

Analyzing citation network based on Web of Science

綜合研究大学院大学 統計科学専攻 張菱軒
中研院統計科學研究所 (台灣) 潘建興
統計数理研究所 中野純司

The Web of Science (WoS) database can be viewed as a big multi-layer, multi-level and dynamic network over years. We first introduce a new quantity called network influence to measure the importance or influence of an article towards not only its neighboring citations but the whole network.

Let $G = (V, E)$ be a citation network, where $V = \{v_1, \dots, v_m\}$ is a set of m articles (nodes) and E is a set of citations (directed edges). We denote $v_i \rightarrow v_j$ in G when v_i is being cited by v_j . We define the “influence range” r as the shortest path among all citation paths in between two articles, i.e.

$$r = \min(r_1, \dots, r_k)$$

where r_1, \dots, r_k are all k possible citation paths between two articles. Then the network influence of an article of interest v_t can be defined as

$$I_t = \sum_{r=1}^{r_M} g(r)(o_t(r) - e_t(r))$$

where r is the influence range from v_t to the rest of the articles in G , r_M is the maximum influence range of interest from v_t , $o_t(r)$ and $e_t(r)$ are the observed and expected number of articles within r from v_t , and $g(r)$ is a weighted function.

Then we undergo a systematic ranking process on all articles in statistics article network. We aim to identify which topics are influential in statistics community in the past ten years. According to our findings, papers related to LASSO have the biggest impact to the statistics research community, followed by other topics like regularization, model selection, hierarchical models and others.

We also propose some ongoing and future works extended from here, including but not limited to, some cross-network study in between article citation networks in other subject areas and statistics.