Developing a Ranking System for Bayesian Kinase Network Models

Alex Joyce, William Ryan, Ali Imami, and Robert McCullumsmith
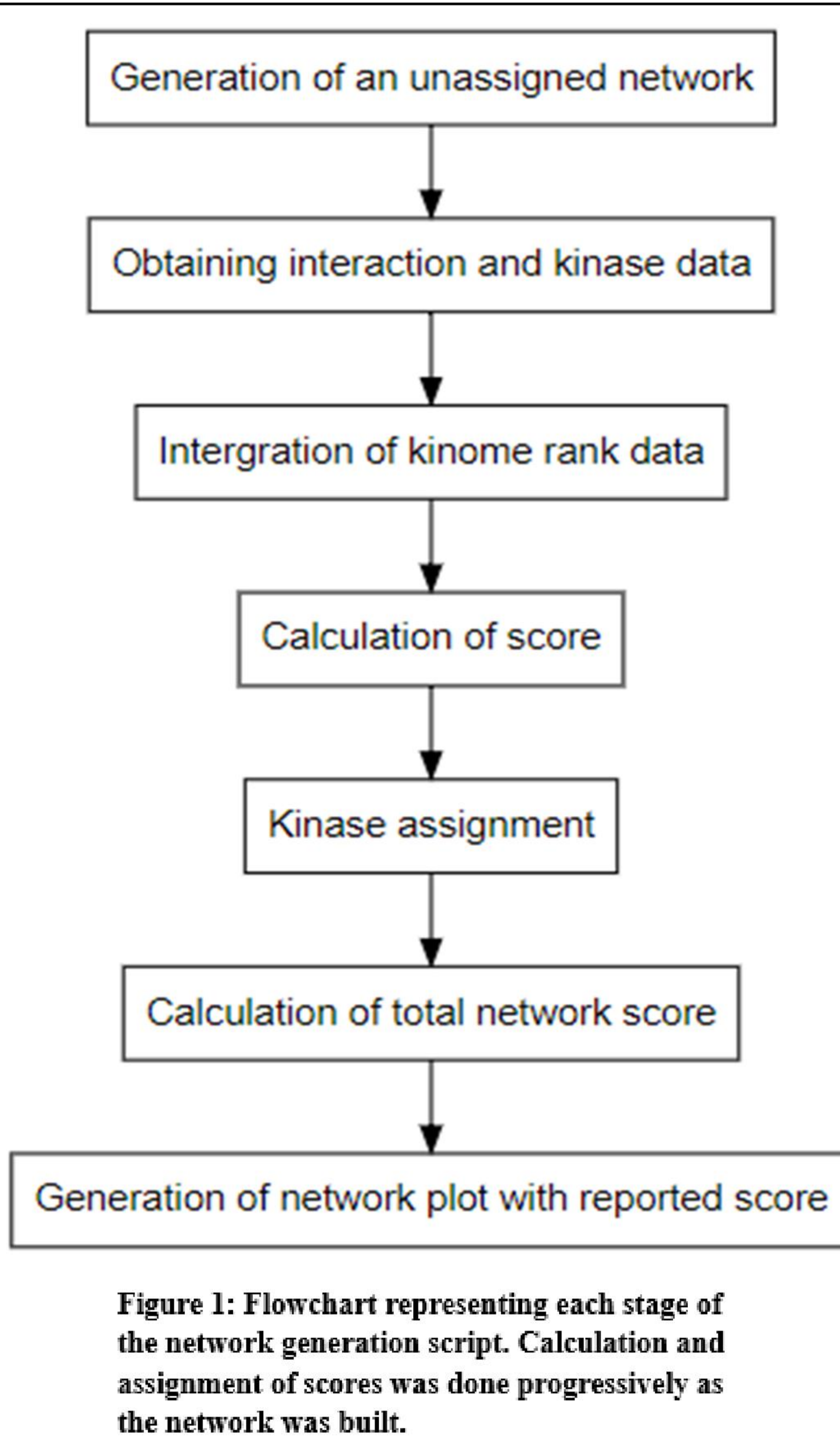### The University of Toledo College of Medicine and Life Sciences Department of Bioinformatics

Introduction

- There has been an increase in regulatory network research over the years.
- KINNET is an R package deigned by the McCullumsmith lab in order to generate Bayesian kinase networks. However, a large number of networks can be generated from any given data set, making analysis difficult.
- Here we present a simple ranking system in order to select the optimal network.

Methods

- Networks were generated using a small script based off of the KINNET package using test data. The workflow for this process is displayed in figure 1.
- Assignment scores were generated using a two-parameter approach:
  - Normalized and scaled percent INDRA statements represent relative knowledge for each kinase as well as network information.
  - Total number of interactions each kinase has per node represent a kinase's relative activity as well as network complexity.
- The score itself (figure 2) was determined by adding one to the knowledge value and raising it to the power of the total interaction number.
- Dark kinases, the bottom third of the kinome in terms of knowledge, had their scores multiplied by negative one in order to differentiate them from light kinases.
- Total network scores were calculated by summing the absolute value of all individual node scores.

Figure 1: Flowchart representing each stage of the network generation script. Calculation and assignment of scores was done progressively as the network was built.

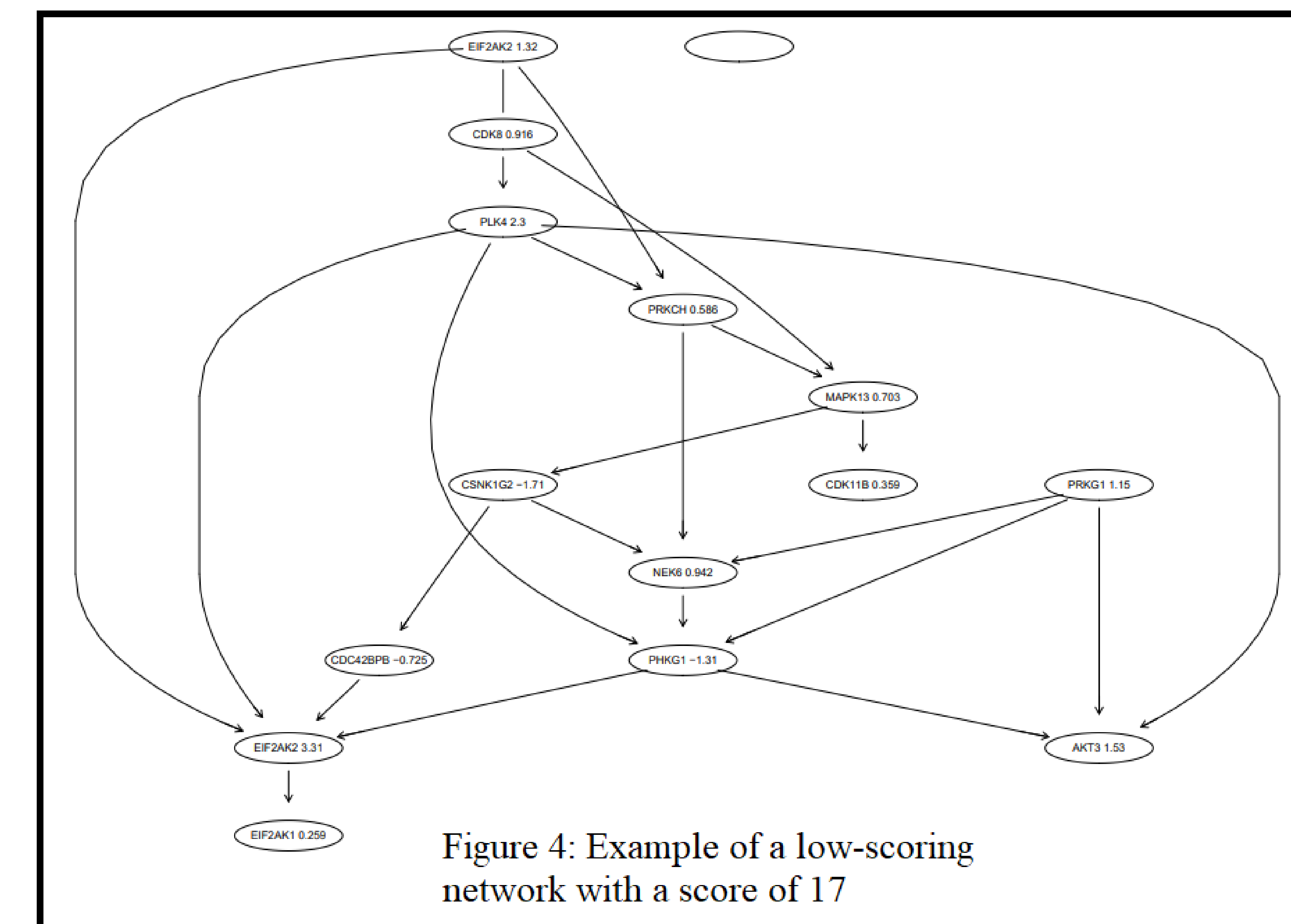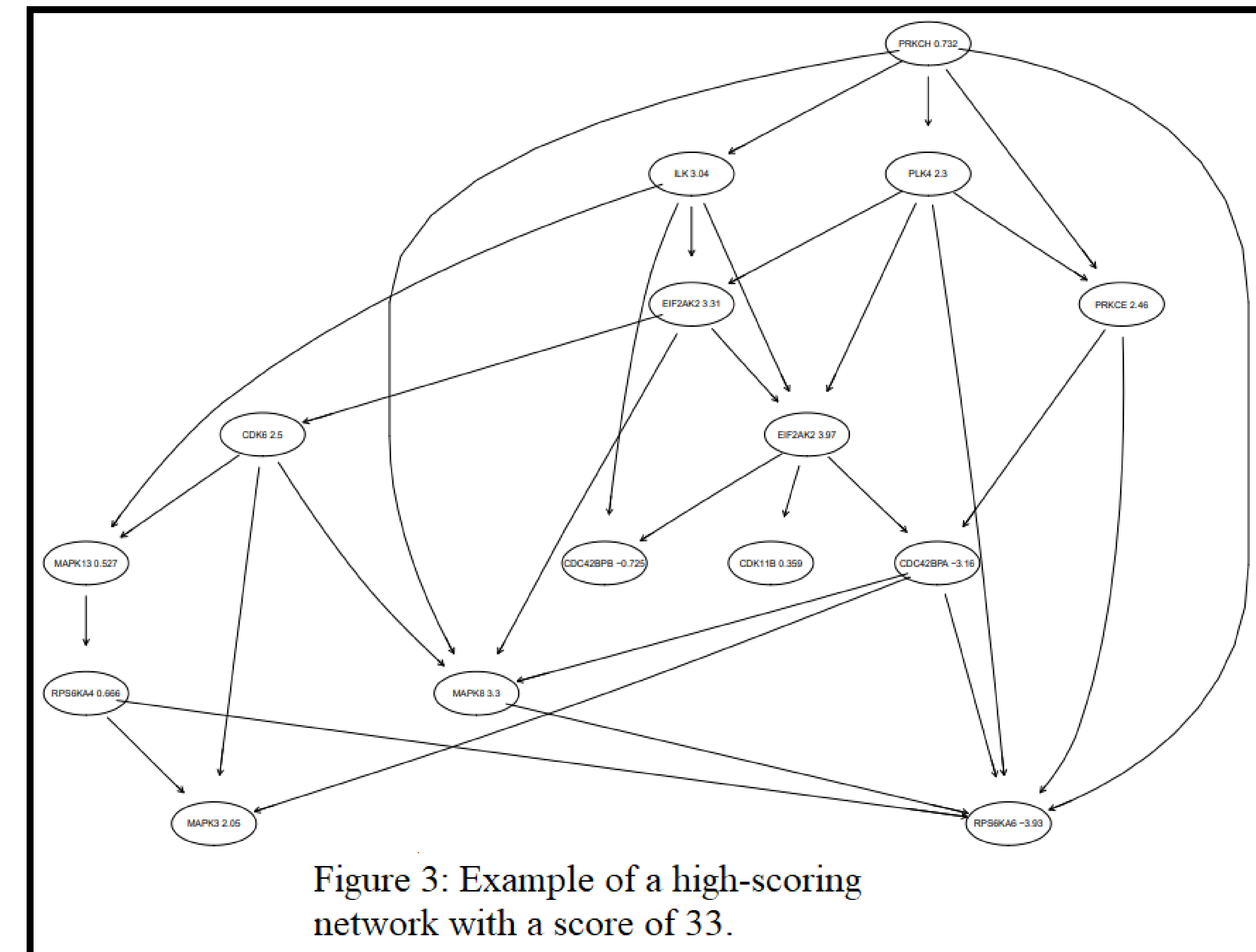$$Score = (rank + 1)^{interactions}$$

Figure 2: Equation used to generate each node's individual score. Scores are equal to one plus the knowledge rank raised to the power of the interaction number.

Results

- Figure 3 is a high scoring network with a score of 33.
- Figure 4 is a low scoring network with a score of 17.
- The primary observation is that the higher scoring network has a higher number of interactions and a greater complexity.
- The higher score also implies a higher level of information (knowledge) contained with the network.

Conclusion

- The system condenses network complexity and information content into a single number, allowing the package to easily select the most optimal network, reducing noise and producing higher quality results.
- Presence of and marking of dark kinases within network scores can be used to potentially discover novel kinase interactions or direct research towards less-understood topics.
- Future goals are to implement a user-specified threshold and extend this scoring system to other types of regulatory network models.

Figure 3: Example of a high-scoring network with a score of 33.

Figure 4: Example of a low-scoring network with a score of 17