Exploring metrics of analyzing citation networks

Invited Paper

Suhas Gundimeda
CH12B026
Indian Institute of Technology Madras

Karnaty Rohit Reddy
EE12B030
Indian Institute of Technology Madras

***Abstract*—We describe some measures and metrics proposed by various papers in scientometrics to analyze citation and co-citation networks.**

I. Introduction

Scientometrics deals with the measuring and analysing science, technology and innovation. In this paper we look into analysis of scientific contributions in peer-reviewed publications using their citation, co-citation and co-authorship networks.

II. Previous Work

A. Citation count

Citation count is the sum of all published citations to a scientific paper. This is the most widely used measure of impact, and is the standard metric against which all of the others are compared.

B. PaperRank

The PageRank algorithm, when applied to citation networks, has the problem of nodes with no inbound edges causing it to have no solution. To solve this problem the algorithm is modified to assign rank of 1 to all initial nodes and reset it to zero after computation.[2]

C. h-index

An author metric whose value is the maximal number h such that the author has at least h papers with at least h citations.

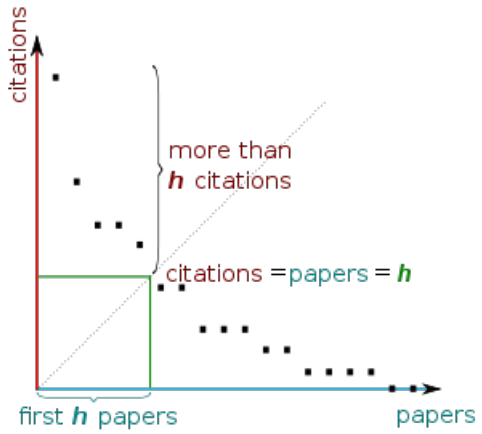

Fig. 1. h-index illustrated

D. PaperRank-Hirsch index

This is a proposed alternative to h-index which uses the calculated PaperRanks. The PaperRanks are normalized to represent more meaningful quantities, namely the most probable number of citations which the node might receive, the Q-value. The index is then defined as the integer number h such that an author has at least h papers with Q-value greater than or equal to h.[2]

E. CoRank

The CoRank algorithm[1], shown by equation, is based upon the principals of the PageRank algorithm, as such it is simply a variation of PageRank algorithm where the input data has changed rather than the algorithm itself. The key difference is that in CoRank, the rank gained from each co-cited publication is the CoRank of that paper CR(p_j) divided by the number of co-citations (Co-Links) CL(p_j) that paper has with other papers

$$CR(n) = \frac{1-\alpha}{|V|} + \alpha \sum_{p_j \in M(p_i)} \frac{CR(p_j)}{CL(p_j)}$$

1) CoRank-LinkCount: CoRank-LinkCount is an algorithm without a weighting factor, meaning that iterative calculation is not required. CoRank-LinkCount looks solely at the number of citations which are received, not by the publication in question, but by all the publications with which the subject paper is co-cited.

Figure 2 shows how the sum of co-citing publications is worked out. Publication n represents the target publication and each p being the co-cited publication; each publication marked c is thus a directly citing publication to n. The total CoRankLinkCount is the sum of all citations towards those papers (p) a publication (n) is co-cited with, including duplicates and c itself

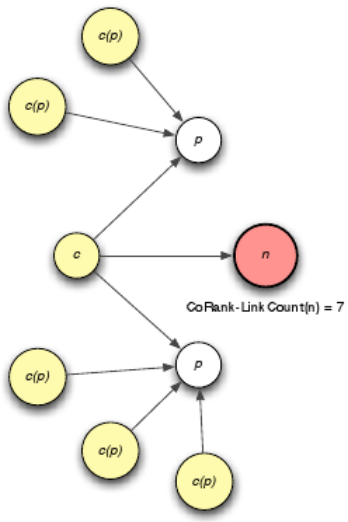

Fig. 2. Network of Co-Cited publications from publication n[1]

F. Betweeness centrality on co-authorship networks

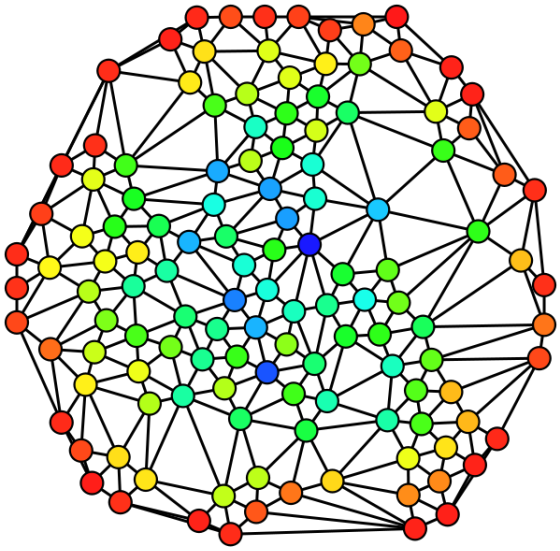

Fig. 3. Hue (from red = 0 to blue = max) shows the node betweenness.

Betweenness centrality is based on the number of shortest paths passing through a vertex. Vertices with a high betweenness play the role of connecting different groups. In the following formula, g_{jik} is all geodesics linking node j and node k which pass through node i. g_{jk} is the geodesic distance between the vertices of j and k.

$$C_B(n_i) = \sum_{j,k \neq i} \frac{g_{jik}}{g_{jk}}$$

In social networks, vertices with high betweenness are the brokers and connectors who bring others together [4]. Being between means that a vertex has the ability to control the flow of knowledge between most others. Individuals with high betweenness are the pivots in the network knowledge flowing. The vertices with highest betweenness also result in the largest increase in typical distance between others when they are removed.[3]

G. Closeness centrality on co-authorship networks

Closeness centrality [5] emphasizes the distance of a vertex to all others in the network through the geodesic distance from each vertex to all others. It is a measure of how long it will take for information to spread from a particular node [4].

$$C_c(n_t) = \sum_{i=1}^{N} \frac{1}{d(n_i, n_j)}$$

Where $d(n_i, n_j)$ is the distance from node i to node j and C_c is the closeness centrality.[3]

H. Group centrality measures

According to [7], the degree centrality of a group is the number of actors outside the group that are connected to the member of that group. Different ties to the same actors by different group members are only counted once. The formula is described below:

$$\begin{aligned} Group\,degree\,centrality &= |N(C)| \\ Normalized\,group\,degree\,centrality &= \frac{|N(C)|}{|V| - |C|} \end{aligned}$$

Where C is a group that is a subset of the set of vertices V, and N(C) denotes the set of all nodes that are not in C but connected to a member of C.

Similarly group closeness and betweeness centralities are explored in [6], [7]. These measures help us analyse the interdisciplinary nature of disciplines.

REFERENCES

[1] DAVID C TARRANT (2011) "A Study of Early Indication Citation Metrics ", University of Southampton, Electronics and Computer Science, PhD Thesis.

[2] Mikalai Krapivin, Maurizio Marchese and Fabio Casati "Exploring and Understanding Scientific Metrics in Citation Networks"

[3] Erjia Yan, Ying Ding, "Applying centrality measures to impact analysis: A coauthorship network analysis"

[4] Yin, L., Kretschmer, H., Hanneman, R. A., & Liu, Z. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. Information Processing and Management, 42 , 1599- 1613

[5] Freeman, L.C. (1979). Centra lity in social networks. Conceptual clarification. Social Networks, 1 , 215-239

[6] Chaoqun Ni, Cassidy R. Sugimoto, and Jiepu Jiang. Degree, Closeness, and Betweenness: Application of group centrality measurements to explore macro - disciplinary evolution diachronically

[7] Everett, M. G., & Borgatti, S. P. (20 05). Extending centrality. In P. J. Carrington, J. Scott & S. Wasserman (Eds.), Models and methods in social network analysis (pp. 57 - 76). New York: Cambridge University Press.