On Parameter Identification Methods for Markov Models Applied to Social Networks

Denis Fedyanin

V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences
Russia, Moscow, 117997, Profsoyuznaya ulitsa, 65

dfedyanin@inbox.ru

Abstract. In this paper we investigate the mutual influence of participants (agents) of a social network on each other using the framework of Markov models. The main objective of this study was to check several hypotheses concerning dependencies between the influence of agents and their impact on several computational models.

Keywords: social network, dissemination of information, Markov model, influence.

1 Introduction

In this paper, we investigate the mutual influence of participants of a social network (we will call them agents in accordance with the terminology used in [1]). We used a general approach taking its roots in previous work [2]. The initial project was divided into several subprojects. The same data analysis methods were used but the input datasets were different. Work [2] used an on-line community consisting of 964 members. In this paper we use well-known methods but apply them to a different on-line community consisting of 2960 members.

Influence is understood as a process of changes in a subject caused by the behavior of other entities, their settings, intentions, views, assessments and their actions during cooperation with them [3]. Observations of psychologists show [4] that agents in a social network often do not have sufficient information for decision-making or are unable to handle available information, which causes that their decisions can be based on the decisions and/or views of other agents (social influence) [5].

Our analysis is based on data of three communities extracted from the Live Journal website. Live Journal (http://www.livejournal.com) consists of blogs, which contain sequences of messages called posts. Additionally we have available event logs, online diaries and other website content including images, multimedia, texts, etc.

The differences between a blog and a traditional diary are caused by the environment: blogs are usually public and involve third-party readers, who may enter into a public debate with the author, by commenting on blogs.

Authors of posts are called bloggers. The majority of posts are available for reading and commenting by other bloggers. Live Journal also provides an opportunity to bloggers to unite in a community, and subscribe to a community to read their blogs. In this case all the new posts in the selected blogs are displayed in a special news feed. The blogger can belong to several communities at the same time.

Information about communities, subscriptions and records themselves in most cases, are open and accessible to any Internet user. For each of the communities anyone can get the list of participants and the list of friends for each participant. Data consists of three tables: the list of communities, the list of bloggers and the list of links between bloggers. Further in this paper the terms "blogger" and "agent "will be used as synonyms. The number of entries in the list of participants who are members of one of the three communities of Live Journal are 964, 2960, 6587, and the number of links are 6359, 49504 and 190427 respectively.

2 Motivation

There are numerous works about properties of Markov models [16], which describe a social network. Traditional, analysis is mostly theoretical [1, 16], but for the successful application of the obtained theoretical results it is necessary to have well-proven algorithms for identification of the model from the observed data. There are some works where these methods are described, for example [1,16].

Despite the high effectiveness of existing computational algorithms, the main disadvantage of a Markov model is the need to build the initial matrix of influence in the infinite or a high degree. In addition there is some uncertainty in the determination of the initial matrix of mutual trust agents and their relationships with other agents.

In this paper preliminary comparison of different methods for determining the influence of agents of the social network without taking into consideration the data of the messages exchanged between the agents, was conducted. The basis for the identification of the network was data of the agents about whose blogs they read. The format of available data and the sample data is presented in table 1.

Table 1. Data fragment

The ID of the connection between agents	The ID of the reading agent	The ID of the agent, a blog which is being read	The ID of the community, to which belong both agents
1	1	2	1
2	3	2	1
3	4	1	2

In future it is worth to use messages exchanged between the agents of the social network.

3 Review of existing mathematical models

In literature, several approaches have been proposed to describe the interaction between participants in a social network: a Markov model or model of De Groot [6], a Linear Threshold Model [7], Independent Cascade Model [8], a filtering and intrusion model, Ising model, cellular automata model, etc. [16]. The models have been investigated from several perspectives: the conditions of convergence of opinions of members of the social network (see [9]), the dynamic of changes of power, the speed of convergence, the condition of the uniqueness of the final opinion (see [10]). In this work, we will use a model, described in detail in the book [1].

In some models, ranking of agents is used, for example, by means of power indices, index of Houde-Bakker [11], calculation of impact-factor of journals, ranking of web pages, PageRank algorithms, as well as the ordering of parameters "betweenness" [13] , "centrality"[14], "clustering" etc. [5,12,15].

4 Abbreviations and definitions

Because of its wide popularity, the description of Markov models in this work for the sake of brevity was not given. Details can be found, for example, in [1]. Note that transitive influence of the *i*-th agent is defined by

$$w_j = \sum_i a_{ij}^{\infty} , \tag{1}$$

where a_{ij}^{∞} is an element of the transitive closure of the matrix of direct influence, can also be computed for the original stochastic matrix of direct influence. In this case we will call it direct influence of *i*-th agent. The common method of agent identification is based on the direct influence matrix which is derived from the adjacency matrix by the formula where a_{ij} is a weight in the matrix of direct influence and b_{ij} is an element of the adjacency matrix.

$$a_{ij} = \frac{b_{ij}}{\sum_i b_{ij}} \tag{2}$$

In some cases, one can try to take into account the impact of the authority of the agent on the strength of the influence. We consider the case when the impact of authority is proportional to the number of friends of the agent, where f_j represents the credibility of the *i*-th agent.

$$a_{ij} = \frac{f_i\left(b_{ij}\right)}{\sum_i f_i\left(b_{ij}\right)} \tag{3}$$

$$f_i(x) = \left(\sum_i b_{ij} \right)^{\beta} \tag{4}$$

5 Hypotheses

1. Direct influence depends on the number of friends of an agent.
2. The number of friends is not correlated with transitive influence.
3. There is a correlation between transitive influences of agents, calculated by different methods taking into account the authority of agents.
4. The direct influence of the agent does not correlate with its transitive influence.
5. Implementation of hypotheses does not depend on the size of the network.

6 Data analysis results

Testing hypothesis 1 reveals that direct influence depends on the number of friends of an agent, and the relationship between them is close to a power-law function as shown in figure 1. The coefficient of correlation is 0.85.

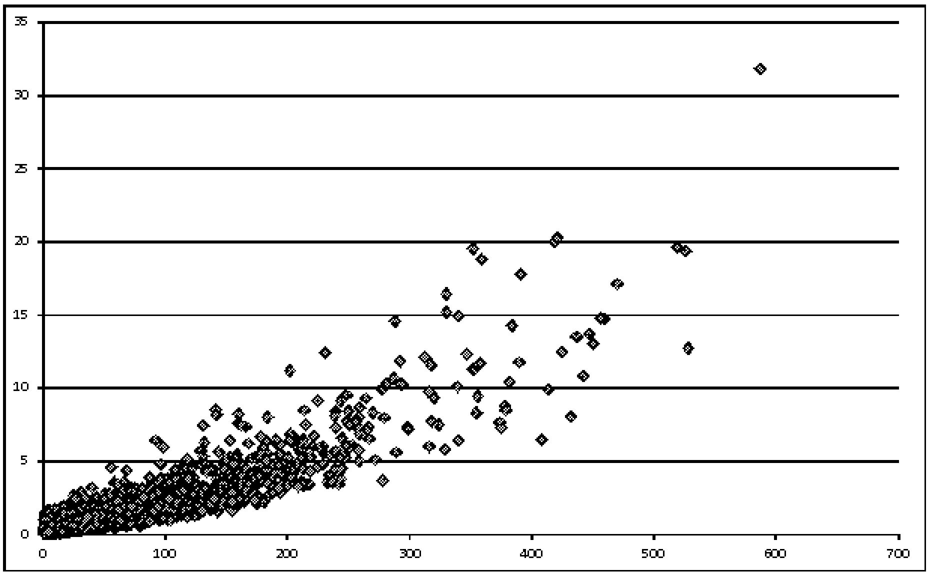

Fig. 1. The dependency between direct influence of agents (vertical) and the number of the agents' friends (horizontal).

Testing hypothesis 2 revealed that the number of friends is not correlated with transitive influence. This is shown in figure 2. The coefficient of correlation is 0.72. In addition to a linear dependence we observe the almost vertical "tail". Its presence means that there are several agents who have a small number of friends, but a substantial influence. In particular, there are three agents for whom the transitive influence exceeds the transitive influence of the agent with the highest number of friends (whose influence can be assumed). The existence of this phenomenon has been theoretically predicted, but validation on real experimental data had not been performed yet. Note that we do not yet have an explanation for the presence of only two main lines in diagram and this issue should be investigated more thoroughly in the future.

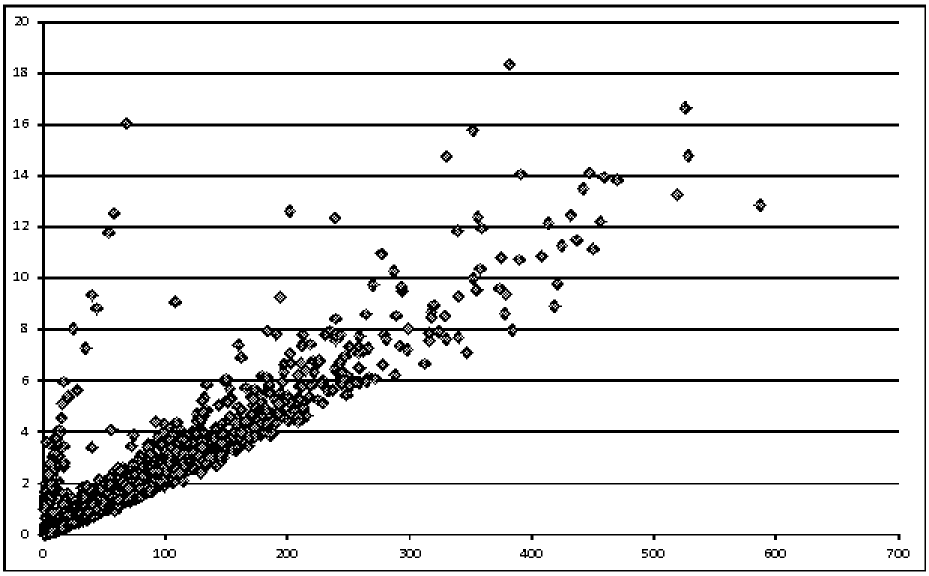

Fig. 2. The dependence of the transitive influence of agents (vertical axis) on the number of agents' friends (horizontal axis).

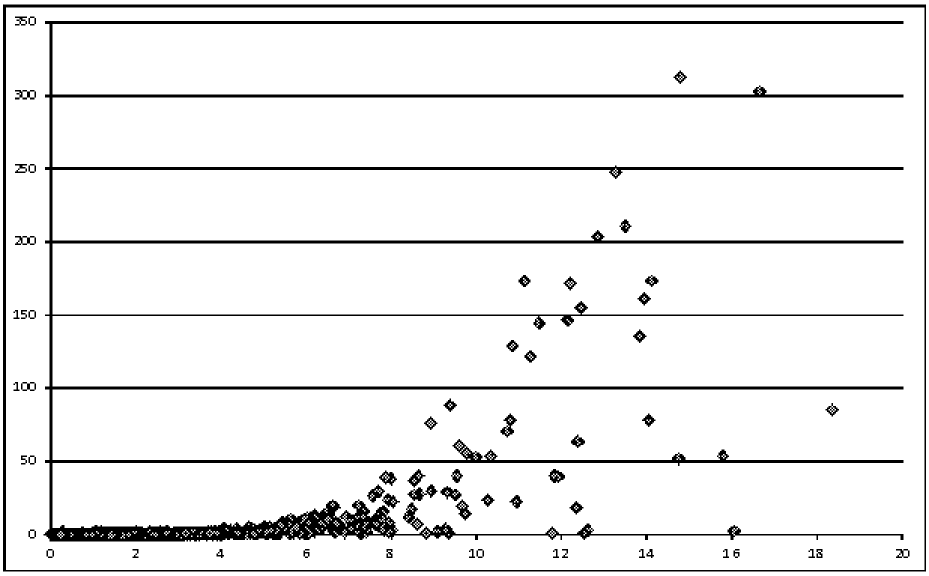

Fig. 3. The transitive influence of agents, calculated without taking into account their authority (horizontal axis) and calculated while taking authority into account (β=4) (vertical axis).

Testing hypothesis 3 revealed that there is no correlation between transitive influences of agents, neither without taking into account the authority, nor when taking authority into account. In figure 3 we see that there is no correlation. The coefficient of the correlation is 0.31. This is an important observation because by making assumptions about the impact of the number of friends of an agent on his credibility, you can get, generally speaking, different results. If we ignore some of the outlier observations, we can once again identify two "tails". The main tail shows a linear dependency, which is not equal to the constant β, and the second tail indicates a non-increasing transitive influence of the agents, despite of the increase in their transitive influence in the case of not taking into account their credibility. Moreover, the figure shows that there are a number of influential agents with low authority. This is consistent with the result that we received in the process of verification of hypothesis 2. So it can be argued that the correlation between transitive influences is complex in nature, and thus, hypothesis 3, cannot be affirmed without additional clarifications.

Testing hypothesis 4 revealed that the direct influence of an agent does not correlate with its transitive visibility. In figure 4 you can see that the linear correlation between direct and transitive influences is not clear. The coefficient of correlation is 0.78. This is also interesting, since we believe that not all agents can make decisions based on the computation of transitive influence, and therefore are forced to use direct influence measures.

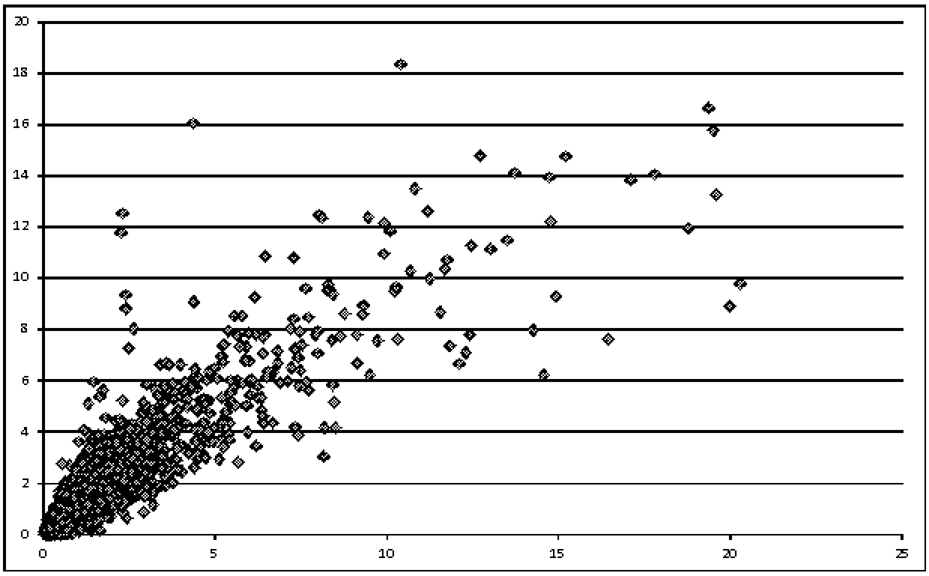

Fig. 4. The dependence of the transitive influence of the agent (vertical) from its direct influence (horizontal), which was calculated without taking into account the authority of agents.

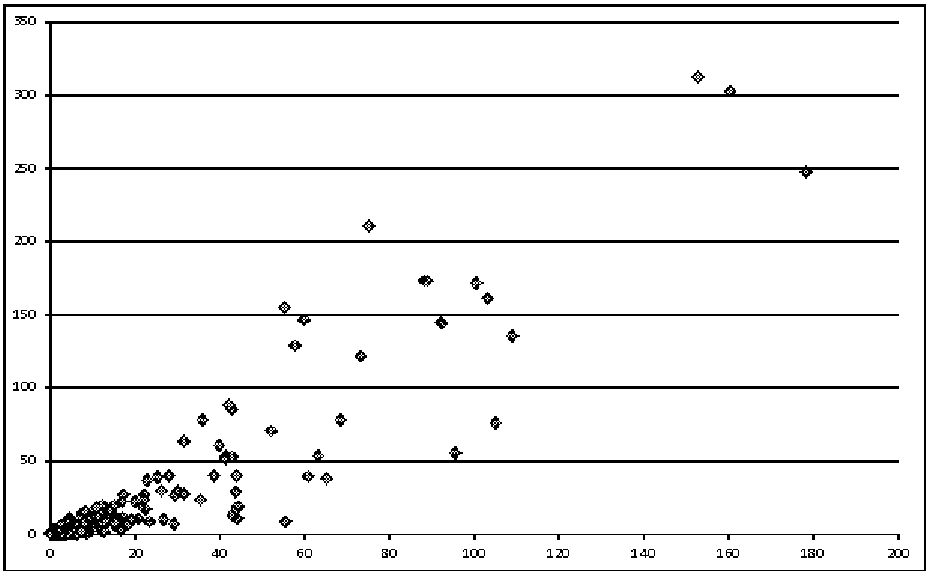

Fig. 5. The dependency of the transitive influence of the agent (vertical) from its direct influence (horizontal), calculated taking into account the authority of agents

Then we come to a rather obvious conclusion, that in real social networks such agents can be mistaken. However, we conclude that there is no ground for hypothesis 4. In the case shown in figure 5, the linear correlation is noticeable. However, it is different for small values of direct influence than for higher values, where you can also identify the correlation. The coefficient of the correlation is 0.92.

Hypothesis 5 states the assumption that the validity of the hypotheses does not depend on the size of the network. This has not yet been verified and is a possible direction for future research.

7 Conclusions and future work

The study showed the presence of a certain number of anomalies and effects that need to be taken into account while identifying optimal Markov model parameters for experimental data. It was shown that the credibility of agents has a significant impact on the influence of agents. It was shown that there is a specific dependence between transitive influence and direct influence. We identified an abnormal cluster of agents, which have a small number of friends, but which have a great transitive influence.

It may be interesting to continue our study by verifying hypothesis 5, as well as including in the analysis the possibility of taking into account the exchange of messages between agents. We also intend to investigate the ranking of agents using methods such as alpha-centrality, the PageRank algorithm, as well as other widely used methods based on direct and transitive influences of agents.

Acknowledgement. The research is supported by the grant 10-07-00129 of Russian Foundation for Basic Research. We would like to express our gratitude to Jonas Poelmans and Dmitry Ignatov for improving the language quality.

References

1. Gubanov, D.A., Novikov, D.A., Chkhartishvili, A. G.: Social network: a model of information influence, control and confrontation. Fizmatlit, 228 p., Moscow (2010) (in Russian)
2. Fedyanin, D.N.: Application of Markov models for the analysis of influence of the participants of the Internet-community. In: Lecture Notes of the all-Russian scientific-practical conference "Analysis of Images, Networks and Texts" (AIST 2012), pp. 132-143. The national Open University "INTUIT", Yekaterinburg (2012)
3. Glossary on Control Theory and its Applications, http://glossary.ru (in Russian)
4. Deutsch, M., Gerard, H.: Study of Normative and Informational Social Influence upon Individual Judgment. In: Journal of Abnormal and Social Psychology. no.51, pp. 629-636. (1995)
5. Zuyev, A.S., Fedyanin, N.A.: Model of management of views agents in co-social networks. In: the Problems of management. № 1. pp. 37-45. ICP RAS, Moscow (2011). (in Russian)
6. DeGroot, M.H.: Reaching a Consensus. In: Journal of American Statistical Association. №69, pp.118-121 (1974)
7. Ganovetter, M.: Threshold Models of Collective Behavior. In: American Journal of Sociology, vol. 83. №6, pp.1420-1443 (1978)

8. Goldberg, J., Libai, B., Muller, E.: Talk of the Networks: A Complex Systems looks at the Underlying Process of Word-of-Mouth. In: Marketing Letters, №2, pp.11-34 (2001)
9. Berger, R.L.: Necessary and Sufficient Conditions for Reaching a Consensus using DeGroot's method. In: Journal of American Statistical Association, vol. 76, pp. 415 – 419 (1981)
10. Golub, B., Jackson, M.O.: Naive Learning in Social Networks: Convergence, Influence and Wisdom of Crowds. Technical Report 64 (2007)
11. Hoede, C., Bakker, R.: A Theory of Decisional Power. In: Journal of Mathematical Sociology, №8, pp. 309-322. (1982)
12. Rusinowska, A., Swart, H.: Generalizing and Modifying the Hoede-Bakker Index. In: Theory and Applications of Rational Structures as Knowledge Instruments. №2. Springer's Lecture Notes in Artificial Intelligence 4342, pp. 60-88. Springer (2007)
13. Freeman, L.: A set of measures of centrality based upon betweenness. In: Sociometry №40, pp. 35–41. (1977)
14. Borgatti, S, Everett, M.: A Graph-Theoretic Perspective on Centrality. In: Social Networks, 28. pp. 466–484. Elsevier (2005)
15. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press (1994)
16. Jackson, M.: Social and Economic Networks. Princeton: Princeton University Press (2008)