《Build System Analysis with Link Prediction》Review

Paper Info

《Build System Analysis with Link Prediction》 Xin Xia, David Lo, Xinyu Wang, and Bo Zhou

Major Contribution

The paper introduces an approach to predict missed dependencies in build system automatically. The major contributions of their work include:

- Predict missed dependencies in a new way: Consider the dependency graph as a sparse network, calculate the similarity between nodes, and predict the uncovered edges based on the similarity
- Perform a detailed comparison when different types of similarity are selected.

The paper is an empirical study and find that Preferential Attachment(PA) achieves the best performance. This result is important when we need to choose a proper method to calculate the similarity between nodes in Concrete Dependency Graph in build system.

Main Work

The authors take advantages of the tool MAKAO to extract the CDG(Concrete Dependency Graph) of Makefile, and regard the CDG as a sparse network. They convert the problem of missed dependency detection to the problem of link prediction in complex network. This is the biggest shinning point of this work.

Then they use similarity-based algorithm in link prediction problem to find the possible missed dependencies. Nine kinds of similarities are considered and tested in the experiment. The result shows that Preferential Attachment(PA) achieves the best performance.

The work gives a new view to solve the problem of missed dependency detection. Some methods in link prediction can be used.

Some Criticism

To my best knowledge, this paper is the first work to predict missed dependencies based on similarity-based algorithms in link predictions of complex network. However, it has several limitations and needs to be discussed and improved.

- The authors perform 9 groups of experiments to find out each similarity can achieve the best performance. This experimental setting is good, but it can not confirm that this kind of similarity is the best choice in all link prediction problems of different kinds of complex networks.
- Although the authors illustrate the definitions of 9 kinds of similarities in detail, they does not give an necessary explanation of similarity-based algorithms in link predictions, which makes me a bit confused.

- The similarity-based algorithm might not be always fit for this problem. It is lack of evidence that the nodes with the similar feature should be linked. It is not always true in CDG of build systems.