2019

Do Metabolic Networks Follow a Power Law? A PSAMM Analysis

Ryan Geib
University of Rhode Island, ryan_geib@my.uri.edu

Lubos Thoma
University of Rhode Island, thoma@math.uri.edu

See next page for additional authors

Our first goal with this analysis was to determine whether we could find support for the power-law hypothesis in metabolic networks. We began by using the *vis* command in PSAMM to generate files from the curated metabolic model folders to represent the *no-fpp* network, which accounts for all element transfers in the metabolic network, and then we used the *primarypairs* command to make files for the *fpp* network, which accounts for the flow of individual elements in the metabolic network. Next, we used both the igraph and poweRlaw packages in R to extract the metabolite degree data from the network files and perform the power-law fitting, respectively. The result of this was that only one out of eight models had a p-value above the 0.1 threshold for the no-fpp representation and only three out of eight models had significant p-values in the fpp representation. Meaning, the power-law hypothesis was not supported for the *no-fpp* representation of metabolic networks. However, we noticed that the networks tended to have distributions which followed the fitting well in the beginning, then rose above the fitting in towards the tail of the distribution. This observation led us to wonder whether there was a more accurate power-law fitting that we could perform on the networks.

In order to address this question, we performed a double fitting analysis on each of the networks. We conducted this analysis by choosing a cutoff value (98%) to segment the data and consider each distribution separately. This means that we created two data sets from the original degree data, one with the bottom 98% of nodes in terms of degree and one with the top 2% of nodes, and then did separate power-law fittings for each data set. We get two p-values from this process: one for the fitting on the bottom 98% and one for the fitting on the top 2%. For the *no-fpp* representation, six out of eight models had a significant p-value for the first fitting and all eight models had a significant p-value for the second fitting. In the *fpp* network, only two out of eight models got a positive result for the first fitting and all eight models had significant p-values for the second fitting. This distribution offers a better characterization of the data, especially in the no-fpp representation, but it will require more work to determine an objective cutoff value that produces good results for metabolic models in general.

As a further question, we wanted to see whether the elemental subnetworks (i.e. only nodes that transfer the specific element in a reaction are considered) followed the power law distribution. To determine this, we used the *element* option in PSAMM to generate *fpp* and *no-fpp* model files for the carbon, nitrogen, phosphorous, and sulfur subnetworks in each model. This approach produced more positive results than fitting the network as a whole, especially for the *fpp* models. For the *fpp* carbon subnetwork, there was support for six out of eight models following a power law distribution. This number was five out of eight for nitrogen, six out of eight for phosphorous, and seven out of eight for sulfur. The *no-fpp* plots of the iNJ661m metabolic model for *M. tuberculosis* are shown to the left with the fittings overlain, as this model and representation had a significant p-value for all subnetworks.