



**HAL**  
open science

## Regulation and signalization graph assembly through Linked Open Data

Marie Lefebvre, Jérémie Bourdon, Carito Guziolowski, Alban Gaignard

► **To cite this version:**

Marie Lefebvre, Jérémie Bourdon, Carito Guziolowski, Alban Gaignard. Regulation and signalization graph assembly through Linked Open Data. JOBIM 2017, Jul 2017, Lille, France. hal-01768420

**HAL Id: hal-01768420**

**<https://hal.science/hal-01768420v1>**

Submitted on 20 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Regulatory and signaling network assembly through Linked Open Data

Marie LEFEBVRE<sup>1</sup>, Jérémie BOURDON<sup>2</sup>, Carito GUZIOLOWSKI<sup>2</sup> and Alban GAINARD<sup>1</sup>

<sup>1</sup> Nantes Academic Hospital, CHU de Nantes, France

<sup>2</sup> LS2N - UMR 6004, University of Nantes, Ecole Centrale de Nantes, France

Corresponding author: marie.lefebvre@univ-nantes.fr, alban.gainard@univ-nantes.fr

## 1 Introduction and problem statement

Nowadays, huge efforts address the organization of biological knowledge through linked open databases. These databases can be automatically queried to reconstruct a large variety of biological networks such as regulatory or signaling networks. Assembling networks still implies manual operations due to (i) source-specific identification of biological entities, (ii) source-specific semantics for entity-entity relationships, (iii) proliferating heterogeneous life-science databases with redundant information and (iv) the difficulty of recovering the logical flow of a biological pathway due to the bidirectionality of chemical reactions. Homogenization of biological networks is therefore costly and error-prone. Existing tools such as the ReactomeFIPlugIn of Cytoscape 3.0[1] or STRING[2] allow to link entities to each other or to identify an entity's membership to a single pathway. Nevertheless, they are still limited in the global modeling aspects (logical rules inferred from the knowledge representation). Here, we present a framework to automate the assembly of regulatory and signaling networks in the context of tumor cells modeling.

## 2 Approach

Our framework is based on Semantic Web technologies. It addresses (i) the uniform identification of multi-source biological entities, (ii) the description of labeled directed graphs through RDF, and (iii) the use of BioPAX [3] as a semantic reference. We consider a list of target gene names or IDs as entry points. The first step consists in retrieving transcription factors (TFs) controlling these target genes. Then, the second step consists in considering the TFs as new entry points for the reconstruction algorithm. The full regulatory network is finally assembled by iteratively applying the second step until no new TFs can be found.

## 3 Demonstration

To assemble networks our algorithm queries PathwayCommons[4] through its SPARQL endpoint and retrieves a graph of TFs associated to target genes. We developed a web tool that displays the biological network assembly step by step, allowing users to interact with the reconstruction process and to visually shape the network. Through this web tool, it is also possible to launch a command line tool (Java) to address larger scale input gene lists. These tools have been deployed on the BiRD Cloud infrastructure. From a list of 1800 targets genes, we were able to assemble in less than 3 minutes a graph of 1474 nodes and 12303 edges.

## 4 Discussion and Conclusion

As future works, we aim at integrating drug-target informations (e.g. KEGGdrug, DrugBank) through SPARQL federated queries to get insights on (i) tumor cells growth and (ii) drug response on patient and cell lines gene expression data. Our tool is freely available at <https://github.com/symetric-group/bionets-demo>

## Acknowledgements

This work was supported by the BiRD bioinformatics facility, the SyMeTRIC project and the GRIOTE project.

## References

- [1] W., Guanming et al. A human functional protein interaction network and its application to cancer data analysis. *Genome Biology*, 11(5):R53, 2010.
- [2] D., Szklarczyk et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, 43:D447–D452, 2015.
- [3] E., Demir et al. The biopax community standard for pathway data sharing. *Nature Biotechnology*, 28:935–942, 2010.
- [4] E.G., Cerami et al. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res.*, 39:D685–90, 2011.