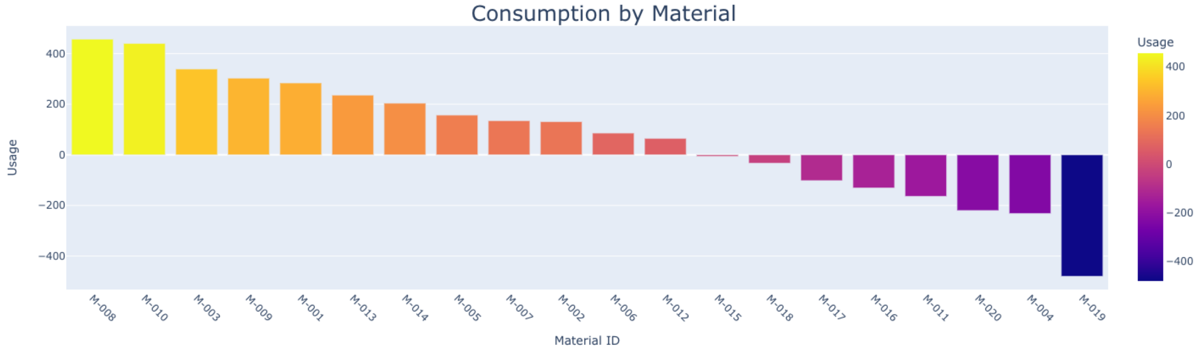

Figure 9 – Representation of the consumption by material.

2. **Convert the dataset to bipartite graph** In this part, for the data set, it was decided to use Neo4j to represent the information in interactive graphics, which also have the facility to create queries that can be represented in a user-friendly way Figure 10. On the other hand, the dataset was prepared and configured to be used and connected to the Noe4j service, sharing the materials, processes, connections and also all the attributes in each node.

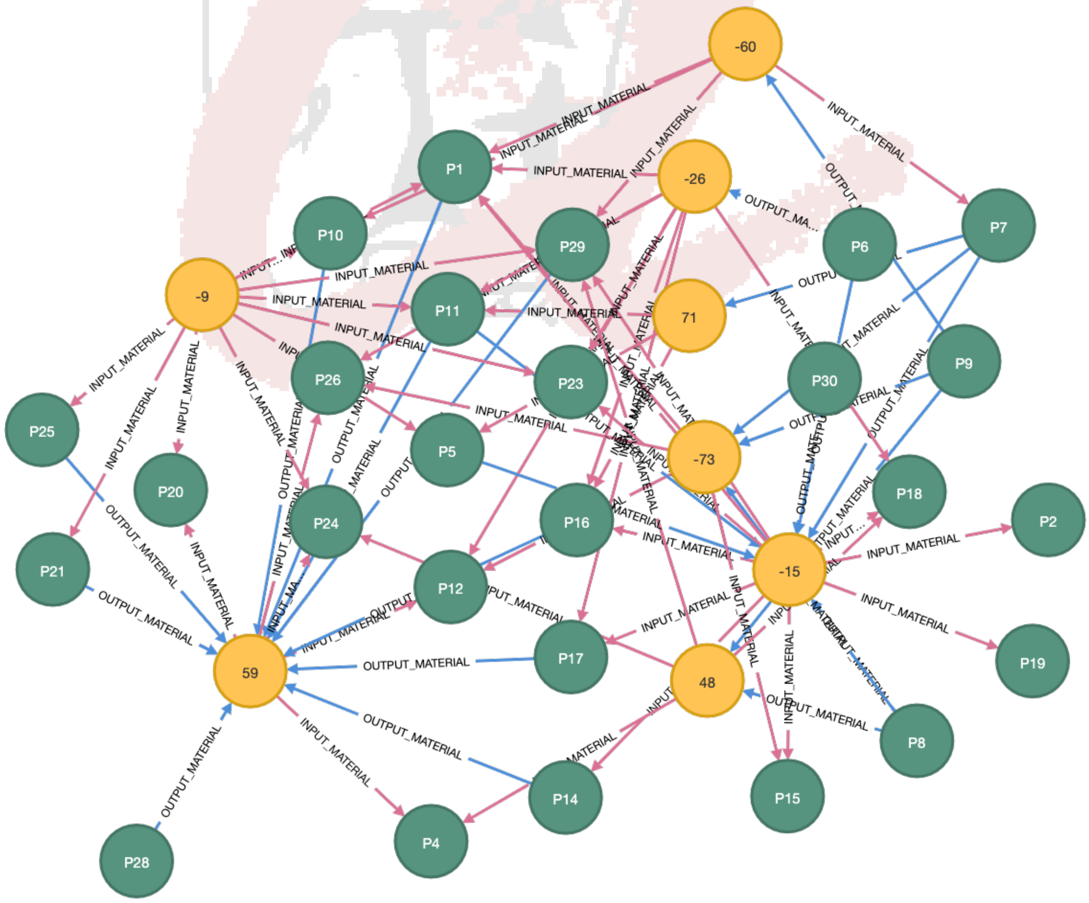

Figure 10 – Representation of the dataset to bipartite graph.

3. **Compute nearest neighbour:** The similarities between each process and materials, are computed and the top K for each element are stored.
4. **Provide recommendations by the nearest neighbour network:** Using the item similarities and the past user interaction, provide recommendations about new process based on the materials that not have connections to other process.
5. **List of recommendations** A set of recommendations are returning by the system in which reusing materials in processes are the main objective of the recommendations.
6. **User feedback:** To validate if the created process meets the needs of the system, the user can validate this new process and if the case met the new assumptions, and in turn only takes materials that are available, it will be stored in the database, otherwise Otherwise, the case will recalculate another possibility to present. for the user
7. **Locally update of the bipartite graphs:** Finally, the system updated the dataset and with it the bipartite graph, updating the new connection for new queries that can be requested from the system. Therefore, the system can add, remove or modify materials and processes with facilities to update and adapt the network.

5 Results

As results of this project we can detail below, a small analysis of descriptors that was carried out on the network that was generated in this problem, as well as its corresponding distribution of degrees. Secondly, the queries that were developed to consult, add or eliminate materials in the process are shown, managing it from our base code and connecting it to Neo4j for its representation and possible consultation of queries from there.

5.1 Descriptors analysis

5.1.1 Degree distribution

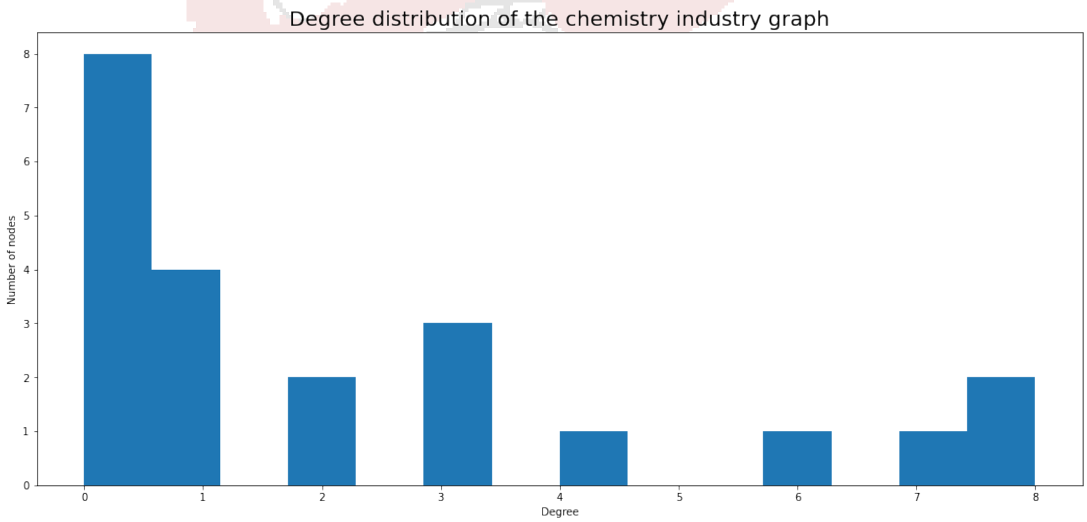

Figure 11 – Degree distribution.

In the descriptor analysis, as shown in the following Figure 12, the maximum number of nodes for this example was 50 while the number of edges was 297. On the other hand, the maximum degree it was 21, otherwise the minimum degree was 7. Likewise, the degree distribution of the network was obtained as shown in the image above11, where it is observed that the number of nodes with degree distribution from 0 to 1 is 12.

```
---------DESCRIPTORS FOR THE GRAPH---------
******************************************
Is the graph directed?            : False
Number of nodes                   : 50
Number of edges                   : 297
Maximum degree                    : ('M-008', 21)
Minimum degree                    : ('P-001', 7)
Diameter                          : 3
Radius                            : 3
Number of connected components    : 1
Number of self loops              : 0
Average path length               : 2.0048979591836735

**************************************************************************************
Clustering coefficient            : {'P-001': 0, 'P-002': 0, 'P-003': 0, 'P-004': 0,
'P-005': 0, 'P-006': 0, 'P-007': 0, 'P-008': 0, 'P-009': 0, 'P-010': 0, 'P-011': 0,
'P-012': 0, 'P-013': 0, 'P-014': 0, 'P-015': 0, 'P-016': 0, 'P-017': 0, 'P-018': 0,
'P-019': 0, 'P-020': 0, 'P-021': 0, 'P-022': 0, 'P-023': 0, 'P-024': 0, 'P-025': 0,
'P-026': 0, 'P-027': 0, 'P-028': 0, 'P-029': 0, 'P-030': 0, 'M-001': 0, 'M-002': 0,
'M-003': 0, 'M-004': 0, 'M-005': 0, 'M-006': 0, 'M-007': 0, 'M-008': 0, 'M-009': 0,
'M-010': 0, 'M-011': 0, 'M-012': 0, 'M-013': 0, 'M-014': 0, 'M-015': 0, 'M-016': 0,
'M-017': 0, 'M-018': 0, 'M-019': 0, 'M-020': 0}

**************************************************************************************
Degree distribution               : [('P-001', 7), ('P-002', 11), ('P-003', 8), ('P-
004', 9), ('P-005', 12), ('P-006', 7), ('P-007', 8), ('P-008', 10), ('P-009', 10),
('P-010', 9), ('P-011', 12), ('P-012', 10), ('P-013', 9), ('P-014', 9), ('P-015', 12),
('P-016', 10), ('P-017', 9), ('P-018', 11), ('P-019', 10), ('P-020', 11), ('P-021',
11), ('P-022', 10), ('P-023', 9), ('P-024', 11), ('P-025', 10), ('P-026', 11), ('P-
027', 12), ('P-028', 10), ('P-029', 11), ('P-030', 8), ('M-001', 15), ('M-002', 11),
('M-003', 15), ('M-004', 16), ('M-005', 13), ('M-006', 14), ('M-007', 13), ('M-008',
21), ('M-009', 14), ('M-010', 16), ('M-011', 17), ('M-012', 14), ('M-013', 12), ('M-
014', 12), ('M-015', 19), ('M-016', 15), ('M-017', 12), ('M-018', 18), ('M-019', 16),
('M-020', 14)]

------------------Descriptors analisys finished------------------
```

Figure 12 – Analysis of chemical industry network descriptors (Processes and materials).

The fully functional database can be accessed with the following link: http://www.overleaf.com https://fa1a273e407fe4d26
$JU_DfQ1Jv30t9MpzNBCW6Ml-jl2gjMpLEsTAtbJp-2CiJ6qupC3QwPtGACLBdQhbHs3lPHrIwFnEGtu-$
$KWy5m_3ntBsV7ilvrsJPUi0cJsYiEJWevL9MuQyMr9_u06WUDDc4D-63hCuodHLxXdHjxxTnO6IOlUzkn2Q5x$
$n3CHYu5XE5P9yRggbYjpYOKxq89-KIYKMtMc3O5ppARc_tJtd046jZ-c9s30y0Vt2_v2XPYc0Ssnw5ewNeo4jDB$

6 Discussion and futures works

- One of the first conclusions that we perceived, was to use complex networks as a solution to a real problem, it means more than just performing a network analysis or how to program a certain type of network, it requires planning since the database is created. , and how a real-world problem can be represented and simulated in something that later becomes a complex network. This, without a doubt, gave us great experience in how to address complex problems and find innovative solutions from a certain approach, which in this case was complex networks.
- Another important fact that we conclude is that at present, there are different and new tools such as Neo4j and recent libraries such as DyNetx - (Dynamic Network library) that help to make representations of complex networks in a more friendly way towards the future and possible final user.
- Due to the time required to continue this work and the long research that was necessary to generate and formulate it from scratch, it is determined that as possible future steps, it remains to make a graphical interface that summarizes all the work and can be used by some expert users to determine the validity of this possible tool, since it was not possible to reach that point of maturity in the project.
- The best application of this model would be as a graph-based recommender system which represents the Complex Network as a bipartite graph. However, as the previous points mentioned, there is a great potential in using these systems for complex and diverse problems.
- Lastly, the development of the framework will be continued as the authors see value in these types of applications, which are not limited to the chemical industry but can also be applied to a plethora of industries.

7 Acknowledgment

We'd like to thank Dr. Luciano Pietronero, for assisting in the research, and analysis of the projects as well as guiding us to the relevant resources to be able to depict a circular economy as bipartite graphs. Dr. Pietronero was met personally by one of the authors during an event, which created this exchange of information.

We'd also like to thank the support team from Neo4j, which answered our ad-hoc queries with minimal response times.

Lastly, we thank Giulio Rosetti for taking the time and explain the issues and possible solutions in implementing Dynamic Networks using his API.

References

[1] Reka Albert, Albert-Laszlo Barabasi. “Statistical mechanics of complex networks”.(2001).

[2] Newman, Mark. ’Networks’. *Second Editon: chapter 1 to 8.* (Newman, Mark. 2018): 13–262.

[3] Haykin, Simon. “Community structure”. *Neural networks and learning machines.*(Prentice Hall/Pearson): 11–17.(2009)

[4] Documentation igraph-Python– for version 0.9.9 https://igraph.org/python/

[5] Albert-László Barabási, “Network Science”.(2016).

[6] ”Circular Economy: Definition, Importance And Benefits — News — European Parliament”. Europarl.Europa.Eu, 2022, https://www.europarl.europa.eu/news/en/headlines/economy/20151201STO05603/circular-economy-definition-importance-and-benefits.

[7] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini, Scientific reports 5, 1 (2015).

[8] Laudati, Dario .. 'The different structure of economic ecosystems at the scales of companies and countries'. 2022. Web.