Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 20 (2013) 439 – 445

Procedia Computer Science

Complex Adaptive Systems, Publication 3
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2013- Baltimore, MD

Analyzing Structural & Temporal Characteristics of Keyword System in Academic Research Articles

Arjun Duvvuru[a], Srinivasan Radhakrishnan[b], Deepali More[b], Sagar Kamarthi[c], Sivarit Sultornsanee[d]*

[a]*Research Institute for Policy Evaluation and Design University of Thai Chamber of Commerce, Bangkok, Thailand*
[b]*Symbiosis Institute of Management Studies, Pune, India*
[c]*Department of Mechanical and Industrial Engineering, Northeastern University, Boston, USA*
[d]*Department of Logistics Management, School of Business, University of Thai Chamber of Commerce, Bangkok, Thailand*

Abstract

Keyword networks, formed from keywords occurring in scholarly articles provide a useful mechanism for understanding academic research trends. In keyword networks, keywords are represented as nodes and a link is formed between a pair of keywords if they appear in the same article. Each link is assigned a weight, representing the number of co-occurrences of the pair in different articles. A statistical and visual analysis of the structural and temporal characteristics of such networks reveals the organizing pattern and the evolution of keywords. In this study we analyse the difference between structured keyword system and unstructured keyword system. We use keywords from two prominent business management journals from USA and India and analyse the corresponding keyword networks. Our results indicate that the network characteristics of structured keyword system are more suitable than unstructured keyword system to analyse research trends and bring forth the emerging areas and popular research methods. The adoption of structured keyword system will aid researchers and funding agencies to optimize their decision on the use of research funding.

Selection and peer-review under responsibility of Missouri University of Science and Technology

Keywords: Complex networks; Keyword networks; Visual analytics; Data analysis; Research trends; Visualization; Business management

1. Introduction

Complex networks have been widely used to analyze and uncover the basic substratum of complex systems. Network perspective enhances the understanding of the structure and behavior of the system. The study in network science proliferated after the discovery of scaling in random networks[1] (scale-free networks). The topological

* Corresponding author. *E-mail address*: sultornsanee@riped.utcc.ac.th

Selection and peer-review under responsibility of Missouri University of Science and Technology
doi:10.1016/j.procs.2013.09.300

characteristics of scale-free networks mainly reflected the statistical abundance of *hubs* or vertices with a high number of connections compared to the average degree of the network $<k>$ [1], and the presence of scale-free vertex degree distributions $P(k)$, where the probability P of finding a vertex of degree k follows a power-law [2] ($P(k) \sim k^{-\gamma}$). Network science has found applications in diverse fields such as sociology, telecommunication, cell biology, genetics, and economics.

In recent years, information available in a range of information rich channels such as newspaper articles, scholarly articles and web-based communities has risen exponentially. Although network science applications are used for information content analysis, they are skewed towards social networking sites and searches generated from keywords. The use of network science to uncover emerging trends in scientific communities is sparse. Existing methods mine for research trends in scholarly articles and are more focused on citation networks [3] [4] [5] (network analysis of citation in scholarly articles). However the keywords appearing in scholarly articles are relatively underused for uncovering research trends. Nascent steps have been taken by researchers to apply network science based measures to keywords appearing in scholarly article and study its topological characteristics [6]. The reasons for ignoring keyword based network approach may stem from lack of keyword organizing standards.

In this study we compare the keyword network topologies of structured keyword system and unstructured keyword system. In a structured keyword system[b], the scientific journal categorizes the keywords under pre-formed categories. In an unstructured keyword system[b] the user defined keywords are not subjected to pre-formed categories. The method and results are discussed in the following sections.

2. Method

2.1. Keyword network

In keyword network, keywords are the nodes and two nodes are connected by a link if they appear in the same article. Links between nodes carry weights. Each time a pair of nodes appears together in subsequent articles in the journal, the weight on the link connecting the nodes increases by one (Fig. 1). Thus the keywords network is an undirected and weighted network represented by an edge list. Each element in this list is a vector $E_k = (i, j, w_{ij})$. Where, i and j are the nodes linked by the *kth* edge E_k and w_{ij} represents the number of times the same keyword pair (i, j) co-occurs in different articles. As shown in Fig. 1, all keywords that appear in the *Keywords* section of articles are linked together, initially forming isolated networks. As more articles are parsed certain keyword pairs (or triples or quadruples) tend to re-appear in other articles. This repetition increases the respective link weights and combines individual components into a network.

2.2. Data

The study focuses on two data sets of keywords. We use keywords from articles appearing in Academy of Management Journal [7] (AMJ) from 2003 to 2012. The keyword data set collected from AMJ appears to follow a structured keyword system. In case of unstructured keyword system we use keywords fom articles appearing in Vikalpa [8] (A management journal from Indian Institute of Management-Ahemdabad) from 2003 to 2012. Both the

---

[b] There is no formal segregation of keyword system as mentioned in this study. We have categorized the two keyword system for the purpose of comparison. We define a structured keyword system where the process, method or sub-domain related keyword is coupled with domain related keyword (example: Decision making, Decision making-research, Decision making-economic aspects, Decision making-mathematical models, Decision making-moral and ethical aspects. Here decision making is the domain keyword and research, economic aspects, mathematical models, moral and ethical aspects are process, method or sub-domain related keyword). Absence of aforementioned structure qualifies as unstructured keyword system (example: A study focusing on talent retention practices using matrix method in company x generates following unstructured keywords, talent retention, company x, and matrix method). The same example under structured keyword system is formed as Human resources-talent retention, Talent retention-mathematical model, Talent retention-matrix method.