Fig. 1. Bipartite Network schematic. A bipartite network (b) made of 2 data sets the “circles”, and the “rectangles”. Projections in the “circle” space (a) and in “rectangle” space (c).

One node can only be connected to a node of the other data type. We used a bipartite networks to construct the relationships of our data. From the bipartite network, one can project the data onto either of the data spaces (Figure ?? a,c). In either single dataset space, the nodes are connected to one another through a vertex of the other space. By ignoring the different types of data, all network properties described above remain valid on the bipartite network (as a single data set network) and on either projection. This type of network gives us three degree-distributions, one for each projection, and one for the bipartite network. Each degree distribution shows how many links each node has. Nodes in a projection of a bipartite network are connected if they share at least one node in the other group. This gives us the ability to visualize connections within a group.

2.3 Human Disease Networks

In recent years there has been a trend toward studying disease through network based analysis of various systems of connections between diseases. The result was the Human Disease Network (HDN) ?? . The nodes in the HDN represent human genetic disorders and the edges represent various connections between disorders, such as gene-gene or protein-protein interactions, to name a few. The HDN is helpful in visualizing connections among human disorders on a large scale. The underlying connections of the HDN contribute to the understanding of the basis of disorders, which in turn leads to a better comprehension of human diseases.

One study by Goh, et al. ??, explored the HDN built on genes shared by different diseases. Another study, which is similar in some ways to ours, by Li et al. ?? traced the SNPs connecting disease traits. In 2009, Silpa Suthram et al. ?? found that when diseases were compared by an analysis of disease-related mRNA expression data and the human protein interaction network, there were significant similarities between some diseases and between some drug treatments. In 2009, Barrenas et al. ?? further studied the genetic architecture of complex diseases by doing a GWAS, and found that complex disease genes are less central than the essential and monogenic disease genes in the human interactome. In the present work, we expand our study to include not only disease traits, but also behaviors and normal variations in humans, such as hair color, and explore large portions of non-coding variants in the human genome. Links between PTs are based on overlapping biological pathways (Section ??).

3 Pathway-based Human Phenotype Network

In this paper, we chose to mesh the methods and results sections, as we present multiple different algorithms (i.e. to build, filter, and identify the modules in the PHPN). Each subsection presents and applies a new method, building on the resulting network of the previous one.

3.1 Building the PHPN

Here we describe our method to construct a network of human phenotypes (traits and diseases) based on shared biological pathways of the associated genes. This is accomplished by linking genes to phenotypes (PTs) from hundreds of GWAS catalogue at NHGRI. Genes were further linked to pathways (PWs) using Reactome. By building these associations, we were able link phenotypes with genes involved in the same pathways. The steps used to build the network are illustrated in Figure ?? and described as follow:

1. From the NHGRI catalog, extract all PTs and link them to their mapped genes;
2. From Reactome, extract all genes in the database and link them to their associated pathways;
3. Match the genes associated to each phenotype to their associated pathways;
4. Connect PTs with overlapping pathways with an undirected edge, setting edge weight as the number of overlapping pathways.