A good phishing site should resemble the target site as much as possible, and it should hide the differences with the target site, at least to the unsuspecting user. This paper leverages this observation to cluster similar suspected phishing sites. Then, instead of crowd-sourcing the verification of a single suspected phishing site, a whole cluster can be verified at once. This is reported to improve both the timeliness and the accuracy of the results on the basis of an experiment with 239 participants. Unfortunately, the control group and the experimental group had a large overlap (174 participants). The authors argue that this does not invalidate the results because of minimal learning effects, but they have no evidence for this.

I believe that the main contribution of the paper is putting forward the idea of clustering similar suspected phishing sites. The paper shows that such clusters abound and that standard techniques (for example, shingling) are effective in discovering those clusters. This suggests important further research not identified in the paper: Is it possible to distinguish suspected phishing sites from genuine sites simply by searching for look-alikes? It would be prudent to keep humans in the loop to avoid liability issues surrounding false positives, and it would be wise to consider the countermeasures that phishers would use to defeat automatic look-alike detection.