Profondeur

Visual Exploration and Analysis of Metabolic Networks

T Cameron Waller

2 December 2016

Course Project
University of Utah
Data Visualization
Alexander Lex
Fall 2016

Abstract

Metabolic networks are large and complex, involving many types of relations between many types of entities. Biologists and clinicians need access to the information from these networks in order to interpret experimental data that inform the diagnosis and treatment of human disease. The goal of this project is to develop methods that enable users to explore the metabolic network. In contrast to many existing technologies, this project employs automatic, dynamic layouts to represent subsets of the network. In future projects, these dynamic layouts will be necessary to represent subsets of the network that the user specifies by custom queries.

Contents

List of Figures

Chapter 1

Proposal

Project Information

Project Title: Profondeur
Project Website: https://tcameronwaller.github.io/profondeur/
Project Repository: https://github.com/tcameronwaller/profondeur
Author Name: Thomas Cameron Waller
Author EMail: cameron.waller@biochem.utah.edu
Author Identifier: 00409400

Introduction

Metabolism sustains all of life's diverse processes, from the release of energy from food to the synthesis of genomes. It is an integral part of the biological system of the cell with extensive connections throughout. Given its central role, it is no surprise that **defects in metabolism drive the pathology of many human diseases[1,2].** These include obesity, diabetes, cardiovascular disease, and cancer. More and more studies in biological research consider metabolic phenotypes and mechanisms, and metabolic profiling is an important component of precision medicine[3,4,5,6,7]. For these reasons **it is necessary to study metabolism from a holistic perspective, in its true context within the entire biological system.**

The vast complexity of metabolism exceeds the ability of human investigators to consider it without supporting methods. As part of the biological system of the cell, metabolism forms an intricate network of relationships between genes, transcripts, proteins, and metabolites. In the human genome, approximately 70,000 genes encode 70,000 transcripts and 30,000 proteins[6]. About 5,000 of these proteins are enzymes or transporters[8,9,10] that either catalyze chemical reactions or transport molecules between cells and sub-cellular compartments. Many more proteins regulate metabolic processes via the expression and activity of relevant genes, transcripts, and proteins. Approximately 30,000 endogenous small molecular metabolites[11] complete the collection. All of these genes, transcripts, proteins, and metabolites relate to each other in multiple ways, forming the metabolic network. This network varies in different tissues throughout the human body. This stunning complexity makes it prohibitive to track all processes without advanced methods to organize and access the information of the metabolic network. We argue that **many studies certainly miss interesting biological trends because the context is too elaborate**.

The author of this project has an interest in the study of metabolism. His primary research

project as a graduate researcher is in the visual analysis of the metabolic network. Indeed this course project is a small-scale start to a bigger research project in collaboration with Alexander Lex, Jared Paul Rutter, and others.

Project Objectives

The biological study of metabolism has potential to benefit the domains of biotechnology, pharmacology, and medicine. Biology's goal is to discover and describe life's functions and processes. As a basic science this domain is primarily informative, and it enables applications in many other domains. Biotechnology alters the metabolic systems of non-human hosts to synthesize hormones, antibiotics, or other drugs. Pharmacology identifies effective drug targets within the biological system, considers the broad effects of drugs on this system, and even exploits subtle side-effects of available drugs for use in new therapies. Knowledge of processes in health enables medicine to recognize perturbations, diagnose diseases, and propose appropriate therapies. **The goal of this project is to develop methods that support the study of metabolism by users in these domains (biology, biotechnology, pharmacology, and medicine).**

In their study of metabolism, a common task in these domains is to **access information about biological entities and the relations between them**. There are many different types of these **entities (genes, transcripts, proteins, metabolites)** and **relations (code, expression, regulation, transport, catalysis, reaction)**. They each have their own specific properties. In complex organisms, genes belong to chromosomes and chromosomal loci. In eukaryotes, proteins and metabolites localize to sub-cellular compartments called organelles. Many proteins have non-protein chemical cofactors that are essential for their functions. Reactions between metabolites belong to metabolic pathways of anabolism or catabolism and also have chemical properties such as rates and reduction or oxidation. A user might need to know any of this information for an entity, a relation, or combinations thereof. For example, a user might need to know which gene encodes a transcript and protein, which protein regulates the transcription of a gene, which metabolites react via a protein's enzymatic catalysis, or which catabolic reactions in the mitochondrial compartment involve proteins that require a specific cofactor. There are also many other possibilities. **Although much of this information is within current knowledge and is available in public databases, the scale of the biological system renders this access nontrivial.**

Another common task in these domains is the **analysis of experimental data that corresponds to entities in the biological system**. Modern -omics technologies describe the biological system at nearly-comprehensive scales. Available information includes mutations and copy variations in genes (genomics), abundance and splice variations of transcripts (transcriptomics), abundance and post-translational modifications of proteins (proteomics), and abundance of metabolites (metabolomics)[6]. It is very important to analyze and interpret these data in the real context of the biological system. For example, biological trends may only be apparent with consideration for network topology. **Correct analysis and interpretation of these experimental data rely on knowledge of the biological system.** This analysis becomes a great challenge at nearly-comprehensive scales.

The goal of this project is to develop computational methods for visual exploration and analysis of the metabolic network. It will address some of the specific needs of these domains. Many of the domain tasks correspond to queries against a database of network information. Analysis of experimental data in the context of network topology will require custom, probabilistic algorithms. Visual design will help the human user to recognize trends and patterns easily.

Summary

Goal: Develop computational methods for visual exploration and analysis of the metabolic network.
Users: Research scientists and engineers in biology, biotechnology, pharmacology, and medicine
Benefit: Advance knowledge of human health and ability to diagnose and treat human disease.
Domain Tasks:
1) Access information about biological entities and the relations between them.
2) Analyze experimental data in the context of the metabolic network.

Data

This project involves two categories of data. The first category includes the current knowledge about entities and relations in the biological system and their properties. This information is available in multiple public databases. This information is enormous and complex, and it is unlikely that all of the information will be available from a single source. To maintain a reasonable scope, this project will use a minor subset of this information. The second category is experimental data from -omics technologies. This sort of data is available in multiple public databases. This sort of data often has technical complications and requires processing. Again, to maintain a reasonable scope, this project probably will not use this category of data.

Data Processing

Some aspects of this project will involve substantial data processing. Information on the metabolic network will come from multiple databases. It will be necessary to organize this information in an appropriate format and assemble an accessible database. Experimental data from -omics technologies is also complex and sometimes requires extensive data processing before it is informative or interpretable.

Visualization Design

The prospective users of this project need to access information about biological entities and the relations between them. This domain task of accessing information involves multiple abstract tasks.

It is necessary to access information that matches the user's interest. With several types of entities, several types of relations, and many instances of each, the information of this network is vast and complex. The user needs to consider subsets of the metabolic system in appropriate detail (according to interest) without neglecting potentially-relevant context. Therein is a challenge, simplifying the information without losing important aspects. This process of selecting a relevant subset of the network will likely require multiple levels of user interaction. **Queries against a comprehensive database will allow the user to define her or his own interest in a fairly broad or general scope.** These queries will need to accommodate many variables in order to provide sufficient specificity. **A visual, interactive interface for dynamic queries will help the user to specify subsets of interest.** Search algorithms will then select relevant subsets of the data, ranking several possibilities according to query criteria. Then the user will be able to select from a list of search results. **A visual interface will help the user to distinguish between search results and select whichever is most appropriate.** After selecting a result from a query search, the user will need to refine the subset further according to her or his specific interest. The user will need to select which types of entities and relations to represent. The user might even need to select specific

entities and relations to represent. **A visual interface for a selection menu and an interactive visualization of the network data will help the user to refine interest subsets.**

It is necessary to communicate the relevant information to the user clearly and efficiently. The user may need to study small or large subsets of the metabolic network at various degrees of detail. These subsets will include a lot of information. From these subsets, users need to recognize connectivity and many properties of entities and their relations. **A visual representation of the network, such as a node-link diagram, will communicate this information to the user.** This diagram will require a dynamic layout. Alternative visual representations of the network might also work. Also, due to the complexity of the metabolic network and the limitations of available visual channels on a two-dimensional display, it will probably be necessary to represent information in multiple views. The user will be able to switch or scroll between views. Animations and other strategies will help to preserve the user's sense of context through these transitions.

It is necessary to associate experimental data with entities and relations from the metabolic network and to identify trends in these data. Keys (such as names of genes, proteins, or metabolites) will associate experimental data to specific entities in the database. Search algorithms will then handle experimental data similarly to other properties or attributes to recognize pathways with significant differences.

Must-Have Features

- interactive selection menu to include or remove specific types of entities and relations
- visual representation for subset of the metabolic network
- interaction with the network representation to include or remove specific elements
- visual representation of multiple types of properties on the network representation

Optional Features

This course project only entails 6 weeks of time for 1 developer. To maintain a reasonable scope, several features will be optional. These features will be more important for a longer-term, more complete project.

- comprehensive database for current knowledge of the metabolic network
- dynamic query interface
- search algorithms to identify pathways between entities matching query criteria
- analytical algorithms to identify network aspects with interesting trends in experimental data
- interface for selection from list of search results
- visual representation of experimental data on the network representation

Project Schedule

This course project entails 6 weeks of time for 1 developer.

Weekly Goals

1. 24 October - 30 October
    - visual design of interface, especially network representation
2. 31 October - 6 November
    - visual design of interface, especially network representation
    - collection, processing, and organization of network data
    - implementation of visual interface
3. 7 November - 13 November
    - implementation of visual interface
    - milestone review
4. 14 November - 20 November
    - implementation of visual interface
    - implementation of interaction
5. 21 November - 27 November
    - implementation of visual interface
    - implementation of interaction
6. 28 November - 2 December
    - screen-cast tutorial video
    - website
    - final submission

Chapter 2

Record

Project Information

Project Title: Profondeur
Project Website: https://tcameronwaller.github.io/profondeur/
Project Repository: https://github.com/tcameronwaller/profondeur
Author Name: Thomas Cameron Waller
Author EMail: cameron.waller@biochem.utah.edu
Author Identifier: 00409400

Introduction

The biological study of metabolism has potential to promote the domains of biotechnology, pharmacology, and medicine. Thereby, this study has potential to benefit human health. Metabolism is a complex system. The experimental study of metabolism involves large and complex data sets that are difficult to interpret. Interpretation requires consideration of the complex metabolic system. Prospective users are investigators in the domains of biology, biotechnology, pharmacology, and medicine. These users have very diverse backgrounds and very variable familiarity and expertise in metabolism, and data analysis.

Users need to access information about biological entities and the relations between them. This information gives relevant context for experimental design and interpretation of experimental results. This information is especially important for interpretation of large-scale data sets that cover the metabolic system broadly.

Current Technology

This section includes a concise depiction of the current technology that is relevant to this project with summaries and evaluations of a set of tools. This set of tools is not comprehensive.

Databases

There are many databases that provide information relevant to the metabolic network. A few notable examples are the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/)[12],

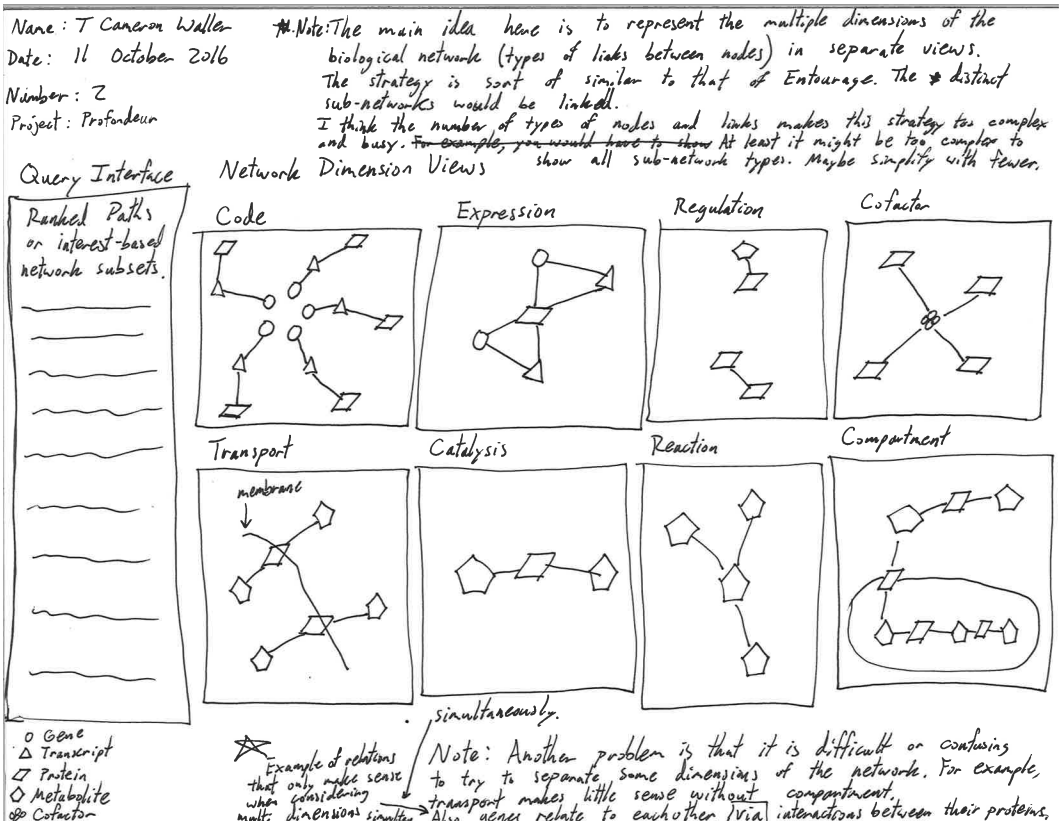

Figure 2.1: Representations for multiple types of relations between multiple types of entities

the MetaCyc Metabolic Pathway Database (http://www.metacyc.org/)[13], and the Universal Protein Resource (UniProt, http://www.uniprot.org/)[14].

Metabolic Models

Systems Biology assembles intricate computational models in order to study cellular functions in a holistic way[15,16]. Communities of scientists thoroughly assemble and maintain computational models that comprise current knowledge of the metabolic systems of multiple species. A current model of high quality is available for human metabolism[10,8].

Visualization and Exploratory Analysis

KEGG Atlas (http://www.kegg.jp/kegg/atlas/)[12] is a web application that enables users to explore metabolic maps from KEGG. The user can select from a large set of available maps for subsets and pathways in metabolism. Maps in the atlas use a static, spatial layout that sometimes distorts the relations between metabolites. Two metabolites may be far apart on the map although they relate by a single reaction. The atlas also does not allow the user to specify custom subsets of the network. This design renders KEGG Atlas inappropriate for visualizing subsets of the metabolic network that traverse typical pathways.

Escher (https://escher.github.io/)[17] is a web application that enables users to draw metabolic maps and visualize experimental data on these maps. It imports information from models of

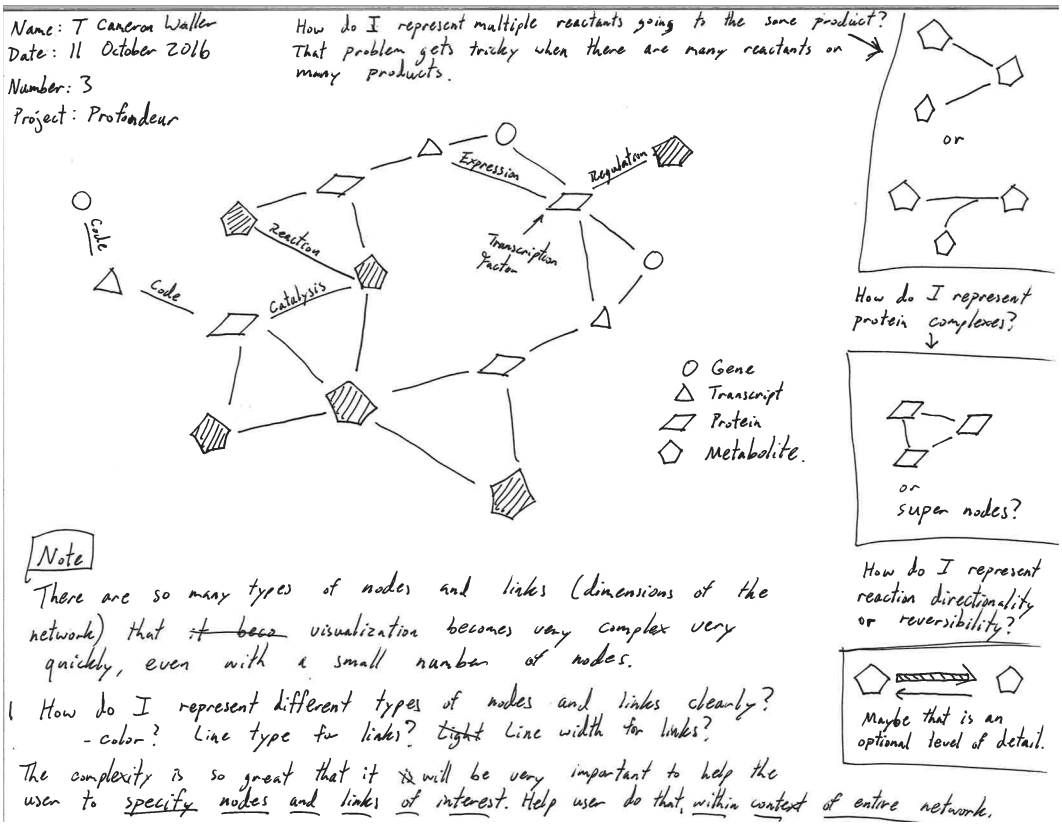

Figure 2.2: Representations for multiple types of relations between multiple types of entities

metabolism that the systems biology community develops and maintains, thereby favoring accuracy, and allows the user to select portions of these models to include in the maps. Escher's fundamental design requires the user to select portions of the metabolic model and position these in a static layout. This design renders Escher inappropriate for a query-based exploration system that would need to represent automatically and dynamically a broad variety of custom subsets of the metabolic network.

Problems/Objectives

Problems

How is it possible to represent the metabolic network accurately and clearly using custom subsets of the network and automatic layouts of node-link graphs?

Users often have sets of multiple entities (metabolites, proteins, transcripts, genes) of interest, such as those for which they have experimental data, in the metabolic system. Users need to know how these entities relate to each other. Many of these relations transcend or traverse typical, common, traditional pathways. There are many dimensions, levels, or layers of relations in the metabolic system. At the most basic level, users need to know how multiple metabolites relate to each other through chemical reactions. These reactions are a single dimension of the metabolic system. Types of entities other than metabolites also participate in this dimension of the metabolic

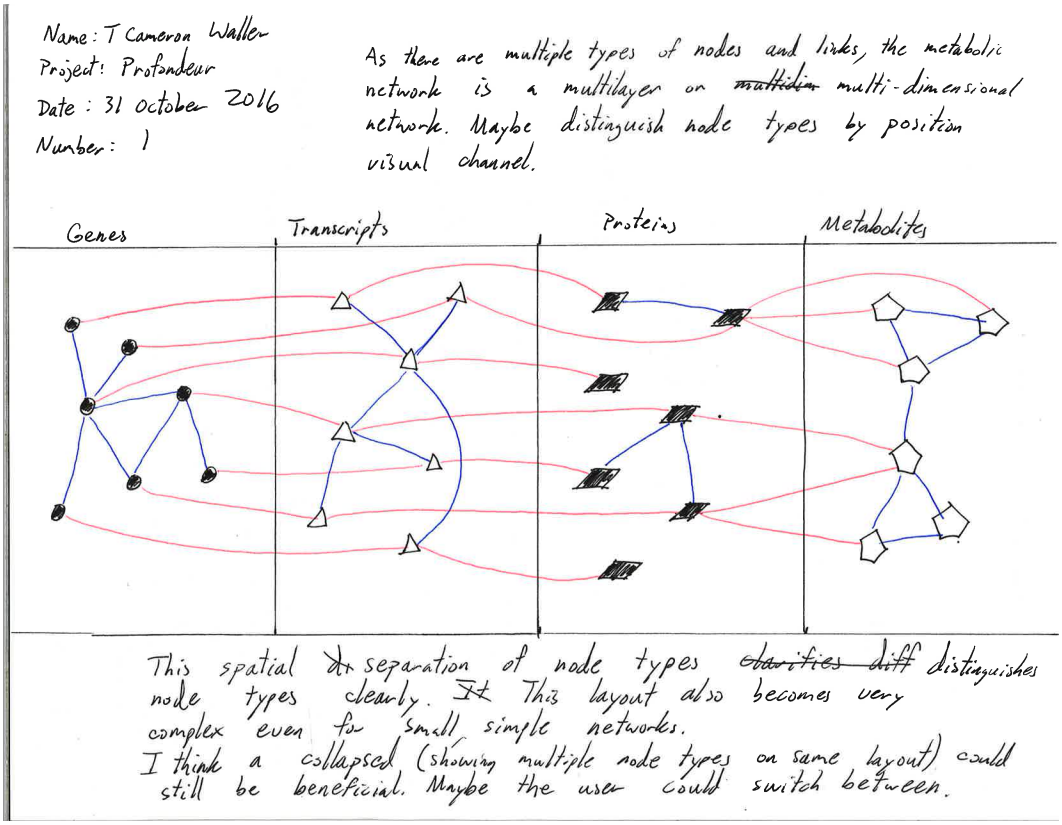

Figure 2.3: Representations for multiple types of relations between multiple types of entities

system. Users need to know the proteins that either catalyze reactions or transport metabolites between cellular compartments. Users need to know the transcripts and genes that encode these proteins.

Users also need to know about properties that might be relevant to relations between entities. Maybe multiple entities share a property. These properties might inform biological interpretation.

Data about the metabolic system describe properties of and relations between genes, transcripts, proteins, and metabolites. Experimental data describe modifications to genes, transcripts or proteins. Experimental data also describe the abundances of transcripts, proteins, or metabolites.

In the most common use scenario, the user needs to interpret data for a single type of entities, such as metabolites. In a higher-level use scenario, the user needs to interpret data for two types of entities simultaneously, such as metabolites and transcripts.

The metabolic system is large and complex. It is impractical, unreasonable, and useless for a human user to attempt to conceptualize the entire system simultaneously. The user needs methods to select subsets of the metabolic system that are relevant to the user's interest (such as an experiment). This selection process itself might involve many criteria and be complex. Methods for this selection are excellent design opportunities. Many possible factors could influence their effectiveness.

Even subsets of the metabolic system might be excessively complex. The user needs methods to explore these subsets. The user needs to recognize important information from the metabolic system easily. The user needs a summary or overview to give context. The user needs details on demand or details according to interest. Methods for this exploration are excellent design opportunities.

Experimental data convey information about entities in the metabolic system. Users need meth-

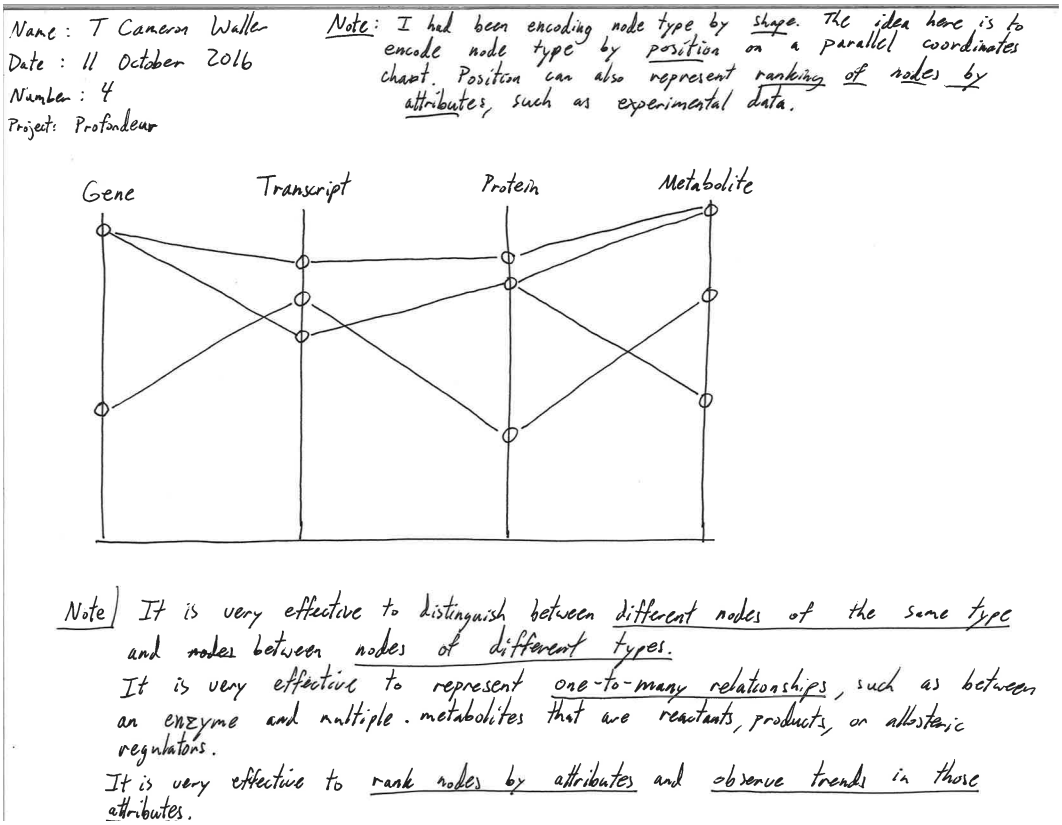

Figure 2.4: Alternative representation with parallel coordinates chart

ods to recognize trends in these data that are dependent on aspects of the metabolic system. Methods for this data communication are excellent design opportunities.

Data

Systems biology is a mature sub-domain of biology that considers and studies biological systems from an holistic perspective. This sub-domain commonly develops intricate mathematical models and uses these in computational simulations. These models include information about metabolites (the small molecules of metabolism), chemical reactions between them, the rates and directionality of these reactions, the cellular compartments where these reactions happen, the protein enzymes that catalyze these reactions, and the transcripts and genes that encode these proteins. The ultimate goal is for these models to represent the entirety of the biological system, and many models are comprehensive collections of current biological knowledge.

An active community in systems biology develops and curates these metabolic models as well as develops and maintains extensive resources for using them. The Systems-Biology-Markup-Language (SBML) (http://sbml.org) relates to the Extensible-Markup-Language (XML) and is an open standard and format for the representation of computational models for biological systems. Robust models in SBML are available for the metabolic systems of many organisms, including *Mus musculus* (mouse) and *Homo sapiens* (human). For this project, I chose to derive data from these metabolic models because they represent the most comprehensive and orderly (mass and charge-balanced reactions,

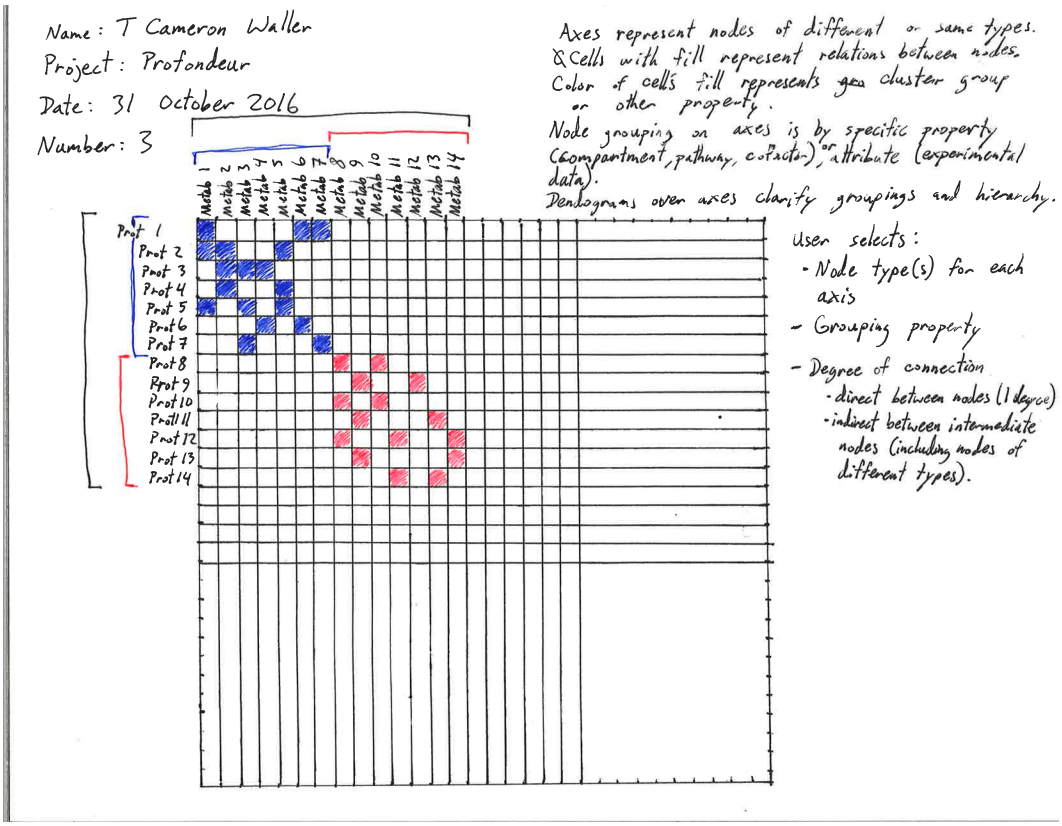

Figure 2.5: Alternative representation with matrix or grid chart

compartments) collections of current biological knowledge on the metabolic system.

COBRApy (https://opencobra.github.io/cobrapy/)[18] is an open-source tool from the community for using metabolic models in Python. I will use COBRApy along with libSBML (http://sbml.org/Software/libSBML)[19] to convert data for metabolic models from SBML format to JavaScript Object Notation (JSON) format. I will use these data in JSON format in the web application, ultimately restructuring it for graph queries and representation in node-link graphs. To pilot the prototype of this web application, I used a small model of central metabolism in *Escherichia coli* (Gram-negative bacterium) that came as a model for demonstration in COBRApy. I modified this model to simplify it and derived the appropriate structure of nodes and links manually.

Exploratory Data Analysis

I am familiar with the metabolic network and visual representations of it from various sources. These include databases and charts that I mentioned in Section "Related Work". I have also drawn many of my own visual representations of the metabolic network using InkScape. I gathered information for these representations from databases such as KEGG, MetaCyc, and UniProt.

For this project, I specifically use data from metabolic models. The structure of these data was less familiar to me. I explored metabolic models to learn about the information that they include and how to interpret this information correctly. I used COBRApy (ref) for this exploration of the raw data from the metabolic model. I also used a visual editor (http://www.jsoneditoronline.org/)

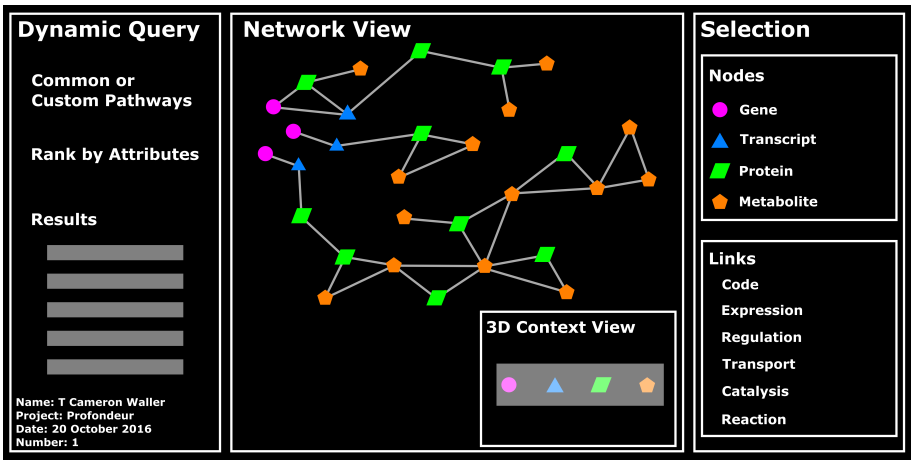

Figure 2.6: Early idea for interface design

to examine the metabolic model in JSON format.

Design

This project's repository (https://github.com/tcameronwaller/profondeur) has a directory with title "design". Within this directory are scans of sketches that I drew or recorded during the design process. Also in this directory is a text file with title "note sketch". This file contains notes on several important or influential sketches and the relevant design concepts.

Network of Multiple Dimensions

I began the design process with a grand idea of the complexity of the biological system and the metabolic network within this system. In reality, the metabolic network involves multiple types of entities (genes, transcripts, proteins, metabolites) and multiple types of relations between them (encoding, expression, regulation, transport, catalysis, reaction). All of these entities and relations have their own properties that may be interesting in a biological context.

I had a multidimensional or multilayer network, and I needed to figure out how to visualize it. I sketched several ideas for how to represent this multidimensional network, as Figures 2.1, 2.2, 2.3, illustrate. I even considered alternative types of charts in order to accommodate the complexity as Figures 2.4 and 2.5 illustrate. I imaged that an interface might look something like Figure 2.6 illustrates.

Network of Reactions between Metabolites

Eventually, on 1 November 2016, I consulted with a domain expert in the biological study of metabolism, Jared P Rutter. Jared encouraged me to begin with a network of reactions between metabolites in order to simplify the problem. I considered designs to illustrate reactions between metabolites.

I decided to represent metabolites as nodes and reactions as links between those nodes. Around that time, as I interacted more with an actual metabolic model, I realized that multiple metabolites can participate as reactants or products of the same reaction. From my background in biology, I

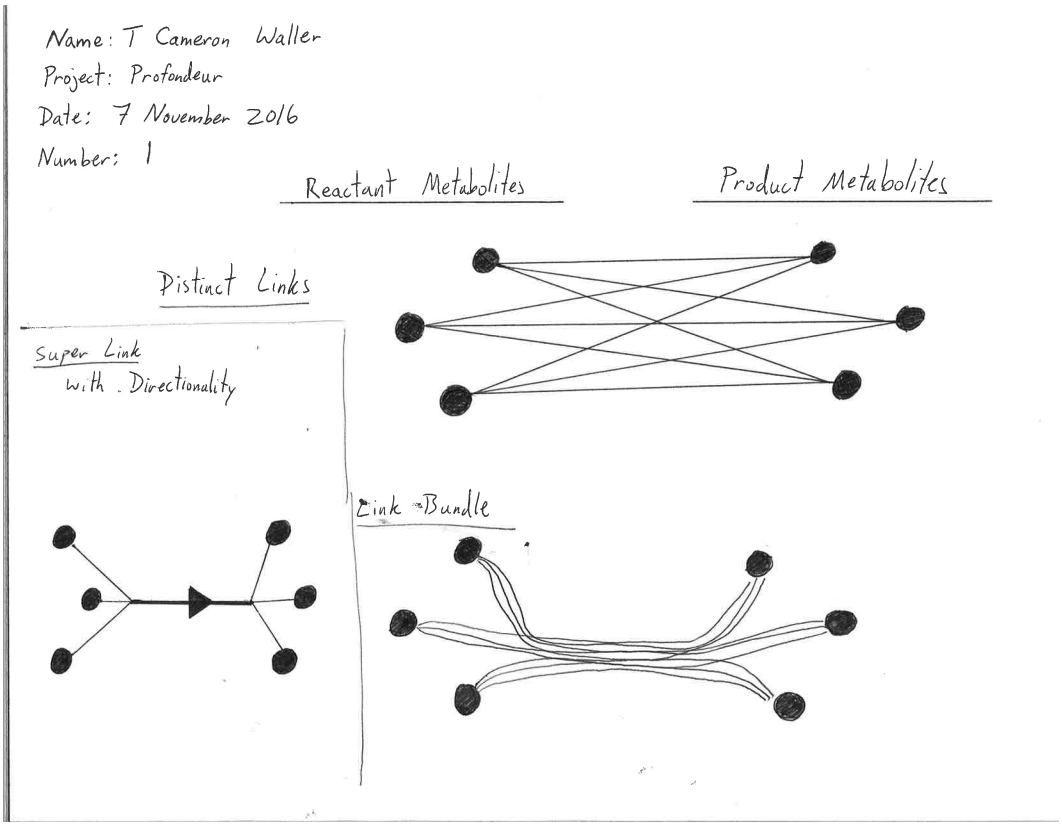

Figure 2.7: Link bundles and super links for reactions between metabolite nodes

already knew that, of course, but I suppose I had not really considered how to accommodate that complexity in the network representation. To simplify, I considered using either link bundles or super links as Figure 2.7 illustrates. I also considered the idea of using super nodes for reactant and product metabolites as Figure 2.8 illustrates, but I realized that super nodes would abolish much of the connectivity of the network as Figure 2.9 illustrates.

Node Replication for Prolific Metabolites

As I worked with the data for the metabolic model more, I realized that the metabolic network can be very complex because some metabolites are very prolific. A single metabolite can participate as a reactant or product in multiple reactions. It can simplify the representation of the network to replicate the nodes for these prolific metabolites, as Figure 2.10 illustrates. In contrast, it is a more accurate representation of the connectivity of the network to preserve the centrality of these metabolites, as Figure 2.11 illustrates. The connectivity of some prolific metabolites (pyruvate) might be more interesting than that of some other metabolites (such as water, proton, carbon dioxide, ADP, ATP, NAD, NADH, NADP, NADPH). I decided that it would be beneficial to allow the user to decide whether or not to replicate nodes for metabolites. My final design sketch, Figure 2.12 illustrates a table interface to allow the user to select metabolites with high-degrees to replicate.

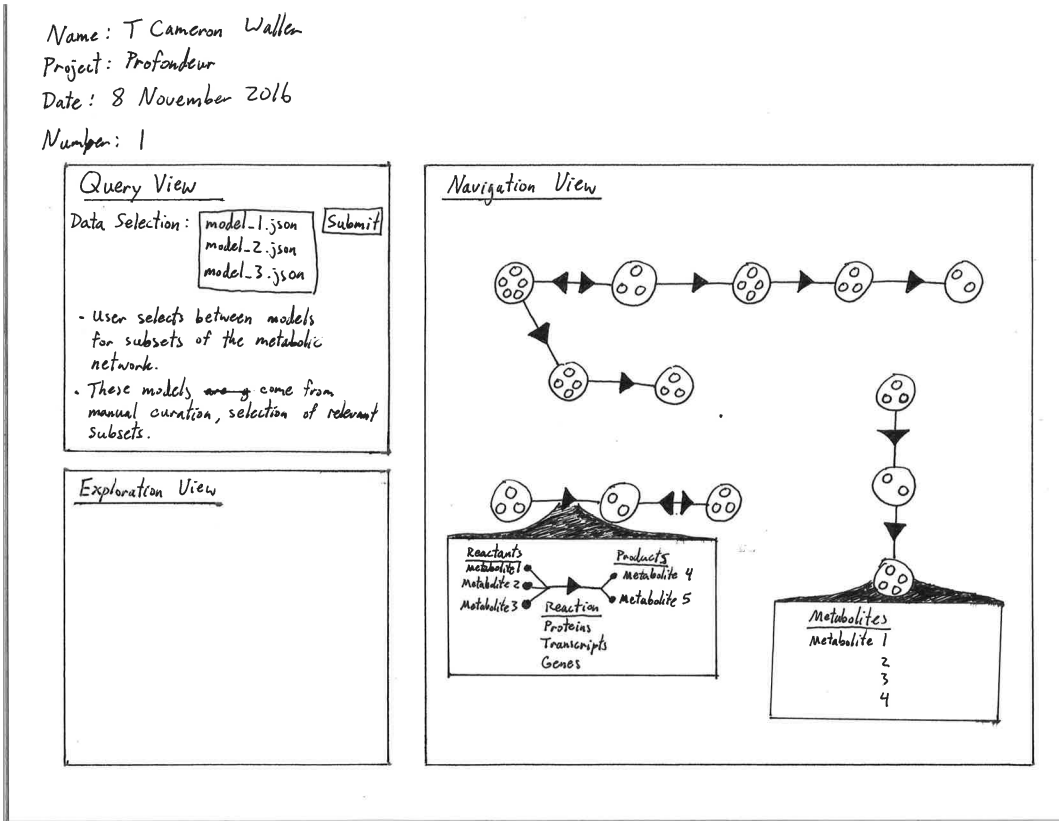

Figure 2.8: Super nodes for reactant and product metabolites

Future Goals

Figure 2.13 illustrates some of my ideas for future goals. The idea is to provide the user with more information about the network according to the user's interest and selections.

Implementation

I implemented the design that Figure 2.13 illustrates (with the exception of showing degrees for metabolites in the entire metabolic model since that is beyond scope). I implemented this interface as a web application using Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript. The code is available in the project repository (https://github.com/tcameronwaller/profondeur), and a demonstration of the interface is available as a website (https://tcameronwaller.github.io/profondeur/).

Evaluation

The prototype demonstrates the feasibility of using an automatic and dynamic layout algorithm to represent node-link diagrams of subsets of the metabolic network. Such layout algorithms are necessary to allow users to visualize custom subsets that they define through queries. The prototype also demonstrates the utility in enabling the user to select which (if any) nodes to replicate in the network. The decision to replicate nodes balances accurate representation of the network against

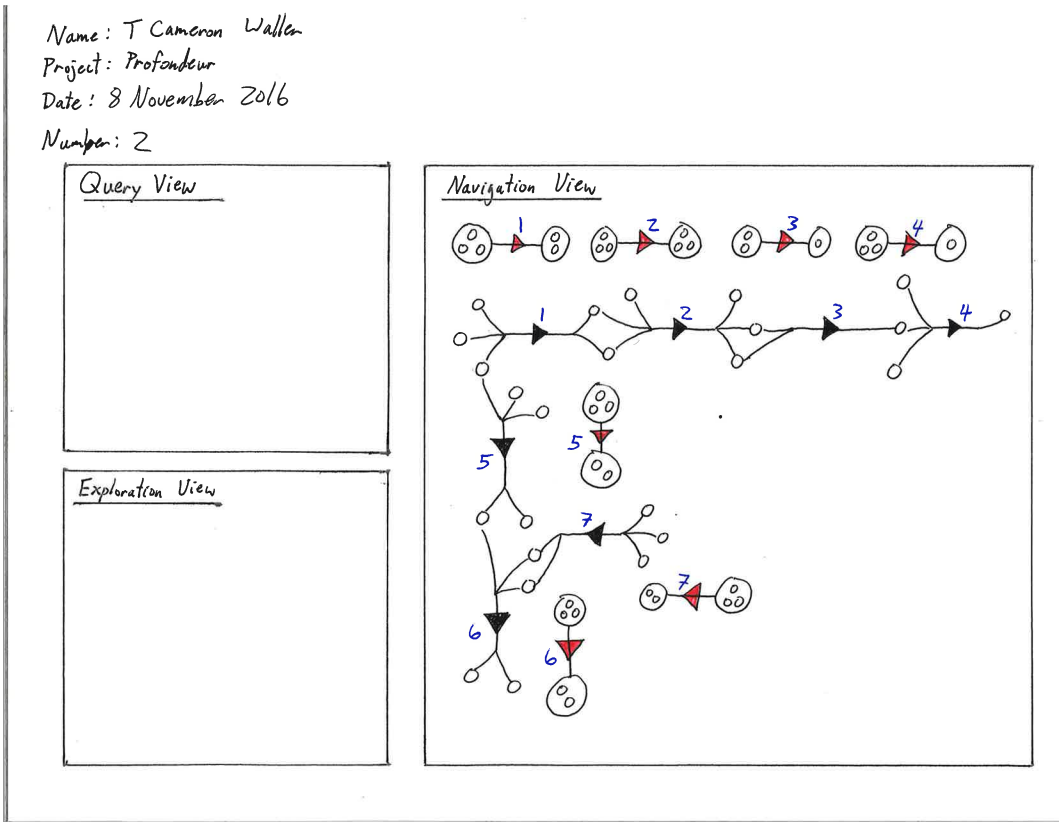

Figure 2.9: Problem with super nodes for reactant and product metabolites

simplification of this representation. The decision really depends on the interest of the user, and I think it is best to leave it to the user.

Within the scope of the project, I consider the prototype to solve some important problems, and to consider some of the most critical requirements of this project. In order to progress further, it was very important to evaluate the feasibility of automatic, dynamic layouts of subsets of the metabolic network from custom queries. Now that this layout is a possibility, the project can progress in the future.

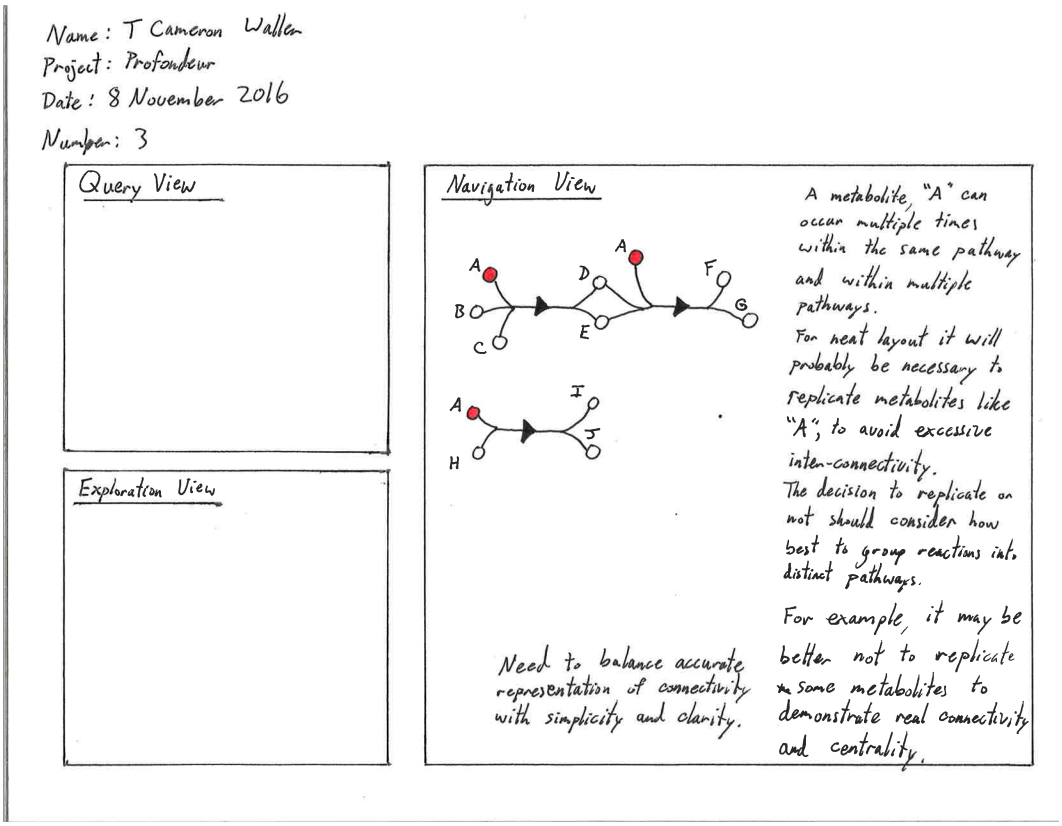

Figure 2.10: Replication of high-degree nodes to simplify network

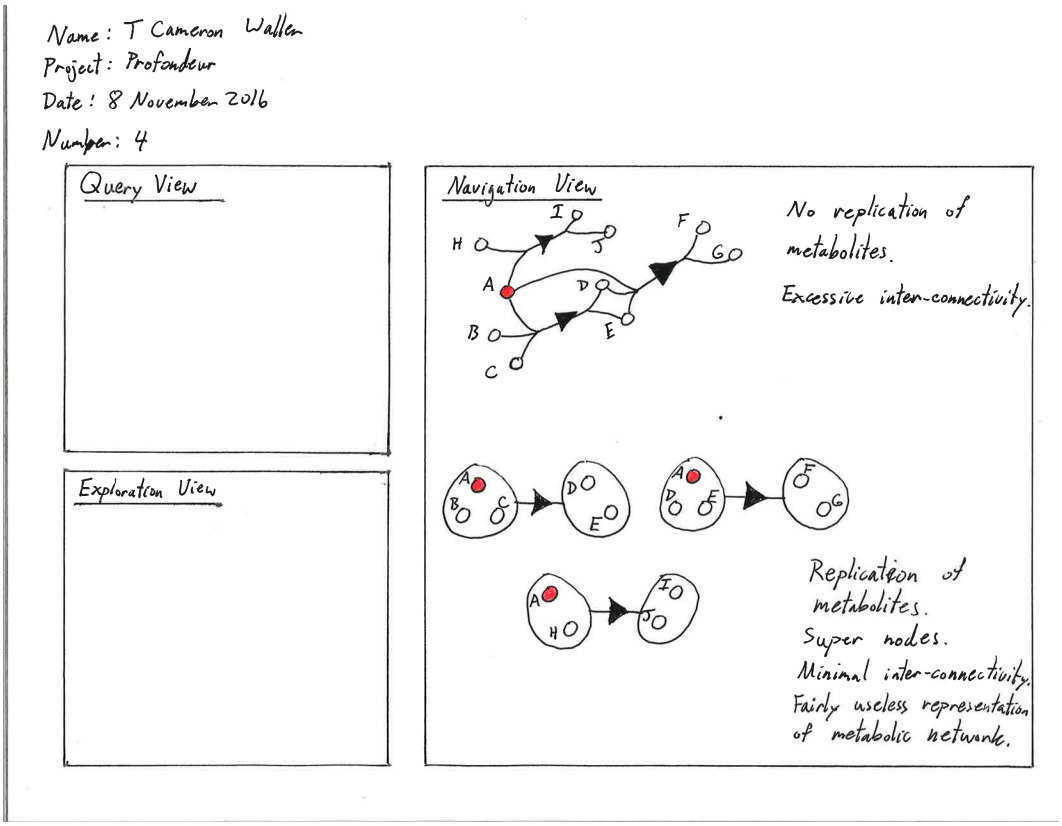

Figure 2.11: Preservation of connectivity in the network

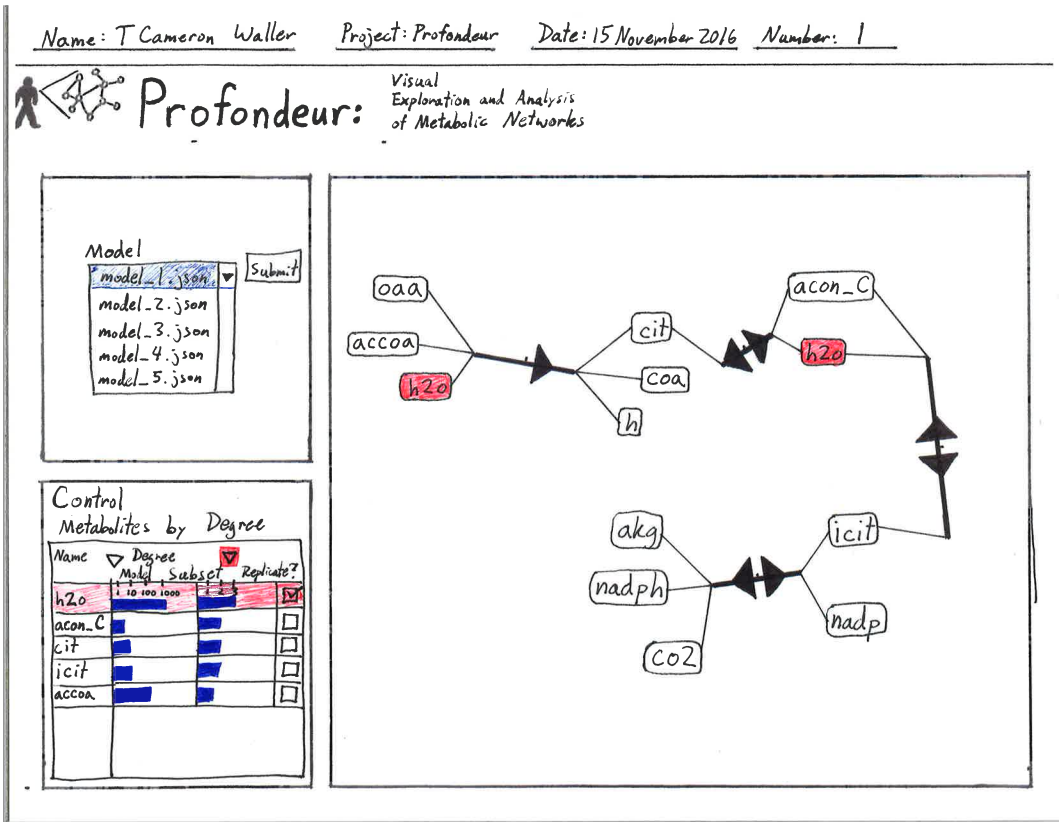

Figure 2.12: User selection of high-degree nodes to replicate

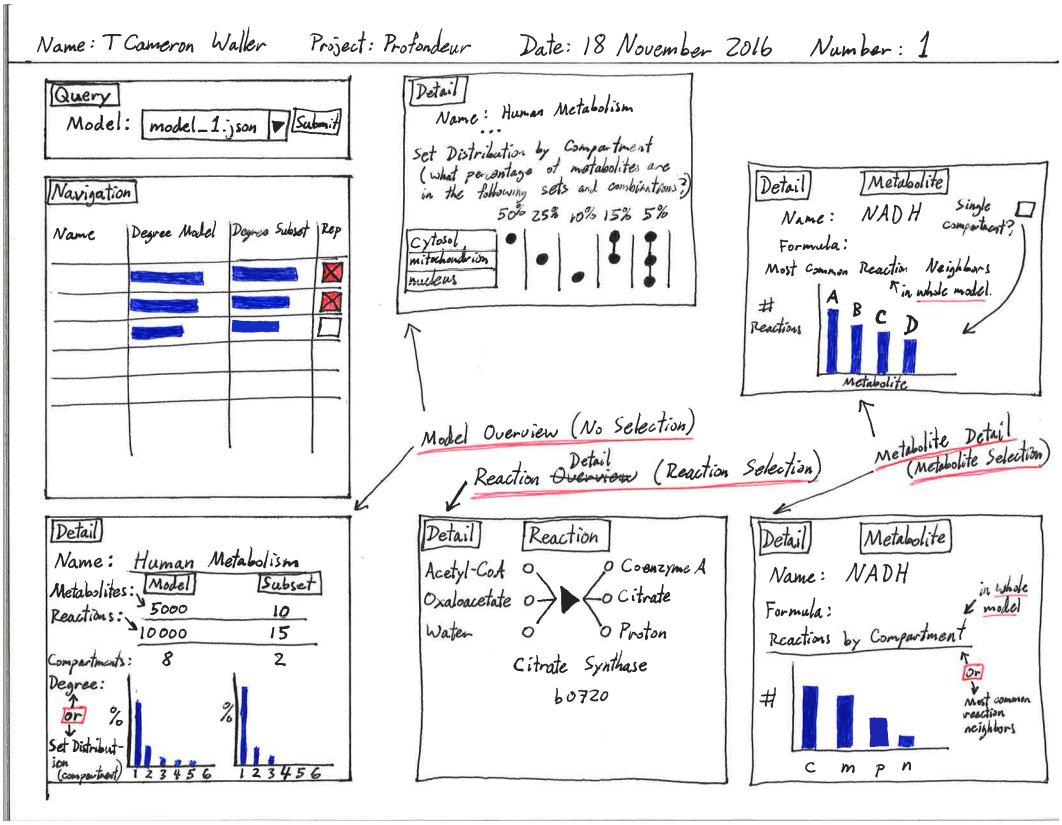

Figure 2.13: Detail views with additional information about the network

Bibliography

[1] Agatha A. van der Klaauw and I. Sadaf Farooqi. "The hunger genes: pathways to obesity". In: *Cell* 161.1 (Mar. 26, 2015), pp. 119–132. ISSN: 1097-4172. DOI: 10.1016/j.cell.2015.03.008.

[2] David W. Haslam and W. Philip T. James. "Obesity". In: *Lancet (London, England)* 366.9492 (Oct. 1, 2005), pp. 1197–1209. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(05)67483-1.

[3] Kirk Beebe and Adam D. Kennedy. "Sharpening Precision Medicine by a Thorough Interrogation of Metabolic Individuality". In: *Computational and Structural Biotechnology Journal* 14 (2016), pp. 97–105. DOI: 10.1016/j.csbj.2016.01.001.

[4] M. Benson. "Clinical implications of omics and systems medicine: focus on predictive and individualized treatment". In: *Journal of Internal Medicine* 279.3 (Mar. 2016), pp. 229–240. ISSN: 1365-2796. DOI: 10.1111/joim.12412.

[5] Francis S. Collins and Harold Varmus. "A new initiative on precision medicine". In: *The New England Journal of Medicine* 372.9 (Feb. 26, 2015), pp. 793–795. ISSN: 1533-4406. DOI: 10.1056/NEJMp1500523.

[6] Xiang Zhang, Jan A. Kuivenhoven, and Albert K. Groen. "Forward Individualized Medicine from Personal Genomes to Interactomes". In: *Frontiers in Physiology* 6 (2015), p. 364. DOI: 10.3389/fphys.2015.00364.

[7] Eric J. Topol. "Individualized medicine from prewomb to tomb". In: *Cell* 157.1 (Mar. 27, 2014), pp. 241–253. ISSN: 1097-4172. DOI: 10.1016/j.cell.2014.02.012.

[8] Neil Swainston et al. "Recon 2.2: from reconstruction to model of human metabolism". In: *Metabolomics: Official Journal of the Metabolomic Society* 12 (2016), p. 109. ISSN: 1573-3882. DOI: 10.1007/s11306-016-1051-4.

[9] Natapol Pornputtapong, Intawat Nookaew, and Jens Nielsen. "Human metabolic atlas: an online resource for human metabolism". In: *Database: The Journal of Biological Databases and Curation* 2015 (2015), bav068. ISSN: 1758-0463. DOI: 10.1093/database/bav068.

[10] Ines Thiele et al. "A community-driven global reconstruction of human metabolism". In: *Nature Biotechnology* 31.5 (May 2013), pp. 419–425. ISSN: 1546-1696. DOI: 10.1038/nbt.2488.

[11] David S. Wishart et al. "HMDB 3.0–The Human Metabolome Database in 2013". In: *Nucleic Acids Research* 41 (Database issue Jan. 2013), pp. D801–807. ISSN: 1362-4962. DOI: 10.1093/nar/gks1065.

[12] Minoru Kanehisa et al. "KEGG as a reference resource for gene and protein annotation". In: *Nucleic Acids Research* 44 (D1 Jan. 4, 2016), pp. D457–462. ISSN: 1362-4962. DOI: 10.1093/nar/gkv1070.

[13] Ron Caspi et al. "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases". In: *Nucleic Acids Research* 44 (D1 Jan. 4, 2016), pp. D471–480. ISSN: 1362-4962. DOI: 10.1093/nar/gkv1164.

[14] UniProt Consortium. "UniProt: a hub for protein information". In: *Nucleic Acids Research* 43 (Database issue Jan. 2015), pp. D204–212. ISSN: 1362-4962. DOI: 10.1093/nar/gku989.

[15] Edward J. O'Brien, Jonathan M. Monk, and Bernhard O. Palsson. "Using Genome-scale Models to Predict Biological Capabilities". In: *Cell* 161.5 (May 21, 2015), pp. 971–987. ISSN: 1097-4172. DOI: 10.1016/j.cell.2015.05.019.

[16] Aarash Bordbar et al. "Constraint-based models predict metabolic and associated cellular functions". In: *Nature Reviews. Genetics* 15.2 (Feb. 2014), pp. 107–120. ISSN: 1471-0064. DOI: 10.1038/nrg3643.

[17] Zachary A. King et al. "Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways". In: *PLoS computational biology* 11.8 (Aug. 2015), e1004321. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004321.

[18] Ali Ebrahim et al. "COBRApy: COnstraints-Based Reconstruction and Analysis for Python". In: *BMC systems biology* 7 (Aug. 8, 2013), p. 74. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74.

[19] Benjamin J. Bornstein et al. "LibSBML: an API library for SBML". In: *Bioinformatics (Oxford, England)* 24.6 (Mar. 15, 2008), pp. 880–881. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btn051.