Paper Summary

The paper “Understanding Website Complexity: Measurements, Metrics, and Implications” talks about a problem that web page load times impact user satisfaction directly, which means that the provides or web owner would like to know how the website complexity affects the user experience. Therefore, the authors do a study on how to measure the complexity of web pages and its effect on performance. Their study focuses on two parts. The first is that they want to quantify the complexity of the web pages with a broad spectrum of metrics. And the second part authors focus on identifying the important complexity metrics that can take most impact on downloading and rendering.

For many different complexity metrics, the author claims that a website’s rank is not a significant indicator of the web complexity, which is that the top 1 website does not means it takes most complexity. Also, the complexity does not depend on the website’s category. However, the *News* sites and *Kid and Teens* sites are a little different. *News* sites load a lot more content than others, and it is from more servers and origins, while *Kids and Teens* sites have more Flash content than others. Meanwhile, authors point out that most websites load a large number of CSS and Javascript objects, and the content from non-origins takes a big fraction of objects and bytes, but their effect on downloading is low.

On the other hand, in the performance part, the authors analyze the question: which metrics have the most impact on performance. The authors state that the top five complexity metrics that determine the complexity are the total number of objects loaded, the number of these objects that are Javascripts, the total webpage size, and the number of servers and origins contacted in loading objects on the page. Also, authors think that the variability of load times is less correlated with metrics, while the number of servers should take the most important reason here.

In conclusion, the authors state that their analysis shows that number of objects are the most important reason for page loading time, while the number of servers are the most important reason for the variability in page loading time. Also, the webpage complexity does not have a lot of effects about its popularity, but it does have effect on its category, just like the *News* sites and *Kids* site cases. On the other hand, even though the non-origins offer many objects and bytes, it does not take a lot of time on loading.

In the whole paper, authors use a lot of real data to support its idea. They take many examples like the top website, and they also build a sparse model for predicting the loading time with error less than 0.1. Furthermore, they use the data to build many graphs in order to make their ideas clear and be more visible for readers, which is the part I like most. The other reason why I like the paper is that the paper gives me an idea that I can choose the website based on its complexity. There are many websites in each area, which makes me cannot decide which one I will use. Now, maybe in the future, I can use this way to decide that. However, it has a point I get confused and do not support the paper. Paper use a lot of data, but it does not give enough evidences to prove the correctness of these data, which makes the paper un-perfect. All in all, I am convinced with the conclusion.