Available online at www.sciencedirect.com
ScienceDirect
Physica A 387 (2008) 661–666
PHYSICA A
www.elsevier.com/locate/physa

Network model of human language

Mária Markošová*

Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, Bratislava, Slovakia

Received 2 July 2007
Available online 29 September 2007

Abstract

The phenomenon of human language is widely studied from various points of view. It is interesting not only for social scientists, antropologists or philosophers, but also for those, interested in the network dynamics. In several recent papers word web, or language as a graph has been investigated [R.F. Cancho, R. Solé, The small world of human language, Proc. R. Soc. London B 268 (2001) 2261–2265; A.E. Motter, P.S. de Moura, Lai Ying-Cheng, P. Dasgupta, Topology of the conceptual network of language, Phys. Rev. E 65 (2002) R 065102; M. Steyvers, J.B. Tenenbaum, The large-scale structure of semantic networks: Statistical analysis and a model of semantic growth, Cogn. Sci. 29 (2005) 41–78].

In this paper I revise recent studies of syntactical word web [R.F. Cancho, R. Solé, The small world of human language, Proc. R. Soc. London B 268 (2001) 2261–2265; S.N. Dorogovtsev, J.F.F. Mendes, Language as an evolving word web, Proc. R. Soc. London B 268 (2001) 2603–2606]. I present a model of growing network in which such processes as node addition, edge rewiring and new link creation are taken into account. I argue, that this model is a satisfactory minimal model explaining measured data [R.F. Cancho, R. Solé, The small world of human language, Proc. R. Soc. London B 268 (2001) 2261–2265; M. Markošová, P. Náther, Language as a graph, in: V. Kvasnička, P. Trebatický, J. Pospíchal (Eds.), Mind, Intelligence and Life, Kelemen, STU Bratislava, 2007, pp. 298–307 (in Slovak)].

PACS: 87.10.+e; 89.20.-a; 89.75.Da; 89.75.Hc

Keywords: Scale-free networks; Word web

1. Introduction

Networks are nowadays very popular targets of investigation. They are good models for various types of interactions, such as social interactions, professional interactions [6], interactions in biology [7], interaction as communication [8,9] to which belong also interactions of people through language [1,2,4,5]. Networks are also interesting objects to study theoretically, because their properties are strongly influenced by the network history and dynamics. Networks can grow with time by node addition, the nodes can become extinct too. Several questions have been asked about the details of the net dynamics. For example, how does the dynamics influence the overall network structure [10,11,13], or what are the dynamics governing real networks [12]?

* Tel.: +421 02 602 95 374; fax: +421 02 65422263.
E-mail address: markosova@ii.fmph.uniba.sk.

doi:10.1016/j.physa.2007.09.027

Network is a collection of nodes interacting through edges. Binary undirected networks are the simplest ones; the edge between two nodes either exists or it does not. Networks are usually characterized by several local and global measures [6]. The most important local measures are clustering coefficient C and the average node degree k. Mathematically the clustering C_i of the node i is a probability that the two neighbours of node i are mutual neighbours as well. Network clustering coefficient C represents an average of all C_i-s. Clustering coefficient is in fact a measure of nontrivial "structure" in the network. By nontrivial is meant, that the network is not a tree or a simple regular lattice with nearest neighbour connections, only. As a global measure, node separation l (average shortest path between randomly chosen sites) is typically used. Separation of nodes shows, how "close" one node is to the other, or, in other words, how well the nodes communicate through edges.

Special type of network is a small world network [6]. Its structure optimises between the local regularity preservation, which tends to enhance node separation l and good global node communication through random shortcuts. In these networks a high clustering coefficient C, is combined with a low node separation l.

As has been mentioned above, networks usually change their size with time [11,13]. Many real networks, such as internet or word web, grow by the continual addition of new nodes. The node addition might be accompanied by node deletion, but the ratio of deleted nodes is often negligible. Therefore the dynamics of real networks are well-captured by the models of growing nets [10,12,11,13].

Many recent studies have shown [7,1,12,14], that real networks, which are created by self-organized processes, have common features. Their static properties are similar to that of small world nets. On the other hand, their degree distribution function, which is influenced by the dynamics, has power law character:

$$P(k) \propto k^{-\gamma}. \tag{1}$$

Such networks are called scale free [10]. The same properties are possessed by the word web [1,3].

In this paper a positional word web is studied [1,4]. Here the words are nodes and the word interaction is defined by the neighbourhood in a sentence. Language is a living phenomenon, developing all the time. Some words are created and some of them fall into disuse. Hence, to understand the word web dynamics, it is important to examine the dynamics of nets with changing number of sites.

This paper is organized as follows: In Section 2 the question of scale-free network structure and dynamics is studied. Section 3 is devoted to the mathematical models of positional word web and in Section 4 my word web model is presented.

2. Scale-free networks

As has been mentioned in the previous section, scale-free structure is a result of self-organized network development. Therefore this process should be natural and simple. The nature of the processes leading to the scale-free structure has been investigated by Barabási and Albert in their fundamental paper [10]. In the Barabási–Albert model (BA model) the growth of net starts from a small bunch of interconnected nodes. Every time unit a node comes and adjoins itself to the old network by m new links. The probability of linking with a certain old node i is proportional to its degree k_i. Such type of linking is called preferential.

How does one describe these processes mathematically? There are several possibilities. In many cases, the most efficient method seems to be a continuous approach of Dorogovtsev and Mendes [13]. Newcoming nodes are labelled by their birth-time s. At time t, node s has, in average, $k(s,t)$ neighbours. The average degree $k(s,t)$, is given by the equation

$$\frac{\partial k(s,t)}{\partial t} = m \frac{k(s,t)}{\int_0^t k(s,t)\mathrm{d}s}. \tag{2}$$

Here the rhs of the Eq. (2) expresses how $k(s,t)$ changes by the preferential linking. To find the solution, the sum of all node degrees expressed as the renormalization integral $\int_0^t k(s,t)\mathrm{d}s$ in the denominator, is to be estimated. It is easy. Every time unit m new edges increase the sum by $2m$. Therefore

$$\int_0^t k(s,t)\mathrm{d}s = 2mt. \tag{3}$$