Accepted Manuscript

Dynamic burstiness of word-occurrence and network modularity in textbook systems

Xue-Mei Cui, Chang No Yoon, Hyejin Youn, Sang Hoon Lee, Jean S. Jung, Seung Kee Han

PII: S0378-4371(17)30641-6
DOI: http://dx.doi.org/10.1016/j.physa.2017.06.002
Reference: PHYSA 18387

To appear in: *Physica A*

Received date : 11 January 2017

Please cite this article as: X. Cui, C.N. Yoon, H. Youn, S.H. Lee, J.S. Jung, S.K. Han, Dynamic burstiness of word-occurrence and network modularity in textbook systems, *Physica A* (2017), http://dx.doi.org/10.1016/j.physa.2017.06.002

Dynamic Burstiness of Word-occurrence and Network Modularity in Textbook Systems

Xue-Mei Cui[1], Chang No Yoon[2], Hyejin Youn[3,4], Sang Hoon Lee[5], Jean S. Jung[2], and Seung Kee Han[2*],

[1]*Normal College, Yanbian University, Yanji 133002, China,*

[2]*Dept. of Physics, Chungbuk Nat'l Univ., Cheongju, Chungbuk, 361-763, Korea,*

[3]*Santa Fe Institute,USA.*

[4]*Institute for New Economic Thinking and Mathematical Institute, University of Oxford, United Kingdom.*

[5]*Integrated Energy Center for Fostering Global Creative Researcher (BK 21 plus) and*

Department of Energy Science,Sungkyunkwan University, Suwon 440—746, Korea

* Correspondence: skhan@chungbuk.ac.kr

We show that the dynamic burstiness of word occurrence in textbook systems is strongly correlated with the modularity of the word association networks. At first, a measure of dynamic burstiness is introduced for the quantification of burstiness of word occurrence in a textbook. The advantage of this measure is that the dynamic burstiness is decomposable into two contributions: one coming from the inter-event variance and the other from the memory effects. As the network modularity of textbook systems is compared with those of surrogate random textbooks without the memory or variance effects, it is shown that the network modularity increases systematically with the dynamic burstiness. The intra-connectivity of individual word representing the strength of a tie with which a node is bound to a module also increases with the dynamic burstiness, which implies that individual words with high burstiness are strongly bound to one module. Based on the frequency and dynamic burstiness, physics terminology is classified into fourcategories: fundamental words, topical words, special words, and common words. In addition, we test the correlation between the dynamic burstiness of word occurrence and network modularity using a two-state model of burst generation.

PACS numbers: 05.45.Xt, 89.75.-k, 89.70.-a, 05.40.Ca

1.Introduction.

Recently, complex network analysis has become one of the most importantresearch topics in biological, social, information, and communication systems[1-4]. The network analysis, in particular, has been also applied to the knowledge systems of human language using very large corpora of written texts of the word lexicon, thesaurus, the Wikipedia, language corpora, and other documents[5-11]. As the knowledge networks are constructed utilizing the relationship of co-occurrence[5] semantics[7], synonym[8], syntactic dependence[9], or the hyperlink, citation, quotation, and any other relationships [10, 11], it isshown that our human language systems exhibit the ubiquitous topological properties observed in complex networks: the small-world phenomena [5],power-law degree distributions [5, 6], hierarchical modular organization [9, 12], and disassortative mixing [9]. Since the community identification in modular network is crucial for understanding the overall structure and functional properties of complex systems [13-17], there also have been efforts to identify communities in the written texts for the classification of knowledge systems[18, 19].

It has been proposed that network modularity is crucial for enhancing robustness, flexibility, and stability[13-17, 20, 21]. Recently, there have been attempts to explain the origin of the modular organization observed in many complex systems. In complex networks, network modularity could be emerged through fluctuations in establishment of links[22], or evolution utilizing the small-preference rules [23]. In a neural network system, the network modularity could be shaped by adaptive dynamics and amplified by the feedback effect [24]. However, it is still unknown what causes the formation of modular organization in various complex systems.

On the other hand, bursting behavior was observed in inter-event interval distribution of human activities like the email correspondence [25]. Similar to the burst firing in neural systems [26], an active state with high firing rate alternates with a quiescent state with a very low firing rate. The bursting behavior was characterized by the large variance in the inter-event interval distribution and also by the long-term memory effect with short inter-event interval is repeated many times before it is interrupted by the quiescent state with long inter-event interval [27-29]. Our human language systems are also shown to evolve in punctuational bursts for a long time scale[30]. Since the outbreak of new topics in the internet is detectable by the burstiness of special word occurrence [31], there have been efforts to analyze the burstiness in written texts [32, 33].

It is expected that the temporal organization of the word occurrence in written texts have strong influence on the topological properties of the word network. For example, the hierarchical organization in a text is shown to induce the long-range temporal correlation in word occurrence [34]. However,studies focused on the relation between the dynamic burstiness and the network properties are very rare. As such an attempt, in this paper, we analyze the dynamic burstiness of word occurrence in physics textbooks and the modular organization of the word-association networks to understand the role of the dynamic burstiness in the formation of modular organization in the word networks. First, we introduce a measure of dynamic burstiness thatcombines both the contributions of inter-event interval variance and inter-event memory effects. Secondly, we compare the network modularity of textbook systems with those of random textbooks that were generated by removing either the variance or memory effects. Finally, we investigate the correlation between the dynamic burstiness and network modularity using a two-state burst generation model [31].

Physics textbooks.

Our analysis is based on three college physics textbooks by GRIFFITH (GR) [35], HEWITT (HE) [36], and KNIGHT (KN)[37]. Physics terminology used in the textbooks is limited to the vocabulary of physics dictionary OXFORD(OX)[38],which contains more than 2400 items. The statistics showing the number of sentences, the number of physics terminology and their total number of occurrences in each textbook and the dictionary are summarized in Table 1. For the analysis of dynamic burstiness word occurrence we