Modeling Business Complexity through Complex Networks and Soft Computing

Mario González Rodríguez

mario.s.gonzlez@gmail.com

(*Dated:* 15 July 2019)

ABSTRACT

The present proposal aims to expose the actual interest of the author on developing a research project in the area of business complexity modeling and analysis using techniques from Graph Theory and Complex Networks Analysis, as well as techniques from Soft and Natural Computing, in order to study the collective phenomena of the of the studied systems. The problems to be dealt with are related to Business/Industrial Process Modeling, e.g. building the process activity/actors interactions as complex networks to analyze the process structure and controllability; Socio-physics modeling and prediction using soft computing techniques (i.e. sustainability/offshoring/gender equality/migration); Business organizational structures, e.g. discover how the interactions of the individuals in an organization facilitate the firm's goals achievement. The goal is to properly use/adapt complex networks/soft computing/data analytics techniques in order to address these topics.

I. INTRODUCTION

Prediction in Business/Economics is an involved task. Treating such problems as a complex system is of exceptional interest now when computer technologies provide enormous possibilities for collecting, storing and processing of information obtained by tracing system behavior. The fundamental interest, is that results from modeling this problems as complex systems, may shed more light on the general aspects of its evolution. I am interested in modeling/understanding the structure and function of such systems. The main tool of analysis presented for this research project will be complex networks modeling (Albert and Barabasi, 2002). Although, such analysis can be used together with soft-computing techniques (i.e. machine learning) to perform classification or regressions tasks for prediction of the system functionality/structure.

Complex structures organized as networks span a wide variety of biological, social, technological and intellectual systems. Complex networks are sets of connected nodes that interact in some way. In the context of network theory, a complex network is a graph (network) with nontrivial topological features. Most social (White, Wellman, and Nazer, 2004; Ahn *et al.*, 2011), biological (Maslov and Sneppen, 2002) and technological networks (Flake *et al.*, 2002) show patterns of connection between its elements that are neither purely regular or purely random, but a mixture of both. One of the most important aspects needed for understanding complex systems is the study of the structure of existing interaction between the elements of the system, and how this structure directly affects the dynamic or functional capacity of the system, as well as emerging behaviors that may appear.

Many measures at different scales can be calculated for a complex network. At the microscopic level one is interested in the properties of the nodes in the network, measuring the node in and out degree, the centrality of the node (betweenness and closeness), the function, position and size of the node, may also be considered. At mesoscopic levels one is interested in the formation of clusters in the network, as well as identifying the number and size of the components in the network. At macroscopic level one is interested in measuring the degree distribution in the network, the connectivity dilution, the clustering coefficient, and average path length of the network. All this measures give different insights about the systems at different levels, and are a powerful technique to approach a variety of problems in Science and Engineering.

Data modeling and analysis using soft computing techniques, such as machine learning, can also shed lights on the aforementioned systems structure and behavior. Well established algorithms in machine learning (Random Forest Ziegler and König (2014), Neural Networks, among others Dreiseitl and Ohno-Machado (2002)) can be used in data analytics classification and regression scenarios helping decision making. Clustering algorithms are useful for discovering underlying structure present in a dataset Bradley, Fayyad, and Reina (1998); Jain (2010). Time series analysis can be used to model and predict temporal behavior of these systems. Deep learning techniques such as Long-short term memory (LSTM) Hochreiter and Schmidhuber (1997) have proved to predict temporal data accurately.

I have worked on a variety of problems on complex systems ranging from sociophysics modeling i.e. sustainability report scoring modeling (González *et al.*, 2015, 2018b), mapping global sustainability González, Alonso-Almeida, and Dominguez (2018), corruption modeling González *et al.* (2018a), gender equality soft computing modeling and prediction Dominguez *et al.* (2019a) human migration analysis Dominguez *et al.* (2019b); to phase transition theory i.e. random field interactions with local/random connectivity (Doria *et al.*, 2015). The insight gained in the aforementioned papers, will be very helpful to deal with Complex Systems, mainly studying the system/network topological structure, and how the topological parameters affect the behavior/function of the system under different circumstances. Also, soft-computing techniques will allow to discover the structure and predict the behavior of this systems from their collected data.

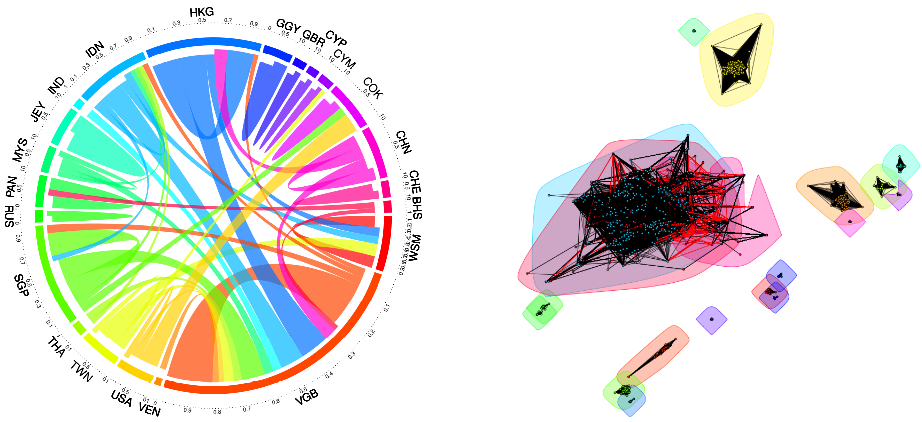

FIG. 1. Left: Top countries in the offshoring network. Right: Microscopic offsoring network.

II. OFFSHORING NETWORK STRUCTURE AND BEHAVIOR

The objective is to analyze the offshoring network constructed from the information contained in the Panama Papers, characterizing worldwide regions and countries as well as their intra- and inter- relationships. The top offshore financial regions and countries of the network will be identified, and their intra- and inter- relationship are mapped and described at different scales: geographic region and country levels.

The offshoring network have been already mapped in a recent work Dominguez, Pantoja, and González (2018). The top regions/countries were aggregated and a fraction of the network was described (1-left). The interest, now, is to identify at the macroscopic level (1-right), the structure of the whole network will be described, as well as the main actors/counties in the network identified and their interaction in the offshoring network.

III. BUSINESS PROCESS MODELING AS COMPLEX NETWORKS

Many problems in Research Operations can be represented and solved using graphs. Minimum spanning tree are build for telecommunications networks design, solving the Travel Salesman Problem, etc. Shortest path algorithms are used for designing facility layout, equipment replacement, etc. Circulation problems such as the maximum flow problem through a single/multiple sources, single/multiples sink flow in a network for airlines crew scheduling, capacity of pipeline transport, etc. PERT and CPM for project Management.

The present work proposes using Complex Networks for BPM analysis. For example the University Community Relations Process (CRP) represented in the Swim-lane diagram in Fig. 2, can be expressed as a graph where the each vertex represents the process activities and the edges the flow of information/materials between activities, see Fig. 3. Once each process

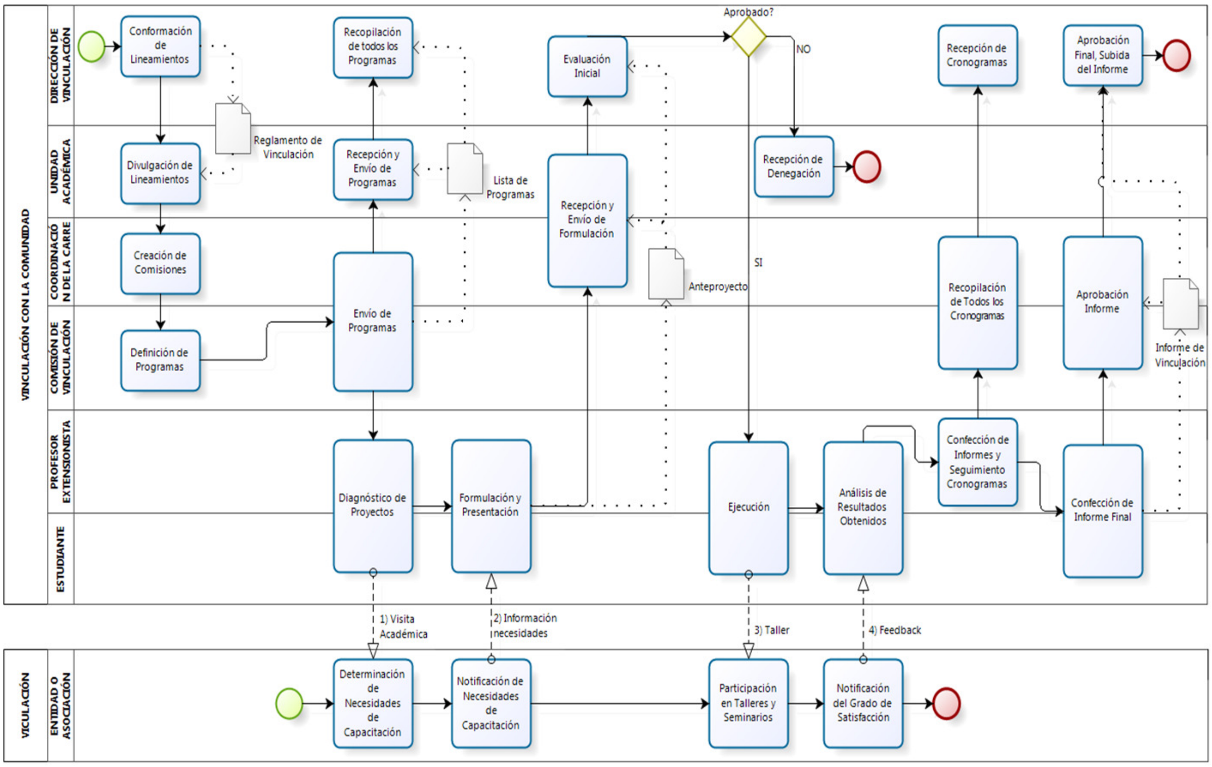

FIG. 2. Swimlane diagram of the University Community Relations Process (CRP).

is represented as a graph, the interrelation between processes can be expressed as a complex network. The metrics of such a network can be measured in order to analyze and optimize the processes. For instance, one can measure and model:

- Graph diameter: the maximum number of activities that must be carried out to end the processes.
- Shortest path: the minimum number of activities that must be carried out to end the processes.
- Average path length and clustering coefficient can give information about the parallelization and recursion degree of the processes.
- Indegree and Outdegree of the process components, identify the process sources and sinks, the ration between them influence the process controllability (Ruths and Ruths, 2014).
- Hubs: Critical activities.
- Clusters: Sets of very interrelated activities, will possible coincide with the designed process and can be used as a process design assessment tool.
- Structural attacks: removing critical activities would make the process collapse? Measure robustness of the process.

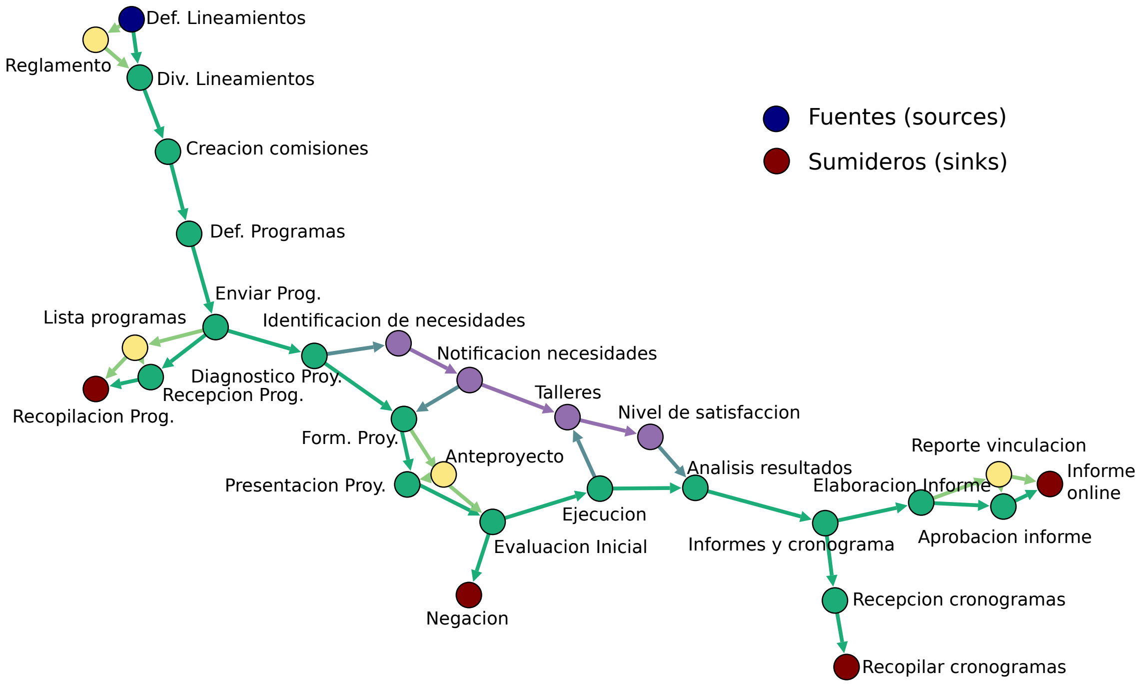

FIG. 3. Graph representation of the CRP presented in Fig. 2. The ratio between sources and sinks influence the process controllability.

IV. ORGANIZATIONAL ANALYSIS

To understand the interaction among individuals in a Business Organization is vital to achieve its goals. A deep understanding of this interactions will provide the manager a lot of flexibility in management and control of the organization. Building the organization sociogram can help to analyze choices or preferences within a group of individuals, and mapping the structure and patterns of group interactions. This can give a useful insight of the power relations within an organization and how to deal with bureaucracy, making the Business more effective.

In Fig. 4 is represented a schematic representation of a network within an organization, the nodes represent the individuals, and the connections between them are given by the interactions through the corporate email. The size of the nodes gives the importance of the individual among the contact list, and the width of the edges give the volume of the communications between individuals. Using the network analysis, one may perform a microscopic analysis for individuals detecting for instance informal leaders (Hubs), a mesoscopic level analysis for identifying groups (clustering), and a macroscopic analysis of the organization (clustering coefficient, average path length) can give an insight of the organization structure as a whole. For instances, Fig. 4-right could represent a network of migration patterns between countries. Countries with similar migration behavior are grouped in clusters, represented by the shaded colored areas.

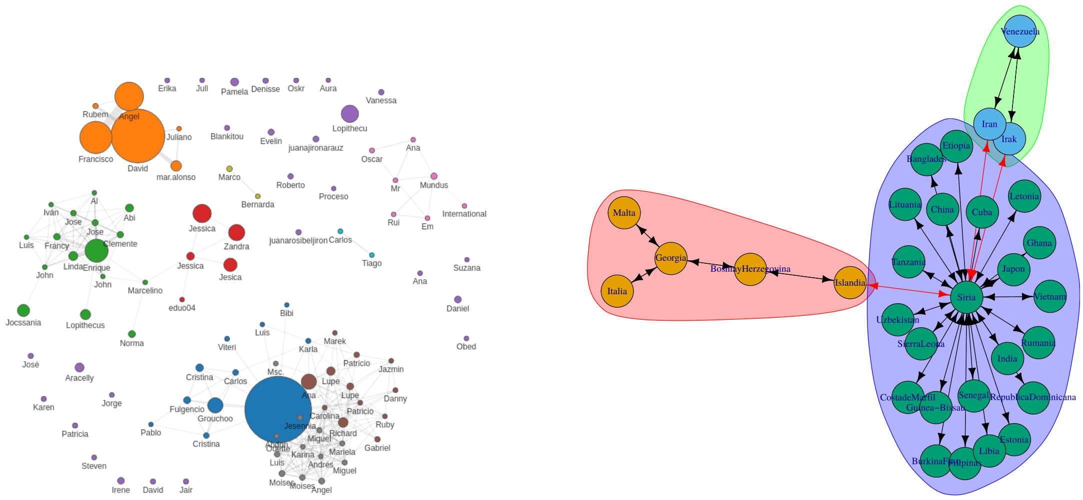

FIG. 4. Left: graph representation of the email interactions in an Enterprise Network. Right: Migration patterns similarity network.

Following the example given in Fig. 4, the network could be build automatically using not only the information about the sender and recipient, the information in the subject and the body of the mail can be also processed using text mining techniques in order to measure not only the structure of interactions but also their functions. These activities belong to the area of Data Analytics techniques. Machine learning on the graph could also be performed, for example to predict new connections to occur in the graph.

V. DISCUSSION

The present proposal aims to apply complex networks techniques to the analysis and understanding of a variety of complex systems from Economics, namely mapping the global offshoring network; Business/Industrial Process Modeling and Optimization; Organizational Analysis. This project aims to increase our understanding of the aforementioned systems through modeling their components interactions as graphs and characterizing their topological structure. There is a whole set of metrics that can be explored, and complex network characterization techniques that can be applied to the proposed problems or variations, that worth to be checked in this project. Finding structure (clustering analysis) in the data and make predictions and forecasting using soft-computing algorithms will also help to understand the systems under study. This project will hopefully make significant contributions to the existing literature on the addressed topics.

REFERENCES

Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., and Barabási, A.-L., "Flavor network and the principles of food pairing," Scientific reports **1** (2011).

Albert, R. and Barabasi, A., "Statistical mechanics of complex networks," Rev. Mod. Phys. **74**, 47 (2002).

Bradley, P. S., Fayyad, U. M., and Reina, C. A., "Scaling EM (Expectation Maximization) Clustering to Large Databases," , 25 (1998).

Dominguez, D., Pantoja, O., De Los Nietos, J., García, H., Sánchez, Á., and González, M., "Soft-computing modeling and prediction of gender equality," in *2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG)* (IEEE, 2019) pp. 242–248.

Dominguez, D., Pantoja, O., and González, M., "Mapping the global offshoring network through the panama papers," in *International Conference on Information Theoretic Security* (Springer, 2018) pp. 407–416.

Dominguez, D., Soria, P., González, M., Rodríguez, F. B., and Sánchez, Á., "A classification and data visualization tool applied to human migration analysis," in *2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG)* (IEEE, 2019) pp. 256–261.

Doria, F., Erichsen Jr, R., Dominguez, D., González, M., and Magalhaes, S., "Random field ising model in a random graph," Physica A: Statistical Mechanics and its Applications **422**, 58–65 (2015).

Dreiseitl, S. and Ohno-Machado, L., "Logistic regression and artificial neural network classification models: a methodology review," Journal of biomedical informatics **35**, 352–359 (2002).

Flake, G. W., Lawrence, S., Giles, C. L., and Coetzee, F. M., "Self-organization and identification of web communities," Computer **35**, 66–70 (2002).

González, M., Alonso-Almeida, M. d. M., and Dominguez, D., "Mapping global sustainability report scoring: a detailed analysis of europe and asia," Quality & Quantity , 1–15 (2018).

González, M., Dominguez, D., Jerez, G., and Pantoja, O., "Periodically diluted begnn model of corruption perception," in *International Conference on Mining Intelligence and Knowledge Exploration* (Springer, 2018) pp. 289–298.

González, M., Dominguez, D., Pantoja, O., Guerrero, C., and Rodríguez, F. B., "Modeling sustainability reporting with ternary attractor neural networks," in *International Conference on Mining Intelligence and Knowledge Exploration* (Springer, 2018) pp. 259–267.

González, M., del Mar Alonso-Almeida, M., Avila, C., and Dominguez, D., "Modeling sustainability report scoring sequences using an attractor network," Neurocomputing **168**, 1181–1187 (2015).

Hochreiter, S. and Schmidhuber, J., “Long short-term memory,” Neural computation **9**, 1735–1780 (1997).

Jain, A. K., “Data clustering: 50 years beyond k-means,” Pattern recognition letters **31**, 651–666 (2010).

Maslov, S. and Sneppen, K., “Specificity and stability in topology of protein networks,” Science **296**, 910–913 (2002).

Ruths, J. and Ruths, D., “Control profiles of complex networks,” Science **343**, 1373–1376 (2014).

White, H. D., Wellman, B., and Nazer, N., “Does citation reflect social structure?: Longitudinal evidence from the globenet interdisciplinary research group,” Journal of the American Society for information Science and Technology **55**, 111–126 (2004).

Ziegler, A. and König, I. R., “Mining data with random forests: current options for real-world applications,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **4**, 55–63 (2014).