ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume LVIII 2 Number 3, 2010

SIMULATING ACTIVATION PROPAGATION IN SOCIAL NETWORKS USING THE GRAPH THEORY

F. Dařena, A. Troussov, J. Žižka

Received: January 29, 2010

Abstract

DAŘENA, F., TROUSSOV, A., ŽIŽKA, J.: *Simulating activation propagation in social networks using the graph theory.* Acta univ. agric. et silvic. Mendel. Brun., 2010, LVIII, No. 3, pp. 21–28

The social-network formation and analysis is nowadays one of objects that are in a focus of intensive research. The objective of the paper is to suggest the perspective of representing social networks as graphs, with the application of the graph theory to problems connected with studying the network-like structures and to study spreading activation algorithm for reasons of analyzing these structures. The paper presents the process of modeling multidimensional networks by means of directed graphs with several characteristics. The paper also demonstrates using Spreading Activation algorithm as a good method for analyzing multidimensional network with the main focus on recommender systems. The experiments showed that the choice of parameters of the algorithm is crucial, that some kind of constraint should be included and that the algorithm is able to provide a stable environment for simulations with networks.

social network, recommender system, graph theory, spreading activation, simulation

INTRODUCTION AND OBJECTIVES

A *social network* (Wellman and Berkowitz, 1988) is a term used for general social structures made of *nodes* that can be individuals, organizations, or any objects. A social network represents relationships and flows between people, groups, organizations, animals, computers, or other information/knowledge processing entities. The nodes, sometimes called as vertices, are interconnected in a general way by *links* that are sometimes called also ties or edges. The links may pass the information either unidirectionally or bidirectionally. The nodes and links can operate at many different levels – from families to close friends, from colleagues to management, from regions to nations, and so like. Social networks are affording us sharing various kinds of information as well as playing a critical role in the management of organizations, decision making, problem solving, collaborating, including facilitating trade and commerce. The results of social networks analysis can be used not only by subjects representing the customers but also by organizations for better targeting their customers and revealing customers' needs.

Several important branches are connected with applying social networks. Naturally, a social network has firstly to be built according to its supposing use. When it is used, such a network is often a very dynamical structure that can change and grow within the time. We can analyze a social network, obtain valuable knowledge about relationships between entities which are modeled as nodes, or about collaborative activities (Freeman, 2006). Analyzers may discover specific expertise, positive and negative interconnections influencing activities of nodes, or hints what and how to improve, remove, add, and so like. The analytic area is also covered by good journals, for example Social Networks: An International Journal of Structural Analysis (four issues per year, Elsevier).

Another interesting and important event is so called *folksonomy* (folk taxonomy), which is a taxonomy created by users of a social network. Folksonomy is useful for social classification, indexing, and tagging both in the private, expert, professional, or commercial life. We can also speak about *collective intelligence*, which may nowadays be one of goals of building social networks. One of the useful social-network applications are Recommender Systems (RS). It is an effective tool to deal with information overload issue. Similar with rating behaviors such as

purchase behavior, click streams, and browsing history, etc., the tagging information implies user's important personal interests and preferences information, which can be used to recommend personalized items to users (Liang et al., 2009). With the increasing popularity of the collaborative tagging systems, tags could be interesting and useful information to enhance RS algorithms. Unlike attributes which are "global" descriptions of items, tags are "local" descriptions of items given by the users (Tso Sutter et al., 2008).

The social-network formation and analysis is nowadays one of objects that are in a focus of intensive research. The objective of the paper is to suggest the perspective of representing social networks as graphs, with the application of the graph theory to problems connected with studying the network-like structures and to study spreading activation algorithm for reasons of analyzing these structures.

MATERIAL AND METHODS

Representing networks by graphs

A social network represents a set of actors which are connected by relationships of different types (Jungnickel, 2007). To be able to analyze the network and to discover characteristics like betweenness, structural cohesion, closeness etc. (Lubbers et al., 2010) which includes large amount of data processing, it is favorable to use computers that can provide results in reasonable time. In order to be able to process social network data on computers, it is necessary to use some formal method. Such an advantageous, mathematically founded approach is the application of graph theory (Bang-Jensen and Gutin, 2008). The advantage of using graphs for representing social networks is also the fact that there exist a significant number of methods that can be readily used for analyzing the data and for solving problems related to social networks (Hanneman, Riddle; 2005). For example, the recommendation process can be seen as a graph search, when the model is used to generate recommendations by finding nodes that have high associations with the starting nodes (Huang, Chung, Chen; 2003).

We can model a multidimensional network (a network containing nodes and links of multiple types) as a *directed graph* (digraph), *G*, which is an object described by a set *V* of *vertices* v_i (called also *nodes*), and a set of *edges*, *E* (called also as *links*). Here, the edges are directed and called *arcs*, creating a set *A* of arcs a_j. Arcs are ordered pairs (coordinates) that define beginnings and endings of each arc. The graph *G* is a *simple graph* (with no loops connecting a vertex directly to itself, and no multiple edges): $G = (V, A)$.

The graphs have the following basic functions available:

- *init(A,V)* … generating initial nodes for the set of arcs,
- *term(A,V)* … generating terminal nodes for the set of arcs, and
- *F(V)* … a real-valued function that sets activation values $\alpha > 0$ of nodes.

Before applying graph-based methods for networks, we need to refine the model of the real world data to make it suitable for particular type of graph-mining technique. Before proceeding to graph-mining by application of a generic algorithm that is not tailored to the domain, we need to pre-process the real-world network taking into account what type of graph-mining will be used.

In order to enable using computer-based tools for domain specialists, that are not necessarily familiar with graph or other theory terminology, we decided to use terminology that is common in social networks, recommender systems and similar domains (see Table I).

I: *Differences between graph and network terminologies*

Network terminology	Graph terminology
node	vertex
link	edge, arc
node and link type	color
node importance	vertex weight

All users and items of a network are represented by *nodes*. A node is characterized by its identifier that enables to identify a node uniquely. It is also used for definition of initial and terminal nodes for links. The identifier (or alternatively a label) might be also useful for network visualization and in explanatory modules. The type of a node enables to distinguish the nodes and classify them into groups (e.g. in tag-aware recommender systems we distinguish actors, resources, tags and instances (assignments) of tagging) which enables e.g. filtering. Node *importance* models the weight of nodes in a graph by assigning numbers from subrange <0, 1>. This number is a result of conversion of physical units to abstract semantic units

A *link* represents a relation between two nodes. The link *weight* quantifies this relation and is used e.g. as a base for calculating the signal decay described by spreading activation algorithm (see below). Value less than 1 means decay of the signal and value greater than 1 amplifies the signal. Following link property, the *reciprocity*, defines whether there exists a reciprocal link that doesn't have to be defined explicitly. The reciprocal link might be created automatically (e.g. when a resource is tagged by an actor, the actor also tags the resource). On the other hand, the possibility not to create the reciprocal link is also kept to have freedom in modeling situations when automatic reciprocity is not undisputable (e.g. a person A, a participant of a social network, indicates that another person – person B – was his teacher, but this person B doesn't indicate A as her pupil).

Spreading Activation method

Spreading Activation is a processing technique on network data structure (Crestani, 1997). It has its roots in psychology (Crestani, Lee; 2000) and is widely used in fields, such as numerical simulation of physics phenomena, epidemic models, information retrieval, and recommender systems. Spreading activation algorithms iteratively propagate the activation from the initial set of nodes referred to as the seed, to the other nodes in a network through outward links. Usually, this propagation is performed until the behavior of the system stabilizes near the so called limit distribution of the algorithm or is stopped by constraints, e.g. the limitation on the total number of iterations (Troussov et al., 2009).

The generic algorithm as described by Troussov, Parra, Brisilovsky (2009) is defined as follows:

1. Initialization (input)

Setting the parameters of the algorithm, network, and a starting list of non-zero valued, "activated" set of nodes $v_i \in V$ with positive values assigned by $F(V)$.

2. Iteration
   a) Expanding the list of nodes activated by their neighbors.
   b) Re-computing activation values of all nodes in the list.
   c) Normalization – scaling linearly up or down the numerical values of all activated nodes to satisfy given preconditions of activation conservation.
   d) Purging the list from nodes having the activity value $\alpha < t$, where t is a threshold value and one of the algorithm parameters.
   e) Checking if the last iteration step (given by a suitable condition like a maximum number of iterations) was reached; if not, returning to the step a).
3. Output

The ranked list of activated nodes with their activity values gained iteratively through $F(V)$.

The Re-computation step in the Iteration phase consists of following actions:

a) Input/Throughput/Output of the links
   - For each activated node v we compute the input signal to each outgoing link e, such that $\text{init}(e) = v$. This computation can be based on the value $F(v)$, the outdegree (number of outgoing links) of a node etc. For instance, if the node v has n outgoing links of the same type, each link e might get input signal:

$$\text{Input}(e) = F(\text{init}(e)) \cdot 1/\text{outdegree}(v)^{\beta},$$

     where β might be equal to 1. It could be also less than one, in which case the node v will propagate more activation to its neighbors than it has.
   - When the signal ("activation") passes through a link e, the activation usually experiences decay by a factor $w(e)$:

$$\text{Output}(e) = \text{Input}(e) \cdot w(e)$$

b) Input/Output of node activation
   - Before the pulse of the activations, the node v has the activation level $F(v)$.
   - Through incoming links v get more activation:

$$\text{Input}(v) = \Sigma\, \text{Output}(e)$$

     for all links e such that $\text{term}(e) = v$.
   - By dissipating the activation through outgoing links, the node v might lose activation:

$$\text{Output}(v) = \Sigma\, \text{Input}(e)$$

     for all links e such that $\text{init}(e) = v$.

c) Computation Of New Level Of Activation
   - The new value of a node v is computed based on the three above mentioned values – $F(v)$, $\text{Input}(v)$ and $\text{Output}(v)$, e.g.

$$F_{new}(v) = a \cdot F(v) + b \cdot \text{Input}(v) + c \cdot \text{Output}(v),$$

     where a, b and c are parameters of the algorithm.

At this moment, two kinds of normalization (calibration) are considered. In both of them, the sum of values $F(v)$ for a set of nodes N defined by some conditions remains constant after each pulse ($\Sigma F(v_i) = \text{const.}, v_i \in N$). In the first case, *Conservation of Initial Activation*, the set N contains nodes that were initially activated. In the second case, *Conservation of total activation*, set N is equal to the set of all activated nodes. The sum of nodes from set N equals to total initial activation (sum of activations of nodes at the moment of initial activation).

RESULTS

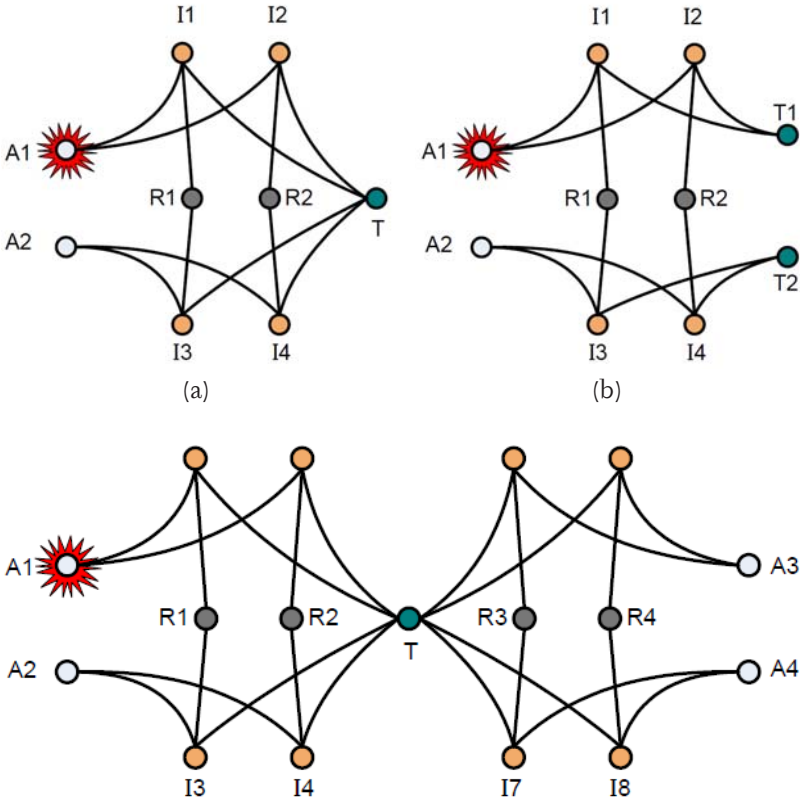
To verify graph based approach and spreading activation methods and to demonstrate the results, several experiments were designed and run. For better understanding the behavior, simple networks (networks with small number of nodes and links) were created. These networks were representing the structure of tag-aware recommender systems where individual nodes represent actors, tag, and the resources (e.g. books, web pages, products) of the system.

The structure of these networks can be seen on Fig. 1. Elements A represent the actors, elements T tags, R are resources and I instances (assignments) of tagging (e.g. I1 means that actor A1 tags the resource R1 by tag T).

The objective of these simulations was to find the relevancy of individual actors to each other base on their usage of tags and their assignments to resources of the system. The initial activation was put on node A1 – the intention was to discover the relevancy of other actor to actor A1.

During the initialization phase, the parameters of the algorithm and the network were set as follows:

- link weights: $w = 0.8$, based on findings of Troussov, Parra, Brusilovsky (2009)



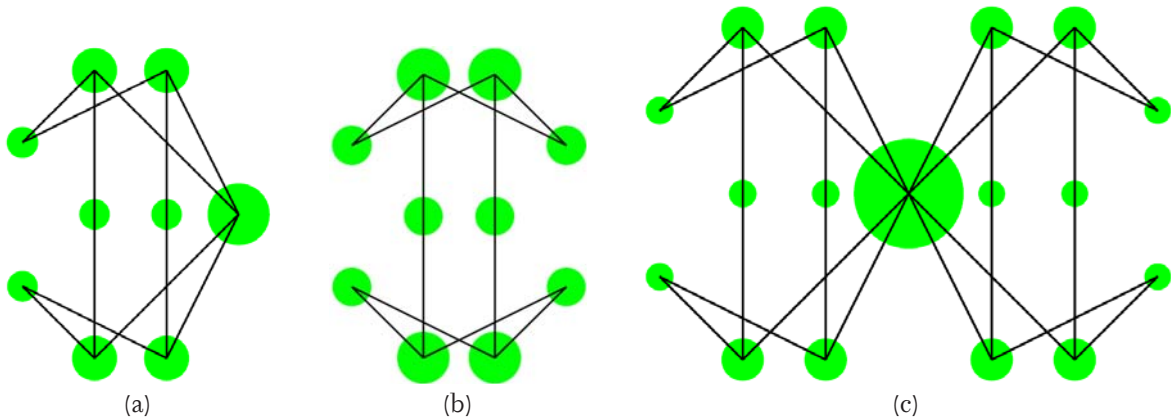
1: *Model networks representing tag aware recommender systems: (a) two actors tagging two resources using the same tag, (b) two actors tagging two resources using different tags, (c) two groups of actors tagging resources using the same tag*

- threshold value necessary for removing nodes from the list of activated nodes: $t = 0.0$
- the iteration process is stopped after limit distribution is reached
- parameter in the formula Input (e) = F(init(e)) · 1/ outdegree$(v)^{\beta}$, used for calculating input activation for links outgoing from nodes: $\beta = 0.5$
- parameters used in the formula $F_{new}(v) = a \cdot F(v) + b \cdot \text{Input}(v) + c \cdot \text{Output}(v)$ for calculating new activation value of a node: $a = 1$, $b = 1$, $c = 0$
- for calibrating values of nodes, *Conservation of Initial Activation* normalization type was used.

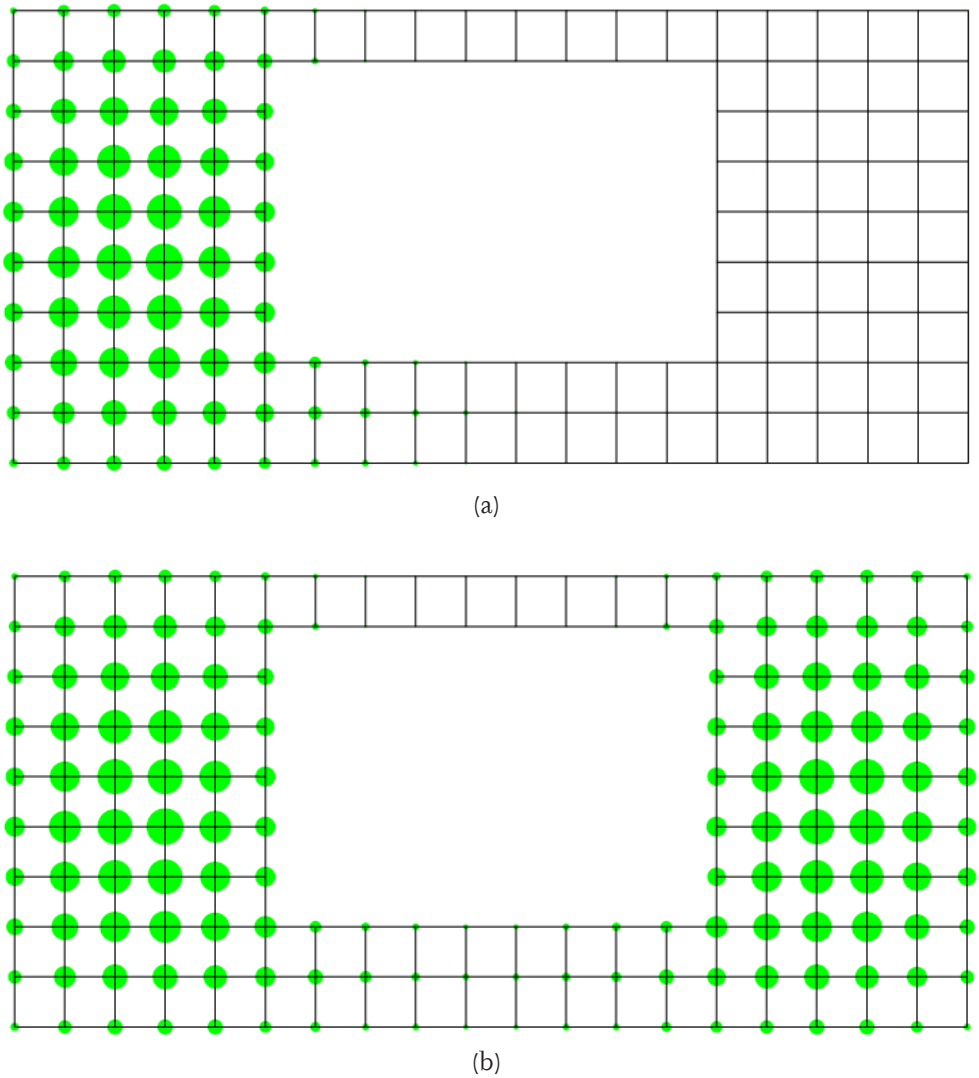
After numerical simulations, result that are visualized on Fig. 2, were obtained.

The same experiments were run with larger networks. These networks had the form of a regular two dimensional grid where some links and nodes were missing and thus splitting the network into several segments (see e.g. Fig 3).

To test the stability of the algorithm, additional activation was provided to several nodes during the experimental iteration process. The findings show that the additional activation during this phase doesn't affect the limit distribution. It causes just a short-



2: *Visualizations of results of numerical simulations with three sample networks*



3: *Visualization of experiment with larger network: (a) a strong cluster after a few dozens of iterations, (b) limit distribution*

time oscillation after which the distribution approaches to the undisturbed process (see Fig. 4). We can thus state that the algorithm is stable to disturbances.

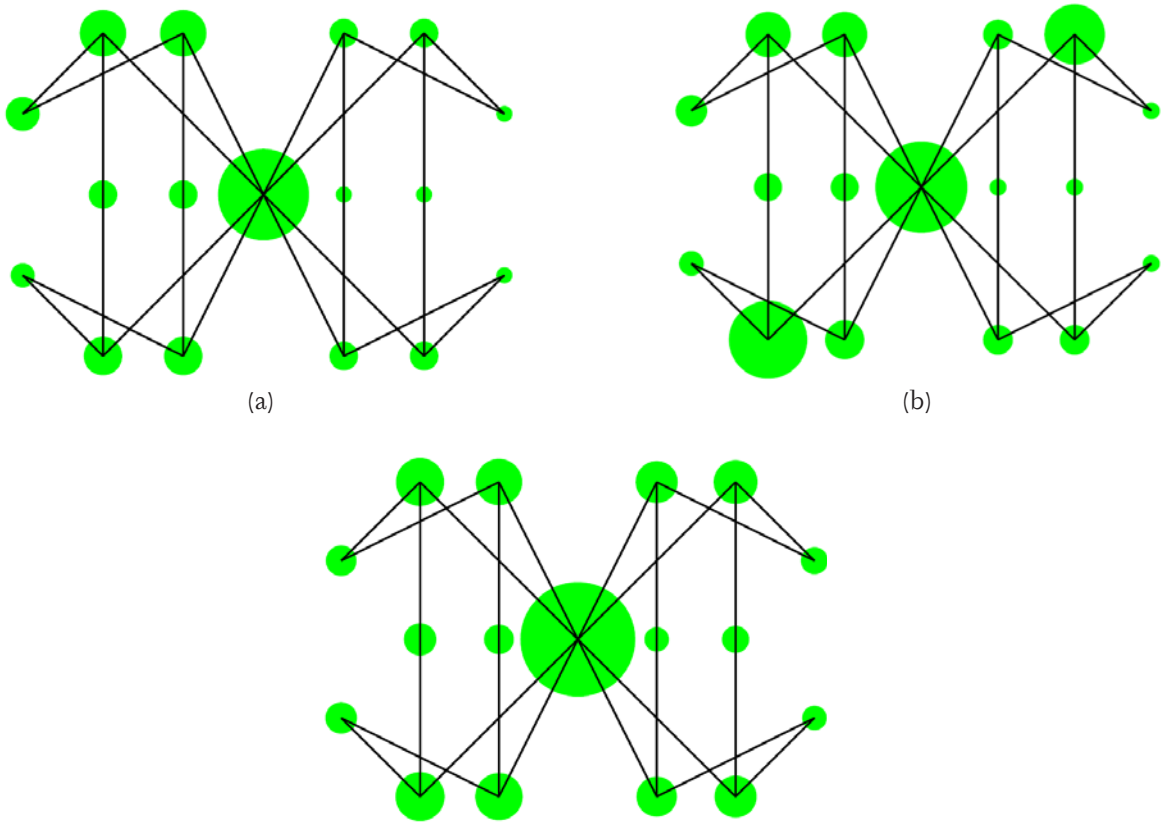
DISCUSSION

It is obvious that the limit distribution depends only on the topology of the network and nodes with more links have higher level of activation. The same results were obtained with different seed (set of initially activated nodes). We can therefore conclude that such results of spreading activation are useful for global ranking for finding the most important nodes in the network.

Big limitation of basic (unconstrained) SA algorithm is the big number of iterations needed for reaching limit distribution. This limitation is obvious mainly in the case of bigger networks. For example, in the experiment from Fig. 3 with initially activated node on the left bottom, the first strong cluster was formed relatively quickly (see Fig. 3a), about after a few dozens of iterations, but the limit distribution of the entire network (see Fig. 3b) was reached after around 15,000 iterations.

Despite of mentioned limitations, the spreading activation algorithm is a good method for analyzing multidimensional network. Because SA is a broad class of algorithms, the choice of parameters of the algorithm is crucial and can be done taking into account the nature of the target application (Troussov, Parra, Brusilovsky; 2009). The basic algorithm could be also modified by adding some kind of constraint that might include distance, fan-out, path or activation constraints (see Crestani, 1997). The results of spreading activation can be also used by other techniques for analyzing the networks that can be domain specific.

At this time, authors are also working on a novel form of spreading activation algorithm that will be able to provide ranking of nodes in a network on different scales – micro, mezzo, and macro (global) levels.



4: *Visualization of experiments with additional activation: (a) the distribution of activation before additional activation, (b) additional activation on nodes I3 and I6, (c) distribution of activation after 5 iterations*

CONCLUSIONS

The paper introduced the approach for representing social networks by graphs and spreading activation algorithm as one of the tools suitable for mining information from multidimensional networks representing e.g. recommender systems (which is one of the promising applications of social networks). Designed experiments showed that the pure, unconstrained form of the algorithm doesn't have very broad utilization (see also Berthold et al., 2009). However, the algorithm is able to provide a stable environment for simulations with networks and there exist a number of possibilities how to modify and/or constrain the algorithm to obtain results more useful for given purpose (e.g. for discovering user similarity or resource relevancy).

Future research will focus on modification of the basic algorithm for obtaining query sensitive results that take into consideration initial conditions of the algorithm and the scale where the search should be performed. The efficiency of the algorithm in term of getting the results in reasonable time will be also important.

SOUHRN

Simulace šíření aktivace v sociálních sítích s využitím teorie grafů

Sociální síť je označením pro struktury složené z lidí, skupin, organizací, počítačů a dalších informací zpracovávajících entit. Tyto sítě umožňují sdílení informací a znalostí a mohou napomáhat při řízení, rozhodování, spolupráci či řešení problémů v různých oblastech činnosti jednotlivců i firem. Vytváření a analýza sociálních sítí je dnes předmětem intenzivního výzkumu. Cílem článku je navrhnout způsob reprezentace sociálních sítí pomocí grafů a využitím teorie grafů na řešení problémů spojených se studiem síťových struktur. Článek popisuje proces modelování multidimenzionálních sítí pomocí orientovaných grafů, včetně jejich nezbytných charakteristik. Pozornost je rovněž věnována metodě šíření aktivace (spreading activation), jako nástroji vhodnému pro analýzu multidimenzionálních sítí se zaměřením na systémy pro doporučování (recommender system). Prezentované experimenty ukazují, že je důležitá vhodná volba parametrů algoritmu, že je nezbytné aplikovat určitý

druh omezení a že je algoritmus schopen poskytnout stabilní prostředí pro simulace prováděné se sociálními sítěmi.

sociální sítě, systém pro doporučování, teorie grafů, šíření aktivace, simulace

Acknowledgements

This paper is supported by the Research program of Czech Ministry of Education number VZ MSM 6215648904/03/03/05.

REFERENCES

BANG-JENSEN, J., GUTIN, G., 2008: *Digraphs: Theory, Algorithms and Applications*. Second Edition. London: Springer Verlag.

BERTHOLD, M. R., BRANDES, U., KÖTTER, T., MADER, M., NAGEL, U., THIEL, K., 2009: Pure Spreading Activation in Pointless. In: *CIKM'09*. Honkg-Kong, ACM. 1915–1918.

CRESTANI, F. 1997: Retrieving Documents by Constrained Spreading Activation on Automatically Constructed Hypertexts. In *Proceedings of the EUFIT 97 – Fifth European Congress on Intelligent Techniques and Soft Computing*, Aachen. 1210–1214.

CRESTANI, F., LEE, P. L., 2000: Searching the Web by constrained spreading activation. *Information Processing and Management*. 36, 4: 585–605. ISSN: 0306-4573.

FREEMAN, L., 2006: *The Development of Social Network Analysis*. Vancouver: Empirical Press.

HANNEMAN, R. A., RIDDLE, M., 2005: *Introduction to social network methods*. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/)

HUANG, Z., CHUNG, W., CHEN, H., 2003: A graph model for E-commerce recommender systems. *Journal of the American Society for Information Science and Technology*. 55, 3: 259–274.

JUNGNICKEL, D., 2007: *Graphs, Networks and Algorithms: Algorithms and Computation in Mathematics*. Third Edition. Springer Verlag.

LIANG, H., XU, Y., LI, Y., NAYAK, R., 2009: Collaborative Filtering Recommender Systems Using Tag Information. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. ISBN: 978-0-7695-3496-1

LUBBERS, M. J., MOLINA, J. L., LERNER, J., BRANDES, U., ÁVILA, J., MCCARTY, C., 2010: Longitudinal analysis of personal networks. The case of Argentinean migrants in Spain. *Social Networks*, 32, 1: 91–104.

TROUSSOV, A., LEVNER, E., BOGDAN, C., JUDGE, J., BOTVICH, D., 2009: Spread of Activation Methods, in Y. Xiang and S. Ali (eds): *Dynamic and Advanced Data Mining for Progressing Technological Development*. IGI Global. ISBN 1-60566-908-3.

TROUSSOV, A, PARRA, D., BRUSILOVSKY, P., 2009: Spreading Activation Approach to Tag-aware Recommenders: Modeling Similarity on Multidimensional Networks. In: *Proceedings of the ACM RecSys'09 Workshop on Recommender Systems & the Social Web*. New York, 2009. 57–62.

TSO SUTTER, K. H. L., MARINHO, L. B., THIEME, L. S., 2008: Tag-aware recommender systems by fusion of collaborative filtering algorithms Export. In: *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*. 1995–1999.

WELLMAN, B., BERKOWITZ, S. D. (Editors), 1988: *Social Structures: A Network Approach*. Cambridge University Press.

Address

Ing. František Dařena, Ph.D., doc. Ing. Jan Žižka, CSc., Ústav informatiky, Mendelova univerzita v Brně, Zemědělská 1, 613 00 Brno, Česká republika, e-mail: frantisek.darena@mendelu.cz, jan.zizka@mendelu.cz, Dr. Alexander Troussov, Ph.D., IBM Dublin Centre for Advanced Studies, Dublin, Ireland, e-mail: atrousso@ie.ibm.com