

Response to reviewer's comments:

1. Research questions should be slightly rephrased or they explanations expanded. From the questions itself it is unclear what are results previously obtained that will be enhance by combining logical dependencies with structural dependencies. The research questions should be slightly reformulated or the explanations expanded.

We rephrased the research questions from section 1.

2. In Section 2.4, some equations and relations are introduced. It would be nice to see the values they return on selected datasets (Ant and Hibernate).

Some of the information is already there, in figures 2, 3, and 4. The support metric is represented as dots, but only the maximum values for each entity are plotted. The average frequency of all the commits is represented as the black line from the figure. Unfortunately, we cannot plot or display these values for the entire system and all the commits, since we have 21886 pairs of entities for Ant, 26902 for Catalina, and 40054 for Hibernate

3. It would be nice to improve the image quality of Figure 2 and Figure 3. Currently, some dots appear solid black, while others appear in different shades of gray, and also their size varies.

We changed the plot settings and right now all the dots have the same color, the size might vary depending on if two dots are close together. In this case, it might appear as a big 'out of shape' dot.

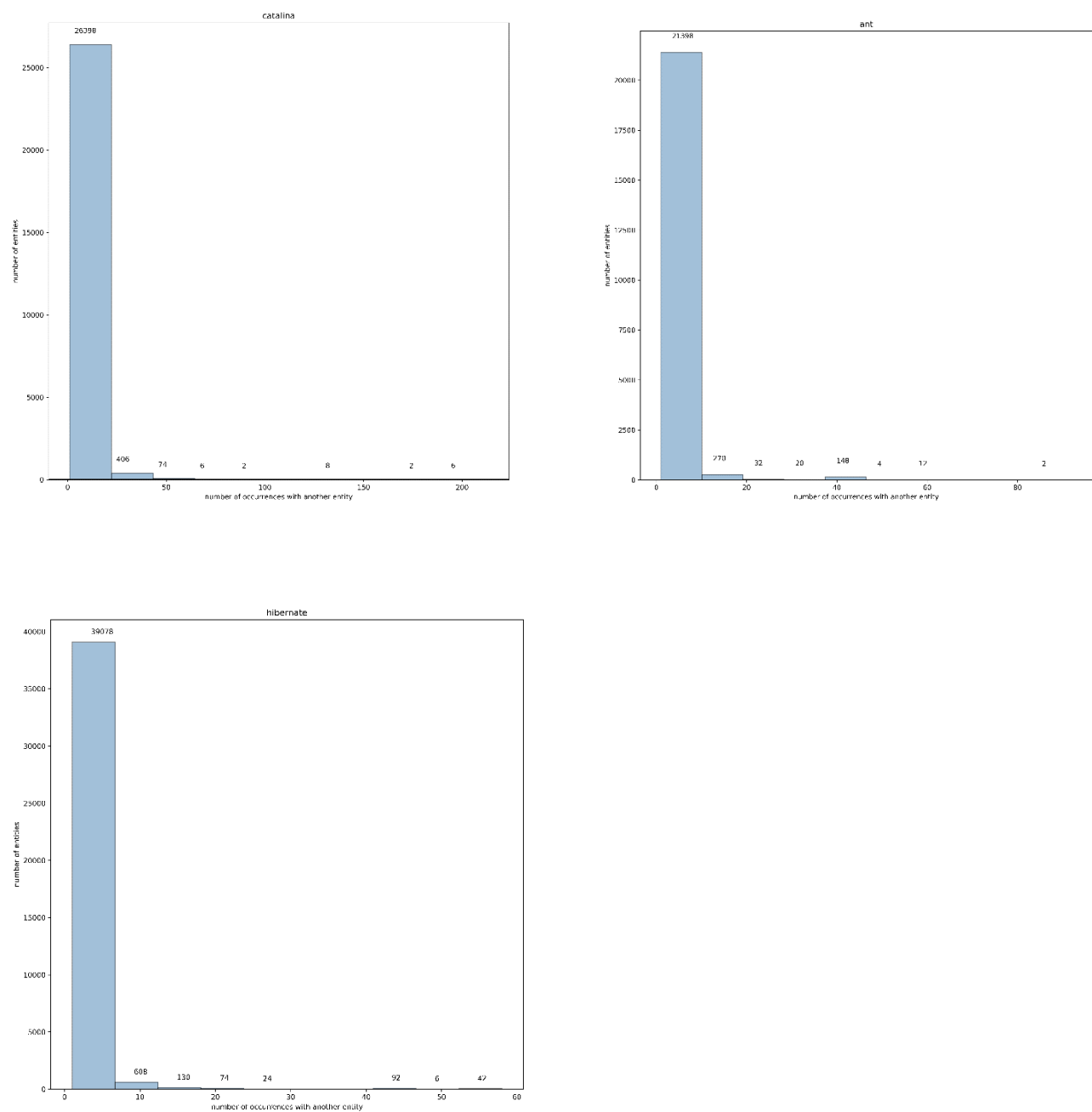
4. It would also be nice to show a pair of histograms corresponding to Figure 2 and Figure 3 that represent all occurrences between entities, not just the maximum ones.

If we plot all the occurrences, the y-axis will increase a lot, and not much information will be visible on the plot. On all the systems, most of the occurrences of an entity with another entity are below 10 (on average). In the figures below this comment, we can observe that we have a big histogram bin (bar) at the beginning, and all the other bars are hard to notice. Because of this, we added the occurrences count above the bar.

We consider that the histogram plot for all occurrences would have been much more conclusive if the numbers of occurrences would have been more evenly distributed on the x-axis.

In figures 2, 3, and 4 from the paper, only the maximum occurrences are plotted, and it can be observed that most of the points are at the bottom of the graphic. It is the same scenario as above, most of the occurrences are below 10 (on average). The only difference is that you can see better the occurrences that have a higher value. Also, plotting all the points wouldn't change the overall picture of the system. The excluded points will only create a line of points even lower at the bottom of the graphic.

We added in the paper an explanation of why we plotted only the max occurrences for all the systems.



5. In the rest of the manuscript three datasets were used and in Section 2.4 data regarding only two datasets were shown. It would be nice to show data regarding Tomcat Catalina also.

We added in the paper the max occurrence figure also for Tomcat Catalina. Figure 4, page 6.

6. It would be nice to name y axis on Figure 4, not only in the text.

We named the y-axis on Figure 5 (by introducing maximum occurrence figure for Catalina, all other figures number shifted).

7. In Section 5.4 it would be nice to have a table concerning Tomcat key classes like tables regarding key classes concerning Ant and Hibernate. If the content of the table is too large, only the top classes can be represented.

We cannot do a top-top key classes table (we cannot choose a criteria of separation), so we added in the table all key classes of Tomcat. We also added in the text examples from the new added table.

8. In Section 5.4 it would be nice to explain how these key classes were identified.

The selected classes from all three tables are the key classes extracted from developer documentation. We added this information in the paper, in chapter 5.4.